PATIENT-CENTERED OUTCOMES RESEARCH INSTITUTE
FINAL RESEARCH REPORT

Testing a Program to Improve the Integration of Behavioral Health Care into Primary Care Settings

Benjamin Littenberg, MD[1]; Jessica Clifton, PhD[1]; Abigail M. Crocker, PhD[1]; Laura-Mae Baldwin, MD[2]; Levi N. Bonnell, PhD, MPH[1]; Ryan E. Breshears, PhD[3]; Peter Callas, PhD[1]; Prama Chakravarti, PhD[4]; Kelly Clark/Keefe, EdD[5]; Deborah J. Cohen, PhD[6]; Frank DeGruy, MD[7]; Lauren Eidt-Pearson, MSW, LICSW[8]; William Elder, PhD[9]; Chet Fox, MD[10]; Sylvie Frisbie[1]; Katie Hekman, RN[11]; Juvena Hitt, MPH[1]; Jennifer Jewiss, EdD[5]; David C. Kaelber, MD, PHD, MPH[12,13]; Kairn Stetler Kelley, PhD[1]; Rodger Kessler, PHD[7]; Jennifer B. O'Rourke-Lavoie, BA[1]; George S. Leibowitz, PhD[14]; C. R. Macchi, PhD[15]; Matthew P. Martin, PhD[15]; Mark McGovern, PhD[16]; Brenda Mollis, MPH, MPA, MA[2]; Daniel Mullin, PsyD, MPH[8]; Zsolt Nagykaldi, PhD[17]; Lisa W. Natkin, PhD[5]; Wilson Pace, MD[18]; Richard G. Pinckney, MD, MPH[1]; Douglas Pomeroy[1]; Paula Reynolds, BA[1]; Gail L. Rose, PhD[1]; Sarah Hudson Scholle, DrPH, MPH[19]; Jeni Soucie, MS[19]; William J. Sieber, PhD[11]; Terry Stancin, PhD[12]; Kurt C. Stange, MD, PhD[20]; Kari A. Stephens, PhD[2]; Kathryn Teng, MD[12]; Elizabeth Needham Waddell, PhD[6]; Constance van Eeghen, DrPH, MHSA, MBA[1]

AFFILIATIONS:

[1] Department of Medicine, Larner College of Medicine, University of Vermont, Burlington
[2] Department of Family Medicine, University of Washington, Seattle
[3] Wellstar Psychological Services, WellStar Health System, Marietta, Georgia
[4] Prama Psychology, PLLC, Sammamish, Washington
[5] College of Education and Social Services, University of Vermont, Burlington
[6] Department of Family Medicine, Oregon Health & Science University, Portland
[7] Department of Family Medicine, University of Colorado School of Medicine, Aurora
[8] Family Medicine and Community Health, University of Massachusetts Chan Medical School, Worcester
[9] Behavioral and Social Sciences, Tilman J. Fertitta Family College of Medicine, University of Houston, Houston, Texas
[10] Department of Epidemiology and Environmental Health, School of Public Health and Health Professions, University at Buffalo, Buffalo, New York
[11] Department of Family Medicine, University of California San Diego, San Diego
[12] Behavioral Health Services, MetroHealth System, Cleveland, Ohio
[13] Internal Medicine-Pediatrics, Case Western Reserve University, Cleveland, Ohio
[14] School of Social Welfare, Stony Brook University, Stony Brook, New York
[15] College of Health Solutions, Arizona State University, Phoenix
[16] Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California
[17] Department of Family and Preventive Medicine, University of Oklahoma College of Medicine, Oklahoma City
[18] DARTNet Institute, Aurora, Colorado
[19] National Committee for Quality Assurance, Washington, DC
[20] Case Western Reserve University, Cleveland, Ohio

To cite this document, please use: Littenberg B, Clifton J, Crocker AM, et al. (2024). *Testing a Program to Improve the Integration of Behavioral Health Care into Primary Care Settings.* Patient-Centered Outcomes Research Institute (PCORI). https://doi.org/10.25302/01.2024.PCS.140924372

Article Information

Original Project Title: Integrating Behavioral Health and Primary Care

PCORI ID: PCS-1409-24372

Date submitted: 1/11/2023
Date accepted: 9/20/2023

TABLE OF CONTENTS

ABSTRACT

Report Abstract

Background: Most patients with behavioral health (BH) conditions such as anxiety, depression, chronic pain, and substance use disorder are seen in primary care settings. Some primary care practices provide integrated BH (IBH) care, in which medical and BH professionals work as a team in 1 location; however, it is difficult to achieve high levels of integration in this setting. We tested a comprehensive, practice-level intervention designed to improve outcomes in adults with multiple chronic medical and BH conditions by increasing BH integration.

Objectives: Our overall objective was to evaluate a practice-level intervention designed to increase BH integration within adult primary care. Aim 1 assessed the effect of the intervention on patient-reported functional status. Aim 2 analyzed the degree of practice integration. Aim 3 explored factors supporting or impeding successful integration. Aim 4 estimated practices' costs of implementing the intervention. Aim 5 investigated the impact of the COVID-19 pandemic on the effect of the intervention. Aim 6 examined the relationship between practice integration and patient outcomes independent of the intervention. Aim 7 modeled the effects of social determinants of health on patients over time. Aims 8 and 9 evaluated the impact of the COVID-19 pandemic on patient outcomes and practice integration. Aim 10 quantified occupational burnout among primary care clinicians and staff before the pandemic.

Methods: For this trial (September 2017 through December 2020), we used a pragmatic, cluster-randomized design. Forty-two primary care practices with on-site BH care were randomly assigned to intervention or usual care groups. The intervention was a 24-month practice change process including an online curriculum, a practice redesign and implementation workbook, remote quality improvement coaching, and an online learning community. Approximately 75 patients with multiple chronic medical and BH conditions from each practice were asked to complete a series of surveys of patient-centered outcomes. Treatment allocation was concealed from patients until the end of the trial. Primary outcomes for aim 1 were changes in the 8 PROMIS®-29 domain scores. Secondary outcomes were changes in medication adherence, self-reported health care use, time lost as a result of disability, cardiovascular capacity, patient centeredness, clinician empathy, and several condition-specific measures. For aim 2, a sample of practice staff completed the Practice Integration Profile (PIP) to estimate the degree of BH integration at their site. Aim 3 used a mixed-methods collective case study of 8 sites. For aim 4, a total of 8 practices estimated the costs of implementing the intervention. For aim 5, an additional patient survey was administered in 2020 and 2021 to model the interaction of the burden of COVID-19 with the intervention's effects. The remaining aims used the data described above in primary and secondary analyses. For each quantitative outcome, we used mixed linear regression models with change in the outcome from baseline to follow-up as the dependent variables. Colocation alone vs colocation plus intervention and potentially confounding variables were modeled as fixed effects. Practice was included as a random intercept.

Results: The final analytic sample had 967 participants in the intervention group and 1459 participants in the usual care group. Participants' average age was 61.9 years, and 65.5% of participants were female. The average participant had 4.4 chronic conditions, most frequently chronic pain (84%), hypertension (82.9%), and depression (46.8%). The 42 participating practices offered family medicine (20/42), internal medicine (7/42), or a mixture of services (15/42). They had an average PIP score of 59.4. The intervention had no effect on any of the primary or secondary patient outcomes (aim 1). Adjusted changes in PROMIS-29 domain scores were all 0.23 or lower on a scale of 0 to 100 points. We used 95% CIs to ensure that we excluded random, clinically meaningful differences for all variables. Subgroup analyses showed no patterns of effect in any population. There was a small effect on the degree of practice integration, reaching statistical significance in 1 of 7 integration domains (aim 2). Factors supporting and impeding integration included leadership commitment, finances, workflow and community systems, and clinicians' shared perspectives (aim 3). The median cost to a practice for implementing the intervention was $20 726 (aim 4). We identified no interactions between the burden of COVID-19 and the effectiveness of the intervention (aim 5). The level of integration was associated with patient outcomes independent of the intervention, both at baseline and longitudinally (aim 6). Specifically, the PIP total score was associated with better physical and mental health function as measured with the PROMIS-29 summary scores and all domain scores. Though some of these associations were statistically significant, the magnitude of the associations suggests the differences are not clinically meaningful at the individual level. Financial, food, and housing insecurities had lasting effects on patients' health, even when earlier insecurities were resolved (aim 7). Patients with the heaviest perceived personal and community burdens as a result of the COVID-19 pandemic had worse health (aim 8). Practices whose final assessments occurred later in the pandemic reported smaller improvements in practice integration: a 0.4 decrease in total PIP score per week (95% CI, −0.70 to −0.04) (aim 9). Occupational burnout varied by role and burnout domain: Medical residents had the highest levels of burnout within each domain (aim 10).

Conclusions: The specific intervention tested in this study was inexpensive and had a small impact on the degree of BH integration but none on patient outcomes; however, practices that had more integration at baseline had better patient outcomes independent of the intervention. Social determinants of health, the burden of COVID-19, and occupational burnout were found to be serious challenges to providing primary care to adults with multiple chronic conditions. Although this particular intervention did not achieve all its expected results, this study should not be interpreted as a failure of the IBH model. Integrated BH care remains an attractive strategy for improving patient outcomes.

Limitations: The design of the study precludes drawing conclusions about the value of IBH distinct from this specific intervention. Generalizability may be limited.

COVID-19 Extension Abstract

Background: The COVID-19 pandemic has had far-reaching consequences for the health care delivery system. Quality improvement efforts, such as those tested in the Integrating BH and Primary Care (IBH-PC) clinical effectiveness trial, were significantly challenged by the changes and restrictions imposed by the pandemic. This might have led to a reduction in effectiveness or, conversely, a greater impact because of greater service needs generated by the pandemic.

Objective: We sought to understand if the COVID-19 pandemic influenced the effect of the IBH-PC intervention on patient outcomes.

Methods: We took advantage of a natural experiment imbedded within the IBH-PC trial. Baseline patient function was collected before the pandemic in 42 primary care practices across 13 states. Follow-up data were collected 2 years later, at which time the pandemic affected most areas of the country with varying levels of intensity. We measured the local burden of COVID-19 by 8 variables including the local incidence and prevalence of COVID-19 infections and patient reports of the personal and community disruption they perceived. The study outcomes were the 8 domains of the PROMIS-29. For each outcome, we built a multivariable, mixed linear regression model with change in the outcome from baseline to follow-up as the dependent variable. Assignment to the randomized IBH-PC intervention (colocation alone vs colocation plus intervention) was modeled as a fixed effect, as were a number of potentially confounding variables. Practice was included as a random intercept. The effect of the pandemic on the intervention was measured as the interaction of COVID-19 burden with the intervention. Because there were 8 predictors and 8 outcomes, we ran a total of 64 models.

Results: A total of 2225 participants were available for analysis. The interaction between the intervention and the burden of COVID-19 reached significance in 5 of the 64 models tested (7.8% [95% CI, 2.6%-17.3%]), similar to the 5% rate of nominal significance expected from random error alone. These 5 models included 5 different outcomes and 4 different COVID-19 markers.

Conclusions: These data provided no evidence that the COVID-19 pandemic was responsible for the apparent lack of effect of the IBH-PC intervention.

Limitations: Our ability to model the burden of COVID-19 was limited by the spatial resolution of epidemiologic data and potential recall bias.

BACKGROUND

The chronic diseases that drive mortality, morbidity, and health care costs in the United States and around the world are largely behavioral in origin and management. Tobacco and alcohol use, diet, physical inactivity, and substance use disorder together account for 38% of all deaths in the United States.[1] Nonadherence to treatment, insomnia, anxiety, depression, and stress contribute to additional morbidity and mortality. Comorbid medical and behavioral concerns drive especially poor outcomes and high costs.[2-4]

Behavioral concerns can often be effectively managed with improved outcomes for patients, their families, and the health care system. Several psychological and behavioral techniques, including cognitive behavioral therapy[5-9] and the Improving Mood—Promoting Access to Collaborative Treatment (IMPACT) study's collaborative care model,[10-12] improve care, especially for depression. Forty percent of primary care patients need behavioral health (BH) services, and the rate is higher among patients with chronic medical conditions.[13-15] Although some primary care professionals (PCPs) are skilled and effective at delivering BH services, many do not have the training, time, or inclination and fail to either identify these problems or seek out other resources to manage them.[14,16] High-quality BH services are the single most difficult area of medical care for patients to access,[14,17] and nearly 70% of the more serious behavioral conditions seen in primary care are neither assessed nor treated,[18,19] with up to 85% of patients not receiving needed mental health care.[20-23]

Behavioral health is shorthand for a wide variety of services that have often been physically, operationally, historically, and even ideologically isolated from one another. These include mental health care, substance use disorder care, health behavior changes, and attention to family as well as other psychosocial factors including management of depression, anxiety, stress, insomnia, overeating, inactivity, smoking, treatment nonadherence, chronic pain, alcohol use, and addiction.[24]

The traditional, and still most prevalent, approach to addressing BH needs in the primary care setting is to refer patients to specialists (psychiatrists, psychologists, counselors, and therapists) in another practice. High-quality BH services are often not available, referrals are difficult, and treatment initiation rates are low in part because of the perceived stigma of seeing a mental health specialist.[14,25,26] Communication between mental health specialists and PCPs is often hampered by confidentiality concerns, legal restraints, and tradition. Many referrals for specialty mental health and substance use disorder services never generate an appointment, and of those patients who do make an appointment, almost half never initiate care.[21,27]

In response, many practices have initiated colocation in which a BH professional (BHP) such as a psychologist or counselor is housed in or near the primary care practice.[28] This eliminates some of the barriers to access, but the BHP functions independently, with separate hours, office space, appointment systems, and medical records. Often, there is inadequate communication among primary care team members and an inability to share progress notes.

A more effective solution may be integrated BH care (IBH) in which the BHP is a full member of the practice, sharing workspace, infrastructure, records, and support systems; participating fully in the life of the practice; and collaborating closely with PCPs in patient assessment and management. In IBH, population management; protocol-driven, evidence-supported care; screening for BH needs; guidelines for external referral; brief visits; and patient engagement are systematized. Integrated BH significantly improves treatment initiation rates compared with referral,[20,27,29] reduces missed and cancelled appointments, improves clinician satisfaction and use,[20,30] and may be associated with improved patient health and well-being.[31,32]

Essentially, IBH is the care a patient receives when PCPs and BHPs collaborate within a system of care rather than focusing on separate approaches.[33] It takes advantage of a wider span of data, goals, and key concerns, including those of the patient and their family. It works to make treatment rational, pick up on missed issues, and boost patient and family engagement. The goals of IBH are to create a system that offers better and more efficient care for mental health and substance use disorder conditions, health behaviors, and chronic medical illnesses; better recognition and management of life stressors, crises, and stress-related physical symptoms; and more thoughtful and efficient health care use. It is important to note that this view applies to the common problems seen in adult primary care and is not meant to isolate other important domains, patient groups, or specialties. For primary care practices to effectively shift from co-located–only services to IBH, a supported practice change process is required.[24,34]

Nearly half of primary care medical home practices now offer some sort of BH services on-site,[16,35] although many still have neither systematic approaches nor specialty personnel.[13] There is also wide variation in implementation models and targets of care.[32,36] Significant obstacles remain to new clinical workflows and financial sustainability.[14,32,37]

There is a gap in evidence to support care and policy decisions when selecting models and elements of IBH. Two Agency for Healthcare Research and Quality (AHRQ) evidence reports[5,38] and a Cochrane systematic review[39] concluded that there is a need for research concerning the effectiveness of different models and elements of IBH. The AHRQ report "Future Research Needs for the Integration of Mental Health/Substance Abuse and Primary Care"[38] specifically declared the need to test variations in delivery models, compare integrated practice with other approaches, investigate diagnoses beyond depression (especially comorbidity), and investigate which patient subgroups are most likely to benefit from integration. Likewise, "A National Agenda for Research in Collaborative Care"[40] (an alternative name for IBH) called for investigations into

- how IBH affects comorbid conditions or disease clusters (eg, diabetes, depression, and coronary artery disease);
- whether patient experiences and clinical outcomes are better in collaborative practices than in usual care; and
- what functional components of collaborative care have the greatest effect on outcomes.

A recent review of primary care BH (PCBH) models indicated that patients, practitioners, and systems seek to understand what works in terms of feasibility and outcomes, for which patients, and under what conditions.[41] The authors of this review also called for a better understanding of the cost of implementation so that decision-makers can be better informed.

For practices that seek fuller integration, there are no well-described, validated strategies for overcoming the myriad issues that arise during the attempt. The Integrating BH and Primary Care (IBH-PC) trial responded to this need by testing the effectiveness of a comprehensive, practice-level intervention—the IBH-PC toolkit, with practice facilitation support, workbooks on process improvement, education modules on IBH, and a learning community—designed to improve outcomes in patients with multiple chronic medical and BH conditions by increasing a practice's degree of BH integration.

Practice facilitation, defined by AHRQ as "a supportive service provided to a primary care practice by a trained individual or team,"[42] is widely documented as effective,[43-47] although there are exceptions that identify its limitations.[48-50] Although an external facilitator (a facilitator who visits the clinic periodically) brings expert knowledge and resources to primary care, such facilitation may suffer from a lack of team-based relationships with clinic members.[51-53] An internal facilitator (a clinic member) can capitalize on strong team relationships, but internal facilitation may suffer from lack of knowledge and supportive resources.[54-56] Successful quality improvement (QI) efforts depend on the ability of practice facilitators to tailor operational expectations to the practice setting,[57] to which Phillips proposed coaches to support facilitation in practice transformation.[58,59] In the IBH-PC study, we chose to combine the strengths of an internal facilitator with those of an external coach.

One of the challenges facing internal facilitators is skill with QI processes and knowledge of the clinical focus, such as integrated care. In response, another kind of resource has emerged over the past 2 decades: toolkits. A toolkit is "an intervention package, or set of tools . . . aimed at quality improvement."[44] Implementation toolkits on their own or with other strategies, however, have mixed results in changing clinician behavior.[44-48] Successful toolkit use appears to be situationally dependent, and the toolkits themselves are sometimes overwhelming to their users.[42-59] We chose to develop a customized toolkit that combined process improvement with integrated care tactics and external coaching for practice facilitators to support locally identified change strategies.

We deployed the IBH-PC toolkit in a broad array of primary care practices and tested its impact on patient functional status as well as on the practices' level of integration. We used a pragmatic randomized trial design to allow our analysis and conclusions to be both rigorous and applicable to a broad range of practices. We sought to understand which groups of patients benefited from it most, which aspects of integration were most valuable, and what contextual factors within practices, such as financial concerns or health care professionals' attitudes, supported or impeded implementation of patient-centered IBH. We also assessed various aspects of clinic investments in time and cost associated with the implementation of the IBH-PC program and the association between integration and patient health independent of the intervention.

During the data collection phase of this study, the COVID-19 pandemic emerged, placing an unprecedented burden on patients, clinicians, and communities. Beyond the direct impact of COVID-19 (eg, respiratory illness, blood-clotting disorders), virtually everyone has been affected by closings, lockdowns, the need to care for family (sick and well), loss of employment, financial disruptions, barriers to care, new anxieties, and changes in the usual social structures that support physical and mental health. These are the very issues that IBH in primary care—a tight coupling of medical and BH services designed to improve patient-reported functional outcomes[60]—is best at managing. In these new circumstances, however, little is known about the continued value of IBH, especially for our most vulnerable patients.

To address the impact of COVID-19, we assembled patient co-investigators, stakeholder representatives, and researchers engaged in the IBH-PC study for a series of group discussions via written and live meetings to understand the data needed to evaluate these issues. Although there was great interest in understanding how patients in general were faring and how primary care was responding to the pandemic, a consensus emerged that the IBH-PC study had special strengths in regard to understanding how the pandemic influenced the effects of BH on patient outcomes. Did the pandemic alter the effectiveness of the IBH-PC intervention? Was an individual’s perceived impact of COVID-19 on themselves and their community related to their health? Were practice staff’s perceptions of the impact of COVID-19 on their practice and community related to their conduct of IBH?

From a policy perspective, if COVID-19 had overwhelmed IBH’s ability to improve outcomes, then IBH may not have been enough; something more or different would be needed. If it hadn’t overwhelmed IBH, though, an even greater emphasis on integrating BH into primary care would be warranted.

Upon completion of the funded project period, we requested an extension of our work to conduct additional analyses of the existing data. Important questions had emerged about the associations between various patient- and practice-level outcomes and other variables in our extensive database, such as practice integration, social determinants of health (SDOH), the burden of COVID-19, and health care worker burnout.

Research Questions

This research addressed 5 main questions:

- Does access to the IBH-PC toolkit, a practice-level intervention, affect patient-centered outcomes in adults with multiple chronic medical and BH conditions?
- Does access to the IBH-PC toolkit affect the degree of practice-level BH care integration in primary care practices?
- What factors support or impede successful integration of BH care into primary care practices?
- What are the costs of implementing the IBH-PC toolkit?
- What is the impact of the COVID-19 pandemic on the effect of the intervention on the patient- and practice-level outcomes collected in aims 1 and 2?

During an extension period, we used the data collected during previously completed activities to analyze 5 additional questions:

- What is the association of practice integration with patient outcomes?
- What is the association between SDOH and patient outcomes?
- What is the impact of the COVID-19 pandemic on patient outcomes?
- What is the impact of the COVID-19 pandemic on practice outcomes?
- What are the moderators and mediators of burnout among PCPs and staff?

To answer the first 5 questions, we designed an intervention in collaboration with our patient advisors that consisted of 4 elements: an educational program to support the skills of all practice members involved (leaders, clinicians, and staff); a protocol-based redesign process in the form of workbooks to guide structured improvement activities and decision points; a coach to support the practice change team in the use of the workbook materials (tactics, references, checklists, protocols, forms, and workflows); and an online learning community for shared problems and solutions. Together, these 4 elements are designed to assist practices in improving IBH given their local environment and constraints.

PARTICIPATION OF PATIENTS AND OTHER STAKEHOLDERS

Participants

We included patients, their family members, BHPs and PCPs, policymakers, and payers in the planning, conduct, analysis, and dissemination of our engagement process. All types of stakeholders were engaged at 3 levels: as investigators on the research team; on the stakeholder advisory group (SAG) to provide the research team with independent advice and feedback on key characteristics of the study, its conduct, and its findings; and on clinical site redesign teams that planned and implemented IBH locally.

Patient Co-Investigators

Three patients representing our target population (patients with chronic, comorbid conditions and their families) were recruited by the principal investigator and researcher co-investigators based on their personal or professional treatment relationships, and they participated as co-investigators on the project and as employees of the University of Vermont (UVM). These patient co-investigators included (1) an actor and artist whose daughter has a long-term chronic disease; (2) a retired elementary school teacher with chronic congestive heart failure, who receives IBH; and (3) a musician, writer, and salesman with chronic neuromuscular disease, depression, and heart disease. The patient co-investigators were paid for their time, and all 3 remained active contributors for the duration of the project.

Stakeholder Advisory Group

Clinicians and BHPs from primary care practices nominated patients and other stakeholders who could provide us with valuable input to serve on the SAG. Membership in our original group of 10 stakeholders grew and evolved to 18 participants who represented active health care professionals, national patient advocacy groups, national clinician groups, policymakers and payers, and the patient voice. The SAG was led by a patient co-investigator. Retention of SAG members was high. We paid them for participating in meetings and other activities, and the leader of the SAG engaged with the members personally. Members found the opportunity to provide input in ways that were directly relevant to the study and that drew on their unique expertise to be the most rewarding.

Clinical Site Redesign Partners

The site redesign teams were encouraged to include 1 to 2 patients alongside clinicians and staff. We provided the redesign teams with a patient partner guide[61] developed by our patient partner co-investigators to facilitate the identification, orientation, and engagement of the patient team members. We also included a manual for the patients themselves to orient them to their role. It was suggested that practitioners identify patients based on their knowledge of that person and their experience with, and understanding of, managing multiple chronic conditions (MCCs); their interest in learning more and contributing to the practice's IBH improvement plans; their ability to consider all team members' perspectives; and their capacity to work virtually as well as on a team.[62] A total of 6 practices out of 20 engaged with and retained patients on their redesign teams.

Principles of Engagement

Reciprocal Relationships

Early on, the research team developed "rules of engagement." For example, the research team agreed to respect the physical and mental health limits of its members, which could change over time. As a result, patient partner co-investigators defined their own boundaries of participation but were always treated as partners. Decisions were generally made by consensus. Input from all team members was expected and supported. As needed, the principal investigator made the final decision for the team.

Co-Learning

Education to support shared understanding of the research process was provided to all patient partners, who were offered the opportunity to participate in IRB training. Notes were kept for all meetings and included the key learning points from each session. These notes were made available to all team members and used to update the study protocol and reinforce future educational sessions. The engagement team oriented new SAG members; they arranged education, technical guidance for participating virtually, and general support as needed and regularly monitored their progress.

Partnership

Patient co-investigators were paid $50 per hour (nominal salary of $100 000 per year if they had been employed full-time) for 4 to 8 hours per week. Members of the SAG were also paid $50 per hour, for approximately 8 to 10 hours per year. We provided transportation to and from meeting sites for the SAG. Accommodations through telecommuting options were made for advisors whose health status or work schedules did not allow travel. A regular meeting time was organized around the schedules and multiple time zones of all SAG members. Members who were unable to attend a meeting were updated on agendas, the outcomes of meetings, follow-up activities, and meeting notes and recordings via Basecamp (37signals), a commercial application to support distributed teamwork and asynchronous communication.

Trust, Transparency, Honesty

The research team used a round-table approach to decision-making, with all members invited to speak from their perspectives. All information about the project was shared and made available through Basecamp. To protect confidentiality, members were able to use a pseudonym for public and external documents.

Stakeholder Input

Planning the Study

The patient co-investigators collaborated with the team by reviewing and contributing to the study design and protocol. For example, they defined what “engagement” should look and feel like for patient advisors and for patient participants on site redesign teams. They reviewed all candidate tools for patient-reported outcomes and assisted in the selection of tools for the SAG’s further review. A key function repeatedly played by these research team members was ongoing clarification of language used in descriptions of the study and in the measurement tools. For instance, the patient advisors recommended some changes in the wording and format of some of the instruments, such as replacing “doctor” with “provider” and eliminating “marijuana” as an example of an illegal substance in the substance use survey.

The SAG reviewed our research question, target conditions, definitions of key terms, choice of outcome measures, and eligibility criteria, providing invaluable advice that informed the study planning and proposal. Starting in June 2014 (2 years prior to obtaining funding), they used a semistructured group discussion process to develop an outline of key measures and, from these, the primary and secondary outcomes. They continued to identify key issues around the use of language, the capacity of the intervention to respond to PCP and BHP bias, and the ability of the study to engage consumer groups in conversations about mental health, BH, and substance use disorder issues. After reviewing the AHRQ definition of IBH, they agreed with the core concepts but argued that it should have more emphasis on outcomes and supplied new wording that became the working language for the entire research project.

Conducting the Study

Patients were involved in all aspects of the conduct of the study. For instance, Jennifer O’Rourke-Lavoie, BA, sat on the executive committee and led the SAG. She oriented new members, arranging education and support as needed and monitoring their progress. She organized quarterly meetings of the SAG, which included progress reviews, presentations of IBH topics, and problem-solving on issues related to study conduct. Paula Reynolds, BA, was instrumental in reviewing patient recruitment and consent documentation to ensure that it communicated well with the target audience. Douglas Pomeroy served on the executive committee and on the education team to ensure that the curriculum was patient-centered. He also served on the publications and presentation team to review all proposed dissemination. He took part in study analysis as a member of the evaluation team. All 3 patient partner co-investigators met weekly throughout the study to plan their assistance with retaining study participants and co-investigator participation through newsletters, study blogging, monitoring, and responding to the active sites’ online learning community activity.

The SAG met as a group each quarter and continued with lively discussions via Basecamp and email. They were a sounding board for the investigators in dealing with difficult or ambiguous decisions during the ongoing operations of the study. For instance, a surprising number of study participants ticked “other” in response to questions about their employment status and provided free-text descriptions of their situation. For instance, “I am disabled, but am a minister at church.” Is this patient employed (perhaps part-time) or disabled? The SAG reviewed these responses and helped us come up with rules to classify them.

The active clinical sites were encouraged to include a paid patient representative on their redesign teams with the support of the patient partner guide written by the patient co-investigators.

Disseminating Study Results

Dissemination was a standing agenda item for the investigators and the SAG, particularly during the last 2 years of the project. The dissemination team, led by 2 representatives from the National Committee for Quality Assurance (NCQA), included patients who provided valuable input on dissemination of materials and messaging. Our patient advisors and other stakeholders played key roles in designing materials to return aggregate results to both patient and clinical study participants. Research team and patient partner co-investigators assisted with publications and designing presentations for scientific conferences. Patient advisors were asked to review all materials so that findings were understandable and usable.

METHODS

Study Overview

We have organized the “Methods” and “Results” sections by study aim. We first present the methods that are relevant for all study aims. Then, methods and results unique to each aim are grouped under the aim title.

We conceptualized IBH-PC as a controlled study of an intervention with treatment assigned at the practice level and outcomes assessed at the patient level, resulting in a cluster-randomized study of patients within practices. Power calculations, analytic approaches, and other aspects of study design were designed to especially optimize aim 1 (patient-level outcomes). To enhance the generalizability of the findings, we aimed to make the study as pragmatic as possible.[63] In other words, we allowed the practices and clinicians the same sort of control over the intervention as would reasonably be expected by typical primary care practices not engaged in research. Other than providing access to electronic health record (EHR) data, the practices had no role in patient selection or outcome data collection. We assigned a “cluster leader” (a senior clinician or faculty member in the region) to liaise between a group of individual clusters (sites) and the study investigators. Each cluster leader was responsible for negotiating data use agreements, access to medical records, and local IRB approvals for several individual sites.

The formal study protocol and detailed research plan were submitted to PCORI at the time of the application. It has been submitted to ClinicalTrials.gov, and all modifications have been documented. The UVM Committee on Human Research in the Medical Sciences and the IRBs of all participating institutions approved the protocol.

Study Setting

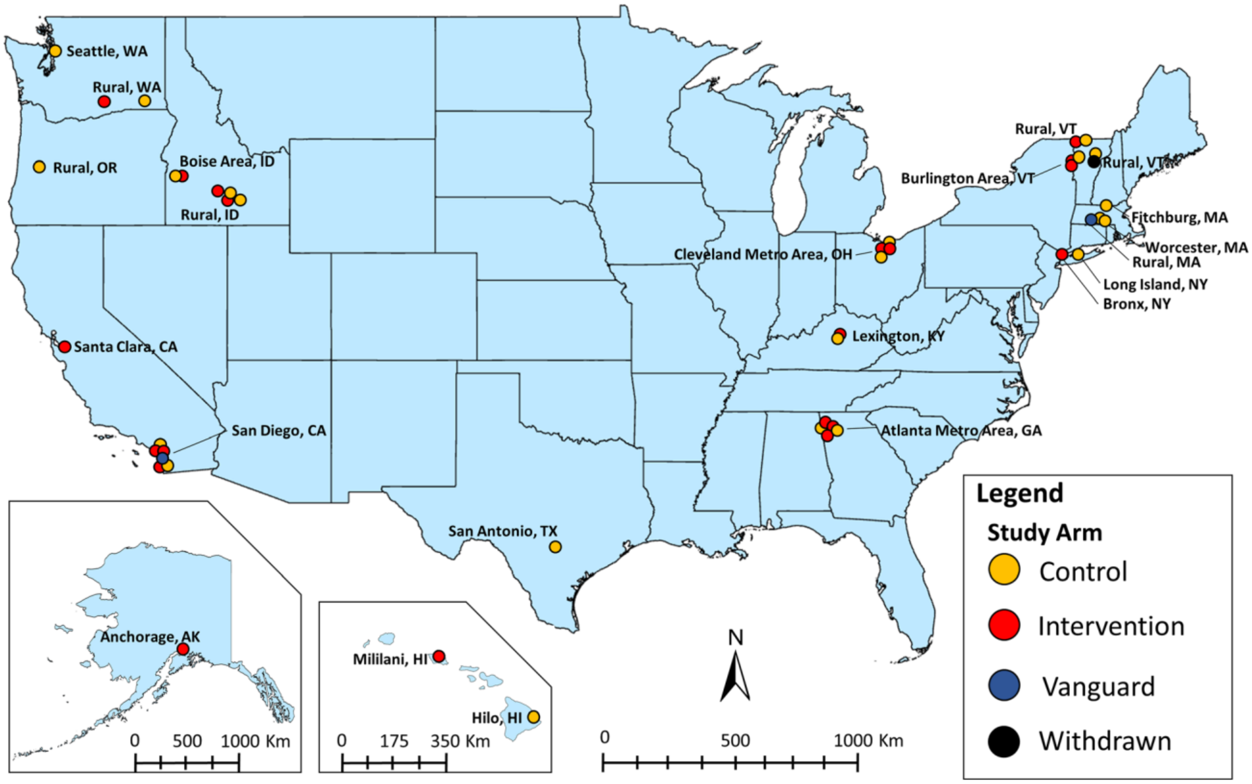
The IBH-PC study enrolled 45 primary care practices across 13 of the United States (Figure 1). Practices represented a diverse distribution of geographic regions, population densities (urban vs rural), patient population sizes, specialties (family vs internal medicine), community health centers, federally qualified health centers, nonprofit and for-profit organizations, resident training sites, and types of ownership (hospital or health system, academic, private). Practices were recruited using professional networks, association listservs, and conference presentations. They were eligible to participate if they

- had at least 1 PCP and at least 1 BHP on-site (co-located);
- had at least 0.5 full-time equivalent BHPs licensed to practice independently;
- were committed to maintaining an on-site BHP for the duration of the study;
- provided the research team with access to EHRs to identify patients with specific medical and BH conditions for recruitment;
- agreed to complete survey instruments periodically throughout the study; and
- agreed to be randomly assigned to either the active or control group.

For practices with more than 10% of their revenue generated by Medicare, at least 1 BHP per practice needed to be eligible to bill Medicare. Practices were ineligible if they were already undergoing, or planned to undergo, another QI initiative similar to the IBH-PC intervention or if they already had an advanced level of BH integration as evidenced by a total Practice Integration Profile (PIP) score above 75. Five practices were excluded for a PIP score above 75. The PIP is a survey of 30 items completed by PCPs and staff about their own practice, used to measure practice-level integration.[64,65]

Study staff randomly assigned the practices to the study groups. Eligible practices were assigned to the active or control group in blocks of 4, using a stratified, randomized approach. Randomization blocks were developed using R (R Foundation for Statistical Computing). Practices were stratified based on degree of BH integration at baseline (total PIP score <50th percentile vs total PIP score ≥50th percentile) and based on geographical area. Treatment allocation was concealed from patients until the end of the trial.

Figure 1. Geographic Distribution of Participating Practices



Participants

Patients were eligible to participate if they

- were at least 18 years old;
- were an active patient of a participating study practice as evidenced by at least 2 visits in a period of 24 months for any purpose, including at least 1 in the most recent 6-month period;
- agreed to complete 3 surveys over 2 years; and
- had both an eligible chronic medical condition and an eligible chronic BH condition, or at least 3 eligible chronic medical conditions. Eligible chronic medical conditions included arthritis; obstructive lung disease, including emphysema, chronic bronchitis, and asthma; nongestational diabetes; and heart disease manifested as heart failure or hypertension. Eligible BH conditions included mood disorder (anxiety or depression); chronic pain (including headache, migraine, neuralgia, fibromyalgia, and chronic

musculoskeletal pain); insomnia; irritable bowel syndrome; and substance, tobacco, or alcohol use disorder.

To determine patient eligibility, medical records were reviewed for 24 months, with the most recent date being within 12 months of each practice's study start date. Within this period, objective evidence that the patient met the minimum age, minimum number of visits, and health conditions criteria was established. These criteria were neither obtained nor reviewed by patient self-report. Eligible patients were selected at random and contacted by mail or phone with an invitation to participate in the study.

Evidence of medical conditions took the form of a specific diagnosis on a patient's personal medical history. Evidence of diabetes may also have been indicated by 3 months of treatment with insulin or other diabetes medications (excepting metformin, which is not specific to diabetes) or by any hemoglobin A_{1C} value greater than 6.4%. Evidence of heart disease may also have been indicated by 3 months of cardiac medications specific to hypertension or heart failure (excepting β-blockers and other medications with broader indications) or by 3 sequential blood pressure measurements with mean systolic pressure greater than 140 mm Hg or mean diastolic pressure greater than 95 mm Hg.

Evidence of BH conditions could take the form of a specific diagnosis, 3 months of certain medications used for behavioral conditions (eg, antidepressants, anxiolytics, opioids, antineuropathic agents, agents for alcohol use disorder or smoking cessation), or persistent inability to attain physiologic control of a medical problem evidenced as systolic blood pressure of greater than 165 mm Hg for 3 months or more or hemoglobin A_{1C} values greater than 9% for 6 months or more.

Interventions and Comparators

The active comparator (the intervention) was access to the IBH-PC toolkit to support a practice-level change process. The control comparator was the colocation of a BHP within or adjacent to the primary care practice (at the same street address), without access to the toolkit.[66]

The intervention, the IBH-PC toolkit, was a set of implementation strategies consisting of 4 components:

- Online education tailored to each clinical and nonclinical role
- Three structured QI workbooks based on Lean methodology, a systematic method of redesigning workflows to improve both care and consumer outcomes[67-69]
- An internet-based learning community
- Remote QI coaches assigned to each clinic, making the toolkit a multifaceted implementation strategy

The toolkit was based on multiple components of the Expert Recommendations for Implementing Change study,[70,71] including education about integrated care, assessment for readiness to implement, technical assistance on implementation, team-based implementation, practice facilitation, patient and family engagement, innovation adaptation, small tests of change, measurement of results, scaling up, and use of a learning collaborative. It was designed to be used by frontline health care workers with little experience in facilitating QI teams and with the support of a remote coach. Portions of the toolkit were iteratively developed in previous studies.[69,72-74] Clinics randomly assigned to the intervention group of the IBH-PC study were given access to the toolkit components through a secure website and could tailor their use based on the needs of the practice and its members. In this highly pragmatic trial, any clinic could end or tailor its use of the toolkit components to best serve its purpose in improving IBH.

The purpose of the IBH-PC toolkit was to guide and support primary care practices through a practice change process as they worked to increase their degree of BH integration through, for example, the improvement of screening, case identification, management, and follow-up. Progress through the toolkit and use of its components were expected to vary among participating practices, and it took practices 9 to 24 months to complete.

The IBH-PC toolkit included an asynchronous, online curriculum about evidence-based concepts of IBH and methods of applying them. The curriculum included an interprofessional course for all practice members and individual courses targeting different practice member roles (practice manager, BHP, PCP, nurse, care manager, staff, IBH-PC facilitator). Courses took approximately 4 to 14 hours to complete depending on the practice member role. The curriculum was made available through Canvas LMS (Instructure, Inc.), an online learning management platform. It was intended, but not required, that all practice members complete the courses appropriate for their roles.

The redesign and implementation process was available as a workbook (PDF format) posted in an online, shared workspace. Practices were encouraged to establish interdisciplinary project teams with an identified champion, a practice member serving as project leader, and an on-site, internal facilitator, among other roles. Team champions' roles in the practice varied and included physicians, BHPs, and administrators. The workbook directed the project team through 4 stages: (1) study start-up and leadership engagement, (2) planning the scope and boundaries of workflow redesign using a Lean management approach,[68,75,76] (3) redesigning workflow with recommended tactics, and (4) implementing those changes in the practice. It was intended, but not required, that the project teams complete all stages sequentially.

To support the project teams through the redesign process, remote QI coaches worked with individual practices. Each study practice was assigned a team of 2 coaches who met with them over the phone or online throughout the practice redesign process (up to 24 months). The remote QI coaches were part of the research team based at UVM. It was intended, but not required, that all project teams would meet with their coaches on a regular basis.

The IBH-PC toolkit included an online forum and learning community created using phpBB forum software, version 2016 (phpBB Limited), an open-source bulletin board system. It provided a platform for practice members and other IBH-PC stakeholders to communicate with one another. Discussion topics varied and were primarily focused on practice members' needs and interests. It was intended, but not required, that all project teams be engaged in the online learning community through regular participation by at least 1 team member.

Study practices randomly assigned to the intervention were encouraged to use the IBH-PC toolkit completely and sequentially as prescribed; however, practices were allowed to use the toolkit out of sequence, modify its sections, or skip them completely. Given the low risk of this practice-level intervention, there were no early stopping rules.

Participating practices were not permitted to participate in other practice change projects similar to the IBH-PC toolkit during the trial. They were, however, allowed to use QI and other strategies to improve any aspect of their practice they saw fit, including improving BH services.

Time Frame for the Study

Practices were enrolled and randomly assigned in a rolling fashion, as their EHRs and patient lists became available. After practices began the intervention, they were afforded up to 24 months of remote coaching (Table 1).

Table 1. Study Timeline[a]

	Baseline	Months 0-1	Months 1-12	Months 21-24	Months 30-33	Months 34-38
Random assignment	X					
Enrollment						
Aim 1 patient recruitment	←X[b]	X				
Aim 2 practice recruitment	←X[b]					
Aim 3 case-study site recruitment					X	
Aim 4 cost-study site recruitment					X	
Interventions						
IBH-PC toolkit—active		X	X	X		
IBH-PC toolkit—control						X→[b]
Patient assessments						
Electronic medical records			X			
Patient-reported outcomes			X	X	X	
COVID-19 burden					X	X
Practice assessments						
Practice integration profile			X	X	X	
Case studies					X	X
Cost assessment					X	X
COVID-19 burden					X	X

Abbreviation: IBH-PC, Integrating Behavioral Health and Primary Care.

[a] Recruitment occurred from April 2016 to March 2018. The IBH-PC toolkit was active from September 2017 to April 2020. Patient-reported outcomes were obtained from March 2018 to December 2020.

[b] Arrows indicate that the activity extended earlier or later.

Changes to the Original Study Protocol

Vanguard Practices

Because of the complexity of both the intervention and the data collection process, we removed 2 practices from the eligibility pool before random assignment and declared them "vanguard" practices. These vanguard practices pilot-tested all aspects of the intervention and its data collection before the randomly assigned sites did, allowing us to adjust study procedures as necessary and to identify a bug in the learning platform used to deliver the online curriculum before the active practices were exposed to it. It also identified confusion over the role of the online community, which was resolved before other practices began.

“Dropped Early” and “Dropped Late” Waves

We had planned to concentrate enrollment of practices into 2 waves (early and late), allowing us to do early preliminary analyses; however, delays in practice start-up related to access to EHR data and local contracting issues disrupted the schedule, forcing us to adopt a continuous enrollment strategy. No preliminary analyses were done.

Dropped Tertiary Outcomes From EHRs

We had originally planned for more extensive data collection from EHRs, including vital signs and laboratory results at multiple time points. Ongoing access difficulties as a result of technical factors and local barriers (eg, strikes, vendor disruptions), however, made this task impossible.

Changes to the Intervention to Increase Pragmatism

In response to feedback from the vanguard practices and the earliest enrolled active sites, we revised the instructions for the intervention to make it clear that the practices were not obligated to take every step in the toolkit and deploy every suggested strategy. Instead, they were advised to select the parts of the intervention that they found most likely to be effective and feasible in their practice setting.

We originally indicated that it would take 6 to 9 months to complete the intervention. In many cases, though, it took 18 months or more. Delays were often a result of local factors such as work stoppages, intervening priorities (new EHR systems), hiring problems, and natural disasters including hurricanes, earthquakes, and even a volcanic eruption.

Aim 1: Patient Surveys

After several months, we modified the recruitment process to include the choice of mailing surveys directly to patients as well as online options. We also moved from obtaining verbal consent from everyone to allowing online and written consent (for mailed surveys). We originally planned to have all surveys that indicated suicidal thinking or another potential emergency to be immediately evaluated by a clinician investigator for prompt referral. The addition of postal surveys made this impractical (the surveys often arrived a week or more after completion), so we modified the surveys to include information on emergency resources such as the National Suicide Prevention Lifeline. In the baseline wave of data collection, we assessed disease-specific outcomes for lung disease, substance use disorder, and alcohol use disorder only in patients with those conditions in their medical histories; however, it became clear that these conditions may have been underdiagnosed in the medical records. In later waves, we requested these items from all respondents.

Aim 3: Contextual Factors That Influenced the Intervention and Outcomes

The COVID-19 pandemic interrupted our plans to make in-person site visits to practices for qualitative data collection, staff interviews, and patient focus groups. We substituted video calls for staff interviews and dropped the patient focus groups in favor of the Patient Centeredness Index (PCI),[77] which was administered with the other survey items to all participants.

Aim 4: Costs of Implementation

This aim was originally proposed as part of aim 2, but it became clear that the degree of detail required, and the special nature of the analysis, warranted a separate aim.

Aim 5: Impact of the COVID-19 Pandemic

This aim was not a part of the original protocol. It was added as part of a COVID-19–related enhancement proposal approved by PCORI in October 2020. It required an additional wave of data collection and new COVID-19-related questions. The additional data collection and survey instruments were approved by the UVM IRB.

Aim 6: The Relationship of Integration to Outcomes, Independent of the Intervention

This analysis, stimulated by the findings of the above analyses, was not included in the original study protocol.

AIM 1: PATIENT-LEVEL IMPACTS

Aim 1 was to test the hypothesis that the IBH-PC toolkit intervention would have an impact on patient-centered functional outcomes.

Study Outcomes

All outcomes were patient-reported. The primary outcomes were the change in mean PROMIS®-29 scores from baseline to follow-up for 8 domains of function: (1) physical function, (2) anxiety, (3) depression, (4) fatigue, (5) sleep disturbance, (6) social functioning, (7) pain intensity, and (8) pain interference.[78] These aspects of functioning and well-being are relevant across many chronic conditions and broadly represent "health-related quality of life." The PROMIS-29 is recommended by a National Institute on Aging working group because of its content validity, reliability, sensitivity to differences in health states, and responsiveness to change.[79] It was selected by our patient advisors, who felt it represented a broad array of features important to patients with behavioral or medical conditions. Although designed as general quality-of-life measures, the PROMIS-29 domain scores have been shown to reliably discriminate patients with various chronic conditions from patients without chronic conditions.[80] For instance, the physical function score is significantly lower in patients with hypertension, depression, anxiety, migraine, asthma, insomnia, diabetes, and chronic obstructive pulmonary disease than in patients without those conditions. The PROMIS instruments have been used in patients with heart failure,[81,82] chronic obstructive pulmonary disease,[83,84] substance use disorder,[85] depression, and anxiety,[86,87] including depression in older patients with cognitive dysfunction.[88] Although minimally important differences (MIDs) were not fully specified, the MIDs for other instruments in the PROMIS series are 2 to 8 points.[89,90]

The secondary outcomes included changes in other mean patient-reported measures of health and quality of care from baseline to follow-up, including the PROMIS-29 physical and mental health summary scores. The patient advisors identified domains of interest and reviewed potential outcome instruments. They were asked to select those that matter to patients and their families in making treatment and care decisions. They identified adherence to medication (measured by the Morisky Green Levine Medication Adherence Scale),[91] engagement in self-care, clinician communication and empathy (the consultation and relational empathy [CARE] measure),[92] use of care (Utilization Patient Report),[93] disability (by restricted-activity days),[94] and physical fitness (the Duke Activity Status Index [DASI])[95] as critical domains.

Markers of disease control included the PROMIS-29 subscales as well as specific measures for depression (the Patient Health Questionnaire-9 [PHQ-9]), asthma (the Asthma Symptom Utility Index [ASUI]),[96] substance use disorder (the Global Assessment of Individual Needs—Short Screener [GAIN-SS]),[97] and problem drinking (the Self-Report Habit Index).[98]

The CARE survey is a 10-item validated self-report measure that assesses patients' perception of clinician empathy.[92] The CARE survey has been shown to have excellent reliability (Cronbach $\alpha = .92$).

The modified, self-reported, medication-taking scale[91] is the Morisky Green Levine Medication Adherence Scale, a 4-item, self-report measure, which assesses overall adherence to the prescription medications a patient may use. The Morisky Green Levine Medication Adherence Scale has been shown to have adequate reliability (Cronbach $\alpha = .61$).[99]

The Utilization Patient Report[93] is a 3-item, self-report measure asking patients to recall their health care use in the past year. Specifically, patients are asked to report visits to the emergency department, overnight stays in the hospital, and outpatient appointments to a health care professional.

Time loss as a result of disability was measured using the restricted-activity days survey,[94] which asks patients to report the restriction of their daily lives as a result of illness and disability.

Physical fitness was measured using the DASI,[95] a 12-item, validated, self-report measurement used to assess functional capacity. The DASI has been shown to have good to excellent reliability in several studies (Cronbach $\alpha = .80$ to .93)[100] and correlates with peak oxygen uptake, as measured by an exercise stress test.

We also included a measure of the degree to which patients perceive their primary care practice as patient-centered. This required the development of a new instrument, the PCI.[77] The PCI contains 14 items about the patient's primary care team such as, "They pay attention to my needs," "They listen to me," and, "We agree on a plan before taking action." Analyses of the current sample indicated the scale is highly internally consistent (Cronbach $\alpha = .96$) and is strongly correlated with a measure of empathy ($r = 0.65$), suggesting convergent validity.

The 7-item Generalized Anxiety Disorder scale (GAD-7)[101] is a self-reported questionnaire developed and validated in a larger primary care patient sample to assess anxiety symptom severity. The GAD-7 has been shown to have excellent reliability in several studies (Cronbach $\alpha = .79$ to .91).[102]

The PHQ-9 is a validated self-report measure of depression symptom severity.[103] The PHQ-9 has good reliability (Cronbach $\alpha = .86$ to .89). Recognizing the sensitive nature of the questions on the PHQ-9 concerning suicidal thoughts and ideation, information on how to access the national suicide helpline was distributed with all paper study materials and was programmed into the electronic survey data capture system.

The ASUI[96,104] is a 10-item, validated, self-report measure assessing control and quality of life as it relates to symptoms of asthma. Reliability for the ASUI is considered good, with a Cronbach $\alpha = .74$.[96]

Alcohol use was assessed using 2 items from the Self-Report Habit Index—Alcohol, which has excellent reliability (Cronbach $\alpha = .94$).[98]

Substance use disorder was assessed using the 5-item Substance Disorders subscale of GAIN-SS,[97] a biopsychosocial screener for individuals presenting with substance use and mental health concerns. The GAIN-SS Substance Use subscale has been shown to have acceptable reliability (Cronbach $\alpha = .77$ to .84).[105]

Covariates

Patients reported their sex (female, male, prefer to self-describe [please describe], prefer not to say); race (White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, other [please describe], prefer not to say); whether they were of Hispanic ethnicity (yes, no, prefer not to say); marital status (never married, married, living as married [common law, civil union], separated, divorced, widowed); level of education (less than 9th grade, 9th to 12th grade without diploma, high school graduate [including GED], associate's degree, bachelor's degree, graduate or professional degree); and total household annual income (less than \$15 000, \$15 000 to \$29 999, \$30 000 to \$44 999, \$45 000 to \$59 999, \$60 000 to \$74 999, \$75 000 to \$99 999, \$100 000 or more).

Because of small sample sizes in some subgroups, for analytic purposes we dichotomized sex as male or female excluding the small number of others, race as White or non-White, marital status as married or living as married vs all others, education level as associate degree or higher vs all others, and income as less than \$30 000 vs higher.

Neighborhood characteristics were calculated for each participant's census tract, including population density in persons per square mile, development status (rural vs urban per the US 2010 Census' rural-urban commuting area codes),[106] and the Social Deprivation Index (a composite of income, education, employment, home ownership, crowded housing, single-parent households, and car ownership variables from the American Community Survey).[107]

Data Collection and Sources

Eligible patients were mailed a letter describing the purpose of the study. If they chose to participate, they were asked to complete a brief set of surveys at 3 time points online, in a mailed paper document, or over the phone. The letter also described the process to opt out of the study and provided a phone number to call for more information on the study. A third-party agency, the DARTNet Institute, followed up with phone calls to eligible patients. Consent was administered and documented by phone, on paper, or online at the time of the baseline set of surveys. Participating patients provided consent to link their EHRs to survey data and could discontinue participation at any time. All study materials were available in both Spanish and English. Patients were blind to their practice's random assignment in the study.

To encourage participation, participants were paid $30 for each survey and sent multiple reminders (via email or first-class mail as they preferred). If they did not respond, we attempted up to 10 phone calls per survey. Baseline surveys were collected between March 2018 and June 2019. Midpoint surveys were collected between June 2019 and March 2020. Final surveys were collected between March 2020 and December 2020. The midpoint surveys were originally intended for an interim analysis that was later determined to be infeasible; however, the surveys also served to maintain contact with participants and are available for secondary analyses.

Sample Size Calculations and Power

To achieve a nominal significance level of $P < .05$ for the hypothesis that any 1 or more of the 8 primary outcomes would achieve significance, we applied the Bonferroni correction,[108] requiring $P < .05/8 = .00625$. All PROMIS-29 domain scales were normalized to a standard deviation of 10 points.[78] We assumed that the within-practice correlations of the PROMIS-29 scales would be 0.03 (similar to the Short Form Health Survey).[109] Although MIDs were not fully specified, the MIDs for other instruments in the PROMIS series range from 2 to 8 points.[89,90] Power calculations conducted prior to study recruitment assumed 75 participants from 40 practices, giving the study 3000 participants and 90% power to detect a difference in any of the 8 PROMIS-29 scales of 2.5 points or more. Given the final sample of 2426 participants in 41 clusters, the study had 92% power to detect differences in the primary outcomes as small as 2.8 points.

Analytical and Statistical Approaches

Analyses for all aims were conducted using the intention-to-treat principle. The outcomes for aim 1 were measured at the patient level, but the intervention was delivered at the practice level. We assumed that the impacts of the various independent variables would be consistent across practices but that each practice may have a different patient function at baseline. The slopes of the independent variables were therefore modeled as constants, but the intercept varied by practice. We built a multivariable, mixed linear regression model for each outcome. The dependent variable was change in the outcome from baseline to follow-up. Group assignment (active vs control) was the main independent variable, modeled as a fixed effect, along with any potentially confounding variables. Practice was included as a random intercept. Because there were a large number of participant and practice characteristics available, we limited potential confounders to characteristics that were associated with both the outcome and the main predictor, with $P < .10$, and that influenced the relationship between the predictor and outcome. We compared the coefficient on the predictor from a mixed model without fixed effect covariates with the coefficient from a model with a single potential confounder. If the 2 coefficients differed by more than 10%, the potential confounder was included in the full multivariable analysis. All analyses were performed in Stata, version 17.0, software (StataCorp LLC).

To assess the degree of treatment heterogeneity across subgroups of patients commonly seen in primary care settings, we performed exploratory subgroup analyses using the same analytic approach described above. In each subgroup, the null hypothesis tested was that the intervention was equally effective as the control in inducing change in patient outcomes over 2 years.

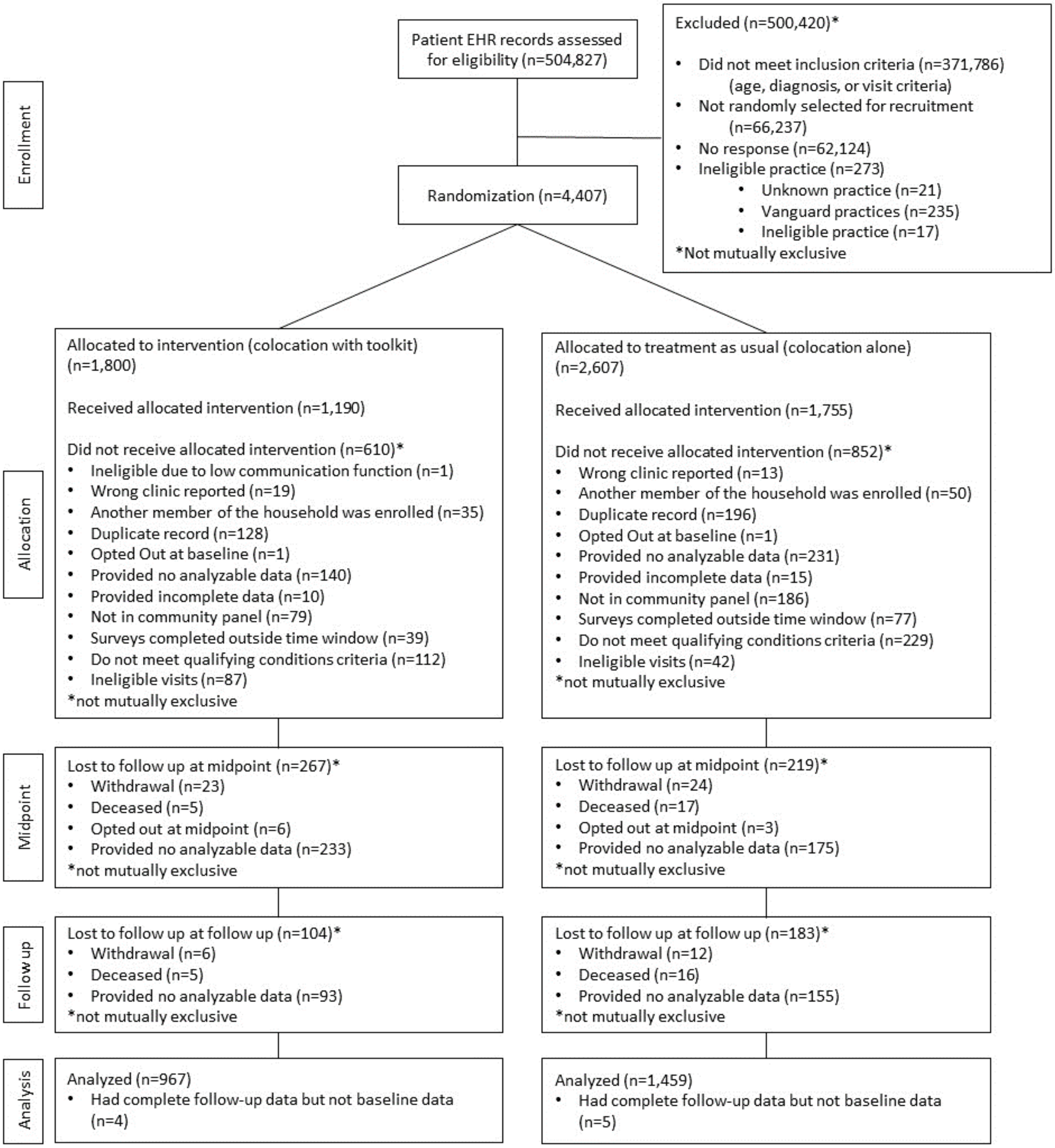
To assess the potential impact of missing data as a result of eligible participants who completed baseline surveys but did not complete final surveys for any reason (lost to follow-up), we compared the baseline characteristics of the 2 groups. We also conducted sensitivity analyses on the PROMIS-29 physical and mental health summaries by comparing the unadjusted effects of the intervention (in simplified analyses without covariates or random effects). A sensitivity analysis was also performed among participants lost to follow-up by assuming an extreme difference in outcomes between participants lost to follow-up in the active and control groups.[110] A difference of 3 times was used because, in our clinical judgment, it is very unlikely that the difference in outcomes between participants in the active and control groups that were lost to follow-up would be greater than this.

Results

Participant Flow

One practice dropped out before randomization, 2 became vanguard sites, and a computer error resulted in no eligible patient surveys from 1 active practice. A total of 2945 eligible participants at 41 participating practices (20 in the active group and 21 in the control group) completed a baseline assessment of the primary outcome measure (PROMIS-29) and were randomly assigned. Of these participants, 2426 completed their final outcome assessments and were eligible for analysis (Figure 2).

Figure 2. CONSORT Participant Flow Diagram



Abbreviation: EHR, electronic health record.

Participant Characteristics

The 2426 participants in the final sample (Table 2) were largely representative of US primary care patients. When compared with percentages in the Centers for Disease Control and Prevention's National Ambulatory Medical Care Survey (NAMCS),[111] a source of objective and reliable information about the provision and use of ambulatory medical care services in the United States, the participants in our study were 66% female (compared with 67% in NAMCS), 77% White (78%), 12% Black (14%), and 7% Latino (12%). The rate of participants residing in rural areas was 20% vs 16% in NAMCS. The average age of our participants (62 years) was higher than the average age of 47 years in NAMCS, which was likely because we recruited participants with multiple medical and BH comorbidities, many of which are age-related. As anticipated, the prevalence of chronic disease was much higher in the study sample than in NAMCS: depression, 47% vs 13%; arthritis, 42% vs 11%; diabetes, 44% vs 15%; hypertension, 83% vs 31%.

Table 2. Characteristics of Final Analytic Patient Sample

	Overall sample (N = 2426)	Intervention group (toolkit + colocation) (n = 967)	Control group (colocation alone) (n = 1459)
Age, mean (SD), y	61.9 (13.1)	61.9 (12.7)	61.9 (13.4)
Sex, No. (%)			
Female	1588 (65.5)	618 (63.9)	970 (66.5)
Male	832 (34.3)	349 (36.1)	483 (33.1)
Prefer to self-describe or not to say	6 (0.2)	0 (0.0)	6 (0.4)
Race, No. (%)			
American Indian or Alaska Native	21 (0.9)	8 (0.8)	13 (0.9)
Asian	71 (2.9)	36 (3.7)	35 (2.4)
Black or African American	281 (11.6)	127 (13.2)	254 (10.6)
Native Hawaiian or other Pacific Islander	37 (1.5)	20 (2.1)	17 (1.1)
White	1859 (76.6)	702 (72.9)	1157 (79.3)
Other or prefer not to say	157 (6.5)	70 (7.3)	83 (5.7)
Ethnicity, No. (%)			
Hispanic	180 (7.4)	84 (8.7)	96 (6.6)
Non-Hispanic	2207 (91.0)	865 (89.5)	1342 (92.0)
Prefer not to say	39 (1.6)	18 (1.8)	21 (1.4)
Marital status, No. (%)			
Never married	395 (16.3)	156 (16.1)	239 (16.4)
Married	1063 (43.8)	450 (46.5)	613 (42.0)
Living as married	63 (2.6)	20 (2.1)	43 (3.0)
Separated	56 (2.3)	16 (1.7)	40 (2.7)
Divorced	523 (21.6)	201 (20.8)	322 (22.1)
Widowed	313 (12.9)	116 (12.0)	197 (13.5)
Prefer not to say	13 (0.5)	8 (0.8)	5 (0.3)
Employment status, No. (%)			
Full time	430 (17.7)	183 (18.9)	247 (16.9)
Part time	161 (6.6)	66 (6.8)	95 (6.5)

	Overall sample (N = 2426)	**Intervention group (toolkit + colocation) (n = 967)**	**Control group (colocation alone) (n = 1459)**
Retired	1009 (41.6)	391 (40.4)	618 (42.4)
Disabled	615 (25.4)	251 (26.0)	364 (25.0)
Homemaker	84 (3.5)	33 (3.4)	51 (3.5)
Student	11 (0.5)	4 (0.4)	7 (0.5)
Unemployed or seeking employment	77 (3.2)	26 (2.7)	51 (3.5)
Other or prefer not to say	39 (1.6)	13 (1.4)	26 (1.7)
Annual household income, No. (%)			
<$15 000	674 (27.8)	269 (27.8)	405 (27.8)
$15 000-$29 999	505 (20.8)	182 (18.8)	323 (22.1)
$30 000-$44 999	294 (12.1)	114 (11.8)	180 (12.3)
$45 000-$59 999	222 (9.2)	84 (8.7)	138 (9.5)
$60 000-$74 999	176 (7.3)	68 (7.0)	108 (7.4)
$75 000-$99 999	195 (8.0)	86 (8.9)	109 (7.5)
≥$100 000	282 (11.6)	136 (14.1)	146 (10.0)
Prefer not to say	78 (3.2)	28 (2.9)	50 (3.4)
Education level, No. (%)			
Less than 9th grade	56 (2.3)	24 (2.5)	32 (2.2)
9th to 12th grade, no diploma	208 (8.6)	85 (8.8)	123 (8.4)
High school graduate (including GED)	1004 (41.4)	396 (41.0)	608 (41.7)
Associate degree	378 (15.6)	156 (16.1)	222 (15.2)
Bachelor’s degree	387 (16.0)	158 (16.3)	229 (15.7)
Graduate or professional degree	344 (14.2)	130 (13.4)	214 (14.7)
Prefer not to say	49 (2.0)	18 (1.9)	31 (2.1)
Chronic conditions, No. (%)			
Arthritis	1023 (42.2)	410 (42.4)	613 (42.0)
Asthma	545 (22.5)	229 (23.7)	316 (21.7)
Chronic obstructive pulmonary disease	330 (13.6)	116 (12.0)	214 (14.7)
Chronic pain	2037 (84.0)	824 (85.2)	1213 (83.1)

	Overall sample (N = 2426)	**Intervention group (toolkit + colocation) (n = 967)**	**Control group (colocation alone) (n = 1459)**
Nongestational diabetes	1075 (44.3)	404 (41.8)	671 (46.0)
Heart failure	188 (7.8)	69 (7.1)	119 (8.2)
Hypertension	2012 (82.9)	805 (83.3)	1207 (82.7)
Irritable bowel syndrome	102 (4.2)	46 (4.8)	56 (3.8)
Anxiety	830 (34.2)	339 (35.1)	491 (33.7)
Depression	1136 (46.8)	439 (45.4)	697 (47.8)
Insomnia	570 (23.5)	231 (23.9)	339 (23.2)
Substance use disorder	559 (23.0)	220 (22.8)	339 (23.2)
Tobacco use	445 (18.3)	171 (17.7)	274 (18.8)
Alcohol use disorder	155 (6.4)	62 (6.4)	93 (6.4)
Chronic conditions, mean (SD), No.	4.4 (1.6)	4.3 (1.6)	4.4 (1.6)
Neighborhood characteristics (home census tract)			
Social Deprivation Index, mean (SD), points[a]	52.6 (27.7)	51.2 (29.0)	53.5 (26.9)
Rural residence, No. (%)	477 (19.7)	136 (14.1)	341 (23.4)
Population density, mean (SD), persons per square mile	3768 (6656)	4980 (6656)	2978 (3314)

[a]The Social Deprivation Index runs from 0 (least deprivation) to 100 (most deprivation).

Primary and Secondary Outcomes

Of the potential covariates tested, only population density of the practice's county, nonprofit status of the practice, the median age of the practice's county, patient practice visits per year, and whether the practice was privately owned altered the association between the intervention and any of the patient outcomes by more than 10% and were included in the final models (Table 3). In the adjusted analyses, the intervention had no significant effects on any of the primary or secondary outcomes (Table 3 and Figure 3).

Table 3. Adjusted Effect of the Intervention on Patient Outcomes

	Sample, No.[a]	Effect[b]	95% CI	*P* value
Primary outcomes				
Anxiety[c]	2426	0.08	−0.53 to 0.69	.79
Depression[c]	2426	0.21	−0.37 to 0.79	.48
Fatigue[c]	2426	0.07	−0.54 to 0.68	.83
Sleep disturbance[c]	2426	−0.05	−0.58 to 0.49	.86
Pain interference[d]	2426	0.19	−0.40 to 0.79	.52
Pain intensity[c]	2426	0.1	−0.08 to 0.28	.27
Social participation[c,e]	2426	−0.06	−0.71 to 0.59	.86
Physical function[e]	2426	0.1	−0.39 to 0.60	.69
Secondary outcomes				
Physical health summary[e]	2426	0.09	−0.39 to 0.56	.72
Mental health summary[c,e]	2426	−0.06	−0.51 to 0.39	.79
CARE total score[c,e]	2383	−0.03	−0.11 to 0.05	.49
MGLS total score[e]	2310	−0.03	−0.11 to 0.05	.46
ED visits in past year[d,f]	2369	−0.01	−0.14 to 0.11	.82
Health care visits in past month[c,d,g,h]	2363	−0.02	−0.29 to 0.26	.90
Hospital days in past year[c]	2363	−0.04	−0.29 to 0.21	.78
Use category[h]	2376	0.07	−0.13 to 0.27	.48
Restricted activity days	2301	0.16	−0.02 to 0.35	.08
Metabolic equivalents (DASI)[c,d,e,g]	2127	0.03	−0.09 to 0.15	.64
PHQ-9 total score[c,f]	2290	0.2	−0.13 to 0.54	.24
PHQ-9 category[c,f]	2290	0.14	−0.04 to 0.32	.12
GAD-7 total score[c]	2317	0.03	−0.27 to 0.32	.86
GAD-7 category[c]	2317	−0.06	−0.25 to 0.13	.53
Asthma Symptom Utility Index[c]	440	−0.02	−0.05 to 0.01	.17
GAIN lifetime SUD screener[d]	376	−0.06	−0.35 to 0.24	.71
GAIN 1-year SUD screener[c,g]	376	−0.15	−0.37 to 0.06	.16
GAIN 3-month SUD screener[c,d,g]	376	0.0	−0.19 to 0.18	.96
GAIN 1-month SUD screener[c,g]	376	−0.04	−0.22 to 0.14	.65
Alcohol use category[c,d,f,g,h]	1747	−0.03	−0.36 to 0.29	.84
Patient Centeredness Index[c,e,f]	2261	−0.8	−2.3 to 0.7	.29

Abbreviations: CARE, Consultation and Relational Empathy; DASI, Duke Activity Status Index; ED, emergency department; GAD-7, General Anxiety Disorder-7; GAIN, Global Appraisal of Individual Needs; MGLS, Morisky Green Levine Medication Adherence Scale; PHQ-9, Patient Health Questionnaire-9; SUD, substance use disorder.

[a] Sample size varies with the availability of covariates.

[b] Effect was calculated by the difference in difference scores and was adjusted for the baseline value of the outcome, for clustering within practice, and by variables where indicated. Values greater than 0 indicate that the intervention was associated with an increase in the outcome measured.

[c] This effect was adjusted based on the variable of population density in the practice's county.

[d] This effect was adjusted based on the nonprofit status of the practice.

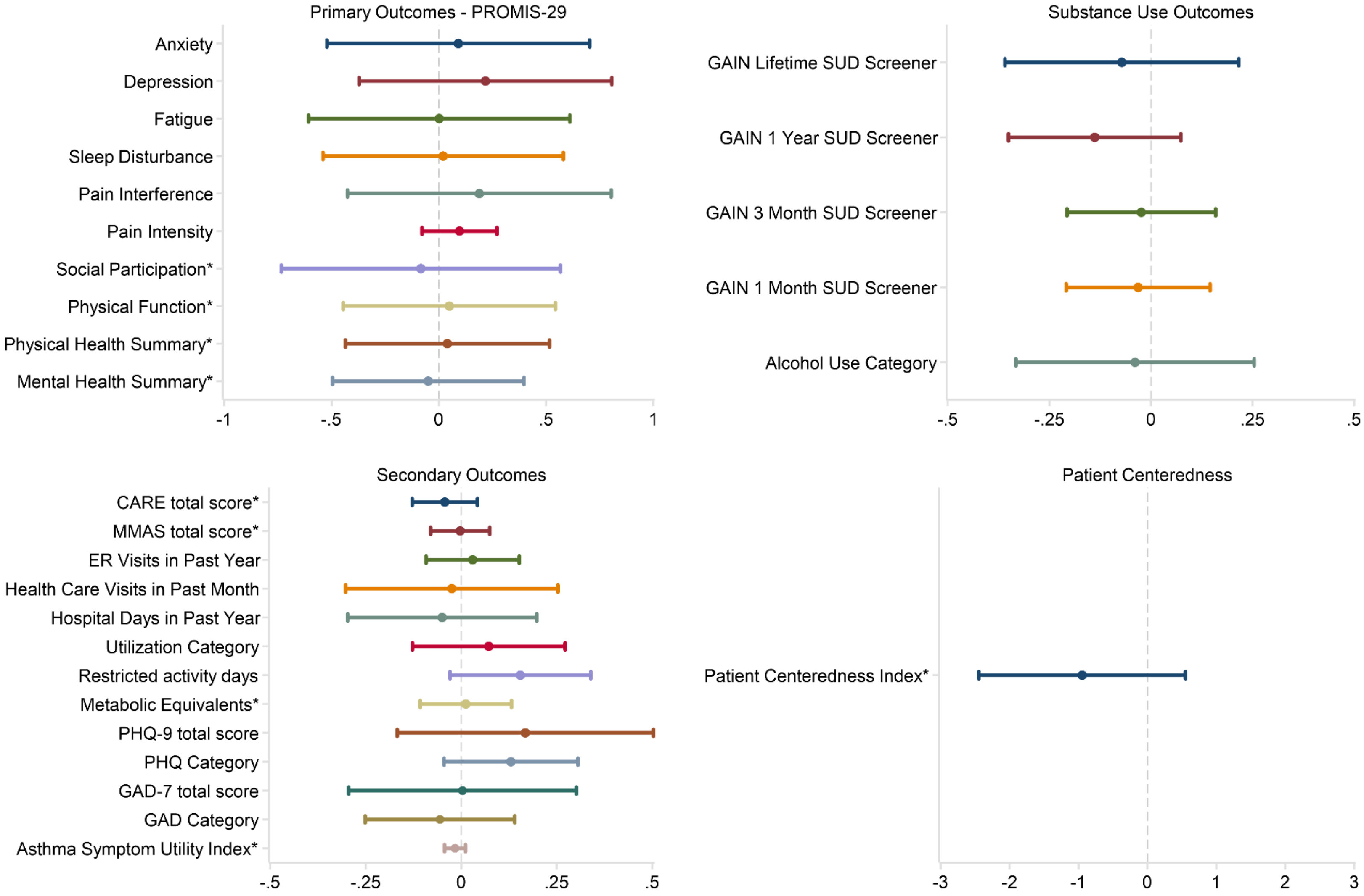
[e] Higher scores on these outcome measures reflect higher functioning. On other measures, higher scores reflect greater impairment.

[f] This effect was adjusted based on the variable of the median age in the practice's county.

[g] This effect was adjusted based on the variable of patient practice visits per year.

[h] This effect was adjusted based on the variable of whether the practice was privately owned.

Figure 3. Effect of the Intervention on Patient Outcomes



Abbreviations: CARE, Consultation and Relational Empathy; ER, emergency room; GAD-7, General Anxiety Disorder-7; GAIN, Global Appraisal of Individual Needs; MMAS, Morisky Green Levine Medication Adherence Scale; PHQ-9, Patient Health Questionnaire-9; SUD, substance use disorder.

The center point of each bar is the mean adjusted effect (the regression coefficient) of the intervention on a specific outcome, with the bars representing the 95% CIs. Values greater than 0 indicate that the intervention was associated with an increase in the outcome measured. All bars include 0, indicating that none were statistically significant (*P* < .05).
*For this variable, a higher score represents better function.

Subgroup Results

We repeated the analyses in subgroups based on

- participant characteristics (age, sex, race, ethnicity, marital status, income level, education level, and employment status);
- the presence of each of the qualifying medical and BH conditions;
- the number of qualifying conditions;
- characteristics of the practice's neighborhood (rurality, Social Deprivation Index score, and population density of their home census tract); and
- practice characteristics (community health center, hospital, private, academic, or nonprofit; specialty; encounters per year; patient panel size; percentage of adult patients on Medicare; BHP hours per week; PCP hours per week; BHP tenure; whether it was a resident training site; total PIP score at baseline; and county median age, sex, race, ethnicity, income level, education level, employment status, rurality, region, population density, and Social Deprivation Index score).

In general, we used census tract data for individual-level covariates and county-level data for practice-level covariates. The spatial scale (ie, census tract vs county-level data) of various covariates was also determined by the availability of data. In all, we examined 29 primary and secondary outcomes in 55 subgroups for 1595 models. Of these, 40 (2.5% [95% CI, 1.8%-3.4%]) demonstrated a statistically significant effect with $P < .05$. On the basis of multiple comparisons alone, we would expect about 1595 × .05 = 80 to achieve significance. There were no patterns among these findings in terms of subgroups or outcomes consistently appearing.

Sensitivity to Missing Data

The participants who completed the follow-up survey were significantly different in several key characteristics than the 18% of participants who were lost to follow-up (Table 4).

Table 4. Participant Characteristics at Follow-Up

	Completed follow-up survey (n = 2426)	Lost to follow-up (n = 519)	*P* value[a]
Age, y	61.9	61.2	.27
Male, %	34.4	42.9	<.001
Race, %			.001
American Indian or Alaska Native	0.9	1.7	
Asian	2.9	4.4	
Black or African American	11.6	14.5	
Native Hawaiian or other Pacific Islander	1.5	1.0	
White	76.6	68.9	
Other or prefer not to say	6.3	9.5	
Hispanic ethnicity, %	7.5	11.6	.005
Married or living as married, %	47.3	42.2	.035
Working (employed, homemaker, or student), %	33.9	27.1	.004
Annual household income <$30 000, %	51.5	63.2	<.001
Any college education, %	46.7	38.8	.001
Chronic conditions, mean, No.	4.4	4.8	<.001
Physical health summary score, No.	45.9	43.7	<.001
Mental health summary score, No.	50.3	49.2	.006

[a] *P* values calculated by χ^2 or by Student *t* test.

We modeled the potential effects of the 2 groups ("completed follow-up survey" and "lost to follow-up") responding differently to the intervention. For the PROMIS-29 physical health summary, the simplified analysis of just the participants who completed follow-up was very similar to the complete model presented above: a small, clinically and statistically insignificant difference. Assuming the physical health of the control group did not worsen and the physical health of the members of the active group lost to follow-up worsened 3 times more than observed, the difference between the 2 groups remained small and statistically insignificant (Table 5). When we repeated the analysis for mental health, which improved slightly over time in both groups, we assumed that the active participants lost to follow-up would have improved 3 times more than observed. The 2 groups nonetheless remained similar.

Table 5. Sensitivity Analyses on Potential Effect of Loss to Follow-Up

Model assumptions	Control group			Active group			*P* value[a]
	No.	Mean	SD	No.	Mean	SD	
Physical health summary							
Follow-up participants only	1459	−0.54	6.07	967	−0.48	6.19	.81
Patients in the control group lost to follow-up don't worsen; patients in the active group lost to follow-up worsen 3 times more than followed active group	1755	−0.45	5.53	1190	−0.66	5.60	.32
Mental health summary							
Follow-up participants only	1459	0.40	5.67	967	0.47	5.82	.79
Patients in the control group lost to follow-up don't improve; patients in the active group lost to follow-up improve 3 times more than followed active group	1755	0.34	5.17	1190	0.64	5.26	.12

[a] *P* values calculated by Student *t* test.

AIM 2: PRACTICE-LEVEL OUTCOMES

Overview

Aim 2 was to test the hypothesis that the change in practice integration over time would differ in the active and control groups.

Study Outcomes

The primary outcome was change in total PIP scores[64,112] from baseline to follow-up. The PIP is a survey of 30 items completed by PCPs and staff about their own practice. It is organized into 6 domains. Practice Workflow includes the policies and procedures that ensure the organizational structure necessary to support consistent delivery of evidence-based services to patients in need. Workspace Arrangement and Infrastructure addresses the physical proximity and use of shared medical records. Integration Methods covers the type and degree of interactions among PCPs and BHPs. Case Identification specifies the practice's procedures for screening and identifying patients who need BH services. Clinical Services covers the BH services available at the practice. Patient Engagement captures the ability of the practice to initiate treatment, involve the patient in developing and delivering the care, and provide support to the patient through ongoing management and follow-up. The domains contain between 2 and 9 questions each and are scored as the average of their item scores. All scores can run from 0 (least degree of integration) to 100 (greatest degree of integration). The total integration score is the unweighted numeric average of the 6 domains and thus ranges from 0 to 100. The PIP was developed with a sound theoretical model.[47] It has face validity, correlates with expert ratings of a series of test scenarios, discriminates among practices with different levels of IBH performance, and has stable test-retest reliability and a stable factor structure.[64,112,113]

Covariates

The predictor was group assignment. Covariates evaluated as potential confounders included practice characteristics (community health center, hospital, private practice, academic, or nonprofit; specialty; encounters per year; patient panel size; percentage of adult patients with Medicare insurance; BHP work hours per week; PCP work hours per week; BHP tenure; and training site) and service area characteristics (county median age, sex, race, ethnicity, income level, education level, employment status, rurality, region, population density, and Social Deprivation Index score).

Sample Size Calculations and Power

Because the unit of analysis was the practice, the sample size was fixed by the number of practices required for aim 1.

Data Collection and Sources

We collected PIP scores from 4 to 5 staff members or clinicians at each practice at each time point. We instructed practices to solicit PIP surveys from 1 PCP, 1 BHP, 1 administrator, and 1 or 2 other practice members of the practice’s choosing. The professional role of each respondent was recorded. For follow-up PIP administration, we asked for the same people to complete the survey, but this was not always possible because of staff turnover. The total and domain PIP scores of the respondents at each practice were then averaged to calculate each practice’s scores. County characteristics were gathered from the US 2010 Census. Practice characteristics were reported by a practice manager at each site.

Analytical and Statistical Approaches

We built multivariable models of change in PIP as a function group assignment, controlling for baseline level of PIP and any identified covariates. Covariates were selected from potential confounders if they changed the strength of association between group and PIP by at least 10%, using the procedure described for aim 1 above. No random effects were modeled, and no correction for multiple comparisons was applied.

Results

Participant (Practice) Flow

We reviewed 129 practices for eligibility. Eighty-four were excluded (4 for a PIP score >75, 21 for other reasons, and 59 withdrew before random assignment). Two practices were selected as vanguard sites to test delivery of the intervention and data collection methods. The remaining 43 sites were randomly assigned, but 1 active site withdrew immediately after random assignment before any data were collected, leaving 22 control practices and 20 active practices for this aim.

Practice Characteristics

The active and control practices were well balanced in terms of their size, specialty, ownership, service area demographics, and other characteristics (Table 6).

Table 6. Practice Characteristics at Baseline

	Control (n = 22)	Active (n = 20)	*P* value[a]
No. of patient visits per year, mean	30 104	23 668	.51
No. of PCPs, mean	9.8	9.8	.62
PCP FTEs, mean	6.1	5.9	.96
Practice panel size, mean, No.	9419	9138	.94
Adult patients with Medicare insurance, %	21.5	21.3	.81
Tenure of on-site BH services, y	6.6	5.9	.86
Practice Integration Profile, median, total score	59.4	59.4	.96
Specialty, No.			.92
Internal medicine	4	3	
Family medicine	11	9	
Mixed	7	8	
Community health center, No.	7	8	.75
Hospital-owned, No.	10	10	>.99
Private ownership, No.	3	1	.61
Nonprofit, No.	18	19	.35
Academic, No.	9	10	.76
Resident training site, No.	7	9	.53
County characteristics			
Median age, y	38.2	36.1	.13
Male sex, %	48.8	48.8	.98
Racial distribution, %			
American Indian or Alaska Native or Pacific Islander	0.7	1.0	.80
Asian	4.5	7.8	.22
Black or African American	7.6	6.8	.33
White	77.4	67.9	.11
Other race	5.1	6.5	.45
Hispanic ethnicity, %	14.6	16.0	.65
Median annual income, $	62 390	65 993	.51
High school graduation rate, %	89.3	89.1	.91
Social Deprivation Index score	43.5	46.4	.76
Population per square mile	775	2540	.35

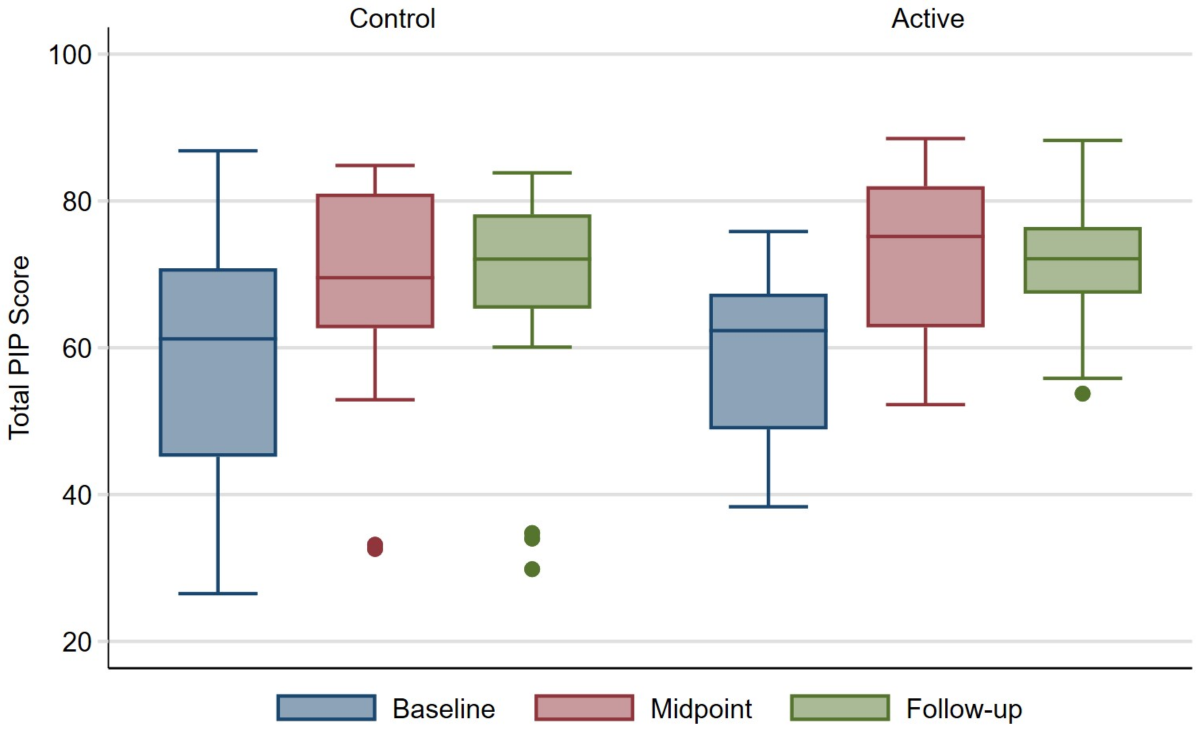
Abbreviations: BH, behavioral health; FTE, full-time equivalent; PCP, primary care professional.

[a] *P* values calculated by Wilcoxon rank sum test or Fisher exact test.

Primary Outcomes

Over the course of the study, total PIP scores improved overall in both study groups, especially between the baseline and midpoint surveys (before the appearance of the COVID-19 pandemic). There was, however, no statistically significant effect of the intervention on total PIP scores (an increase of 3.8 points [95% CI, −3.6 to 11.0]; *P* = .30) (Figure 4).

Figure 4. Total PIP Scores Over Time



Abbreviation: PIP, Practice Integration Profile.
Each box-and-whisker plot represents a distribution of total PIP scores. The middle line of each box represents the median value with the box including the IQR observations. The whiskers represent the range with outliers noted by individual symbols. Midpoint and follow-up occurred approximately 1 and 2 years after baseline.

After adjustment for potential confounders, all 6 PIP domain scores improved in the active group compared with the control group. The change in the Workflow domain was statistically significant (an increase of 9.3 points [95% CI, 1.7-16.9]; *P* = .02) (Table 7 and Figure 5).

Table 7. Adjusted Effect of the Intervention on Practice Integration

PIP domain[a]	Effect[b]	95% CI	*P* value
Total	+3.8	−3.6 to 11.2	.30
Workflow	+9.3	1.7-16.9	.02
Clinical services	+2.6	−6.3 to 11.5	.56
Workspace[c]	+3.5	−7.0 to 14.1	.50
Integration	+3.6	−7.9 to 15.1	.53
Patient identification[d]	+3.1	−3.6 to 9.8	.36
Patient engagement	+3.2	−5.3 to 11.6	.45

Abbreviation: PIP, Practice Integration Profile.

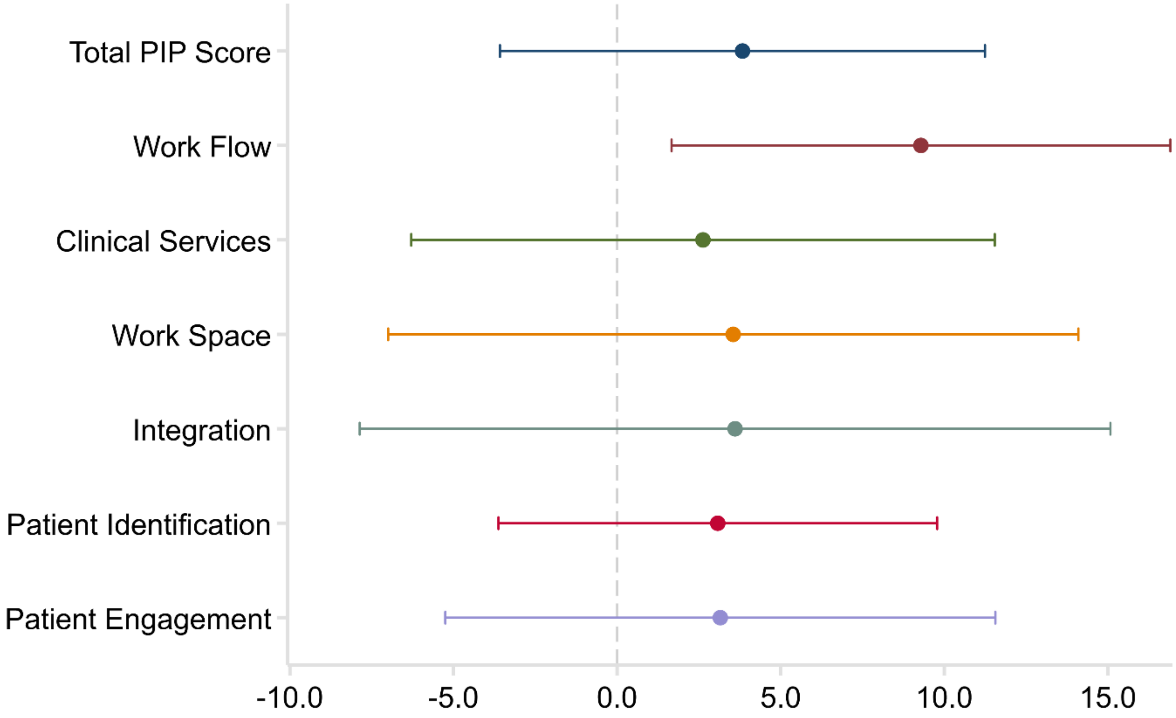
[a] Domain and total scores range from 0 to 100.

[b] Effect was calculated by the difference in difference scores and adjusted for the baseline value of the outcome, clustering within practice, and some variables where indicated.

[c] This effect was adjusted by the variable of the median age of the practice's county.

[d] This effect was adjusted by the variable of the population density of the practice's county.

Figure 5. Effect of the Intervention on Practice Integration



Abbreviation: PIP, Practice Integration Profile.

The center point of each bar is the mean adjusted effect (the regression coefficient) of the intervention on a specific outcome, with the bars representing the 95% CIs. Values greater than 0 indicate that the intervention was associated with an increase in the outcome measured. Bars that do not cross 0 indicate statistical significance with *P* < .05.

AIM 3: CONTEXTUAL FACTORS THAT INFLUENCED THE INTERVENTION AND OUTCOMES

Overview

Aim 3 was to assess the contextual factors that may have supported or impeded successful integration of BH into primary care practices.

Study Design

In this collective case study, we used multiple cases (ie, participating primary care practices) to examine contextual factors that influence integration.[114] A case study approach was selected because it provides an opportunity to explore complex system change initiatives, such as integrating BH into primary care. Our qualitative approach was grounded in a constructivist research paradigm, assuming that interviewees constructed their perspectives of integrating BH based on their lived experiences and how they make meaning of the world.[115] We used quantitative data as our primary case study site selection criterion; after our selection process was complete, this aim focused on qualitative data.

To gain an understanding of the IBH-PC project and potential case study sites, we reviewed project documents such as coaches' notes, learning community forum postings, and intervention materials. We administered a note-taking tool (see Appendix A) quarterly to cluster leaders, which was designed to provide an efficient way to gather information about contextual factors that might be influencing their practices. At several critical junctures, we engaged with key stakeholders to gather feedback to inform our study design and share emerging findings. For example, we shared updates and facilitated discussions with the SAG and with attendees of the IBH-PC annual meeting of co-investigators.

Review of the literature and consultation with subject matter experts informed our development of the data collection instruments. The semistructured interviews featured open-ended questions and probes that interviewers could use to elicit further detail as needed. The interview questions addressed contextual factors at 4 levels: (1) the practice, (2) the larger organization, (3) the community, and (4) the broader external environment. To help communicate these conceptual levels, we developed a Venn diagram of the 4 overlapping levels of potentially influential contextual factors along with examples at each level (see Appendix B). We used this graphic during interviews to help visually communicate the complexities of our research question with the goal of data elicitation.[116] We developed a series of interview guides that aligned with the variety of interviewee roles and a bank of potential interview questions. We conducted a pilot site visit to field-test the interview guides and associated procedures, then refined the question wording and sequencing accordingly.

Sampling Strategy

We purposefully selected 8 primary care practices to vary on key characteristics.[117] We identified tiers of "high-change" and "low-change" practices, where change was defined as change in median total PIP score from the baseline to midpoint assessment. Our site selection was based on midpoint PIP score so that we would have time to begin data collection activities before final PIP assessments. In consultation with the IBH-PC executive committee, 8 participating practices (6 active; 2 control) were purposefully selected to include high-change and low-change sites that varied on characteristics such as geographic location, size, and practice type. Table 8 outlines key characteristics of the 8 selected sites.

Table 8. Characteristics of Selected Sites for Contextual Factors Analysis

PIP score									FTEs	
Baseline	Midpoint	Change	Practice type	Specialty	US region	Patient visits per year, No.	Patient panel, No.	Total staff, No.	PCPs	BHPs
45	84	39	Hospital owned, nonprofit	FM	West	18 984	8752	47	4.3	2.3
42	77	34	Hospital owned, academic, nonprofit	FM	Mountain	4264	5775	57	4.0	0.6
68	89	20	FQHC, CHC, nonprofit	FM/IM	New England	11 651	3992	19	3.8	2.8
61	64	3	FQHC, CHC, nonprofit	FM	South	31 528	8356	63	6.3	6.3
73	76	3	Private, for-profit	FM/IM	Central	12 000	3600	14	3.4	0.5
60	61	1	FQHC, CHC, nonprofit	FM	New England	51 727	12 041	6	11.8	0.8
55	85	30	Private, for-profit	FM/IM	South	42 201	20 000	8	14.3	1.0
40	64	24	Hospital owned, nonprofit	FM/IM	South	37 799	18 900	10	8.6	1.0

Abbreviations: BHP, behavioral health professional; CHC, community health center; FM, family medicine; FQHC, federally qualified health center; FTE, full-time equivalent; IM, internal medicine; PCP, primary care professional; PIP, Practice Integration Profile.

Data Collection Methods

The IBH-PC project was initiated a few years before the assembly of the qualitative investigator team. Existing IBH-PC staff had established relationships with 1 or more key staff members at each of the selected sites. These relationships helped to establish a connection with a site liaison at each practice, who helped schedule interviews and distribute the IRB consent materials. We contacted potential sites through emails and virtual meetings held between March and September 2020. With recruitment assistance from cluster leaders, all 8 selected sites agreed to participate. Potential interviewees at each site were contacted by a research staff member or their practice's IBH-PC liaison.

Prior to each interview, an IBH-PC staff member sent the interviewee a 2-page information sheet. The information sheet described the IBH-PC project and the purpose of this qualitative component of the research. At the beginning of each interview, the interviewer introduced herself to the participant and responded to any questions the interviewee had about the study.

We conducted semistructured interviews with staff in primary care practices, which generated the primary data for our analyses. We interviewed practice staff representing a variety of roles (triangulation of data sources) to gather multiple perspectives on contextual factors that influenced integration in each site. We also interviewed IBH-PC staff serving in 2 roles: (1) coaches providing technical assistance to support practices' integration efforts and (2) researchers assisting with study activities. Interviews with IBH-PC staff elicited background information on the selected practices and helped identify appropriate interviewees.

In total, we conducted 82 semistructured interviews, an average of 10.6 interviews per site (including 3 interviews with individuals whose roles positioned them to address contextual factors in 2 sites, but who were only counted once). Across the 8 sites, interviewees included 14 practice leaders, such as medical directors and owners of private practices; 13 BHPs; 16 PCPs; 9 nurses and medical assistants; 5 front desk staff; and 14 others, including data analysts and care coordinators.

At the first selected practice, we conducted interviews on-site in early March 2020. Our university restricted business travel shortly after the emergence of the COVID-19 pandemic. We pivoted to virtual data collection and conducted the remaining interviews via Zoom or telephone. Most interviews ran for approximately 60 minutes. All the interviews were conducted by a researcher with doctoral-level qualitative research training and were audio-recorded and transcribed. During most interviews, only the participants and researchers were present. In a small number of instances, the interviewee participated in the interview from a location within the primary care office, and other staff passed through during the course of the interview. No repeat interviews were conducted. Field notes were not taken during interviews. Interviewers sometimes jotted notes to assist with the interview process, such as topics to be probed further.

The researchers discussed data saturation throughout the study process. The team used the pilot site visit as an initial means of gauging the number of interviews that might be necessary to reach an appropriate level of data saturation. As the data collection proceeded in the 8 selected sites, the researchers conducting the interviews and analyzing the interview data discussed saturation on a site-by-site basis to inform decisions about whether to pursue additional interviews at a given site. For each practice, we determined that we had gathered enough data to develop a robust understanding of the contextual factors influencing integration at that site, given that little additional information with substantive bearing on the research question surfaced in the final interviews.

In addition to interviews, we asked a few practice staff who had used the IBH-PC intervention toolkit to complete an intervention use questionnaire. The intervention use questionnaire was designed to gather information about the practice’s use of the online curriculum, practice redesign and implementation workbook, remote QI coaching services, and an online learning community. Respondents were asked to indicate the degree to which their practice used the toolkit components on a scale ranging from a great degree to not at all. Respondents were also given the opportunity to provide comments about the toolkit’s components.

Data Analysis

Our analytic team was comprised of 3 researchers with doctoral-level qualitative research training. One team member was responsible for tracking and managing data. Once interview transcripts were received from the professional transcriptionist, files were saved in password-protected locations. To minimize risks to participant confidentiality, transcripts were not released to individuals outside the research team. During the interviews, the researchers often restated key points articulated by interviewees to confirm that the comments were being interpreted correctly. Team members who led interviews also conducted analysis helping to ensure transcriptions were consistent with their experience of the interviews.

We organized the data using a site- and role-specific naming protocol to facilitate cross-case analysis. Transcripts were imported into Dedoose qualitative data analysis software (SocioCultural Research Consultants) to aid our cyclical coding process.[118] The analysis included progressive cycles of coding and categorizing the transcribed interview data to generate themes and subthemes that responded to the aim 3 research question. We conducted an in-depth, thematic analysis of the 997 pages of interview transcripts.[119] Our coding process began using a code list consisting of 58 terms and definitions that were informed by topics in the interview guides and graphic. During our coding process, we created memos to communicate emerging thoughts about potential refinements to our code list. In weekly meetings, we discussed our emerging list of themes building toward consensus on the findings.

Two team members were assigned to each transcript, independently reading and applying codes from the list to data excerpts. Three members of the research team carried out the analysis (analyst triangulation) and employed other validity-enhancing techniques, including assessing intercoder consistency at several junctures and honing coding procedures accordingly. Using Dedoose blind coding and testing center features, we periodically tested our consistency in applying codes to the data. All 3 researchers blind-coded the same 3 transcripts with an average intercoder agreement of 26%. Using the training center, we tested portions of our code list throughout our analytic process, cumulating in a test of 20 of the highest applied codes. Our average Cohen K pooled score across the 3 researchers was 0.72, which is categorized as good agreement.[120,121] In weekly meetings, we discussed our code-list revision memos, our consistency testing, our coding style, and our interpretations of each code. These discussions strengthened our collaborative analytic process.

Results

The following 4 themes frame the findings regarding contextual factors that supported or impeded integration at the selected primary care practices: (1) leadership's commitment to integration, (2) financial considerations, (3) workflow and communication systems, and (4) clinicians' perspectives on integration and team-based health care. Each theme is discussed below (and presented in Table 9), accompanied by illustrative quotes, beginning with supportive factors and followed by impeding factors.

Table 9. Quotes Illustrating Themes Supporting and Impeding IBH

Interviewee's role (and ID)[a]	Illustrative quote
Theme: leadership commitment to IBH	
Supporting factors	
Director of medicine (D1)	I have tried to make it known that integrated behavioral health is just part of primary care's team, part of the resources to support our patients all together.
Behavioral health leader (B4)	There was no integrated behavioral health in primary care within the system until we got the green light from our CEO. Dr [X] and I had submitted this program through budget for I think 3 consecutive cycles, and the previous medical president had not approved it. And then I sat down with the CEO and said, "Look, this is what we're trying to do." And she said, "What do you need?" and I said, "This amount of money." She said, "Great, let's do it." So, it was really her support that was the catalyst for getting the clinical program off the ground.
Impeding factors	
Director of medicine (D3)	The biggest challenge is demonstrating the value in the same way we demonstrate value in other ways in health care. It's very difficult to show downstream savings. . . . The biggest struggle we have is [that] executive leaders have traditionally thought a little bit differently.
Primary care professional (P8)	The larger organization . . . was a big barrier. . . . We were charged with this thing, but . . . we've dealt with significant pushback from different parts of the leadership.
Theme: financial considerations	
Supporting factors	
Director of medicine (D1)	As an FQHC . . . we are reimbursed at a slightly higher rate than other private practices. This is a complicated patient population that has more medical and behavioral health comorbidities that are not necessarily captured in regular office bills.
Director of billing (M1)	If our patients are more compliant and they have less hospitalization, then we get incentivized through our ACOs. The ACOs see the benefits that behavioral health has had on our patients because they are not going to the hospitals or ER as often as they used to.
Impeding factors	
BHP (B1)	The reimbursement rates are super crappy, and the health and behavior codes just changed this year. The health and behavior codes used to be anything above 6 minutes. I can guarantee that I spent 6 minutes with the person. Now it is 16 minutes . . . and I cannot guarantee that.

Interviewee's role (and ID)^{a}	Illustrative quote
Director of medicine (D2)	We need state and federal health care policies that better support integration and care teams. If you're a recognized behavioral health provider, at any level that is appropriate to take care of patients, you should be able to bill any payer. Having to navigate who can see who based on their insurance is just a logistical nightmare that impedes your ability to take care of patients.
Theme: workflow and community systems	
Supporting factors	
BHP (B5)	When they're [BHPs] sitting right there, it's a whole lot easier for them [BHPs] to talk to the patient for a few minutes. We are trying to work on a process where we do huddles in the morning between the behavioral health and care teams . . . so that we have more awareness about who is on the schedule for that day and what kinds of needs they have.
BHP (B1)	The group text is something we've been doing for a few years. And we started doing that because our behavioral health team grew, and we had 2 clinic sites and most all of our counselors have shared days at both clinics. It was hard for people to keep up with who's working where today. We came up with the solution of doing the group text so that they have more immediate access.
Impeding factors	
Director of medicine (D1)	We have a lot of staff turnover. From a workflow standpoint, once you get somebody trained, then they leave, and you've got a new person coming in.
Chief compliance officer (M2)	We do have a Spanish-speaking behavioral health provider, but just 1. More than 50% of our patients speak Spanish. . . . You lose some of the connection when you need an interpreter. It's not just Spanish-speaking patients. There's Kinyarwanda and Swahili and Arabic. One Wednesday, we had 27 different languages on the language line.
Theme: clinicians' perspectives on integration and team-based health care	
Supporting factors	
Medical director (D4)	If you have various levels of practitioners . . . that are very much a part of the team and identify their patients as part of their team as well . . . that really extends our ability to move beyond a specific diagnosis, a specific treatment, and look at what is going to be successful for this individual person. . . . When that team functions well with the patient at the center . . . it allows everyone to change their language and change their expectations. That allows for a much more positive intervention experience across the board.

Interviewee's role (and ID)[a]	Illustrative quote
Primary care professional (P2)	It's very easy to feel like you're in a situation with no great answers. So working together with other people who have some other ideas can help—not always, but it's still good to work with other people that are willing to think about these problems with you collaboratively.
Impeding factors	
Primary care professional (P11)	We have all been trained to address health and provide health care in a very particular social construct. People who come from different ways of engaging in health care, [and] then intersect with the way we do it [IBH]—it doesn't always work well.
Director of medicine (D1)	We have a lot of medical practitioners and behavioral health clinicians that are part-time, so it makes it harder to have consistency with the same providers working with the same behavioral health clinicians frequently and developing smooth processes of communication.

Abbreviations: ACO, accountable care organization; BHP, behavioral health professional; CEO, chief executive officer; ER, emergency room; FQHC, federally qualified health center; IBH, integrated behavioral health; ID, identification.

[a] The ID codes in parentheses identify the specific interviewees.

Leadership Commitment to Integration

Interviewees often mentioned practice leaders' commitment to increasing access to BH care as essential to supporting integration. Practice staff and leaders identified 2 important means of making this commitment operational. First, leaders needed to champion integration in clear, consistent, and prominent ways. As 1 medical director stated, "I've tried to make it known that IBH is just part of primary care's team, part of the resources to support our patients." Second, leaders needed to invest in staffing to coordinate the practice's integration efforts and ensure that BHPs' schedules allowed time for medical colleagues and new patients to access their services. A BH leader described repeated attempts to gain funding for staffing to support integration. When a new CEO came on board, she quickly allocated the necessary financial resources: "Her support was the catalyst for getting the clinical program off the ground."

In contrast, interviewees noted certain leadership stances impeded integration efforts. Some practice leaders did not appreciate the BH staffing levels required to coordinate patient care, including "warm handoffs," in which a PCP introduces a patient to an IBH clinician as part of a referral. Some leaders pressured BHPs to maximize billable activities to an extent that substantially diminished their availability to provide team-based care. Leaders' financial concerns are necessarily influenced by payment models (addressed in the next section); however, value judgments were also a factor. One medical director asserted that "the biggest challenge is demonstrating the value [of integration] in the same way we demonstrate value in other [aspects of] health care. It's very difficult to show downstream savings." In his experience, executive leaders often focus on traditional cost-benefit analyses that define value too narrowly.

Financial Considerations

Many interviewees emphasized that integration requires flexible financial support. Participants noted that alternatives to fee-for-service models, such as capitated payment models, offer financial flexibility to support integration. For instance, some of the selected practices received monthly payments from accountable care organizations for portions of their patient panels. The billing director at one site explained, "[T]he ACOs [accountable care organizations] see the benefits that BH has had on patients because they're not going to the hospitals or emergency rooms as often." Capitated payment models alleviate some pressure for BHPs to maximize billable activities and allow them to be available for consultations with medical colleagues and patients via warm handoffs. The Health Resources and Services Administration mandates that federally qualified health centers offer BH services to their patients. The medical director at a federally qualified health center indicated that his site receives "bonus payments as one way of repaying [the practice] for maintaining a higher level of patient care." In addition to highlighting the role of capitated and value-based payment models, interviewees stressed the importance of having knowledgeable and persistent personnel to navigate the complicated, ever-shifting insurance billing requirements and to secure reimbursement for billable services. Of note, some practices decided to absorb some of the costs of providing IBH because they strongly believe integration is beneficial to their patients.

As suggested, interviewees often cited fee-for-service models as impediments to integration, given that they do not reimburse for activities such as warm handoffs and consultations between PCPs and BHPs. In addition, interviewees noted that some insurance policies will not reimburse for patients seeing a PCP and a BHP on the same day. Interviewees also stated that reimbursement rates for BH services are so low that they jeopardize the long-term sustainability of integration. Insurers also limit which types of BHPs can bill for services provided. A medical director explained that any clinician “should be able to bill any payer. Having to navigate who can see who based on insurance is a logistical nightmare that impedes your ability to take care of patients.”

Workflow and Communication Systems

The participating practices developed customized workflow and communication systems to support their delivery of integrated care. Interviewees underscored the importance of clear, consistent, user-friendly systems for collaborating on patient care and determining when BHPs are available. Physical workspaces that allowed PCPs and BHPs to work in close proximity facilitated collaboration, particularly for warm handoffs and spur-of-the-moment consultations. Interviewees outlined an array of intentional communication strategies used to support the delivery of integrated care. A BHP described “morning huddles” that convened PCPs and BHPs to increase “awareness about who is on the schedule for that day” and anticipate needs for collegial consultations and warm handoffs. Some BHPs invited colleagues to knock on their door, even if it meant interrupting a patient visit, to allow for a brief consultation or warm handoff. A few sites arranged blocks of time in which a BHP was on call (rather than in scheduled sessions) to maximize their availability and enhance the timeliness and fluidity of integrated care. Practice members also used telecommunications, including messaging through the EHR system, sending group texts, telephoning, and videoconferencing, to coordinate integrated care. In addition to developing systems for staff communications, practices recognized the need to help patients understand how BH services could enhance their care. Clinicians often found it beneficial to explain to patients that working with a BHP does not necessarily entail long-term, in-depth psychotherapy. In many cases, patients received the support and practical guidance they needed to address health goals in a relatively short timeframe. A nurse practitioner tells her patients that working with a BHP can help them “learn coping skills to cope with their situation better.”

Challenges to establishing clear and consistent workflow and communication systems included high staff turnover. A medical director also explained that having lots of part-time staff in the practice made it “harder to have consistency, with the same providers working with the same BHPs, and develop smooth processes for communication.” Many sites found it difficult to establish effective systems for identifying when BHPs were available for patient consults, particularly if BHPs’ offices were not located nearby. Challenges also resulted from limited BHP staffing levels that impeded patients’ timely access to services. Interviewees emphasized the importance of enabling patients to meet with a BHP when they are ready. As a medical assistant observed, “You might lose the momentum of the patient because it’s a scary thing to admit you need help.” Stigmas about BH services and “what it means” to see a counselor dissuaded some patients from accessing care. Good communication regarding the role of the BHP from the person doing the handoff helped to reduce this reluctance. Other barriers to providing integrated care involved cultural and language differences. Many practices used online translation or interpreter services, as illustrated by an interviewee’s observation that, “one Wednesday, we had 27 different languages on the language line.” Though these services were necessary and helpful, the “back-and-forth” with an interpreter sometimes made it more difficult to build a trusting relationship between patients and clinicians.

Clinicians’ Perspectives on Integration and Team-Based Care

Interviewees often described integration as a means of putting patients’ needs at the center of primary care services. Although efforts to enhance integration tend to focus on the processes required to deliver team-based care, many interviewees emphasized that integration helps ensure that patients feel cared for when receiving primary care services. Most of the participating clinicians expressed the belief that integration is vital to improving patient outcomes, and this conviction appeared to fuel their efforts. As 1 interviewee explained:

> *If you have various practitioners . . . that are very much a part of the team and identify their patients as part of their team . . . that extends our ability to move beyond a specific diagnosis, a specific treatment, and look at what is going to be successful for this individual person. . . . When that team functions well with the patient at the center . . . it allows everyone to change their language and expectations. It allows for a much more positive intervention experience across the board.*

In addition, many interviewees valued the benefits that result for clinicians. They appreciated that integration enhances the level of support they receive from colleagues, informs their clinical decision-making through consultation with clinicians offering different areas of expertise, builds their professional knowledge and skills over time as a result of these interactions, and protects them from feeling overwhelmed and burned-out. A physician described the tendency to “feel like you’re in a situation with no great answers.” He explained that “working with other people who have some other ideas can help—not always, but it’s still good to work with other people that are willing to think about these problems with you collaboratively.” Some interviewees noted that the COVID-19 pandemic underscores the need for integration, given the magnitude and complexity of the related challenges faced by patients and clinicians alike.

Along with clinician perspectives that support integration, interviewees identified barriers that may result from different disciplinary backgrounds and professional training in a practice. As one health care professional stated:

> *We’ve all been trained to address health and provide health care in a particular social construct. People who come from different ways of engaging in health care and then intersect with the way we do it—it doesn’t always work.*

A lack of shared understanding or appreciation of integration among clinicians created obstacles in some settings. Interviewees from a few sites acknowledged that physicians varied considerably in the degree to which they embraced integration. Some physicians maximized opportunities to engage with the practice's BHPs and to provide integrated care for their patients whereas other physicians took a minimalist approach.

AIM 4: IMPLEMENTATION COSTS

Overview

Our objective for this aim was to describe the costs of implementing the IBH-PC toolkit intervention borne by the practices.

Study Outcomes

We conceptualized costs to include staff time spent on planning meetings and implementation activities, supplies, and capital costs, but we did not include the ongoing operational costs of providing integrated services.

Sample Size Calculations and Power

No formal power estimates were employed for this descriptive analysis. We judged that 8 practices would provide a representative sample of the active sites.

Data Collection and Sources

Eight practices were selected for the aim 4 analysis from the 21 intervention practices. Selection of aim 4 practices was independent of selection of aim 3 practices. For aim 4, practices were purposefully selected to be reasonably representative of all the intervention practices in terms of (1) the number, type, and efforts of its medical professionals; (2) the practice’s patient panel size; (3) the practice’s patient visit volume; (4) the number and efforts of its BHPs; (5) the practice’s NCQA Patient-Centered Medical Home status; (6) the practice’s ratio of adult patients on Medicare; (7) the type of practice (eg, community health center, health system–owned, clinician-owned, academic); and (8) the location of the practice (urban, rural). One practice declined to participate after being selected because of competing priorities and was replaced with an alternative practice that had closely matching characteristics.

We developed a cost assessment tool (CAT) to collect data from selected staff members at each of the 8 practices according to their specific practice role (eg, clinician, BHP, staff, office manager). The CAT captured close estimates (in minutes) of time-effort data related to implementing the IBH-PC program at those clinic sites.

Analytical and Statistical Approaches

Fields in the CAT spreadsheet obtained data on the amount of time spent by practice members (in minutes) doing activities outlined in the IBH-PC toolkit, by each practice role and by the number and type of practice participants performing these activities. The CAT summed the time each practice member spent completing all program steps across 4 project components, including IBH-PC workbooks, practice member coaching, educational curricula, and learning community activities, over a 24-month period.

The CAT generated estimates of time-effort and cost in US dollars using median total compensation rates reflective of each practice's zip code and the participants' professional roles. Missing or erroneous information was corrected by working with QI team facilitators at each clinic site.

Results

The 8 selected econometric study practices were located across the United States, including Hawaii. Their patient panels ranged from 2000 to 16 000; they saw between 4200 and 32 000 patients per year; their NCQA Patient-Centered Medical Home status ranged from no certification to level 3; they employed 3 to 19 PCPs and 1 to 8 BHPs. Four of the practices represented community health centers of various types, 4 practices included family medicine residencies (academic and nonacademic), and 1 practice was managed by a large private health system. Half of the practices were located in rural areas. Two practices were urban.

There were no significant differences between the selected econometric study practices and the rest of the practices in geographical distribution (rurality), patient panel size and number of patient visits, clinician workforce (PCP and BHP FTE), specialty mix, adult Medicare ratio, or level of baseline IBH. The 8 practices represented a moderately higher ratio of academic institutions (24% vs 10%), resident training sites (50% vs 34%), and community health centers (50% vs 28%), but these differences were not statistically significant. Similarly, no significant differences were observed between the econometric study patient subsample and the rest of the IBH-PC study in the characteristics described in Table 10.

Table 10. Characteristics of the Econometric Study Sites

	Econometric study practices (n = 8)	All other IBH-PC practices (n = 33)	*P* value[a]
Practice characteristics			
Patient panel size, mean (SD), No.	7861 (1451)	9608 (5344)	.39
Patient encounters per year, mean (SD), No.	19 644 (10 958)	28 469 (21 034)	.26
Adult patients on Medicare, mean (SD), No.	23.0 (18.7)	20.7 (13.3)	.69
Primary care professionals, mean (SD), No.	8.3 (3.0)	10.1 (6.3)	.43
BHPs, mean (SD), No.	2.6 (2.3)	2.8 (2.7)	.85
Total PIP score at baseline, mean (SD)	60.1 (11.2)	58.9 (15.1)	.83
Community health centers, No. (%)	4 (50.0)	10 (30.3)	.29
Sites with residency training, No. (%)	4 (50.0)	11 (33.3)	.38
Academic sites, No. (%)	2 (25.0)	3 (9.1)	.22
NCQA PCMH < level 2, No. (%)	4 (50.0)	19 (57.6)	.69
Patient characteristics			
Age, mean (SD), y	64.2 (4.9)	65.7 (4.4)	.39
Aged 65 years or older, mean (SD), No.	47.7 (16.8)	51.8 (15.6)	.51
Percentage White race, mean (SD)	78.8 (27.0)	76.9 (19.0)	.82
Percentage male sex, mean (SD)	37.3 (9.9)	36.5 (6.8)	.80
Percentage Hispanic ethnicity, mean (SD)	11.3 (8.2)	11.7 (12.2)	.92
Percentage married, mean (SD)	52.6 (6.6)	46.6 (14.5)	.32
Percentage employed, mean (SD)	30.9 (9.9)	28.1 (8.5)	.41
Percentage low income, mean (SD)	50.5 (29.7)	48.3 (20.8)	.81
Percentage with some college education, mean (SD)	44.8 (18.8)	48.6 (17.9)	.60

Abbreviations: BHP, behavioral health professional; IBH-PC, Integrating Behavioral Health and Primary Care; NCQA, National Center for Quality Assurance; PCMH, Patient-Centered Medical Home status; PIP, Practice Integration Profile.

[a] *P* values calculated by Student *t* test or Fisher exact test.

The median estimated total cost was $20 726. The estimated total costs ranged from $12 381 to $60 427. The median estimated cost for the planning stage was $10 258, and the median estimated cost for the implementation stage was $9208.

AIM 5: THE IMPACT OF THE COVID-19 PANDEMIC

Overview

Aim 5 was a later enhancement of the original study. It sought to determine if the perceived burden of the COVID-19 pandemic influenced the effect of the intervention on functional outcomes. We suspected that in communities with a high COVID-19 burden, the efforts of practice staff to address and cope with the burden might divert resources needed to effectively implement the intervention. To accomplish this aim, we modeled the interaction of COVID-19 burden and the intervention (group assignment) upon function status. There were no changes to the original design of the study or the intervention beyond the additional survey items described in the sections that follow.

Study Outcomes

The outcome variables were the PROMIS-29 scores for the 8 domains. See aim 1 for more details.

Covariates

We defined 8 potential markers of the burden of the COVID-19 pandemic on patients and practices:

- Incidence was defined as the cumulative number of cases reported per 100 000 residents for the home county of the patient as of the date of their final survey.
- Prevalence was defined as the number of new cases diagnosed per 100 000 residents in the patient's county in the 14 days before they completed their final survey.
- Deaths were quantified as the cumulative number of deaths from COVID-19 reported per 100 000 residents for the home county of the patient as of the date of their final survey.
- County-level unemployment rates as well as COVID-19 prevalence, incidence, and mortality rates based on state and local health agency reports were gathered from the *New York Times* data.[122]

Perceived COVID-19 burden was assessed using 4 questions (Table 11). These questions were created by the study team because no validated measure of COVID-19 burden had been published in the literature.

Table 11. Perceived COVID-19 Burden Survey Items

Domain	Question	Answer options
Personal burden	How much has life changed for you personally because of COVID-19?	Not at all A little Some A lot Everything is different now
Community burden	How much has life changed for your community because of COVID-19?	
Personal diagnosis	Were you personally diagnosed with COVID-19 by a doctor or other health care professional?	Yes No
Housemate or family diagnosis	Was anyone in your household or immediate family diagnosed with COVID-19 by a doctor or other health care professional?	

Potential patient-level confounders included age, sex, race, ethnicity, marital status, income level, education level, employment status, urban residence, census tract Social Deprivation Index score, and census tract population. Potential practice-level confounders included community health center status, hospital affiliation, private ownership, nonprofit status, specialty, encounters per year, patient panel size, percentage of adults covered by Medicare, BHP staffing in hours per week, PCP staffing in hours per week, duration of on-site BH services in years, presence of resident trainees, and baseline degree of integration (by PIP score). Additional potential confounders included the median age, racial breakdown, ethnic breakdown, income level, education level, employment status, Social Deprivation Index score, and population density of the county of the practice as well as the date of the survey.

Sample Size Calculations and Power

The sample size was fixed by the time this analysis was conceived. The sample size and power calculations were therefore those described for aim 1.

Time Frame

Surveys were completed between April 29, 2020, and January 8, 2021, approximately 6 months after the third wave of aim 1 surveys.

Data Collection and Sources

A fourth wave of patient outcomes was collected using the same methods described for aim 1. This wave included the same data as the surveys collected for aim 1 as well as the items related to the perceived burden of COVID-19 (Table 11).

Analytical and Statistical Approaches

Using the same analytic approach used for aim 1, we built mixed linear models of the effect of the interaction between the intervention and the burden of COVID-19 on patient outcomes, with random intercepts for practice, adjusting for potential confounders. If the interaction term was associated with the outcome with a $P < .05$, it was considered evidence that COVID-19 influenced the effect of the intervention. We repeated the regression for each of the 8 markers of COVID-19 burden and each of 8 patient outcomes for a total of 64 models.

Results

Of the 2460 eligible participants who completed the third wave of surveys (aim 1), 235 either did not respond to multiple attempts to contact them or opted out of the fourth wave of surveys, leaving 2225 participants who completed the COVID-19 survey items. The participants who completed the fourth wave of surveys did not significantly differ from the participants who completed the first 3 waves in terms of demographic or clinical characteristics (Table 12) and were largely representative of US primary care patients. They were 66% female (compared with 66% in waves 1-3), 76% White (77%), 12% Black (12%), and 7% Latino (7%). Participants who resided in rural areas made up 20% of wave 4 participants compared with 20% in waves 1 through 3. There were also minor differences between survey waves 1 through 3 and survey wave 4 in marital status, income level, employment status, and history of chronic conditions. The data point of education level was not collected during survey wave 4.

Table 12. Participant Characteristics of the Final Analytic Sample vs the Wave 4 Sample

Characteristic	Survey waves 1-3 (n = 2426)[a]	Survey wave 4 (n = 2225)[a]
Age, mean (SD), y	61.9 (13.1)	63.6 (13.0)
Sex, No. (%)		
Female	1588 (65.5)	1466 (65.9)
Male	832 (34.3)	753 (33.8)
Prefer to self-describe or not to say	6 (0.2)	6 (0.3)
Race, No. (%)		
American Indian or Alaska Native	21 (0.9)	21 (1.0)
Asian	71 (2.9)	68 (3.1)
Black or African American	281 (11.6)	255 (11.5)
Native Hawaiian or other Pacific Islander	37 (1.5)	37 (1.7)
White	1859 (76.6)	1699 (76.4)
Other or prefer not to say	157 (6.5)	145 (6.3)
Ethnicity, No. (%)		
Hispanic	180 (7.4)	164 (7.4)
Non-Hispanic	2207 (91.0)	2026 (91.1)
Prefer not to say	39 (1.6)	35 (1.5)
Marital status, No. (%)		
Never married	395 (16.3)	376 (16.9)
Married	1063 (43.8)	963 (43.3)
Living as married	63 (2.6)	60 (2.7)
Separated	56 (2.3)	49 (2.2)
Divorced	523 (21.6)	481 (21.6)
Widowed	313 (12.9)	284 (12.8)
Prefer not to say	13 (0.5)	12 (0.5)
Employment status, No. (%)		
Full time	430 (17.7)	387 (17.4)
Part time	161 (6.6)	152 (6.8)
Retired	1009 (41.6)	943 (42.4)
Disabled	615 (25.4)	568 (25.5)
Homemaker	84 (3.5)	83 (3.7)
Student	11 (0.5)	11 (0.5)
Unemployed or seeking employment	77 (3.2)	73 (3.3)

Characteristic	Survey waves 1-3 (n = 2426)[a]	Survey wave 4 (n = 2225)[a]
Other or prefer not to say	39 (1.6)	8 (0.4)
Annual household income, No. (%)		
<$15 000	674 (27.8)	637 (28.6)
$15 000-$29 999	505 (20.8)	475 (21.4)
$30 000-$44 999	294 (12.1)	275 (12.4)
$45 000-$59 999	222 (9.2)	205 (9.2)
$60 000-$74 999	176 (7.3)	156 (7.0)
$75 000-$99 999	195 (8.0)	176 (7.9)
≥$100 000	282 (11.6)	259 (11.6)
Prefer not to say	78 (3.2)	42 (1.9)
Chronic conditions, No. (%)		
Arthritis	1023 (42.2)	956 (43.0)
Asthma	545 (22.5)	500 (22.5)
Chronic obstructive pulmonary disease	330 (13.6)	305 (13.7)
Chronic pain	2037 (84.0)	1875 (84.3)
Nongestational diabetes	1075 (44.3)	997 (44.8)
Heart failure	188 (7.8)	179 (8.0)
Hypertension	2012 (82.9)	1851 (83.2)
Irritable bowel syndrome	102 (4.2)	98 (4.4)
Anxiety	830 (34.2)	761 (34.2)
Depression	1136 (46.8)	1047 (47.1)
Insomnia	570 (23.5)	530 (23.8)
Substance use disorder	559 (23.0)	524 (23.6)
Tobacco use	445 (18.3)	412 (18.5)
Alcohol use disorder	155 (6.4)	146 (6.6)
Chronic conditions, mean (SD), No.	4.4 (1.6)	4.4 (1.6)
Neighborhood characteristics (home census tract)		
Social Deprivation Index, mean (SD), points[b]	52.6 (27.7)	53.1 (27.8)
Rural location, No. (%)	477 (19.7)	445 (20.0)
Population density, mean (SD), people per square mile	3768 (6656)	3800 (6811)

[a] Percentages may not add up to 100.0% due to rounding.

[b] The Social Deprivation Index runs from 0 (least deprivation) to 100 (most deprivation).

Some aspects of the burden of COVID-19 differed by experimental condition (Table 13). More participants in the intervention group had positive COVID-19 diagnoses (3.7%) compared with the control group (2.2%); however, housemate or family member diagnosis of COVID-19 and perceived personal and community burden did not differ between groups. The prevalence and incidence rates of COVID-19 and unemployment rates were higher in counties where control participants were located compared with counties where active participants were located.

Table 13. COVID-19 Burden Survey Responses, by Study Group

	Intervention group (toolkit) (n = 877)	Control group (no toolkit) (n = 1348)	*P* value
Personal COVID-19 diagnosis, No. (%)	32/877 (3.7)	30/1348 (2.2)	.05
Housemate or family member with COVID-19 diagnosis, No. (%)	64/877 (7.3)	92/1348 (6.8)	.66
Perceived personal COVID-19 burden, No. (%)			.89
Not at all	52/870 (6.0)	78/1343 (5.8)	
A little	127/870 (14.6)	213/1343 (15.9)	
Some	242/870 (27.8)	380/1343 (28.3)	
A lot	244/870 (28.1)	355/1343 (26.4)	
Everything is different	205/870 (23.6)	317/1343 (23.6)	
Perceived community COVID-19 burden, No. (%)			.77
Not at all	23/870 (2.6)	29/1343 (2.2)	
A little	69/870 (7.9)	69/1343 (5.1)	
Some	179/870 (20.6)	179/1343 (13.3)	
A lot	386/870 (44.4)	386/1343 (28.7)	
Everything is different	212/870 (24.4)	212/1343 (15.8)	
County COVID-19 burden, mean (SD)			
Prevalence rate, per 100 000	136 (191)	215 (283)	<.001
Incidence rate, per 100 000	1114 (1151)	1474.6 (1373)	<.001
Mortality rate, per 100 000	31.5 (47.6)	34.7 (35.9)	.06
Unemployment rate, %	7.8 (4.5)	8.6 (4.8)	<.001

The interaction between the intervention and the burden of COVID-19 reached significance in 5 of the 64 models tested (7.8% [95% CI, 2.6%-17.3%]), similar to the 5% rate of nominal significance expected from random error alone. The models with significant interaction terms were (1) sleep disturbance and COVID-19 deaths, (2) anxiety and perceived community burden, (3) pain intensity and perceived community burden, (4) physical function and personal diagnosis of COVID-19, and (5) pain interference and housemate or family member with a diagnosis of COVID-19. These 5 models included 5 different outcomes and 4 different COVID-19 markers, suggesting no particular pattern.

AIM 6: THE RELATIONSHIP OF INTEGRATION TO OUTCOMES

Overview

Aim 6 was to investigate the relationship between practice integration and patient outcomes independent of the role of the intervention. Because the results of aims 1 and 2 were negative and we lacked a concurrent control without integration activities, we sought to understand if the intervention failed because it did not induce enough integration beyond those aims in the control group or because integration was achieved but was ineffective. We therefore performed a post hoc exploratory analysis, setting aside the intervention and treating the data as a large, observational cohort.

Study Outcomes

The outcome variables were the 8 PROMIS-29 domains as well as the physical and mental health summary scores.

Covariates

The predictor was the median of the total PIP score for each practice. Patient, practice, and neighborhood characteristics collected for aims 1 and 2 were considered potential confounders.

Sample Size Calculations and Power

The sample size was fixed by the needs of aim 1.

Data Collection and Sources

See aims 1 and 2.

Analytical and Statistical Approaches

This analysis was conducted at the patient level. We built a mixed linear model with functional status as the dependent variable, the median of the total PIP score for the respondent's practice as a fixed independent effect, and practice as a random intercept. Patient, practice, and neighborhood characteristics were included as fixed effects if they changed the strength of the association between PIP score and the outcome by more than 10%. We performed the analysis for each of the 8 PROMIS-29 domains as well as the physical and mental health summary scores at both baseline and follow-up. These post hoc analyses were not part of the original study design and are exploratory and descriptive rather than hypothesis-testing.

Results

Regardless of assignment to the active intervention or the control group, we observed a positive association between degree of integration (total PIP score) at baseline and patient mental health function (PROMIS-29) at baseline and 2 years. After adjusting for potential confounding and clustering within practice, all 8 baseline PROMIS-29 domain scores and both physical and mental health summary scores were associated with baseline total PIP score in the anticipated direction (higher integration associated with better function). Anxiety, sleep disturbance, social participation, and the mental health summary score were statistically significantly improved (Table 14 and Figure 6). Confounders included in each model are indicated in Table 14. Total PIP scores at baseline ranged from 27 to 87 with a median (IQR) score of 61 (45-67). Comparing outcomes at the lower-quartile score with outcomes at the upper-quartile score revealed differences on the order 1 point in PROMIS-29 scores, which is not generally considered clinically significant, although they may be important at the population level.

We also examined the association of PIP total score at baseline with patient outcomes independent of any potential effect of the intervention at follow-up in longitudinal analyses. As was the case at baseline, after adjustment, all 8 domains and both summary scores were associated with total PIP score at baseline in the anticipated direction (higher integration associated with better function). Anxiety, social participation, and mental health summary scores were statistically significantly improved (Table 14 and Figure 6). As was the case at baseline, the changes observed were small and probably not clinically significant for individuals.

Table 14. Association of Practice Integration and Patient Outcomes

	Baseline			Follow-up		
PROMIS-29 domain	**Effect**	**95% CI**	***P* value**	**Effect**	**95% CI**	***P* value**
Anxiety	−0.06[a,b,c,d]	−0.12 to 0.00	.05	−0.05[a,b,c,d]	−0.09 to −0.01	.01
Depression	−0.04[b,c,d]	−0.10 to 0.01	.10	−0.04[a,b,c,d]	−0.08 to 0.01	.13
Fatigue	−0.03[a,b,c,d,e]	−0.07 to 0.01	.10	−0.03[a,b,c,d]	−0.07 to 0.01	.19
Sleep disturbance	−0.05[a,c,d]	−0.10 to −0.01	.03	−0.02[b,c,f]	−0.07 to 0.02	.30
Pain interference	−0.01[a,b,e,f,g]	−0.06 to 0.04	.64	−0.03[a,b,c,d,f,g]	−0.07 to 0.02	.30
Pain intensity	−0.01[b,c,d,f,g]	−0.03 to 0.01	.52	−0.01[b,c,d,f,g]	−0.02 to 0.01	.45
Social participation	+0.05[b,c,d]	0.01-0.09	.02	+0.06[a,b,c,d]	0.02-0.10	.01
Physical function[h]	+0.04[a,b,c,d,f,g]	−0.01 to 0.08	.12	+0.02[a,b,d,f,g]	−0.03 to 0.07	.39
Physical health summary[h]	+0.04[a,b,c,d,e,f,g]	−0.01 to 0.09	.12	+0.02[a,b,d,f,g]	−0.03 to 0.07	.35
Mental health summary[h]	+0.05[a,b,c,d]	0.00-0.09	.05	+0.04[a,b,c,d]	0.00-0.09	.04

[a] This effect was adjusted for the variable of sex.

[b] This effect was adjusted for the variable of a practice being privately owned.

[c] This effect was adjusted for the variable of a practice’s behavioral health professional work hours per week.

[d] This effect was adjusted for the variable of the following employment status answer options: employed, homemaker, and student.

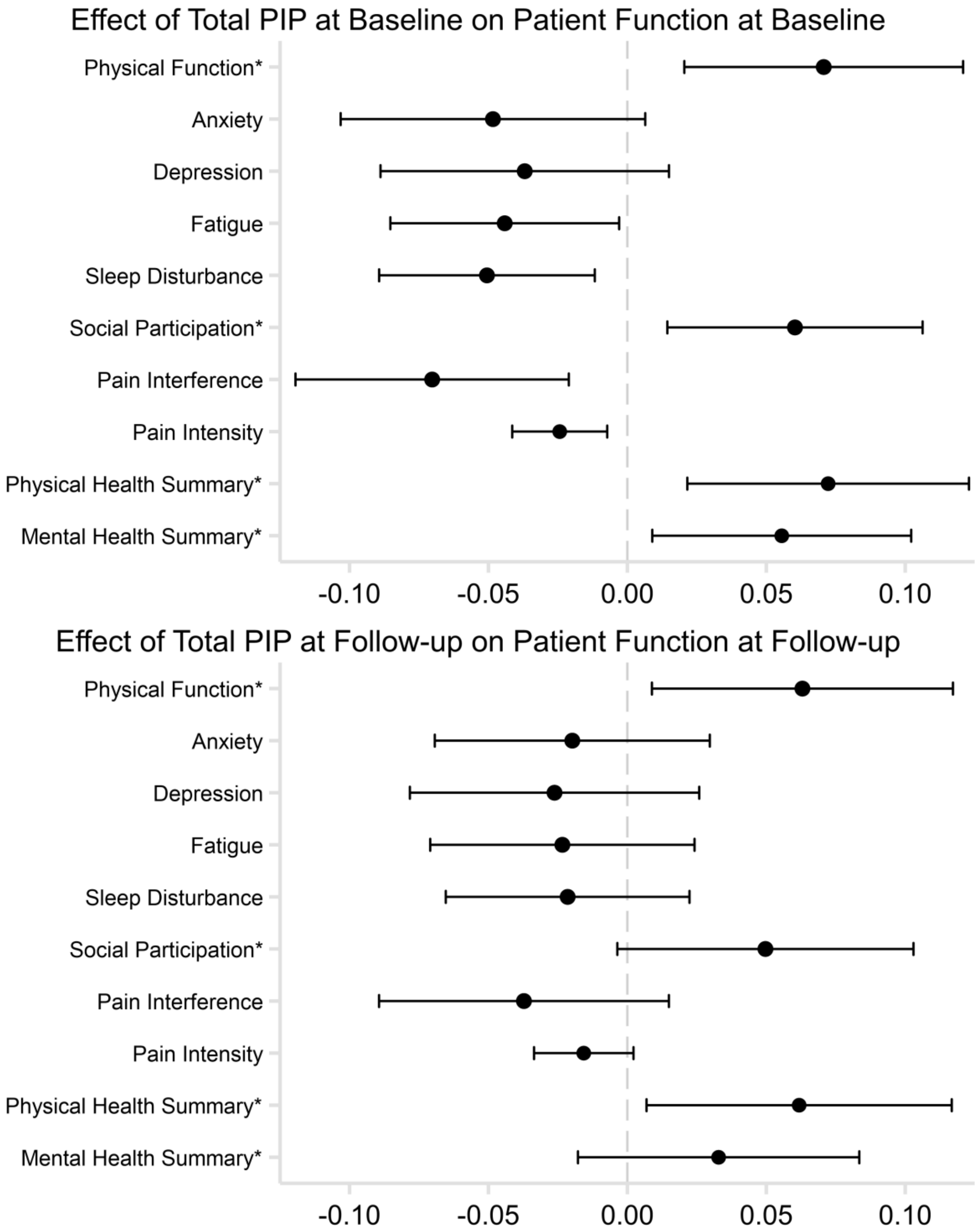
[e] This effect was adjusted for the variable of patient practice visits per year.

[f] This effect was adjusted for the variable of the practice county’s Social Deprivation Index score.

[g] This effect was adjusted for the variable of the practice’s county-level employment rate.

[h] Higher scores indicate improved function and have been adjusted for the baseline value of the outcome, for clustering within practice, and by specific variables where indicated.

Figure 6. Association of Practice Integration and Patient Outcomes



Effect of Total PIP at Baseline on Patient Function at Baseline

Effect of Total PIP at Follow-up on Patient Function at Follow-up

Abbreviation: PIP, Practice Integration Profile.
*For this score, higher values have better outcomes.
The center point of each bar is the mean adjusted association (the regression coefficient) of the baseline median total PIP score with specific patient outcomes, adjusting for potential confounders and employing a random intercept for each practice. The bars represent the 95% CIs. Values greater than 0 indicate that integration was associated with an increase in the outcome measured. Bars that do not cross 0 indicate statistical significance (*P* < .05).

AIM 7: WHAT IS THE ASSOCIATION BETWEEN SOCIAL DETERMINANTS OF HEALTH AND PATIENT OUTCOMES?

Overview

We sought to characterize the relationship of SDOH over time on patient-reported health outcomes in chronically ill adults and to understand this relationship in subgroups of interest: sex, race, ethnicity, and education level.

Study Outcomes

Patient-reported health outcomes were measured using the PROMIS-29 physical health summary T-score and mental health summary T-score at follow-up (survey wave 3).

Covariates

Social determinants of health were measured using 4 yes-or-no questions related to food, housing, and financial insecurities. Food insecurity was considered present if someone responded yes to the question: “In the past 12 months, did you ever eat less than you felt you should because there wasn’t enough food?” Housing insecurity was considered present if someone responded yes to at least 1 of the following 2 questions: “In the past 12 months, did you ever have to sleep at a friend or family member’s house because you had no place else to live?” and “In the past 12 months, did you ever have to sleep outside, in a public place, or in a shelter because you had no place else to live?” Financial hardship was considered present if someone responded yes to the following question: “In the past 12 months, did you ever not have enough money to pay your basic living expenses (mortgage, rent, utilities, medicines, etc)?” Social determinants of health were assessed at both baseline (survey wave 1) and follow-up (survey wave 3).

Sample Size Calculations and Power

The sample size was fixed by the needs of aim 1.

Data Collection and Sources

See aims 1 and 2.

Analytical and Statistical Approaches

Each SDOH predictor (food, housing, and financial insecurity) was analyzed individually with participants being classified as having no SDOH predictors, SDOH predictors at survey wave 1 only, SDOH predictors at survey wave 3 only, or SDOH predictors at both waves.

Average physical and mental health summary T-scores were estimated using 95% CIs for each of the 3 SDOH predictors across the 4 SDOH categories (SDOH predictors not present, present at baseline only, at follow-up only, and at both baseline and follow-up). Subgroup analyses were also conducted to understand this relationship by sex (male or female), race (White or non-White), ethnicity (Hispanic or non-Hispanic), and education level (college or no college).

Results

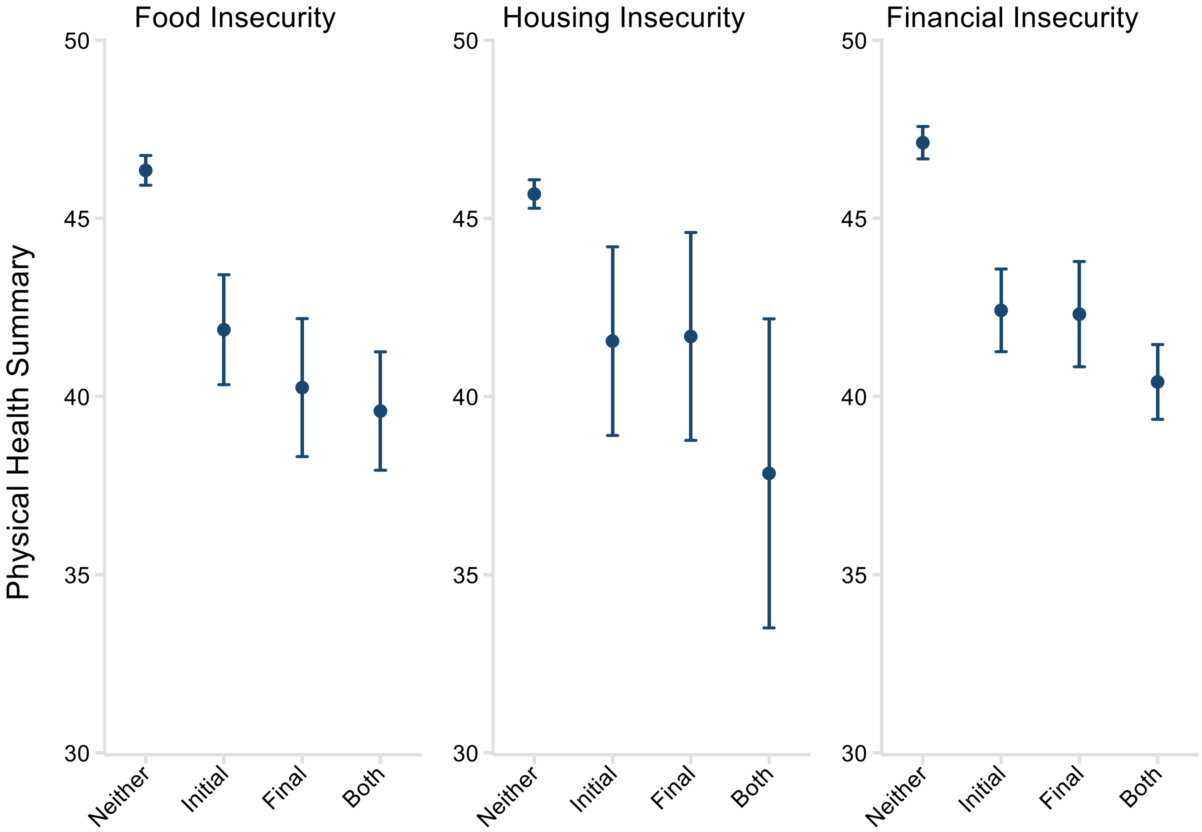
Physical health summary T-scores as a function of the persistence of SDOH predictors are presented in Table 15 and Figure 7. Mental health summary T-scores as a function of the persistence of SDOH predictors are presented in Table 16 and Figure 8. Health was the best (T-scores were highest) for individuals with no SDOH insecurities and the worst (T-scores were lowest) for individuals with SDOH insecurities present at both survey waves. Individuals with an SDOH insecurity at wave 1 only had better health than individuals whose SDOH insecurity persisted over both waves. These patterns were consistent for both physical and mental health outcomes and across all subgroups.

Table 15. Physical Health Summary as a Function of the Persistence of Social Determinants of Health

		Physical health summary T-score, mean (95% CI)			
Social determinant	**Group**	**Neither wave**	**Wave 1 only**	**Wave 3 only**	**Both waves**
Food insecurity	All participants	46.3 (45.9-46.8)	41.9 (40.3-43.4)	40.2 (38.3-42.2)	39.6 (37.9-41.3)
Housing insecurity	All participants	45.7 (45.3-46.1)	41.6 (38.9-44.2)	41.7 (38.8-44.6)	37.8 (33.5-42.2)
Financial insecurity	All participants	47.1 (46.7-47.6)	42.4 (41.3-43.6)	42.3 (40.8-43.8)	40.4 (39.4-41.5)
Food insecurity	Female	45.8 (45.3-46.3)	40.7 (38.8-42.6)	38.2 (35.9-40.5)	39.1 (37.2-41.1)
	Male	47.4 (46.7-48.1)	44.0 (41.4-46.7)	45.2 (41.7-48.7)	41.1 (37.9-44.3)
Housing insecurity	Female	45.0 (44.5-45.5)	39.8 (36.4-43.2)	40.5 (37.2-43.8)	35.7 (29.4-41.9)
	Male	47.0 (46.3-47.6)	44.7 (40.4-49.0)	47.0 (40.8-53.3)	39.8 (33.8-45.7)
Financial insecurity	Female	46.4 (45.9-47.0)	41.7 (40.3-43.1)	42.6 (40.8-44.4)	39.9 (38.6-41.1)
	Male	48.4 (47.6-49.1)	44.1 (42.0-46.2)	41.8 (39.1-44.4)	41.6 (39.7-43.5)
Food insecurity	Non-White	45.3 (44.3-46.2)	41.0 (37.9-44.2)	39.9 (36.2-43.6)	40.3 (36.8-43.8)
	White	46.6 (46.1-47.1)	42.0 (40.2-43.8)	40.8 (38.5-43.2)	39.4 (37.5-41.3)
Housing insecurity	Non-White	44.7 (43.8-45.6)	41.8 (37.5-46.1)	35.2 (27.5-42.9)	37.3 (29.6-45.0)
	White	45.9 (45.5-46.4)	41.7 (38.3-45.0)	43.1 (39.9-46.3)	38.1 (32.9-43.3)
Financial insecurity	Non-White	46.3 (45.2-47.4)	43.0 (41.1-45.0)	43.4 (40.8-46.1)	39.7 (37.7-41.8)
	White	47.3 (46.8-47.8)	42.1 (40.7-43.5)	41.6 (39.8-43.4)	40.7 (39.5-41.9)
Food insecurity	Non-Hispanic	46.4 (45.9-46.8)	41.9 (40.3-43.6)	40.5 (38.4-42.5)	39.7 (37.9-41.4)
	Hispanic	45.9 (44.3-47.5)	40.8 (35.9-45.8)	33.3 (25.0-41.6)	39.3 (34.1-44.4)
Housing insecurity	Non-Hispanic	45.7 (45.3-46.2)	41.4 (38.5-44.3)	41.7 (38.8-44.6)	37.5 (32.8-42.2)

		Physical health summary T-score, mean (95% CI)			
Social determinant	**Group**	**Neither wave**	**Wave 1 only**	**Wave 3 only**	**Both waves**
	Hispanic	44.7 (43.3-46.2)	41.5 (34.4-48.7)	—	39.8 (28.9-50.7)
Financial insecurity	Non-Hispanic	47.2 (46.7-47.6)	42.2 (40.9-43.4)	42.0 (40.5-43.6)	40.5 (39.4-41.6)
	Hispanic	46.8 (44.9-48.6)	44.1 (40.5-47.7)	43.8 (39.4-48.2)	38.8 (35.5-42.1)
Food insecurity	No college	44.6 (44.0-45.1)	40.6 (38.6-42.5)	38.9 (36.6-41.2)	39.0 (36.9-41.0)
	College	48.2 (47.6-48.8)	44.1 (41.6-46.6)	43.3 (39.9-46.8)	40.9 (38.1-43.7)
Housing insecurity	No college	43.8 (43.2-44.3)	41.8 (38.7-45.0)	40.0 (36.0-43.9)	38.3 (32.9-43.6)
	College	47.8 (47.3-48.4)	41.0 (36.3-45.6)	43.6 (39.4-47.7)	37.4 (29.8-44.9)
Financial insecurity	No college	45.2 (44.5-45.8)	41.8 (40.4-43.2)	41.7 (39.8-43.5)	39.9 (38.6-41.2)
	College	48.9 (48.3-49.6)	43.9 (41.8-46.0)	43.3 (41.0-45.6)	41.4 (39.7-43.2)

Figure 7. Relationship of Physical Health to Social Determinants of Health Over Time



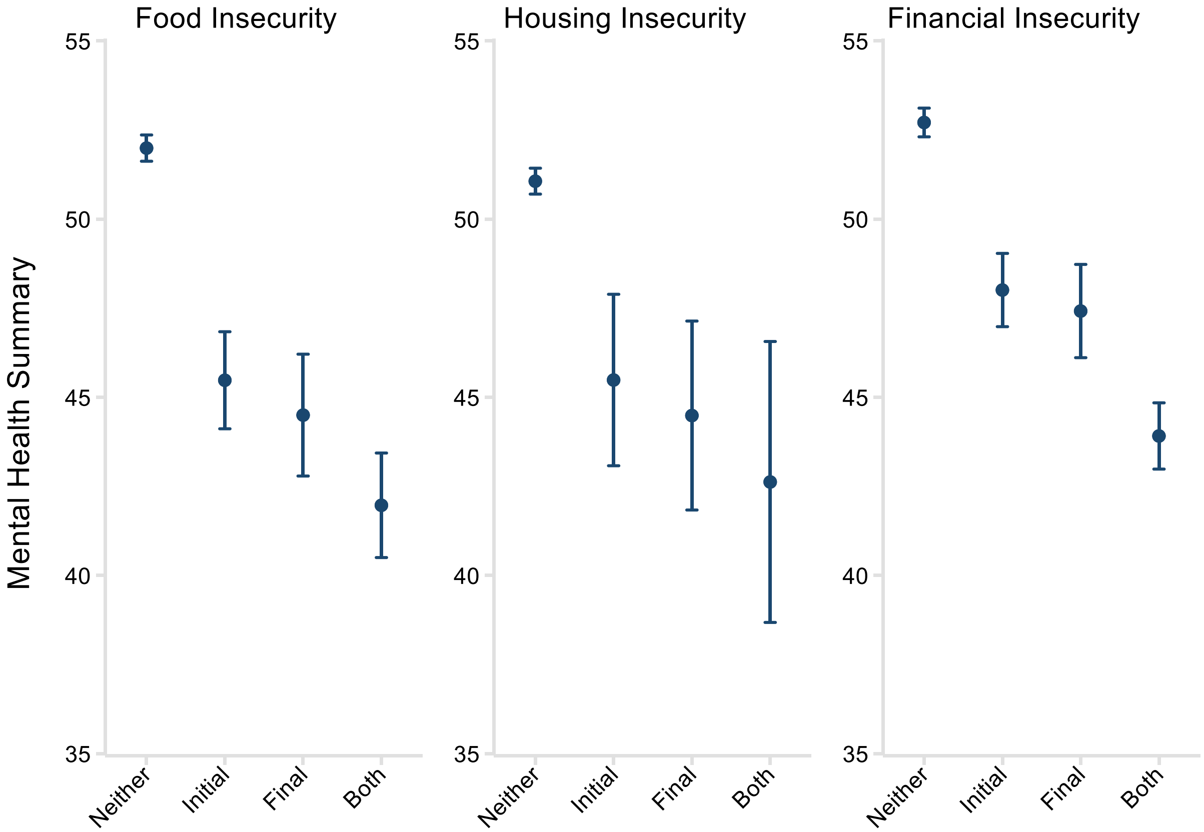
Each point represents the mean PROMIS-29 physical health summary score for the subgroup of respondents indicated on the horizontal axis. The tails represent the 95% CIs. On the *x*-axes, the label “Neither” indicates that the insecurity was not present at initial survey or follow-up; “Initial,” the insecurity was present at the initial survey and resolved over time; “Final,” the insecurity was not present initially but developed over time; and “Both,” the insecurity was present at both surveys over time.

Table 16. Mental Health Summary as a Function of the Persistence of Social Determinants of Health

		Mental health summary T-score, mean (95% CI)			
Social determinant	**Group**	**Neither wave**	**Wave 1 only**	**Wave 3 only**	**Both waves**
Food insecurity	All participants	52.0 (51.6-52.4)	45.5 (44.1-46.8)	44.5 (42.8-46.2)	42.0 (40.5-43.4)
Housing insecurity	All participants	51.1 (50.7-51.4)	45.5 (43.1-47.9)	44.5 (41.8-47.1)	42.6 (38.7-46.6)
Financial insecurity	All participants	52.7 (52.3-53.1)	48.0 (47.0-49.0)	47.4 (46.1-48.7)	43.9 (43.0-44.8)
Food insecurity	Female	51.6 (51.1-52.0)	44.4 (42.7-46.1)	44.1 (42.1-46.2)	41.3 (39.5-43.0)
	Male	52.8 (52.2-53.4)	47.5 (45.2-49.8)	46.0 (42.9-49.0)	44.2 (41.3-47.0)
Housing insecurity	Female	50.6 (50.1-51.0)	44.5 (41.5-47.6)	42.7 (39.7-45.7)	39.8 (34.1-45.5)
	Male	52.0 (51.4-52.7)	47.3 (43.4-51.2)	52.9 (47.2-58.6)	45.2 (39.8-50.6)
Financial insecurity	Female	52.3 (51.8-52.8)	47.7 (46.5-49.0)	47.0 (45.4-48.5)	43.4 (42.3-44.5)
	Male	53.5 (52.8-54.2)	48.7 (46.8-50.5)	48.6 (46.2-50.9)	45.4 (43.7-47.1)
Food insecurity	Non-White	51.6 (50.8-52.4)	46.1 (43.3-48.9)	44.2 (41.0-47.5)	42.5 (39.4-45.6)
	White	52.1 (51.6-52.5)	45.2 (43.6-46.7)	45.2 (43.1-47.3)	41.9 (40.2-43.5)
Housing insecurity	Non-White	50.8 (50.0-51.6)	44.4 (40.5-48.3)	42.3 (35.3-49.3)	37.8 (30.8-44.8)
	White	51.1 (50.7-51.5)	46.1 (43.0-49.2)	45.4 (42.5-48.3)	44.9 (40.1-49.6)
Financial insecurity	Non-White	52.9 (51.9-53.9)	48.5 (46.8-50.3)	48.9 (46.6-51.3)	43.9 (42.1-45.7)
	White	52.7 (52.2-53.1)	47.7 (46.5-49.0)	46.7 (45.1-48.3)	43.9 (42.8-45.0)
Food insecurity	Non-Hispanic	52.0 (51.6-52.4)	45.4 (43.9-46.8)	44.9 (43.1-46.7)	41.9 (40.3-43.4)
	Hispanic	51.4 (50.0-52.8)	46.6 (42.2-50.9)	40.3 (33.0-47.6)	42.9 (38.4-47.5)
Housing insecurity	Non-Hispanic	51.1 (50.7-51.5)	45.1 (42.4-47.7)	44.5 (41.8-47.1)	43.0 (38.7-47.3)

		Mental health summary T-score, mean (95% CI)			
Social determinant	**Group**	**Neither wave**	**Wave 1 only**	**Wave 3 only**	**Both waves**
	Hispanic	50.2 (48.9-51.6)	47.7 (41.2-54.2)	—	40.5 (30.6-50.4)
Financial insecurity	Non-Hispanic	52.7 (52.3-53.1)	48.0 (46.9-49.1)	47.4 (46.0-48.8)	44.0 (43.0-45.0)
	Hispanic	53.1 (51.4-54.7)	48.7 (45.5-51.8)	48.1 (44.2-52.0)	42.3 (39.4-45.2)
Food insecurity	No college	51.1 (50.6-51.6)	44.8 (43.1-46.5)	43.2 (41.1-45.2)	41.5 (39.7-43.3)
	College	52.9 (52.4-53.5)	46.7 (44.5-48.9)	47.5 (44.4-50.6)	42.9 (40.4-45.4)
Housing insecurity	No college	49.9 (49.4-50.4)	46.5 (43.6-49.4)	43.1 (39.4-46.7)	42.6 (37.7-47.5)
	College	52.4 (51.9-52.9)	43.3 (39.0-47.6)	46.0 (42.2-49.8)	43.6 (36.6-50.5)
Financial insecurity	No college	51.8 (51.2-52.3)	47.9 (46.7-49.1)	46.9 (45.2-48.5)	43.4 (42.2-44.5)
	College	53.6 (53.0-54.2)	48.3 (46.4-50.1)	48.3 (46.2-50.3)	44.9 (43.4-46.5)

Figure 8. Relationship of Mental Health to Social Determinants of Health Over Time



Each point represents the mean PROMIS-29 mental health summary score for the subgroup of respondents indicated on the horizontal axis. The tails represent the 95% CIs. On the *x*-axes, the label "Neither" indicates that the insecurity was not present at initial survey or follow-up; "Initial," the insecurity was present at the initial survey and resolved over time; "Final," the insecurity was not present initially but developed over time; and "Both," the insecurity was present at both surveys over time.

AIM 8: WHAT IS THE IMPACT OF THE COVID-19 PANDEMIC ON PATIENT OUTCOMES?

Overview

This analysis sought to explore the association between self-perceived personal and community changes as a result of the COVID-19 pandemic and health among vulnerable primary care patients experiencing MCCs independent of the IBH-PC intervention. In contrast, aim 5 examined the impact of the pandemic upon the intervention's effectiveness.

Study Outcomes

The outcome measures were the PROMIS-29 physical and mental health summary scores, GAD-7, PHQ-9, and DASI, collected at both baseline (before COVID-19) and follow-up (during COVID-19).

Covariates

COVID-19 burden was assessed using 4 questions (Table 11). The primary predictor was perceived personal change as a result of COVID-19, measured on a 5-point Likert scale. Secondary predictors were perceived community change as a result of COVID-19 (also on a 5-point Likert scale) as well as a personal diagnosis of or a close contact being diagnosed with COVID-19 (each coded as binary variables).

For each of the 5 health outcomes, the respective baseline measure of health was included as a covariate in the model, effectively assessing change in health.

Potential confounders included participant age, sex, race, ethnicity, marital status, employment status, annual household income, education level, and number of qualifying chronic conditions; the existence of financial, housing, or food insecurity; county-level estimates of unemployment; county-level incidence of COVID-19; and Social Deprivation Index score.

Information on changes in income; disruptions in schooling; delay of medical care; and loss of employment, childcare, and insurance as a result of COVID-19 was collected as binary yes-or-no questions for descriptive purposes.

Sample Size Calculations and Power

The sample size was fixed by the needs of aim 1.

Data Collection and Sources

See aims 1, 2, and 5.

Analytical and Statistical Approaches

We summarized patient information by the 5 levels of self-perceived personal and community changes as a result of the COVID-19. χ^2 tests and analysis of variance were used to assess bivariate relationships. Relative differences in health outcomes were calculated between the highest and lowest quintiles of self-perceived personal and community changes as a result of COVID-19. Multivariable regression models accounting for baseline mental and physical health were used to assess how self-perceived personal and community changes as a result of COVID-19 were associated with mental and physical health during the COVID-19 pandemic. Linear regression was used to assess perceived personal and community changes and mental health; however, the bivariate relationship between self-perceived personal and community changes as a result of COVID-19 and physical health violated the linearity assumption for linear regression, so we dichotomized self-perceived personal and community changes as a result of COVID-19 for physical health ("Not at all," "A little," "Some," vs "A lot," "Everything is different now") and used logistic regression. A likelihood ratio test was performed between the clustered model and the nested model and found not to be significant. Clustering was therefore deemed not necessary for this particular aim.

Results

The sample for this analysis included 2213 participants. The mean age was 64 years. The majority of participants identified as women (66%), non-Hispanic (93%), White (78%), low-income (50%), unmarried (53%), and unemployed or retired (71%). The median (IQR) number of chronic conditions was 4 (3-5). The most common chronic conditions were heart disease (86%), chronic pain (84%), mood disorder (63%), and diabetes (44%). A total of 22% of participants reported a decrease in income as a result of COVID-19, 5% reported losing their jobs, 4% reported that they could not get health care, 2% reported losing their health insurance, 1% reported interruption of schooling, and 1% reported loss of childcare. The mean PROMIS-29 mental health summary score (51) was similar to that of the average US population while the physical health summary score (45) was below average. The average GAD-7 and PHQ-9 scores and metabolic equivalent tasks were 4, 6, and 6, respectively. Demographic information, neighborhood characteristics, and outcomes varied by self-perceived personal and community changes as a result of COVID-19. Importantly, the difference in mental and physical health scores ranged dramatically between the highest and lowest quintiles of self-perceived personal and community changes as a result of COVID-19 (Table 17).

In most cases, the PROMIS-29 mental health summary, GAD-7, and PHQ-9 scores worsened with each increase of perceived personal and community change as a result of COVID-19. Multivariable models found significant negative impacts of self-perceived personal and community changes as a result of COVID-19 on all 3 measures of mental health. A personal diagnosis of COVID-19 was also significantly related to a decrease in mental health functioning as measured by the PROMIS-29 but not with worsening depression or anxiety. A family diagnosis of COVID-19 was not associated with any measure of mental health.

Personal changes as a result of COVID-19 were related to a decrease in physical health summary scores, with borderline statistical significance. No other measures of self-perceived personal and community changes as a result of COVID-19 were related to either measure of physical health (Table 18).

Table 17. Characteristics of Participants at Follow-Up, Stratified by Self-Perceived Personal Changes as a Result of COVID-19

		Self-perceived personal changes as a result of COVID-19[a]				
	Overall (N = 2213)[a]	**Not at all (n = 130)**	**A little (n = 340)**	**Some (n = 622)**	**A lot (n = 599)**	**Everything (n = 522)**
Demographic information						
Age, mean (SD), y	64 (13)	64 (13)	64 (13)	64 (13)	63 (13)	63 (13)
Female sex, No. (%)	1462 (66)	86 (67)	208 (61)	395 (64)	395 (66)	378 (72)
Male sex, No. (%)	745 (24)	43 (33)	131 (39)	227 (36)	203 (34)	141 (27)
Other sex, No. (%)	6 (0)	1 (1)	1 (0)	0 (0)	1 (0)	3 (1)
White race, No. (%)	1692 (78)	91 (71)	281 (84)	497 (81)	470 (80)	353 (70)
Hispanic, No. (%)	163 (7)	13 (10)	21 (6)	39 (6)	40 (7)	50 (10)
Low household income (<$30 000), No. (%)	1102 (50)	77 (62)	161 (48)	280 (46)	275 (47)	309 (61)
Married or living as married, No. (%)	1020 (46)	54 (42)	164 (49)	291 (47)	297 (50)	214 (41)
Employed, homemaker, or student, No. (%)	634 (29)	41 (32)	101 (30)	194 (31)	174 (29)	124 (24)
Education (less than a college degree), No. (%)	1156 (47)	87 (70)	182 (55)	287 (47)	293 (50)	307 (60)
Number of chronic conditions, median (IQR)	4 (3-5)	4 (3-5)	4 (3-5)	4 (3-5)	4 (3-5)	4 (3-6)
Neighborhood characteristics, mean (SD)						
Census tract Social Deprivation Index[b]	53 (28)	60 (27)	51 (26)	50 (28)	52 (22)	58 (29)
County COVID-19 prevalence rate	189 (259)	221 (240)	226 (306)	174 (260)	177 (248)	186 (237)
County COVID-19 incidence rate	1367 (1311)	1760 (1307)	1460 (1351)	1225 (1289)	1272 (1301)	1488 (1293)
County COVID-19 mortality rate	33 (41)	36 (42)	32 (37)	31 (38)	32 (42)	39 (45)
County unemployment rate	8 (5)	7 (4)	8 (5)	8 (5)	8 (5)	8 (5)
Other COVID-19 changes, No. (%)						
Income decreased	476 (22)	20 (15)	50 (15)	118 (19)	145 (24)	143 (27)

		Self-perceived personal changes as a result of COVID-19[a]				
	Overall (N = 2213)[a]	**Not at all (n = 130)**	**A little (n = 340)**	**Some (n = 622)**	**A lot (n = 599)**	**Everything (n = 522)**
Lost job	106 (5)	3 (2)	8 (2)	19 (3)	39 (7)	37 (7)
Couldn’t get medical care	95 (4)	1 (1)	9 (3)	26 (4)	35 (6)	24 (5)
Lost health insurance	38 (2)	3 (2)	1 (0)	7 (1)	11 (2)	16 (3)
Lost childcare	17 (1)	0 (0)	0 (0)	2 (0)	10 (2)	5 (1)
School interrupted	31 (1)	0 (0)	3 (1)	6 (1)	9 (2)	13 (3)
Outcomes, mean (SD)						
Mental health summary score[b]	51 (8)	53 (10)	53 (9)	52 (8)	50 (9)	48 (9)
GAD-7	4 (5)	3 (5)	3 (4)	3 (4)	5 (5)	5 (6)
PHQ-9	6 (6)	5 (6)	4 (5)	5 (5)	6 (6)	7 (6)
Physical health summary score[b]	45 (10)	45 (10)	47 (10)	46 (10)	46 (10)	43 (10)
DASI (METs)[b]	6 (2)	6 (2)	7 (2)	7 (2)	7 (2)	6 (2)

Abbreviations: DASI, Duke Activity Status Index; GAD-7, Generalized Anxiety Disorder-7; MET, metabolic equivalent task; PHQ-9, Patient Health Questionnaire-9.

[a] Rows may not sum to 100.0% due to rounding. Some cells may have a slightly smaller n value than the rest of the column due to missing data.

[b] This outcome has a reverse scale in which higher scores are better.

Table 18. Multiple Regression Coefficients of Self-Perceived Personal and Community Changes as a Result of COVID-19 on Health Outcomes

	Mental health summary score[a]	GAD-7	PHQ-9	Physical health summary score[a]	METs[a]
	Standardized regression coefficients (95% CI)				
Personal changes as a result of COVID-19	−0.55 (−0.72 to −0.37)[b]	0.28 (0.16-0.39)[b]	0.35 (0.22-0.47)[b]	−0.44 (−0.88 to 0.00)[c]	−0.05 (−0.16 to 0.05)
Community changes as a result of COVID-19	−0.57 (−0.79 to −0.36)[b]	0.28 (0.14-0.42)[b]	0.33 (0.17-0.48)[b]	−0.20 (−0.67 to 0.28)	−0.02 (−0.13 to 0.09)
Personal COVID-19 diagnosis	−1.32 (−2.58 to −0.11)[c]	0.44 (−0.36 to 1.26)	0.63 (−0.26 to 1.52)	−0.49 (−1.82 to 0.84)	−0.4 (−0.33 to 0.26)
Family COVID-19 diagnosis	0.39 (−0.40 to 1.19)	0.13 (−0.39 to 0.65)	−0.35 (−0.93 to 0.22)	0.30 (−0.56 to 1.16)	0.11 (−0.89 to 0.31)

Abbreviations: GAD-7, Generalized Anxiety Disorder-7; MET, metabolic equivalent task; PHQ-9, Patient Health Questionnaire-9.

[a] Higher scores indicate better health.

[b] *P* < .01.

[c] *P* < .05.

AIM 9: WHAT IS THE IMPACT OF THE COVID-19 PANDEMIC ON PRACTICE OUTCOMES?

Overview

This analysis sought to explore the association between practice and community changes as a result of COVID-19 and the degree of integration achieved by the practices independent of the IBH-PC intervention. In contrast, aim 5 examined the impact of the pandemic upon the intervention's effectiveness.

Study Outcomes

The outcome variable was total PIP score at the final survey (after COVID-19).

Covariates

Items assessed at the final survey (after COVID-19) were the predictors. These included perceptions of changes in the number of patient visits (total, BH, face-to-face) and changes in staffing (medical professionals, nurses, BHPs, and nonclinical staff) measured on a 3-point scale of "less," "the same," or "more." Perceptions of the degree to which COVID-19 impacted the delivery of IBH and the degree to which COVID-19 affected the ability of the practice to improve IBH (the purpose of the original clinical trial) were measured on a 5-point scale of "greatly impeded," somewhat impeded," "no impact," "somewhat enhanced," or "greatly enhanced." An open-ended item asked for a description of the influence COVID-19 had on the delivery of IBH. County-level COVID-19 prevalence and incidence based on state and local health agency reports were gathered from *New York Times* data.[122] Duration of the pandemic at the time of the survey varied for each practice and was recorded as weeks since March 18, 2020.

Sample Size Calculations and Power

The sample comprised all participating sites described earlier.

Data Collection and Sources

See aim 5.

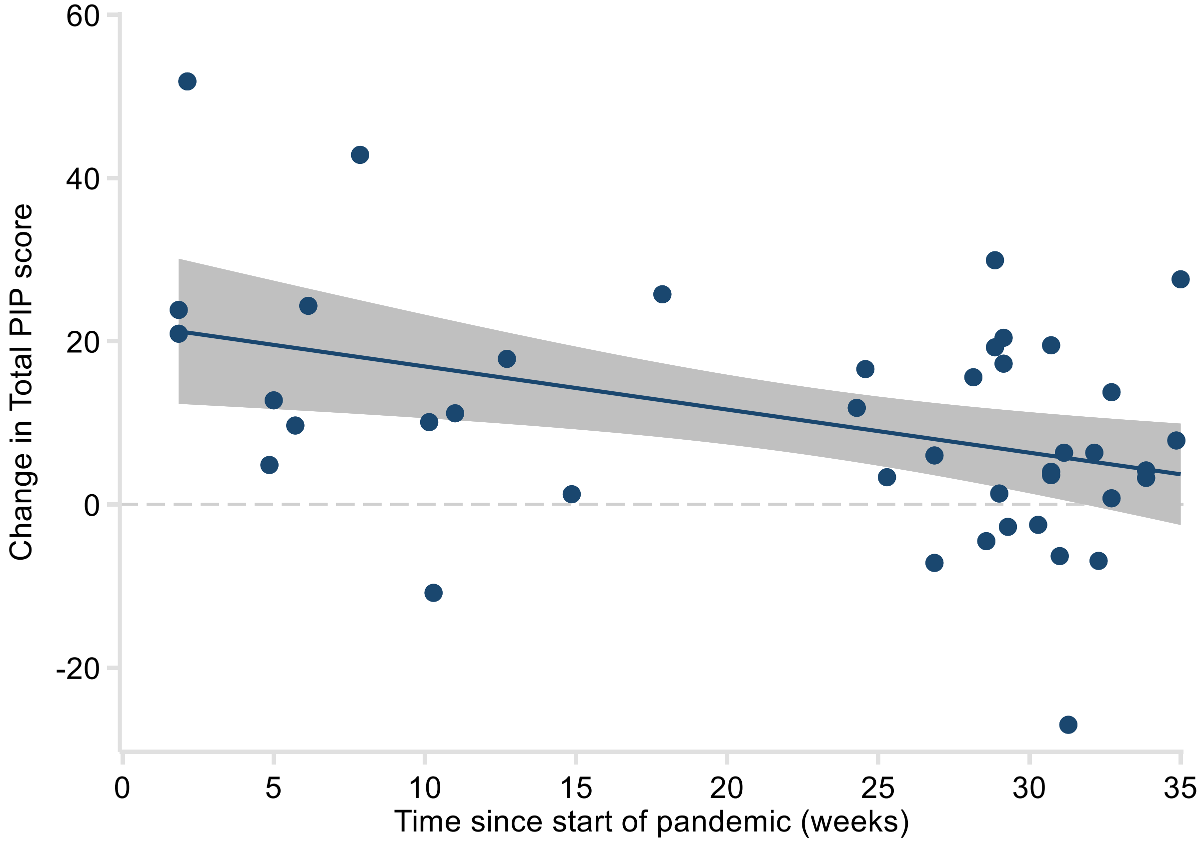
Analytical and Statistical Approaches

Multivariable regression models accounting for baseline total PIP score were used to assess how each measure of COVID-19 burden was associated with changes in IBH.

Results

The sample included 42 primary care practices in the United States (Table 6). We found that over the course of 3 years, PIP scores improved from a mean of 59.4 to a mean of 69.6, a change of 10.2 points (95% CI, 5.7-14.7; *P* < .001). The number of weeks into the pandemic was associated with a decrease in the final total PIP score of 0.4 (95% CI, −0.70 to −0.04; *P* = .028). Figure 9 shows each practice's final PIP score plotted against the number of weeks since March 18, 2020.

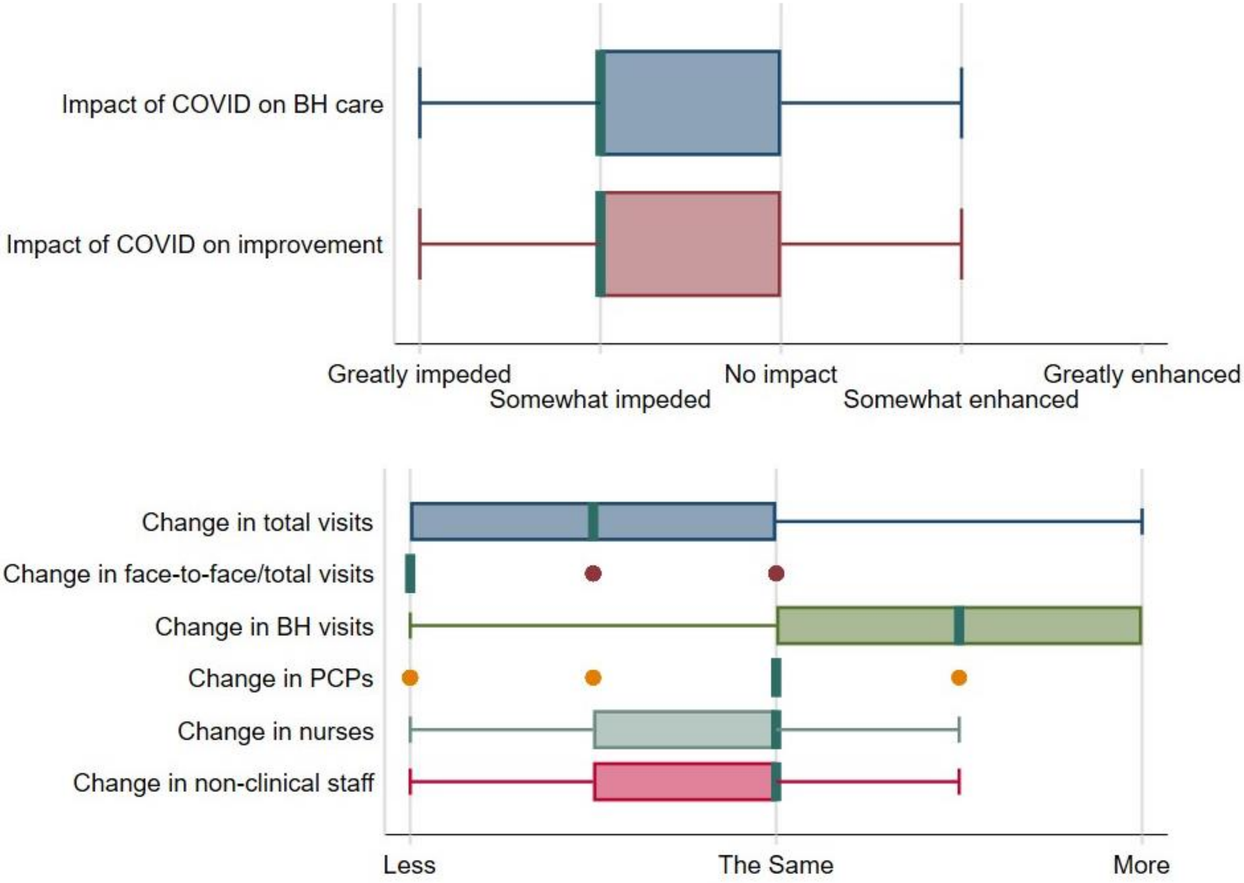
Figure 9. Change in Practice Integration Over Time



Abbreviation: PIP, Practice Integration Profile.
Each dot represents the median total PIP score for 1 practice measured at the final survey (after COVID-19). The line was fit by unadjusted linear regression. The gray area represents the 95% CI around the line.

Staff perceptions were that the COVID-19 pandemic tended to impede the delivery of BH care as well as QI efforts around IBH. It was also thought to be responsible for reductions in total visits to the practice, face-to-face visits, nursing staff, and nonclinical staff, but less so for PCP staffing. Behavioral health visits were thought to have increased (Figure 10). None of these values was associated with a change in IBH as measured by total PIP score. Neither the county-specific prevalence nor incidence of COVID-19 was associated with changes in total PIP score.

Figure 10. Perceived Effects of COVID-19 on Practice Characteristics



Abbreviations: BH, behavioral health; PCP, primary care professional.
Each box-and-whisker plot represents a distribution of staff perceptions. The darkened line of each box represents the median value with the box including half the observations from the 25th to the 75th percentile. Note that when all values are the same in a box plot (excluding outliers), the result is a single line representing the median. The whiskers represent the range with outliers noted by individual symbols.

AIM 10: WHAT ARE THE MODERATORS AND MEDIATORS OF BURNOUT AMONG PRIMARY CARE PROFESSIONALS AND STAFF?

Note: *Parts of the material presented in this section previously appeared in the following peer-reviewed publication: Clifton J, Bonnell L, Hitt J, et al. Differences in occupational burnout among primary care professionals.* J Am Board Fam Med. *2021;34(6):1203-1211. doi:10.3122/jabfm.2021.06.210139*

Overview

The purpose of this analysis was to understand the variations in burnout domains (emotional exhaustion, depersonalization, and personal accomplishment) among PCPs across the United States.

Study Outcomes

Degree of occupational burnout was assessed using an adaptation of the Maslach Burnout Inventory, a validated, 9-item measure with 3 domains: depersonalization, emotional exhaustion, and personal accomplishment.

Covariates

The predictor was professional role, classified as PCP, medical resident, BHP (psychologists, social workers, and counselors), nurse, other clinical staff (eg, medical assistants and care and referral coordinators), and nonclinical professionals. Potential confounders included personal, practice-level, and county-level factors. Personal variables included age, sex, race, ethnicity, education level, and years working in the field. Practice variables included setting (academic, hospital, community health center, private); nonprofit status; specialty (family medicine, internal medicine, mixed); professional full-time equivalents; tenure of BH services; degree of IBH; training of medical residents; patient panel size; number of patient encounters; and proportion of patients on Medicare. Publicly available, county-level variables were matched with the location of the practice. These included the county's region, urban or rural status, population density, Social Deprivation Index score, and median age, sex, race, ethnicity, income level, and level of education.

Sample Size Calculations and Power

The sample size was determined by the number of staff at the participating practices at the time of the survey. All were invited to participate. An estimated 1100 practice members received at least 1 email invitation from a practice delegate (practice manager, medical director, or other professional) to complete an anonymous online survey.

Data Collection and Sources

A delegate from each of the 42 participating practices (eg, a practice manager, BH or medical director) was asked to distribute a survey invitation email to all practice staff. The online, anonymous survey was distributed separately from the other study assessments between June 2018 and October 2019. The survey was anonymous to reduce practice members' concerns about their supervisors or other practice colleagues learning their opinions of a sensitive workplace issue. To maintain anonymity, we relied on the practice delegate to distribute the survey, and thus the target sample of 1100 is an estimate. No compensation was provided to survey respondents.

Analytical and Statistical Approaches

We used multivariable linear regressions to assess the relationships of each domain to each professional role. Each role was dummy-coded and included in each model. Practice was included as a random intercept to account for the correlation of personal-level measures within worksites. If a potential covariate changed the coefficient of any professional role on burnout by more than 10% in a model containing only professional role as the predictor, it was included in the final model. Adapted Maslach Burnout Inventory subscales that were missing items were removed from the related analyses. Fifty-two participants were missing at least 1 item. Kruskal-Wallis tests were performed to compare continuous data across groups. χ^2 tests were performed to compare categorical data across groups. All tests were 2-tailed, with $P < .05$ as the statistical threshold for significance. All analyses were conducted using Stata, version 16, software.

Results

A total of 687 participants from 41 practices participated. On average, 7 participants (range, 2-44) from each practice completed the anonymous survey. More than half identified as non-Hispanic White women under 45 years of age, with more than 10 years of experience in their occupation. Nearly half of participants completed graduate school. Primary care professionals made up 24% of the sample (Table 19).

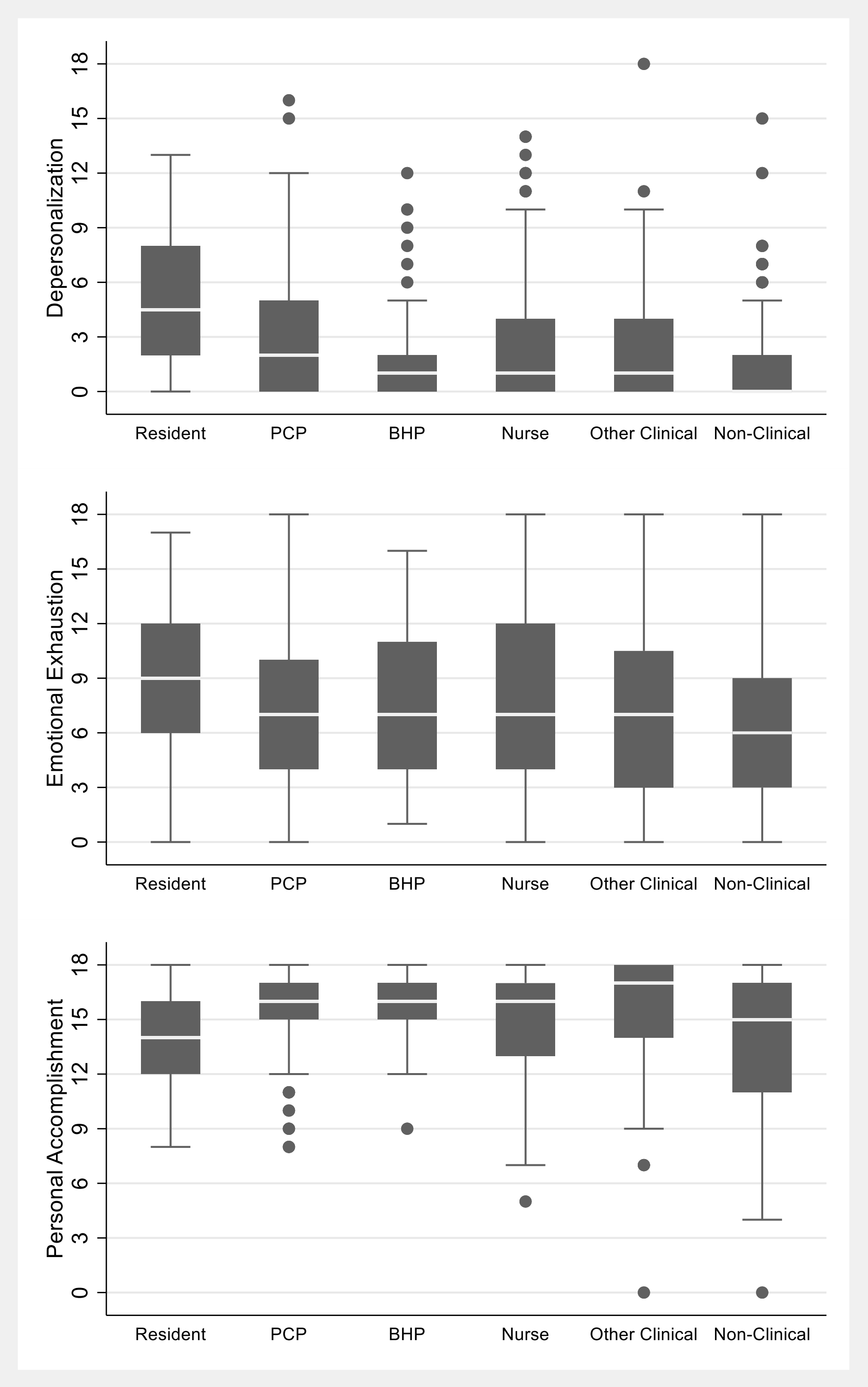
Table 19. Demographic Characteristics of Burnout Survey Participants

Characteristics	Primary care professionals (n = 167)	Medical residents (n = 56)	Behavioral health professionals (n = 80)	Nurses (n = 122)	Other clinical staff (n = 108)	Nonclinical staff (n = 152)	All (N = 685)
Age in years, %							
Under 25	0.0	0.0	1.3	5.7	10.2	10.5	5.1
25-34	30.9	85.7	28.8	23.0	37.0	24.3	28.8
35-44	24.2	12.5	36.3	29.5	28.7	23.0	27.0
45-54	18.4	1.8	18.8	18.9	13.9	21.1	18.4
55-64	20.2	0.0	12.5	18.9	10.2	18.4	17.1
≤65	6.3	0.0	2.5	4.1	0.0	2.6	3.7
Female sex, %	65.5	57.1	81.0	90.9	87.8	91.3	81.0
White race, %	81.4	73.2	90.8	67.2	66.7	67.9	74.8
Hispanic or Latino ethnicity, %	5.5	3.6	5.2	14.1	18.6	26.4	13.6
Attended graduate school, %	92.4	100	97.5	12.9	10.2	7.9	47.0
Years in occupation, mean (SD)	13.0 (12.1)	1.9 (0.9)	7.9 (8.2)	12.4 (12.2)	8.0 (7.9)	7.9 (8.6)	10.4 (10.7)

Burnout scores varied by role and domain (Figure 11). Unadjusted analyses revealed significant differences by role for all 3 domain scores (Table 20). Medical residents had the highest levels of burnout within each domain. Nonclinical staff experienced the least depersonalization and emotional exhaustion. Primary care professionals had the best personal accomplishment score.

Multilevel linear regression analyses provided further evidence that the adapted Maslach Burnout Inventory subscales differed by role. Of the 35 potential covariates tested, only race, education, age, years working in the field, and practice training site status altered the association between role and burnout by more than 10% and were included in the final models. Adjusting for the 5 confounders reduced the strength of association for all outcomes, but many of the differences across roles remained significant in all 3 domains. Medical residents still had the most burnout in all 3 domains, followed by PCPs for depersonalization, nurses for emotional exhaustion, and nonclinical staff for personal accomplishment (Table 21).

Figure 11. Burnout by Job Role for Each Domain



Abbreviations: BHP, behavioral health professional; PCP, primary care professional.
Scores range from 0 to 18. For both the depersonalization and emotional exhaustion domains, higher scores indicate worse burnout; for the personal accomplishment domain, lower scores are worse.
Reproduced from Clifton et al. *J Am Board Fam Med*. 34(6), 1203-1211. Reprinted with permission from *J Am Board Fam Med* (Copyright ©2021). All Rights Reserved.

Table 20. Burnout by Job Role

Maslach Burnout Inventory domain[a]	**All**	**Primary care professional**	**Medical resident**	**Behavioral health professional**	**Nurse**	**Other clinical staff**	**Nonclinical staff**	***P* value**
Depersonalization (n = 633)								
Median (IQR)	1 (0-4)	2 (0-5)	4.5 (2-8)	1 (0-2)	1 (0-4)	1 (0-4)	0 (0-2)	<.001
Moderate or severe (≥4), %	28	34	61	18	29	26	15	<.001
Exhaustion (n = 685)								
Median (IQR)	6 (3-9)	7 (4-10)	9 (6-12)	7 (4-11)	7 (4-12)	7 (3-11)	6 (3-9)	.002
Moderate or severe (≥7), %	52	56	70	51	52	52	44	.039
Personal accomplishment (n = 634)								
Median (IQR)	16 (13-17)	16 (15-17)	14 (12-16)	16 (15-17)	16 (13-17)	17 (14-18)	15 (11-17)	<.001
Moderate or severe (≤14), %	33	15	46	21	35	29	46	<.001

[a] Domain scores range from 0 to 18. Higher scores for depersonalization and emotional exhaustion domains indicate worse burnout; lower scores for the personal accomplishment domain are worse.

Adapted from Clifton et al. *J Am Board Fam Med*. 34(6), 1203-1211. Used with permission from *J Am Board Fam Med* (Copyright ©2021). All Rights Reserved.

Table 21. Multivariable Regression Analyses of Burnout by Job Role With Covariates[a]

	Depersonalization		Exhaustion		Personal accomplishment	
Role	***b***	**95% CI**	***b***	**95% CI**	***b***	**95% CI**
Resident (base case)	—		—		—	
PCP	−1.36[b]	−2.47 to −0.26	−0.60	−2.11 to 0.91	1.21[b]	0.18-2.25
BHP	−2.73[b]	−3.89 to −1.56	−0.99	−2.59 to 0.62	1.49[b]	0.41-2.58
Nurse	−1.80[b]	−3.24 to −0.36	−0.47	−2.37 to 1.43	1.47[b]	0.13-2.81
Other clinical staff	−2.59[b]	−4.03 to −1.16	−1.20	−3.11 to 0.71	1.93[b]	0.61-3.26
Nonclinical staff	−2.98[b]	−4.42 to −1.54	−2.1[b]	−4.02 to −0.20	0.17	−1.16 to 1.51

Abbreviations: *b*, unstandardized independent regression coefficient when accounting for all the other variables in the model; BHP, behavioral health professional; PCP, primary care professional.

[a] Multivariable regression analyses with practice-level clustering, adjusted for race, graduate school attendance, age, years working, and training status. Higher scores for depersonalization and emotional exhaustion indicate worse burnout; lower scores for personal accomplishment are worse.

[b] $P < .05$.

DISCUSSION

Summary of Results

This was a negative study. The intervention tested was not effective in changing any of the primary or secondary patient outcomes studied. There was weak evidence that the intervention induced a small change in one aspect of practice integration (workflow), consistent with the process management orientation of the intervention. The costs of implementing the intervention were modest at a median of $20 726 per practice. Although the COVID-19 pandemic had substantial impacts on patient outcomes and practice integration, we could not detect an interaction between the pandemic and the intervention. In other words, there is no evidence that the pandemic is responsible for the intervention's lack of effectiveness. We do not know if local responses to the pandemic (eg, lockdowns, loss of employment, loss of childcare) muted the effect of the intervention.

Mixed-methods analyses identified potential contextual factors that may have impeded or supported practice integration, including leadership commitment, funding considerations, workflow and communication systems, and professionals' attitudes toward integration and team-based health care.

The IBH-PC toolkit is just 1 possible intervention to accelerate the integration of BH and primary care services. Although ineffective in this study, there are many approaches that may be useful in improving the integration of care, the benefits of which are already confirmed.

Additional analyses of this dataset demonstrated valuable findings. We learned that chronically ill patients in practices with greater levels of integration experienced better health than chronically ill patients in practices with lower levels of integration regardless of whether the practice was in the active or control group. This finding was true at both baseline and follow-up and when analyzed both cross-sectionally and longitudinally. Though this association is not evidence of causality, it is consistent with prior literature and with our interpretation that the null hypothesis we observed in the main study were a function of an ineffective toolkit, poor implementation of the toolkit, or contamination of the control group, not a failure of integration per se.

Chronically ill patients with housing, income, or food insecurities experienced poorer physical and mental health particularly if those insecurities were observed at both measurement time periods instead of just 1.

Among adults with MCCs, self-perceived personal and community changes as a result of the COVID-19 pandemic were associated with poor health. This vulnerable population may have been particularly susceptible to the negative effects of COVID-19. Our analysis established a baseline of epidemiological data on COVID-19 burden and health among primary care patients with MCCs.

Practices generally perceived reductions in total patient and face-to-face visits and increases in IBH visits during the pandemic, but these were unrelated to changes in integration. Perceived staffing of all types fell, and COVID-19 was generally seen as impeding IBH delivery and IBH improvement or sustainment. None of these factors was associated with integration as measured by the PIP. The intensity of COVID-19, as measured by county prevalence and incidence, was not associated with integration; however, the number of weeks into the pandemic was associated with a significant decrease in integration at the final assessment.

Before COVID-19, most professionals in these practices were experiencing occupational burnout. Clinicians, especially medical residents and PCPs, may have been experiencing aspects of burnout more intensely than their nonclinical colleagues. Based on these variations, interventions to mitigate burnout may need to be tailored by professional role.

Results in Context and Lessons Learned

The literature indicates that IBH is effective for patients in need of BH care, resulting in better self-management and management of care, higher patient satisfaction and emotional well-being, lower mean depressive scores, decreased wait times for treatment,[123] and increased likelihood of engaging in care and attending visits. Particular attention has been paid to positive reviews of the Chronic Care Model, although some studies were inconclusive.[124] Nonetheless, there is substantial evidence that primary care clinics have had difficulty integrating BH services at an effective level.

Several studies have examined implementation strategies for effective integration, identifying that establishing continuum-of-care pathways; attending to patient transitions (eg, referrals); co-locating a workforce; engaging physicians; establishing a mission focus; and having motivation to change, care management techniques, and team-based approaches are all important. Toolkits, as a particular kind of implementation strategy for integration, have been successful in some settings and are generally well received; however, they are not fully supported by robust study designs reporting on patient outcomes.

The intervention applied here was somewhat less costly than some other implementation strategies designed to improve mental health services. For instance, Lang and Cornell assessed a yearlong, statewide dissemination of trauma-focused cognitive behavioral therapy in 10 child outpatient clinics at community mental health agencies in Connecticut through learning collaboratives. The average incremental implementation cost was $89 575, with a range of $34 697 to $130 063. A study of implementing primary care and mental health integration at 8 clinics in 2 Veterans Affairs medical center networks found that, over the course of 28 months, the cost of salary support for facilitation activities alone was $236 263 in network A and $208 314 in network B, each of which had 4 clinics to support.

The major lessons learned by the research team were (1) patient-centered data collection can bypass the difficult processes of collecting outcomes in the clinic setting, allowing for remarkably complete data; (2) the intervention we tested, the IBH-PC toolkit, was not intense or precise enough to garner the change needed to improve patient outcomes during a pandemic beyond what was done by control practices; but (3) the mixed-methods (aim 3) and post hoc (aim 6) analyses suggested a great need for, and the value of, better BH services in primary care settings.

Potential to Affect Health Care Decision-Making

These data suggest that the particular intervention tested should not be widely promoted or disseminated as applied in this study; however, they should not be used to dismiss other efforts to address the BH needs of primary care patients. A large body of research supports the IBH model.

Generalizability

The patients and practices studied are highly representative of primary care in the United States, except for the fact that the practices were all committed to supporting a BHP on-site, something that is not a universal feature of US primary care. The descriptive analyses (aims 3, 4, and 6-10) are therefore probably not generalizable to other primary care settings. The experimental results (aims 1, 2, and 5) apply to the intervention tested and not to other maneuvers to improve BH integration.

Subgroup Analyses and Heterogeneity of Treatment Effects

We explored many prespecified subgroups of patients based on participants' personal characteristics, diagnoses, neighborhood characteristics, and their practices' characteristics. We found little convincing evidence for heterogeneity of treatment effect. The results were essentially similar in all subgroups examined.

Study Limitations and Challenges

For patient-level outcomes (aim 1), this study was quite large. It had 90% power to detect differences as small as 2.5 points in any of the 8 PROMIS-29 domain scales. Although MIDs are not specified for the PROMIS-29, MIDs for other PROMIS instruments range from 2 to 8 points. The largest difference observed in this study was 0.2 points, which is neither clinically nor statistically significant (*P* = .47). The smallest *P* value observed for any of the outcomes was .09. Although a few combinations of outcome and subgroup had nominal *P* values less than .05, given their lack of consistency, they were more likely a result of multiple comparisons and type I error than of a valid reproducible association. It is unlikely that this study was negative because of low sample size.

Subgroup analyses are limited by the smaller sample sizes involved and by the exploratory, rather than hypothesis-testing, nature of the analyses.

For practice-level analyses, however, such as the impact of the intervention on changes in PIP score (aim 2), the sample size was only 42 practices. With a standard deviation of 14.2 points, the study had only 14% power to detect the observed effect of 3.8 points with $P < .05$. It had 80% power for an effect of 12.7 points. It is nonetheless striking that the total PIP score and all 6 of the PIP subdomains moved in the anticipated direction, suggesting the intervention had a small beneficial effect on BH integration. Despite the small sample size, the intervention appeared to be particularly effective in the workflow domain.

Other aspects of the design, such as preregistration of the protocol, randomization, high levels of follow-up, and use of self-reported outcomes rather than investigator observations minimized bias and confounding. Measurement bias at the patient level seems unlikely given that patients did not know which group of the study their practice was in; however, staff who reported the degree of integration were aware of their random assignment, which raises the possibility of social desirability response bias for the practice-level assessments.

The patient population studied was older and sicker than average primary care patients but representative of patient with MCCs. The practices were unusual, however, in that they all had a strong commitment to BH care. All had on-site BHPs at the start of the trial and had developed a funding plan to support those professionals independent of the research project. It is unclear how common this scenario is in US health care today, but it is certainly not universal.

Practice-to-practice variability in the degree of integration at baseline and frequent change over time are major features of primary care and were present in this sample. For instance, 2 different practices were temporarily shut down by natural disasters; 5 suffered major work stoppages because of strikes; and several medical professionals and BHPs, managers, and other staff left in the middle of the study, causing the clinics to sometimes have lengthy periods of no or reduced services as new staff were recruited and trained. Although the

COVID-19 pandemic affected all sites, the intensity of the pandemic differed dramatically over time and among sites.

The pragmatic design of the study may have contributed to the null hypothesis. The intervention at the active sites was designed to be flexible and customizable. Although all active sites participated in the online curriculum and coaching to some degree, they determined the degree of engagement that suited their local values, resources, capacities, and needs. Practices varied in how long it took to start and complete the intervention. For practices that took longer, the opportunity to observe postintervention changes was decreased. Each practice chose different practice improvement tactics to implement, and they achieved varying levels of success in executing them. This is exactly what we would expect when introducing an intervention into any heterogeneous set of environments. This real-world, pragmatic assessment is an important strength of the study, one that allowed us to test the intervention in authentic, generalizable settings.

The pragmatic nature of this study places some limitations on the degree to which a formal process evaluation can be conducted on this complex health intervention, thereby diminishing the ability to describe fidelity to the planned intervention. Further analysis of the effect of intervention timing, use of the intervention, and the varying needs of patient populations for the intervention is warranted and is possible. The design of the study did not allow us to explore whether patients who actually received BH services fared differently than others.

Although the lost-to-follow-up rate was low (18% over 2 years), the participants lost to follow-up were significantly sicker and more socially vulnerable than participants who completed all surveys. We believe this is because patients with poorer health and greater social challenges are more likely to lose insurance (and thus leave care) and to lose housing (becoming harder to trace), to become even more sick (perhaps less willing or able to participate), and to die. Even when we assumed that the intervention was substantially more effective among lost participants, however, the treatments groups did not differ, suggesting it is very unlikely that missing data were responsible for the null hypothesis seen.

Although the control sites did not have access to the intervention itself, they were free to undertake any QI activities they deemed appropriate, and they may have used material similar to some elements of the intervention, possibly masking an intervention effect.

Did we study the right outcome? The PROMIS-29 has face validity, correlates well with other outcomes, and has been widely used across a broad range of conditions. To avoid missing an effect, we examined 19 other outcome measures across a wide range of functional domains, but none of them showed a significant effect.

Could confounders have obscured a true treatment effect? Although randomization mitigates this phenomenon, it is still possible; however, none of the potential confounders proposed on theoretical grounds proved to be an important confounder in multivariable models. Neither was the intervention associated with a consistent difference in outcomes in any of the subgroups examined.

Did we study the right comparator? A randomized design using sites without BH services of any kind was considered but was unfeasible because funds to provide BH services were not available. When the study was designed, it appeared that the field of primary care was committed to BH integration, at least to the degree of having BHPs co-located with PCPs. We therefore used colocation without the intervention as the comparator condition. This unfortunately means that although this study addresses the utility of the specific intervention to achieve high levels of integration, it does not directly address whether practices with simple colocation are better than primary care with no BH services on-site.

Improving quality of care for BH concerns is difficult. It is possible, even likely, that the intervention as used in the active sites was too weak to have the desired effect. Most practices were not able to afford a full complement of BHPs adequate for their total patient panel's needs, and it's possible that patients surveyed for the study did not have adequate access to BH services, even at active sites. Health care professionals were often overwhelmed by acute problems that interfered with their capacity to provide IBH. Some clinicians never fully endorsed the concepts of IBH. Practices often had fewer team meetings and opted to apply fewer specific tactics than we anticipated. Importantly, the practices faced a remarkable array of intervening factors that limited their ability to make improvements, such as natural disasters, staffing and leadership changes, work stoppages, ownership changes, computer outages, and, of course, a global pandemic. We were often impressed by the resiliency of these health care professionals and staff in their ability to provide any level of care, even if it did not always achieve all we hoped for.

Perhaps the intervention was essentially underpowered because it relied on BHPs with a variety of training backgrounds, including both master's-prepared counselors and doctoral-level psychologists, although there are no data to indicate if either preparation is better suited to providing BH care. We believe that hiring doctoral-level BHPs and psychiatrists is not feasible in most primary care settings. Perhaps, in spite of online education, the BHPs and medical professionals in the intervention practices were ill prepared because they had originally been trained to deliver or refer to traditional mental health services and found it difficult to learn new habits and approaches required to treat primary care patients with medical and behavioral conditions. Perhaps the BHPs at these practices did not employ evidence-based practices in spite of the education provided by the study; however, it was not possible to recruit specially trained counselors to replace the existing professionals in the practices.

A major limitation of the study is that it studied only 1 possible intervention and provided little information on the possible effectiveness of other strategies for improving BH care.

The mixed-methods case study (aim 3) involved a small number of practices and is not generalizable. We interviewed a cross-section of roles represented in practices with varying characteristics; however, we did not speak with patients about their experiences with integration services. We were not able to visit 7 of the case study sites as originally planned because we had to adjust our data collection activities as a result of COVID-19. Without being on-site, we had a limited understanding of the how the physical layout of a practice influenced its workflow. In addition, being on-site would have given us the opportunity to observe aspects of the practice culture and working relationship that are not easily noticed in a virtual environment.

Although we drew from a broad array of practice structures, the implementation cost study (aim 4) collected data from only 8 practices, limiting generalizability. Data were subject to errors of memory, although the CAT was structured with multiple prompts to mitigate these.

The analyses of the effects of COVID-19 (aims 5, 8, and 9) are limited by timing; the pandemic has evolved substantially, and results from 2020 and 2021 may not apply to current experience. There was no validated measure of COVID-19 burden, though, so our questions were created by the study team. Aims 6 through 10 are nonexperimental and not well-suited to demonstrating causality, although they do yield valuable observations and raise questions not answerable with our dataset. For example, it is possible the intervention was associated with improvements in patient outcomes, but these were not captured in our analyses because they were nullified by COVID-19.

In the end, the single most important limitation to keep in mind when considering the main results of this study is that it evaluated an intervention to improve integration of BH and primary care, not whether integration or, for that matter, BH and primary care are themselves valuable.

Future Research

Future research should be directed at better ways to deliver BH care to all patients, including new interventions to improve integration with primary care. Although the impact of the IBH-PC toolkit was not as robust as expected, it should be examined to see which parts may be retained for future interventions, using different research designs or targets of change. For example, the intervention appeared to have a small impact on the workflow aspect of integration. The aim 6 findings that more practice integration is associated with better patient outcomes warrant prospective investigation in other settings.

Future designs might incorporate methods for determining which patients received BH services and whether they were evidence-based. Future practice-based research should also examine relationships related to practice engagement in the intervention steps and other factors to understand how to improve outcomes through integrated medical and BH services for patients with MCCs.

CONCLUSIONS

The specific intervention tested in this pragmatic trial was inexpensive and had only a small impact on the degree of BH integration and none on patient outcomes. We were unable to identify any important effects among numerous secondary outcomes and subgroups of participants, nor did we detect evidence that the burden of the COVID-19 pandemic was responsible for the findings. Practice leaders’ commitment, financial considerations, workflow and communication systems, and health care professionals’ perspectives on integration and team-based health care can be both supporting and impeding contextual factors. In post hoc analyses, practices that had more integration at baseline had better patient outcomes independent of the intervention.

Although this particular intervention was ineffective as reported in this study, IBH remains an attractive strategy for improving patient outcomes. We believe that pragmatic clinical trials are a viable approach to evaluating the effectiveness of interventions. The unplanned modifications and disruptions to delivering the intervention might have destroyed a traditional randomized controlled trial, but they only increased our understanding of real-world factors influencing practices’ barriers to providing integrated care.

REFERENCES

1. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA*. 2004;291(10):1238-1245. doi:10.1001/jama.291.10.1238

2. Druss BG, Walker ER. Mental disorders and medical comorbidity. *Synth Proj Res Synth Rep*. 2011;(21):1-26.

3. Kathol RG, McAlpine D, Kishi Y, et al. General medical and pharmacy claims expenditures in users of behavioral health services. *J Gen Intern Med*. 2005;20(2):160-167. doi:10.1111/j.1525-1497.2005.40099.x

4. Melek S, Norris D. *Chronic Conditions and Comorbid Psychological Disorders*. 2008. Milliman Research Report; 2008. https://www.cbhc.org/wp-content/uploads/2015/11/chronic-conditions-and-comorbid-RR07-01-08.pdf

5. Butler M, Kane RL, McAlpine D, et al. Integration of mental health/substance abuse and primary care. *Evid Rep Technol Assess (Full Rep)*. 2008;(173):1-362.

6. Knoerl R, Lavoie Smith EM, Weisberg J. Chronic pain and cognitive behavioral therapy: an integrative review. *West J Nurs Res*. 2016;38(5):596-628. doi:10.1177/0193945915615869

7. Trauer JM, Qian MY, Doyle JS, Rajaratnam SMW, Cunnington D. Cognitive behavioral therapy for chronic insomnia: a systematic review and meta-analysis. *Ann Intern Med*. 2015;163(3):191-204. doi:10.7326/M14-2841

8. Hall J, Kellett S, Berrios R, Bains MK, Scott S. Efficacy of cognitive behavioral therapy for generalized anxiety disorder in older adults: systematic review, meta-analysis, and meta-regression. *Am J Geriatr Psychiatry*. 2016;24(11):1063-1073. doi:10.1016/j.jagp.2016.06.006

9. Weitz ES, Hollon SD, Twisk J, et al. Baseline depression severity as moderator of depression outcomes between cognitive behavioral therapy vs pharmacotherapy: an individual patient data meta-analysis. *JAMA Psychiatry*. 2015;72(11):1102-1109. doi:10.1001/jamapsychiatry.2015.1516

10. Katon WJ, Lin EHB, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med*. 2010;363(27):2611-2620. doi:10.1056/NEJMoa1003955

11. Reed SJ, Shore KK, Tice JA. Effectiveness and value of integrating behavioral health into primary care. *JAMA Intern Med*. 2016;176(5):691-692. doi:10.1001/jamainternmed.2016.0804

12. Blackmore MA, Carleton KE, Ricketts SM, et al. Comparison of collaborative care and colocation treatment for patients with clinically significant depression symptoms in primary care. *Psychiatr Serv*. 2018;69(11):1184-1187. doi:10.1176/appi.ps.201700569

13. Kathol RG, Degruy F, Rollman BL. Value-based financially sustainable behavioral health components in patient-centered medical homes. *Ann Fam Med*. 2014;12(2):172-175. doi:10.1370/afm.1619

14. Crowley RA, Kirschner N; Health and Public Policy Committee of the American College of Physicians. The integration of care for mental health, substance abuse, and other behavioral health conditions into primary care: executive summary of an American College of Physicians position paper. *Ann Intern Med*. 2015;163(4):298-299. doi:10.7326/M15-0510

15. American Hospital Association. The state of the behavioral health workforce: A literature review. 2016. Accessed December 12, 2023. https://www.aha.org/system/files/hpoe/Reports-HPOE/2016/aha_Behavioral_FINAL.pdf

16. Kessler R, Miller BF, Kelly M, et al. Mental health, substance abuse, and health behavior services in patient-centered medical homes. *J Am Board Fam Med*. 2014;27(5):637-644. doi:10.3122/jabfm.2014.05.140021

17. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. *Health Aff (Millwood)*. 2009;28(3):w490-w501. doi:10.1377/hlthaff.28.3.w490

18. Kessler RC, Demler O, Frank RG, et al. Prevalence and treatment of mental disorders, 1990 to 2003. *N Engl J Med*. 2005;352(24):2515-2523. doi:10.1056/NEJMsa043266

19. Substance Abuse and Mental Health Services Administration. *Key Substance Use and Mental Health Indicators in the United States: Results From the 2018 National Survey on Drug Use and Health*. HHS Publication No. PEP19-5068, NSDUH Series H-54. August 2019. Accessed December 12, 2023. http://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf

20. Dollar KM, Kearney LK, Pomerantz AS, Wray LO. Achieving same-day access in integrated primary care. *Fam Syst Health*. 2018;36(1):32-44. doi:10.1037/fsh0000327

21. Diez-Canseco F, Toyama M, Ipince A, et al. Integration of a technology-based mental health screening program into routine practices of primary health care services in Peru (the Allillanchu Project): development and implementation. *J Med Internet Res*. 2018;20(3):e100. doi:10.2196/jmir.9208

22. Possemato K, Shepardson RL, Funderburk JS. The role of integrated primary care in increasing access to effective psychotherapies in the Veterans Health Administration. *Focus (Am Psychiatr Publ)*. 2018;16(4):384-392. doi:10.1176/appi.focus.20180024

23. Wray LO, Pikoff E, King PR, Hutchison D, Beehler GP, Maisto SA. Veterans' mental health beliefs: facilitators and barriers to primary care-mental health use. *Fam Syst Health*. 2016;34(4):404-413. doi:10.1037/fsh0000231

24. Baird M, Blount A, Brungardt S, et al; Working Party Group on Integrated Behavioral Healthcare. Joint principles: integrating behavioral health care into the patient-centered medical home. *Ann Fam Med*. 2014;12(2):183-185. doi:10.1370/afm.1633

25. Colon-Gonzalez MC, McCall-Hosenfeld JS, Weisman CS, Hillemeier MM, Perry AN, Chuang CH. "Someone's got to do it"—primary care providers (PCPs) describe caring for rural women with mental health problems. *Ment Health Fam Med*. 2013;10(4):191-202.

26. Parcesepe AM, Cabassa LJ. Public stigma of mental illness in the United States: a systematic literature review. *Adm Policy Ment Health*. 2013;40(5):384-399. doi:10.1007/s10488-012-0430-z

27. Kessler R. Mental health care treatment initiation when mental health services are incorporated into primary care practice. *J Am Board Fam Med*. 2012;25(2):255-259. doi:10.3122/jabfm.2012.02.100125

28. Heath B, Wise Romero P, Reynolds K. *A Standard Framework for Levels of Integrated Healthcare*. SAMHSA-HRSA Center for Integrated Health Solutions. April 2013. Accessed January 25, 2024. https://www.thenationalcouncil.org/wp-content/uploads/2020/01/CIHS_Framework_Final_charts.pdf?daf=375ateTbd56

29. 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. *Transl Behav Med*. 2012;2(3):337-344. doi:10.1007/s13142-012-0141-8

30. Young J, Gilwee J, Holman M, Messier R, Kelly M, Kessler R. Mental health, substance abuse, and health behavior intervention as part of the patient-centered medical home: A case study. *Transl Behav Med*. 2012;2(3):345-354. doi:10.1007/s13142-012-0148-1

31. Manderscheid R, Kathol R. Fostering sustainable, integrated medical and behavioral health services in medical settings. *Ann Intern Med*. 2014;160(1):61-65. doi:10.7326/M13-1693

32. Ramanuj P, Ferenchik E, Docherty M, Spaeth-Rublee B, Pincus HA. Evolving models of integrated behavioral health and primary care. *Curr Psychiatry Rep*. 2019;21(1):4. doi:10.1007/s11920-019-0985-4

33. Korsen N, Narayanan V, Mercincavage L, al e. *Atlas of Integrated Behavioral Health Care Quality Measures*. AHRQ Publication No. 13-IP002-EF. Agency for Healthcare Research and Quality; 2013.

34. 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(1):200. doi:10.1186/s13063-021-05133-8

35. Richman EL, Lombardi BM, Zerden LD. Mapping colocation: using national provider identified data to assess primary care and behavioral health colocation. *Fam Syst Health*. 2020;38(1):16-23. doi:10.1037/fsh0000465

36. Vana B. *Vermont Colocation of Primary Care and Mental Health Survey*. 2014. https://1library.net/document/zkon0gey-vermont-colocation-of-primary-care-and-mental-health-survey-bud-vana-md-spring-2014.html

37. Laderman M, Mate K. Integrating behavioral health into primary care. *Healthc Exec*. 2014;29(2):74, 76-77.

38. Carey TS, Crotty KA, Morrissey JP, et al. *Future Research Needs for the Integration of Mental Health/Substance Abuse and Primary Care: Identification of Future Research Needs From Evidence Report/Technology Assessment No. 173*. Agency for Healthcare Research and Quality; 2010. Report no.: 10-EHC069-EF. Accessed December 12, 2023. https://www.ncbi.nlm.nih.gov/books/NBK51247/

39. Bower P, Knowles S, Coventry PA, Rowland N. Counselling for mental health and psychosocial problems in primary care. *Cochrane Database Syst Rev*. 2011;2011(9):CD001025. doi:10.1002/14651858.CD001025.pub3

40. 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.* Mullican C, ed. Agency for Healthcare Research and Quality; 2011. AHRQ publication no. 11-0067. https://www.indianacouncil.org/wp-content/uploads/2021/10/AHRQ-Collaborative-Care-Research-Agenda.pdf

41. Hunter CL, Funderburk JS, Polaha J, Bauman D, Goodie JL, Hunter CM. Primary Care Behavioral Health (PCBH) model research: current state of the science and a call to action. *J Clin Psychol Med Settings*. 2018;25(2):127-156. doi:10.1007/s10880-017-9512-0

42. Knox L, Brach C. The Practice Facilitation Handbook: Training Modules for New Facilitators and Their Trainers. Agency for Healthcare Research and Quality; 2013. AHRQ publication no. 13-0046-EF. https://www.ahrq.gov/sites/default/files/publications/files/practicefacilitationhandbook.pdf

43. Baskerville NB, Liddy C, Hogg W. Systematic review and meta-analysis of practice facilitation within primary care settings. *Ann Fam Med*. 2012;10(1):63-74. doi:10.1370/afm.1312

44. Berta W, Cranley L, Dearing JW, Dogherty EJ, Squires JE, Estabrooks CA. Why (we think) facilitation works: insights from organizational learning theory. *Implement Sci*. 2015;10:141. doi:10.1186/s13012-015-0323-0

45. Dickinson WP, Dickinson LM, Nutting PA, et al. Practice facilitation to improve diabetes care in primary care: a report from the EPIC randomized clinical trial. *Ann Fam Med*. 2014;12(1):8-16. doi:10.1370/afm.1591

46. Michaels L, Anastas T, Waddell EN, Fagnan L, Dorr DA. A randomized trial of high-value change using practice facilitation. *J Am Board Fam Med*. 2017;30(5):572-582. doi:10.3122/jabfm.2017.05.170013

47. Wang A, Pollack T, Kadziel LA, et al. Impact of practice facilitation in primary care on chronic disease care processes and outcomes: a systematic review. *J Gen Intern Med*. 2018;33(11):1968-1977. doi:10.1007/s11606-018-4581-9

48. Gold R, Bunce A, Cowburn S, et al. Does increased implementation support improve community clinics' guideline-concordant care? Results of a mixed methods, pragmatic comparative effectiveness trial. *Implement Sci*. 2019;14(1):100. doi:10.1186/s13012-019-0948-5

49. Liddy C, Hogg W, Singh J, et al. A real-world stepped wedge cluster randomized trial of practice facilitation to improve cardiovascular care. *Implement Sci*. 2015;10:150. doi:10.1186/s13012-015-0341-y

50. Seers K, Rycroft-Malone J, Cox K, et al. Facilitating Implementation of Research Evidence (FIRE): an international cluster randomised controlled trial to evaluate two models of facilitation informed by the Promoting Action on Research Implementation in Health Services (PARIHS) framework. *Implement Sci*. 2018;13(1):137. doi:10.1186/s13012-018-0831-9

51. Ruhe MC, Carter C, Litaker D, Stange KC. A systematic approach to practice assessment and quality improvement intervention tailoring. *Qual Manag Health Care*. 2009;18(4):268-277. doi:10.1097/QMH.0b013e3181bee268

52. Rogers ES, Cuthel AM, Berry CA, Kaplan SA, Shelley DR. Clinician perspectives on the benefits of practice facilitation for small primary care practices. *Ann Fam Med*. 2019;17(Suppl 1):S17-S23. doi:10.1370/afm.2427

53. Ye J, Zhang R, Bannon JE, et al. Identifying practice facilitation delays and barriers in primary care quality improvement. *J Am Board Fam Med*. 2020;33(5):655-664. doi:10.3122/jabfm.2020.05.200058

54. Bauer MS, Miller C, Kim B, et al. Partnering with health system operations leadership to develop a controlled implementation trial. *Implement Sci*. 2016;11:22. doi:10.1186/s13012-016-0385-7

55. Ritchie MJ, Dollar K, Miller CJ, et al. *Using Implementation Facilitation to Improve Healthcare: Implementation Facilitation Training Manual, Version 3*. Veterans Health Administration, Behavioral Health Quality Enhancement Research Initiative; 2020. Accessed December 12, 2023. https://www.queri.research.va.gov/tools/Facilitation-Manual.pdf

56. Young AM, Keller HH, Barnes R, Bell JJ. Clinicians as novice facilitators: a SIMPLE case study. *J Health Organ Manag*. 2019;33(1):78-92. doi:10.1108/JHOM-03-2018-0071

57. Cohen DJ, Sweeney SM, Miller WL, et al. Improving smoking and blood pressure outcomes: the interplay between operational changes and local context. *Ann Fam Med*. 2021;19(3):240-248. doi:10.1370/afm.2668

58. Phillips RL. The need for coaches in the clinical world. *Ann Fam Med*. 2021;19(3):194-195. doi:10.1370/afm.2700

59. Phillips RL Jr, Cohen DJ, Kaufman A, Dickinson WP, Cykert S. Facilitating practice transformation in frontline health care. *Ann Fam Med*. 2019;17(Suppl 1):S2-S5. doi:10.1370/afm.2439

60. Peek CJ; the National Integration Academy Council. *Lexicon for Behavioral Health and Primary Care Integration: Concepts and Definitions Developed by Expert Consensus*. Agency for Healthcare Research and Quality; 2013. AHRQ publication no. AHRQ 13-IP001-EF. https://integrationacademy.ahrq.gov/sites/default/files/2020-06/Lexicon.pdf

61. Patient partner guide. Integrating Behavioral Health in Primary Care Patient Partner Team. Updated April 27, 2022. Accessed December 24, 2022. https://sites.google.com/view/ibhpc/workbooks/patient-partnering

62. Institute for Patient- and Family-Centered Care. *Partnering With Patients and Families to Enhance Safety and Quality: A Mini Toolkit*. 2013. Accessed December 12, 2023. https://ipfcc.org/resources/Patient-Safety-Toolkit-04.pdf

63. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ*. 2015;350:h2147. doi:10.1136/bmj.h2147

64. 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(4):342-356. doi:10.1037/fsh0000227

65. Macchi CR, Kessler R, Auxier A, et al. The practice integration profile: rationale, development, method, and research. *Fam Syst Health*. 2016;34(4):334-341. doi:10.1037/fsh0000235

66. Kessler R, Miller BF, Kelly M, et al. Mental health, substance abuse, and health behavior services in patient-centered medical homes. *J Am Board Fam Med*. 2014;27(5):637-644. doi:10.3122/jabfm.2014.05.140021

67. Carlson B. Getting patients in the door faster can boost satisfaction, outcomes. *Manag Care*. 2002;11(3):52-54.

68. Jimmerson C, Weber D, Sobek DK. Reducing waste and errors: piloting lean principles at Intermountain Healthcare. *Jt Comm J Qual Patient Saf*. 2005;31(5):249-257. doi:10.1016/s1553-7250(05)31032-4

69. van Eeghen C, Littenberg B, Holman MD, Kessler R. Integrating behavioral health in primary care using lean workflow analysis: a case study. *J Am Board Fam Med*. 2016;29(3):385-393. doi:10.3122/jabfm.2016.03.150186

70. Perry CK, Damschroder LJ, Hemler JR, Woodson TT, Ono SS, Cohen DJ. Specifying and comparing implementation strategies across seven large implementation interventions: a practical application of theory. *Implement Sci*. 2019;14(1):32. doi:10.1186/s13012-019-0876-4

71. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci*. 2015;10:21. doi:10.1186/s13012-015-0209-1

72. van Eeghen C, Edwards M, Libman BS, MacLean CD, Kennedy AG. Order from chaos: an initiative to improve opioid prescribing in rheumatology using Lean A3. *ACR Open Rheumatol*. 2019;1(9):546-551. doi:10.1002/acr2.11078

73. van Eeghen C, Kennedy AG, Pasanen ME, MacLean CD. A new quality improvement toolkit to improve opioid prescribing in primary care. *J Am Board Fam Med*. 2020;33(1):17-26. doi:10.3122/jabfm.2019.01.190238

74. van Eeghen CO, Littenberg B, Kessler R. Chronic care coordination by integrating care through a team-based, population-driven approach: a case study. *Transl Behav Med*. 2018;8(3):468-480. doi:10.1093/tbm/ibx073

75. Anderson JB, Marstiller H, Shah K. Lean thinking for primary care. *Prim Care*. 2019;46(4):515-527. doi:10.1016/j.pop.2019.07.009

76. Womack JP, Jones DT. *Lean Thinking: Banish Waste and Create Wealth in Your Corporation*. Simon & Schuster; 1996.

77. Rose GL, Bonnell LN, O'Rourke-Lavoie JB, et al. Development and validation of the patient centeredness index for primary care. *J Clin Nurs*. 2022;31(23-24):3485-3497. doi:10.1111/jocn.16177

78. Cella D, Riley W, Stone A, et al; PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63(11):1179-1194. doi:10.1016/j.jclinepi.2010.04.011

79. Working Group on Health Outcomes for Older Persons with Multiple Chronic Conditions. Universal health outcome measures for older persons with multiple chronic conditions. *J Am Geriatr Soc*. 2012;60(12):2333-2341. doi:10.1111/j.1532-5415.2012.04240.x

80. Rothrock NE, Hays RD, Spritzer K, Yount SE, Riley W, Cella D. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol*. 2010;63(11):1195-1204. doi:10.1016/j.jclinepi.2010.04.012

81. Fischer HF, Klug C, Roeper K, et al. Screening for mental disorders in heart failure patients using computer-adaptive tests. *Qual Life Res*. 2014;23(5):1609-1618. doi:10.1007/s11136-013-0599-y

82. Flynn KE, Dew MA, Lin L, et al. Reliability and construct validity of PROMIS® measures for patients with heart failure who undergo heart transplant. *Qual Life Res*. 2015;24(11):2591-2599. doi:10.1007/s11136-015-1010-y

83. Irwin DE, Atwood CA Jr, Hays RD, et al. Correlation of PROMIS scales and clinical measures among chronic obstructive pulmonary disease patients with and without exacerbations. *Qual Life Res*. 2015;24(4):999-1009. doi:10.1007/s11136-014-0818-1

84. Lin FJ, Pickard AS, Krishnan JA, et al; CONCERT Consortium. Measuring health-related quality of life in chronic obstructive pulmonary disease: properties of the EQ-5D-5L and PROMIS-43 short form. *BMC Med Res Methodol*. 2014;14:78. doi:10.1186/1471-2288-14-78

85. Johnston KL, Lawrence SM, Dodds NE, Yu L, Daley DC, Pilkonis PA. Evaluating PROMIS® instruments and methods for patient-centered outcomes research: patient and provider voices in a substance use treatment setting. *Qual Life Res*. 2016;25(3):615-624. doi:10.1007/s11136-015-1131-3

86. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D; PROMIS Cooperative Group. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment*. 2011;18(3):263-83. doi:10.1177/1073191111411667

87. Vilagut G, Forero CG, Adroher ND, Olariu E, Cella D, Alonso J; INSAyD investigators. Testing the PROMIS® Depression measures for monitoring depression in a clinical sample outside the US. *J Psychiatr Res*. 2015;68:140-150. doi:10.1016/j.jpsychires.2015.06.009

88. Levin JB, Aebi ME, Smyth KA, et al. Comparing Patient-Reported Outcomes Measure Information System Depression Scale with legacy depression measures in a community sample of older adults with varying levels of cognitive functioning. *Am J Geriatr Psychiatry*. 2015;23(11):1134-1143. doi:10.1016/j.jagp.2015.04.004

89. 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(1):104-107. doi:10.1136/annrheumdis-2013-204053

90. Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System-Cancer scales in advanced-stage cancer patients. *J Clin Epidemiol*. 2011;64(5):507-516. doi:10.1016/j.jclinepi.2010.11.018

91. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care*. 1986;24(1):67-74. doi:10.1097/00005650-198601000-00007

92. 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(6):699-705. doi:10.1093/fampra/cmh621

93. 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(6):532-544. doi:10.1191/1740774504cn051oa

94. Adams PF, Hendershot GE, Marano MA; Centers for Disease Control and Prevention/National Center for Health Statistics. Current estimates from the National Health Interview Survey, 1996. *Vital Health Stat 10*. 1999;(200):1-203.

95. 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(10):651-654. doi:10.1016/0002-9149(89)90496-7

96. Bime C, Wei CY, Holbrook JT, Sockrider MM, Revicki DA, Wise RA. Asthma symptom utility index: reliability, validity, responsiveness, and the minimal important difference in

adult asthmatic patients. *J Allergy Clin Immunol*. 2012;130(5):1078-1084. doi:10.1016/j.jaci.2012.07.058

97. Dennis ML, Chan YF, Funk RR. Development and validation of the GAIN Short Screener (GSS) for internalizing, externalizing and substance use disorders and crime/violence problems among adolescents and adults. *Am J Addict*. 2006;15 Suppl 1(Suppl 1):80-91. doi:10.1080/10550490601006055

98. Morean ME, DeMartini KS, Foster D, et al. The Self-Report Habit Index: assessing habitual marijuana, alcohol, e-cigarette, and cigarette use. *Drug Alcohol Depend*. 2018;186:207-214. doi:10.1016/j.drugalcdep.2018.01.014

99. Venturini F, Nichol MB, Sung JC, Bailey KL, Cody M, McCombs JS. Compliance with sulfonylureas in a health maintenance organization: a pharmacy record-based study. *Ann Pharmacother*. 1999;33(3):281-288. doi:10.1345/aph.18198

100. Chialà O, Vellone E, Klompstra L, Ortali GA, Strömberg A, Jaarsma T. Relationships between exercise capacity and anxiety, depression, and cognition in patients with heart failure. *Heart Lung*. 2018;47(5):465-470. doi:10.1016/j.hrtlng.2018.07.010

101. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. 2006;166(10):1092-1097. doi:10.1001/archinte.166.10.1092

102. Dear BF, Titov N, Sunderland M, et al. Psychometric comparison of the generalized anxiety disorder scale-7 and the Penn State Worry Questionnaire for measuring response during treatment of generalised anxiety disorder. *Cogn Behav Ther*. 2011;40(3):216-227. doi:10.1080/16506073.2011.582138

103. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA*. 1999;282(18):1737-1744. doi:10.1001/JAMA.282.18.1737

104. Revicki DA, Leidy NK, Brennan-Diemer F, Sorensen S, Togias A. Integrating patient preferences into health outcomes assessment: the multiattribute Asthma Symptom Utility Index. *Chest*. 1998;114(4):998-1007. doi:10.1378/chest.114.4.998

105. Conway KP, Green VR, Kasza KA, et al. Co-occurrence of tobacco product use, substance use, and mental health problems among adults: Findings from Wave 1 (2013-2014) of the Population Assessment of Tobacco and Health (PATH) Study. *Drug Alcohol Depend*. 2017;177:104-111. doi:10.1016/j.drugalcdep.2017.03.032

106. Economic Research Service. 2010 rural-urban commuting area (RUCA) codes. US Department of Agriculture. Accessed February 20, 2016. http://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation.aspx

107. Butler DC, Petterson S, Phillips RL, Bazemore AW. Measures of social deprivation that predict health care access and need within a rational area of primary care service delivery. *Health Serv Res*. 2013;48(2 Pt 1):539-559. doi:10.1111/j.1475-6773.2012.01449.x

108. Abdi H. The Bonferroni and Šidák corrections for multiple comparisons. In: Salkind N, ed. *Encyclopedia of Measurement and Statistics*. Sage; 2007:103-107.

109. Littenberg B, MacLean CD. Intra-cluster correlation coefficients in adults with diabetes in primary care practices: The Vermont Diabetes Information System field survey. *BMC Med Res Methodol*. 2006;6:20. doi:10.1186/1471-2288-6-20

110. Browner WS, Newman TB, Cummings SR, et al. *Designing Clinical Research*. 5th ed. Wolters Kluwer; 2022.

111. National Ambulatory Medical Care Survey. Centers for Disease Control and Prevention. Updated December 30, 2021. Accessed October 25, 2021. http://www.cdc.gov/nchs/ahcd/about_ahcd.htm

112. 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(2):274-284. doi:10.1007/s10880-021-09806-z

113. Mullin DJ, Hargreaves L, Auxier A, et al. Measuring the integration of primary care and behavioral health services. *Health Serv Res*. 2019;54(2):379-389. doi:10.1111/1475-6773.13117

114. Stake RE. *The Art of Case Study Research*. Sage Publications; 1995.

115. Creswell JW, Poth CN. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. 4th ed. Sage Publications; 2018.

116. Manja V, Nrusimha A, MacMillan H, Schwartz L, Jack S. Use of ecomaps in qualitative health research. *The Qualitative Report*. 2021;26(2):412-442. doi:10.46743/2160-3715/2021.4565

117. Patton MQ. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. 4th ed. Sage Publications; 2014.

118. Saldaña J. *The Coding Manual for Qualitative Researchers*. 2nd ed. Sage Publications; 2013.

119. Glesne C. *Becoming Qualitative Researchers: An Introduction*. 5th ed. Pearson; 2014.

120. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971;76(5):378-382. doi:10.1037/h0031619

121. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.

122. A repository of data on coronavirus cases and deaths in the US. GitHub. Updated March 24, 2023. Accessed November 1, 2021. https://github.com/nytimes/covid-19-data

123. Possemato K, Johnson EM, Beehler GP, et al. Patient outcomes associated with primary care behavioral health services: a systematic review. *Gen Hosp Psychiatr*. 2018;53:1-11. doi:10.1016/j.genhosppsych.2018.04.002

124. Hopman P, de Bruin SR, Forjaz MJ, et al. Effectiveness of comprehensive care programs for patients with multiple chronic conditions or frailty: a systematic literature review. *Health Policy*. 2016;120(7):818-832. doi:10.1016/j.healthpol.2016.04.002

RELATED PUBLICATIONS

Published

Bonnell LN, Clifton J, Rose GL, Waddell EN, Littenberg B. Urban-rural differences in mental and physical health among primary care patients with multiple chronic conditions: a secondary analysis from a randomized clinical trial. *Int J Environ Res Public Health*. 2022;19(23):15580. doi:10.3390/ijerph192315580

Bonnell LN, Clifton J, Wingood M, Gell N, Littenberg B. The relationship between mental and physical health and walking during the COVID-19 pandemic. *J Am Board Fam Med*. 2022;jabfm.2022.AP.220090. doi:10.3122/jabfm.2022.AP.220090

Bonnell LN, Crocker AM, Kemp K, Littenberg B. The relationship between social determinants of health and functional capacity in adult primary care patients with multiple chronic conditions. *J Am Board Fam Med*. 2021;34(4):688-697. doi:10.3122/jabfm.2021.04.210010

Bonnell LN, Littenberg B. Nonlinear relationships among the natural environment, health, and sociodemographic characteristics across US counties. *Int J Environ Res Public Health*. 2022;19(11):6898. doi:10.3390/ijerph19116898

Bonnell LN, Troy AR, Littenberg B. Exploring non-linear relationships between neighbourhood walkability and health: a cross-sectional study among US primary care patients with chronic conditions. *BMJ Open*. 2022;12(8):e061086. doi:10.1136/bmjopen-2022-061086

Clifton J, Bonnell L, Hitt J, et al. Differences in occupational burnout among primary care professionals. *J Am Board Fam Med*. 2021;34(6):1203-1211. doi:10.3122/jabfm.2021.06.210139

Crocker A, Natkin LW, Callas P, et al. Intraclass correlation and variance in the characteristics of primary care patients managing chronic medical and behavioral conditions. *Cureus*. 2022;14(11):e30970. doi:10.7759/cureus.30970

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(1):200. doi:10.1186/s13063-021-05133-8

Ma KPK, Mollis BL, Rolfes J, et al. Payment strategies for behavioral health integration in hospital-affiliated and non-hospital-affiliated primary care practices. *Transl Behav Med*. 2022;12(8):878-883. doi:10.1093/tbm/ibac053

Rose GL, Bonnell LN, Clifton J, Natkin LW, Hitt JR, O'Rourke-Lavoie J. Outcomes of delay of care after the onset of COVID-19 for patients managing multiple chronic conditions. *J Am Board Fam Med*. 2022;35(6):1081-1091. doi:10.3122/jabfm.2022.220112R1

Rose GL, Bonnell LN, O'Rourke-Lavoie JB, et al. Development and validation of the patient centeredness index for primary care. *J Clin Nurs*. 2022;31(23-24):3485-3497. doi:10.1111/jocn.16177

van Eeghen C, Hitt JR, Pomeroy DJ, Reynolds P, Rose GL, O'Rourke-Lavoie J. Co-creating the Patient Partner Guide by a multiple chronic conditions team of patients, clinicians, and researchers: observational report. *J Gen Intern Med*. 2022;37(Suppl 1):73-79. doi:10.1007/s11606-021-07308-0

Bonnell LN, Clifton J, Natkin LW, Hitt JR, Littenberg B. The association of self-perceived changes due to COVID-19 with mental and physical health among adult primary care patients with multiple chronic conditions: a US-based longitudinal study. *J Multimorb Comorb*. 2024;14:26335565231222148. doi:10.1177/26335565231222148

Jewiss J, Natkin LW, Clark/Keefe K, Crocker A, Welkowitz JA. Integrating behavioural health and primary care: qualitative findings on contextual factors that influence integration. *Fam Pract*. 2023;cmad029. doi:10.1093/fampra/cmad029

Nagykaldi Z, Littenberg B, Bonnell L, et al. Econometric evaluation of implementing a behavioral health integration intervention in primary care settings. *Transl Behav Med*. 2023;13(8):571-580. doi:10.1093/tbm/ibad013

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(6):483-495. doi:10.1370/afm.3027

Ma KPK, Mollis BL, West II, et al. Integrated behavioral health in primary care residency and nonresidency practices. *Fam Med*. 2023;55(8):530-538. doi:10.22454/FamMed.2023.715036.

Under Revision

McGovern MP, Dunn J, Bonnell LN, et al. The association between depression and substance use among primary care patients with comorbid medical and behavioral health conditions. *J Prim Care Community Health*. 2023;14:21501319231200302. doi:10.1177/21501319231200302

Waddell EN, Leibowitz GS, Bonnell LN, Rose GL, McGovern M, Littenberg B. Practice-level documentation of alcohol-related problems in primary care. *JAMA Netw Open*. 2023;6(10):e2338224. doi:10.1001/jamanetworkopen.2023.38224

ACKNOWLEDGMENTS

We would like to acknowledge the many contributions to this project, including those of partners, stakeholders, and the patients and clinicians who answered our calls for assistance through their survey responses, interviews, conversations, and many other thoughtful additions to this work. Lacking full knowledge of the extent to which so many helped us, we would like to recognize the following colleagues for their explicit efforts to guide us in patient-centered research: Andrea Auxier, Sarah Bartlett, Dana Basiliere, Andrew Bertagnolli, Georgia Brown, Anne Donahue, Florence Fee, Walter Grant, James Hester, Gene (Rusty) Kallenberg, Katie MacDonald, Susan McDaniel, Ricka McNaughton, John Muench, and Andrew Pomerantz.

Data Sharing Plan

All de-identified, patient-level data appear on a publicly accessible server, along with a data dictionary and metadata. The data are released to the public domain under a Creative Commons CC0 1.0 license (https://creativecommons.org/publicdomain/zero/1.0/), which allows downloaders to copy, modify, distribute, and perform the work without additional permission. Data that could not reliably be de-identified, such as interview transcripts and other qualitative information, were excluded.

APPENDICES

Appendix A. Cluster Leader Note-Taking Tool

Appendix A: Cluster Leader Note Taking Tool

Cluster Leader Note Taking Tool

This tool is designed to provide an efficient way to gather your observations about each of your practices. Please document your observations **to date.** (Hereafter, we'll simply ask you to note changes you observe on a quarterly basis).

We welcome as much detail as you are able to provide. We will review your notes to enhance our understanding of contextual factors occurring at practices. To the extent possible, please provide enough detail for us to follow up with you or practice staff, as needed, to delve deeper into the identified contextual factors.

If a description applies to multiple sites, no need to re-type your response for each practice. Please write: "same as practice X." Your 5 practices are addressed on the following pages:

- Practice X –Page 1
- Practice Y – Page 2

Based on your interactions with the ***Practice X*** to date, please describe any contextual factors that may influence the way behavioral health services are delivered at the following levels: practice, larger organization, local community, and broader external environment.

The practice (i.e., staffing composition and dynamics, workflow, practice structure and culture, EHR updates, leadership, physical office space, training, etc.)

The larger organization (i.e., management structures, financial structures, corporate vision, leadership, etc.)

The local community (i.e., behavioral health resource availability, natural disasters, demographic characteristics, etc.)

The broader external environment (i.e., health care policies, payment processes, financial incentives for behavioral health, etc.)

Is there anything else you would like to share?

Based on your interactions with the ***Practice Y*** to date, please describe any contextual factors that may influence the way behavioral health services are delivered at the following levels: practice, larger organization, local community, and broader external environment.

The practice (i.e., staffing composition and dynamics, workflow, practice structure and culture, EHR updates, leadership, physical office space, training, etc.)

The larger organization (i.e., management structures, financial structures, corporate vision, leadership, etc.)

The local community (i.e., behavioral health resource availability, natural disasters, demographic characteristics, etc.)

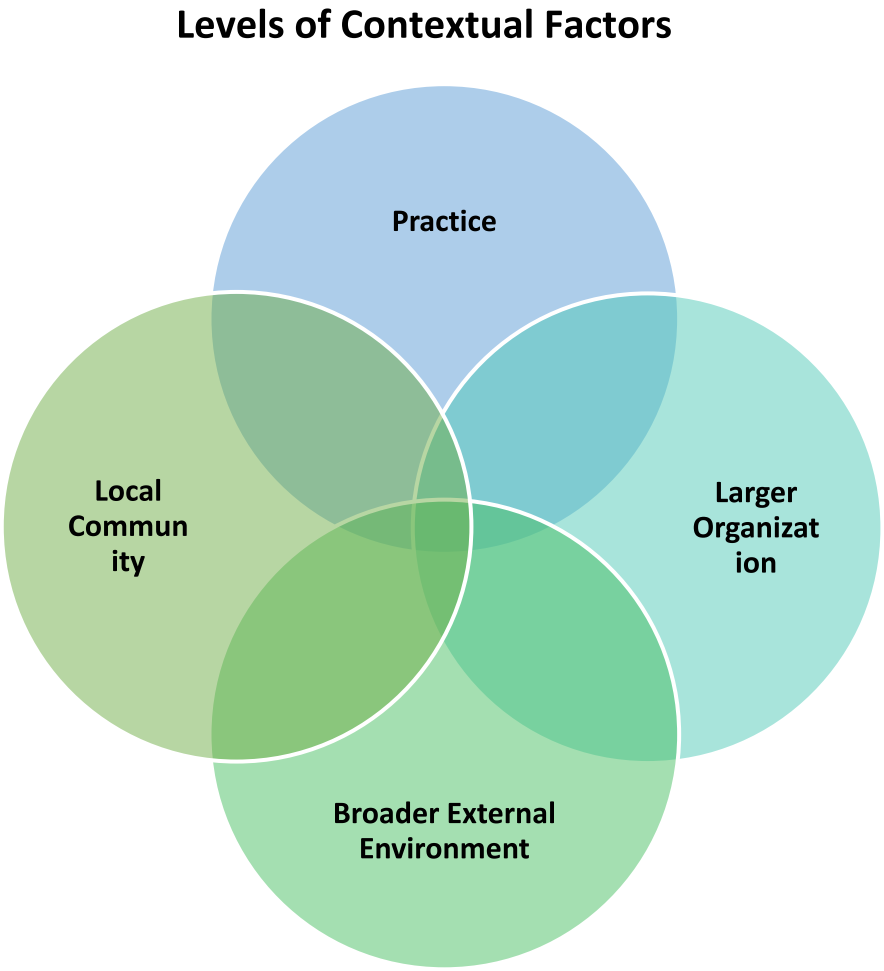
The broader external environment (i.e., health care policies, payment processes, financial incentives for behavioral health, etc.)

Is there anything else you would like to share?

Appendix B. Interview Graphic

Appendix B. Interview Graphic

Research Question: What factors support or impede successful integration of behavioral health into Primary Care practices, and how?



Examples of Factors that May Support or Impede BH Integration

Practice Level

- Intervention implementation
- Staffing composition, training, & dynamics
- Workflow
- Practice structure, culture, & change capacity
- Physical space
- Primary care model
- Characteristics of patient population
- Major events
- COVID-19 Pandemic

Larger Organization

- Management structures, responsiveness, & adaptibility
- History with BH & quality improvement intiatives
- Finanical structures
- Alignment of corporate vision/mission with IBH & IBHPC objectives
- Leadership engagement

Local Community

- Availability of behavioral health resources
- Demographic characteristics
- Geographic characteristics
- Behavioral health issues
- Culture & ethnic composition
- Natural disasters

Broader Environment

- State & federal health care policies
- Payment processes
- Financial incentives & funding for BH
- Insurance policies
- Data sharing regulations

Acknowledgment:
Research reported in this report was funded through a Patient-Centered Outcomes Research Institute® (PCORI®) Award (PCS-1409-24372).
Further information available at: https://www.pcori.org/research-results/2015/testing-program-improve-integration-behavioral-health-care-primary-care-settings#section_covid

PATIENT-CENTERED OUTCOMES RESEARCH INSTITUTE
RESEARCH SUMMARY

PROJECT INFORMATION

December 2023

Testing a Program to Improve the Integration of Behavioral Health Care into Primary Care Settings

Principal investigator
Benjamin Littenberg, MD

Organization
University of Vermont and State Agricultural College

What was the research about?

When primary care clinics offer mental health care on-site, it's called co-located care. But care teams at clinics with co-located care may not work together or share information about patients' care. In integrated behavioral health, or IBH, mental health staff, like counselors and psychologists, and primary care clinicians, like doctors and nurse practitioners, work together to treat patients. They also share records and data systems.

In this study, the research team looked at whether a program to help clinics use IBH improved well-being among patients with physical and mental health conditions. They also wanted to see if the program improved how well primary and behavioral health care teams worked together. The research team compared clinics that did and didn't receive the program.

What were the results?

After two years, patients in clinics that did and didn't receive the IBH program didn't differ in:

- Anxiety, depression, or fatigue
- Ability to fall and stay asleep
- Disability, physical fitness, and ability to do activities
- Pain and how pain interfered with their life
- How often they took their medicine as directed
- Self-care activities and social functioning
- Use of care

Compared with clinics that didn't receive the program, clinics that did:

- Were better at tracking patient information and care
- Didn't differ in provider empathy and ability to communicate effectively with patients

Who was in the study?

The study included 2,426 patients receiving care in one of 41 primary care clinics with co-located care across the United States. Patients had at least one long-term health problem and one long-term mental health condition or at least three long-term health problems. Of these patients, 77 percent were White, 12 percent were African American, 3 percent were Asian, and 2 percent were Native Hawaiian or Pacific Islander. Also, 1 percent were American Indian or Alaska Native and 7 percent were another race or preferred not to say; 7 percent were Hispanic. The average age was 62, and 66 percent were women.

What did the research team do?

The research team assigned clinics by chance to receive the IBH program or not. Staff at clinics assigned to the program could learn about IBH using an online course and a workbook. They received access to an online learning community and coaching services.

At the start of the study and one and two years later, the research team surveyed patients and clinic staff.

Patients, family members, clinicians, and behavioral health staff helped design the study.

What were the limits of the study?

The research team let clinics decide when to use the program and which parts to use. Results may have differed if the team required clinics to use the program within a set amount of time or in a certain way.

Future studies could look at other ways to encourage IBH in primary care.

How can people use the results?

Primary care clinics can use the results when looking for ways to improve how well mental health and primary care providers work together in patient care.

To learn more about this project, visit www.pcori.org/Littenberg020.