

Association of organizational culture and climate with variation in the clinical outcomes of collaborative care for maternal depression in community health centers

Implementation Research and Practice

Volume 4: Jan-Dec 2023 1–14

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/26334895231205891

journals.sagepub.com/home/irp



Nathaniel J. Williams¹, Joan Russo², Melinda Vredevoogd² ,
Tess Grover², Phillip Green³, Enola Proctor⁴, Amritha Bhat²,
Jürgen Unützer² and Ian M. Bennett^{2,5} 

Abstract

Background

Organizational factors may help explain variation in the effectiveness of evidence-based clinical innovations through implementation and sustainment. This study tested the relationship between organizational culture and climate and variation in clinical outcomes of the Collaborative Care Model (CoCM) for treatment of maternal depression implemented in community health centers.

Method

Organizational cultures and climates of 10 community health centers providing CoCM for depression among low-income women pregnant or parenting were assessed using the organizational social context (OSC) measure. Three-level hierarchical linear models tested whether variation in culture and climate predicted variation in improvement in depression symptoms from baseline to 6.5-month post-baseline for $N = 468$ women with care ± 1 year of OSC assessment. Depression symptomology was measured using the Patient Health Questionnaire (PHQ-9).

Results

After controlling for patient characteristics, case mix, center size, and implementation support, patients served by centers with more proficient cultures improved significantly more from baseline to 6.5-month post-baseline than patients in centers with less proficient cultures (mean improvement = 5.08 vs. 0.14, respectively, $p = .020$), resulting in a large adjusted effect size of $d_{adj} = 0.78$. A similar effect was observed for patients served by centers with more functional climates (mean improvement = 5.25 vs. 1.12, $p < .044$, $d_{adj} = 0.65$). Growth models indicated that patients from all centers recovered on average after 4 months of care. However, those with more proficient cultures remained stabilized whereas patients served by centers with less proficient cultures deteriorated by 6.5-month post-baseline. A similar pattern was observed for functional climate.

¹School of Social Work, Boise State University, Boise, ID, USA

²Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, USA

³Center for Behavioral Health Research, University of Tennessee, Knoxville, TN, USA

⁴Brown School of Social Work, Washington University, Saint Louis, MO, USA

⁵Department of Family Medicine, University of Washington School of Medicine, Seattle, WA, USA

Corresponding author:

Ian M. Bennett, Department of Family Medicine, Research Section, University of Washington, Box 354696, Seattle, WA 98195-4696, USA.
Email: ibennett@uw.edu



Conclusions

Variation in clinical outcomes for women from historically underserved populations receiving *Collaborative Care* for maternal depression was associated with the organizational cultures and climates of community health centers. Implementation strategies targeting culture and climate may improve the implementation and effectiveness of integrated behavioral health care for depression.

Plain Language Summary: While many implementation theories espouse the importance of organizational culture and climate for the successful implementation of evidence-based practices in primary care, there is little research that tests this hypothesis. Since there are interventions which can improve organizational culture and climate, having more evidence that these factors are associated with implementation would support efforts to modify these aspects of a community health center as a means of improving implementation. This study showed that the extent to which patients clinically benefitted from the evidence-based Collaborative Care Model for maternal depression was related to the prevailing culture and climate of community health centers where they received treatment. Specifically, women seen at centers in which the staff and providers indicated that their organizations prioritize responsiveness to patients' needs over competing organizational goals and maintain competence in up-to-date treatment models (referred to as proficient culture), and understand their role in the organization and receive the cooperation and support they need from colleagues and supervisors to perform their job well (functional climate) were associated with sustained improvements in depression symptoms. This benefit was independent of other factors already known to be associated with these outcomes. Implementation strategies that target organizational culture (i.e., priorities and expectations for staff) and climate (i.e., quality of working environment) may improve the clinical outcomes of integrated collaborative care models for depression and reduce the commonly seen variation in outcomes across health centers.

Keywords

Organization, culture, climate, maternal, depression, primary care, integrated behavioral health, collaborative care

Introduction

Globally, depression is the leading cause of disabled years lived and accounts for more than \$200 billion in annual costs from lost productivity and treatment expenses in the United States alone (Greenberg et al., 2015; Whiteford et al., 2013). Maternal depression, which includes mood disorders occurring during pregnancy and parenting, is the leading cause of health burden among women of reproductive age from high, middle, and low-income countries (Ferrari et al., 2014; Whiteford et al., 2013). Its impact is compounded by the myriad negative effects that maternal depression has on infants and families (Bennett et al., 2016; Pilowsky et al., 2006; Stein et al., 2014). In the United States, maternal depression disproportionately affects low-income and racial/ethnic minority women, with 9%–15% experiencing it during the perinatal period and as much as 25% experiencing it during child-rearing (Gaynes et al., 2005; Melville et al., 2010; Vesga-López et al., 2008). The majority of depression care in the United States is delivered in primary care settings, which have been called the “de facto US mental health system” (Regier et al., 1993). This is particularly true for low-income patients, who face access barriers to specialty mental health care (Bennett et al., 2009, 2010).

The Collaborative Care Model (CoCM) for mental disorders is a complex intervention with significant clinical benefit and cost-effectiveness for the treatment of adult

depression in primary (Gilbody, Bower, Fletcher, et al., 2006; Gilbody, Bower, and Whitty, 2006; Katon et al., 2010; Unützer et al., 2008) and perinatal care (Melville et al., 2014; Wells et al., 2002). This multidisciplinary team model has been identified as critical to improving the mental health of the US population by the Centers for Medicare & Medicaid Services, which have promulgated billing codes for public insurance to support CoCM implementation (Liao et al., 2019). Despite the high level of evidence for the effectiveness of CoCM, it has not been widely implemented (Katon et al., 2010; Katon & Unützer, 2006; Unützer, 2008). Moreover, the success of CoCM implementation and sustainment for general adult populations, measured by fidelity and clinical patient benefit, is highly variable (Solberg et al., 2013; Unützer et al., 2020). The identification of tools and strategies that can increase the rate of effective implementation of this evidence-based model for depression in primary care settings may improve population behavioral health.

Organizational culture and climate have been proposed to influence the effective implementation and clinical outcomes of complex clinical interventions within agencies providing mental health services (Glisson & Williams, 2015; Woolston, 2005) including primary care health centers (Kramer et al., 2017). Glisson and colleagues define *organizational culture* as the shared expectations and behavioral norms that guide and direct providers' care-related behaviors within the organization (Glisson

et al., 2008). *Organizational climate* is defined as providers' shared perceptions of the impact of the work environment on their personal well-being. Prior research has linked these dimensions of culture and climate to service outcomes in social service agencies (Glisson & Green, 2011; Williams & Glisson, 2014) and has shown that interventions targeting these dimensions are associated with improved social service outcomes (Glisson et al., 2010, 2013). Studies in specialty mental health settings have shown that therapists' use of evidence-based psychotherapy techniques is higher in clinics with more proficient cultures and more functional climates (Beidas et al., 2019; Williams et al., 2018) and that these dimensions of culture and climate are related to practitioners' level of fidelity to evidence-based practice models in schools (Williams et al., 2019). Studies link variation in organizational culture and climate to variation in implementation outcomes, such as adoption of evidence-based practices and fidelity to evidence-based practice models (Li et al., 2018; Proctor et al., 2011; Smith et al., 2018; Williams & Glisson, 2020).

Despite these findings, no studies we can identify have tested whether variation in organizational culture and climate explain variation in the clinical outcomes of evidence-based practices implemented in primary care (Li et al., 2018). This is an important knowledge gap given the centrality of clinical outcomes to the goals of implementation science (Eccles & Mittman, 2006). Assessing the clinical importance of culture and climate as constructs related to care improvement and their independence from other factors influencing implementation and outcomes is needed to guide further work in this area. Given evidence that culture and climate vary across primary care health centers which serve historically underserved populations, we wished to test whether this variation could help explain variation in patient clinical outcomes of the evidence-based CoCM for maternal depression (Kramer et al., 2017). We hypothesized that patients served by centers with more proficient cultures and more functional climates would experience superior clinical outcomes of CoCM. Evidence for this hypothesis would support organization-level strategies for the implementation and sustainment of complex, team-based interventions for common mental disorders and other chronic illnesses.

Method

Study Setting

This study made use of clinical and survey data collected from July 2013 to July 2015 in the course of routine care of women participating in the High Risk Mothers' program, a part of the Mental Health Integration Program of Washington State in the United States. Funded by the State of Washington and King County and administered by the Community Health Plan of Washington State, this program provides behavioral health services for women

in pregnancy and with children in 13 community health centers (Federally Qualified Health Centers and Look-Alikes) in King County. Details of the program are described elsewhere, but briefly, women in pregnancy or with children who are identified in participating health centers with elevated depression symptomatology are offered the team-based CoCM model (Bauer et al., 2012; Huang et al., 2012). The interdisciplinary team caring for the patient consists of the primary care provider, a care manager, and a psychiatric consultant. In accordance with the CoCM model and quality metrics for depression care, treatment typically lasts 6–8 months and occurs in phases (Croghan et al., 2006; National Committee for Quality Assurance, n.d.). During the initial diagnosis and treatment decision phase, treatments are selected and implemented based on the clinical assessment. During the acute care phase, treatment trials of different therapeutic modalities (non-pharmacologic and pharmacologic), treatment doses, and pharmacologic agents are applied until a significant improvement is seen. Response to treatment is monitored by assessing depression symptoms at intake and at every subsequent contact in order to guide care.

Clinical data used in this study (i.e., depression symptomatology as described below) were collected via an internet-based electronic care management software system designed to support the delivery of CoCM (Unützer et al., 2002). This system is utilized by care managers, and data from the system have been used to evaluate outcomes in several previous studies of CoCM (Bauer et al., 2012; Huang et al., 2012, 2013; Ratzliff et al., 2013). The present study was reviewed and determined exempt by the Institutional Review Board of the University of Washington, review numbers 46029 and 48940.

Independent and Dependent Variables

Organizational culture and climate. The organizational culture and climate of participating health centers were assessed using the 105-item organizational social context (OSC) measure (Glisson et al., 2008). The reliability, factor validity, criterion-related validity, and predictive validity of the OSC have been supported in numerous studies including two national studies in specialty mental health and child welfare settings, respectively, as well as randomized controlled trials (Glisson et al., 2008; Glisson et al., 2013, 2016a, 2016b; Glisson & Green, 2011; Glisson, Green, et al., 2012; Olin et al., 2014; Williams et al., 2017; Williams & Beidas, 2018; Williams & Glisson, 2013, 2014).

Organizational culture is defined by the OSC as the norms and expectations that govern the way work is done in an organization and is assessed along three dimensions of proficiency, rigidity, and resistance (Glisson et al., 2008; Glisson & Green, 2011). Proficient organizational cultures are characterized by norms and expectations that clinicians will be responsive to client needs and maintain competence in up-to-date treatment practices. Rigid cultures

are characterized by norms and expectations that clinicians will have limited flexibility and discretion in carrying out work tasks, minimal input into important management decisions, and high compliance with rules and regulations. Resistant cultures expect clinicians to maintain the status quo.

Organizational climate is defined by the OSC as clinicians' shared perceptions of the impact of the work environment on their personal well-being and is assessed along three dimensions of engagement, functionality, and stress (Glisson et al., 2008; Glisson & Green, 2011). Clinicians in engaged climates perceive that they are able to sustain personal concern for their clients and accomplish many personally meaningful goals in their work. Clinicians in functional climates perceive that they receive the support and cooperation they need from colleagues to effectively do their job and have a clear understanding of their role and how it fits into the organization. Clinicians in stressful climates perceive that they are overloaded in their work role and pulled in conflicting directions (role conflict). Alpha reliabilities for the six OSC dimensions in the present sample were as follows: proficiency ($\alpha = 0.90$), rigidity ($\alpha = 0.68$), resistance ($\alpha = 0.77$), engagement ($\alpha = 0.80$), functionality ($\alpha = 0.92$), and stress ($\alpha = 0.92$).

The OSC conceptualizes culture and climate as the shared perceptions of treatment team members; consequently, the perceptions of individuals who are directly involved in delivering care are the most informative for understanding how culture and climate influence treatment delivery. Thus, individual OSC surveys were administered to members of the clinical teams involved with the High Risk Mothers' program at each participating health center. Participation in the data collection sessions was optional and voluntary. In order to ensure candid responses to the surveys, supervisors were not present when care team members completed surveys. Responses were collected by study staff who were not known to care team members personally, and surveys were placed in sealed envelopes before leaving the survey locations.

In accordance with best practices, care team members' individual responses to the OSC were aggregated (i.e., averaged) to the health center level in order to measure culture and climate in the 10 participating health centers (Chan, 1998; Klein et al., 1994; Rousseau, 1985). Aggregation is guided by the use of composition models that describe how items must be worded in order to justify aggregation and establish empirical criteria for evaluating the construct validity of aggregated variables (Chan, 1998). Aggregation of individual scores to the center level is justified when there is a high level of interrater agreement on each culture/climate dimension within each health center. We evaluated interrater agreement using the $r_{wg(j)}$ index, which ranges from 0 to 1 (higher values indicate greater agreement) (James et al., 1993). A cutoff of 0.7 on the $r_{wg(j)}$ index is recommended to support aggregation of individual culture/climate scores to the unit level (Bliese, 2000; LeBreton & Senter,

2008). Examination of the $r_{wg(j)}$ values for all 10 health centers in the sample on all six OSC dimensions indicated no values fell below $r_{wg(j)} = .91$. This provided a high level of support for clinicians' agreement on their perceptions of culture and climate within each center. Aggregate raw scores were converted to T scores based on a national sample of 100 mental health centers with a $\mu = 50$ and $\sigma = 10$ (Glisson et al., 2008).

Depression symptom level. Patients' depression symptom level at intake and at each subsequent contact was assessed during usual clinical care using the Patient Health Questionnaire-9 (PHQ-9), a widely used instrument that is valid for use in pregnancy as well as a wide range of adult populations for screening and monitoring depression symptom change over time (Gjerdingen et al., 2009; Kroenke et al., 2001; Spitzer et al., 2000). Total scores are calculated by summing the nine items. Using standard conventions, we made use of the cut score of 10 or above to identify women with likely depression and with moderate to severe symptom levels (Kroenke et al., 2001).

Control Variables

Because of known associations with the primary independent and dependent variables, and in order to adjust models for differences across health centers (Hofmann & Gavin, 1998; Raudenbush & Bryk, 2002), we included the following variables as covariates in all analyses: center size (number of patients served), patient ethnicity (Latina vs. not Latina), patient race (Black/African American vs. not Black/African American), patient age at enrollment into care, and patient history of prior psychiatric treatment as reported by the patient (yes/no) (Huang et al., 2012). Because level of implementation support is associated with variation in patient outcomes at the site level among the clinics, we included type of implementation support (i.e., Basic vs. Enhanced) as a control variable. While this is described in detail elsewhere, briefly *Basic support* included in person team training and access to implementation tools while *Enhanced support* added prolonged external practice facilitation (Solberg et al., 2013).

Statistical Analysis

As a preliminary analysis, we tested whether there was variation in rates of clinically significant improvement in PHQ-9 scores from baseline to 6.5-month post-baseline across health centers. This time period corresponds to standard quality metrics for depression care (National Committee for Quality Assurance, n.d.). Clinically significant improvement was defined using the Jacobson and Truax reliable and clinically significant change criterion c that has performed well in other research (McMillan et al., 2010). Patients achieved clinically significant improvement if their PHQ-9 total score improved by 6.76 points or more from baseline to their last follow-up

and if their final follow-up score was below the clinical cutoff of 10. Analysis of variance (ANOVA) tested whether health centers differed on the mean of patients who achieved clinically significant improvement.

The relationships between health center culture and climate and growth in patients' depression symptoms from baseline to 6.5-month post-baseline were tested using three-level hierarchical linear models (HLMs) (Hedeker & Gibbons, 2006; Raudenbush & Bryk, 2002). HLMs are ideal for longitudinal data that are clustered within health centers because they make use of all available measurements, generate unbiased parameter estimates when unobserved values are missing at random, adjust standard errors for nesting at both the patient and health center levels, and do not require uniform spacing of observations across patients (Verbeke & Molenberghs, 2009). In the analyses, patients' PHQ-9 total score at intake and at each subsequent observation point was modeled as a function of growth parameters (time and time-squared) at level 1, patient characteristics at level 2, and health center characteristics (including culture or climate) at level 3. Models incorporated random effects at the patient and health center levels to account for the nesting of time within patient and patient within health center. All patient and center covariates were grand mean centered to facilitate model interpretation and to adjust for differences across health centers in case mix (Hofmann & Gavin, 1998). There were no missing data on health center characteristics or patient covariates included in the models. Because this was a clinical dataset, patients had different numbers of outcome (PHQ-9) assessments and different assessment schedules. Missingness on the outcome was addressed using maximum likelihood estimation under the assumption data were missing at random. To increase the plausibility of the missing at random assumption, we included covariates in the models that predicted the total number of observations for each patient (ethnicity, age at entry, and clinic size). Models were estimated in HLM 8 software (Raudenbush, 2004). The models estimate the unique effect of each of the six dimensions of culture and climate on growth in patients' symptom scores for depression from baseline to 6.5-month post-baseline, over and above the effects of patient-level characteristics (i.e., patients' ethnicity, race, age at entry, and prior history of treatment), variation in patient characteristics across sites (i.e., case mix), center size, and type of implementation support (Raudenbush & Bryk, 2002).

Separate models were estimated for each of the six dimensions of culture and climate in order to identify which dimensions were most closely associated with variation in clinical outcomes. The Benjamini-Hochberg false discovery procedure was used to control for multiple comparisons (Benjamini & Hochberg, 1995). Concerns regarding multicollinearity prohibited estimation of a single model with multiple dimensions of culture and climate included simultaneously (Yu et al., 2015); the average correlation among the culture and

climate dimensions was $r = .70$ in this sample (see Table 2 for a correlation matrix). As a sensitivity analysis, we tested the effects of each culture/climate dimension on growth in patients' depression symptoms with and without control variables, and the results were substantively similar; consequently, only analyses with control variables are presented. To facilitate interpretation of the model results, culture and climate dimensions were scaled so that model coefficients represented the change in PHQ-9 score for a 10-point change (one population standard deviation) in the culture/climate dimension.

Preliminary analyses confirmed that there was significant clustering of PHQ-9 depression scores at the health center level ($ICC(1) = .09$, $p < .001$) and patient level ($ICC(1) = .36$, $p < .001$) and that growth in depression symptoms was best represented by a quadratic growth model at level 1 rather than by a linear growth model ($\chi^2 = 155.29$, $df = 1$, $p < .001$). In quadratic growth models, change in symptoms is described by two parameters: (1) the instantaneous growth rate (represented by the time parameter), which differs at each time point and indicates the direction and magnitude of change at that time (e.g., at baseline), and (2) the rate of acceleration (represented by the time-squared parameter), which remains constant across all time points and indicates the magnitude and direction of change in the rate of change across the entire study period. Significant associations between culture/climate and the rate of acceleration indicate that culture/climate explains variation in patients' overall change in depression from baseline to the study endpoint (6.5-month post-baseline). Significant associations between culture/climate and the instantaneous rate of change at a specific time point provide insight into when culture/climate begin to differentially influence depression outcomes (e.g., at 4-month post-baseline).

To facilitate interpretation of the model results, effect sizes were calculated representing the standardized mean difference in PHQ-9 total scores at 6.5-month post-baseline between patients served by health centers with high versus low values of the culture/climate dimension. The value for "high" culture/climate was defined as the mean value for clinics in the upper quartile on the culture/climate dimension; "low" was defined as the mean value for clinics in the lower quartile on the dimension. The standardized mean difference at 6.5 months is reported as Cohen's d (Cohen, 2013), which is commonly interpreted as small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$). In addition, we calculated an adjusted Cohen's d , d_{adj} (Durlak, 2009), which represents the standardized mean difference between the groups in their change in PHQ-9 total score from baseline to 6.5-month post-baseline (Feingold, 2019).

Results

OSC surveys were collected from 13 FQHCs (Federally Qualified Health Center) from June 2014 to September

2014. The survey results from clinical teams at 10 FQHCs were included in the analyses because three teams did not have clinical outcome information during the study period. There were no differences in care team member characteristics between FQHCs included in the analyses versus those excluded ($p > .05$). Furthermore, there were no differences in care team member characteristics among the 10 clinics included in the study sample ($p > .05$). Table 1 presents characteristics of the FQHCs and care team members who were included in the study. Across

FQHCs, the average number of care team members who responded to the OSC was 5.4 (range 3–9). Culture and climate scores for participating primary care health centers were consistent with national validation studies (see Table 2).

Change in depression symptoms, measured using the PHQ-9 total score, was assessed for 468 women in pregnancy or with children enrolled in the High Risk Mothers' program in participating FQHCs from July 2013 to July 2015 (a year prior to administration of the OSC to 1 year after). Within the study period, a total of $N = 675$ women had one or more episodes of care for depression at a participating FQHC. Of these, $n = 207$ were excluded from analyses because their intake PHQ-9 score for the episode was less than the cutoff score of 10, resulting in a total sample of $N = 468$ women with $N = 1,455$ PHQ-9 observations (average N of observations per participant = 3.1, $SD = 2.4$). The average number of patients included in the analyses per health center was $n = 46.8$ ($SD = 37.6$). As is shown in Table 1, the sample was comprised of women (mean age = 33 years) of diverse race/ethnicity. The mean and range of initial depression symptomatology level (mean = 16.2, $SD = 4.69$; range = 10–27), included patients in the moderate (PHQ-9 score of 15–19) and severe (20–27) range; a high percentage of women reported suicidal ideation (28.2%). There were significant differences among the 10 participating health centers on patients' mean age at baseline, proportion of patients of Latina ethnicity, proportion of patients of Black/African American race, and proportion of patients with a history of prior treatment ($p < .05$); all of these were included as control variables in the analyses to adjust for differences in case mix.

Table 1

Characteristics of Participating FQHCs, Care Team Members, and Patients

Characteristic	<i>M</i> (<i>N</i>)	<i>SD</i> (%)	Min	Max
Care team members ($N = 54$)				
Years of experience	19.06	11.39	2	40
Years in current health center	9.41	7.75	1	33
Age in years	46.73	12.55	25	67
Position				
Direct care provider	28	51.9		
Supervisor	8	14.8		
Management	4	7.4		
Other (e.g., quality improvement officer)	10	18.5		
Highest education completed				
High school graduate	3	5.6		
Associate degree	6	11.1		
Bachelor's degree	11	20.4		
Master's degree	15	27.8		
Doctoral degree	19	35.2		
Race				
Asian	7	13.0		
Black or African American	2	3.7		
White	33	61.1		
Other	9	16.6		
More than one race	3	5.6		
Hispanic/Latino(a)	12	22.2		
Female	46	85.2		
Patients ($N = 468$)				
Age in years	33.47	8.39	15	69
PHQ-9 total score at baseline	16.21	4.69	10	27
Endorsed suicidal ideation (PHQ-9)	132	28.2		
Prior history of psychiatric treatment	135	28.8		
Hispanic/Latina	223	47.6		
Race				
American Indian/Alaskan Native	2	0.4		
Asian	44	9.4		
Black or African American	47	10.0		
Native Hawaiian or Other Pacific Islander	5	1.1		
White	109	23.3		
More than one race	40	8.5		
Unknown or not reported	221	47.2		

Variation in Clinical Outcomes Across Health Centers

On average, 42% of patients achieved clinically significant improvement in their PHQ-9 total score from baseline to their last follow-up based on the clinically significant improvement criterion described above. There was significant variation across health centers in their rates of clinically significant improvement, $F = 2.39$, $df = 9$, 313, $p = .012$, with a range of 17%–57% of patients improved at each health center.

Association of Health Center Culture and Climate With Improvement in Depression

The results of the HLMs testing the relationships between each dimension of culture/climate and growth in mothers' depression symptoms, adjusted for center characteristics, type of implementation support, and patient characteristics, indicated that two dimensions of culture/climate were related to significantly greater improvement in depression

Table 2
Organizational Social Context (OSC) Measures for Health Centers (N = 10)

OSC dimension	M (N)	SD (%)	Min	Max	Correlations					
					1	2	3	4	5	
Culture										
1. Proficient culture	54.32	12.05	30.70	70.67						
2. Rigid culture	56.26	6.40	44.57	65.16	-.22					
3. Resistant culture	58.59	11.00	40.22	73.64	-.62	.84				
Climate										
4. Engaged climate	53.92	14.31	26.84	74.08	.82	-.59	-.83			
5. Functional climate	64.99	11.81	43.99	82.21	.73	-.46	-.67	.68		
6. Stressful climate	52.15	9.64	38.63	66.69	-.66	.82	.96	-.89	-.65	

Note. Culture and climate scores are scaled as *T* scores with a population mean of 50 and standard deviation of 10 ($\mu = 50, \sigma = 10$). Correlations $>|.63|$ are statistically significant at $p < .05$.

symptoms from baseline to the end of the 6.5-month period (see Table 3). Patients served by FQHCs with more proficient cultures improved significantly more from baseline to 6.5-month post-baseline than patients served by FQHCs with less proficient cultures ($B = -0.12, SE = 0.05, p = .020$), resulting in a large effect size of $d = 0.95$ (95% CI = 0.28–1.62) when comparing patients served by centers with highly proficient cultures (average of upper quartile) versus centers with less proficient cultures (average of lower quartile) at 6.5 months (see Figure 1). On average, patients served by health centers with highly proficient cultures improved by 5.08 points from baseline to 6.5-month post-baseline, compared to an average improvement of 0.14 points for patients served by centers with less proficient cultures. This represents a large difference in change scores of $d_{adj} = 0.78$ (95% CI = 0.10–1.45).

Figure 1 shows the covariate-adjusted mean PHQ-9 total scores from baseline to 6.5-month post-baseline for patients served by centers with high versus low proficiency cultures. As is shown in the figure, on average, patients served by centers with highly proficient cultures recovered after 4 months (i.e., PHQ-9 total score < 10) and remained stabilized below the clinical cutoff of 10 at 6.5-month post-baseline ($M = 9.71$); however, patients served by centers with less proficient cultures improved after 4 months but then deteriorated to well above the clinical cutoff by 6.5-month post-baseline ($M = 15.74$). The instantaneous rate of change coefficients (B_1) in Table 3 show exactly *when* outcomes began to diverge for patients served by clinics with more and less proficient cultures. Specifically, at 4-month post-baseline, patients served by clinics with *less* proficient cultures began *deteriorating* significantly faster than patients served by clinics with more proficient cultures ($B = -0.44, SE = 0.17, p = .041$)—a trend that continued to the end of the study and resulted in the large difference in outcomes.

Patients served by FQHCs with more functional climates also improved significantly more from baseline to 6.5-month post-baseline than patients served by FQHCs

with less functional climates ($B = -0.08, SE = 0.04, p = .044$), resulting in a large effect size of $d = 0.81$ (95% CI = 0.32–1.31) when comparing patients served by centers with low versus high functionality climates at 6.5-month post-baseline (see Figure 1). As shown in Figure 1, after adjusting for covariates, patients served by centers with high-functionality climates improved by 5.25 points on average compared to an improvement of 1.12 points among patients served by centers with low-functionality climates. This resulted in a medium difference in change scores of $d_{adj} = 0.65$ (95% CI = 0.16–1.15) at 6.5-month post-baseline. Similar to the results for proficient culture, patients served by centers with both high and low functionality climates improved on average from baseline to ~4-month post-baseline; however, at 4-month post-baseline, patients served by centers with less functional climates began experiencing significantly greater deterioration in depression symptoms compared to patients served by centers with more functional climates ($B = -0.35, SE = 0.13, p = .037$). By 6.5-month post-baseline, patients served by centers with highly functional climates remained below the clinical cutoff ($M = 9.56$) whereas patients served by centers with less functional climates were well above the clinical cutoff ($M = 14.74$). Both the effects of proficient culture and functional climate remained statistically significant after adjusting for multiple comparisons using the Benjamini-Hochberg procedure with a false discovery rate of 0.15. Additional File 1 presents plots of change in depression symptoms for the culture and climate dimensions that were not significantly related to this outcome.

Discussion

In this study of clinical outcomes for low-income patients with maternal depression treated by CoCM in 10 primary care health centers, we found that an assessment of health center culture and climate was associated with variation in patients' improvement in clinical depression from

Table 3
Effects of Culture and Climate on Change in Depression Symptoms (PHQ-9 Total Score)

Culture/climate dimension	Effect on instantaneous growth rate at time t_i			Effect on rate of acceleration across times $t_0 - t_4$			Adjusted effect size at time t_i	
	B_1	SE	p	B_2	SE	p	d_{adj}	95% CI
Proficient culture								
Baseline (t_0)	0.50	0.27	.118	-0.12*	0.05	.020	-0.17	-0.38, 0.03
2 months (t_1)	0.03	0.11	.791				0.24	-0.09, 0.58
4 months (t_2)	-0.44*	0.17	.041				0.05	-0.32, 0.43
6 months (t_3)	-0.90*	0.35	.043				-0.57	-1.12, -0.01
6.5 months (t_4)	-1.01*	0.40	.044				-0.78	-1.45, -0.10
Rigid culture								
Baseline (t_0)	0.56	0.42	.228	-0.05	0.08	.542		
2 months (t_1)	0.37	0.16	.059					
4 months (t_2)	0.18	0.26	.504					
6 months (t_3)	-0.00	0.54	.995					
6.5 months (t_4)	-0.05	0.61	.941					
Resistant culture								
Baseline (t_0)	0.16	0.27	.574	0.02	0.05	.750		
2 months (t_1)	0.22	0.10	.075					
4 months (t_2)	0.28	0.16	.124					
6 months (t_3)	0.35	0.34	.348					
6.5 months (t_4)	0.36	0.38	.385					
Engaged climate								
Baseline (t_0)	0.14	0.22	.552	-0.06	0.04	.123		
2 months (t_1)	-0.11	0.09	.249					
4 months (t_2)	-0.36*	0.13	.037					
6 months (t_3)	-0.60	0.28	.076					
6.5 months (t_4)	-0.66	0.32	.083					
Functional climate								
Baseline (t_0)	0.31	0.22	.223	-0.08*	0.04	.044		
2 months (t_1)	-0.02	0.09	.830					
4 months (t_2)	-0.34*	0.13	.037					
6 months (t_3)	-0.67	0.28	.053					
6.5 months (t_4)	-0.75	0.31	.055					
Stressful climate								
Baseline (t_0)	0.08	0.30	.789	0.03	0.06	.548		
2 months (t_1)	0.22	0.12	.111					
4 months (t_2)	0.35	0.19	.111					
6 months (t_3)	0.48	0.39	.265					
6.5 months (t_4)	0.51	0.44	.290					

Note. Coefficients were estimated using three-level hierarchical linear models with time points ($N = 1,455$) nested within patients, ($N = 468$) nested within health centers ($K = 10$). All models control for center size, level of implementation support, patient age at baseline, race, ethnicity, and prior history of psychiatric treatment. Culture and climate are scaled so that coefficients represent the change in PHQ-9 score for a 10-point change (one population standard deviation) in the culture/climate dimension. Statistically significant B_1 coefficients indicate that culture/climate had an effect on patients' instantaneous growth rate in depression at that specific time t_i ; negative coefficients indicate greater improvement in depression. Statistically significant B_2 coefficients indicate that culture/climate had an effect on patients' overall acceleration, or change, in depression across the entire study period; negative coefficients indicate greater improvement from baseline to the study endpoint. Effects sizes (d_{adj}) represent the standardized mean difference between groups on mean change in PHQ-9 total score from baseline to time t_i (i.e., Cohen's d ; Feingold, 2019).

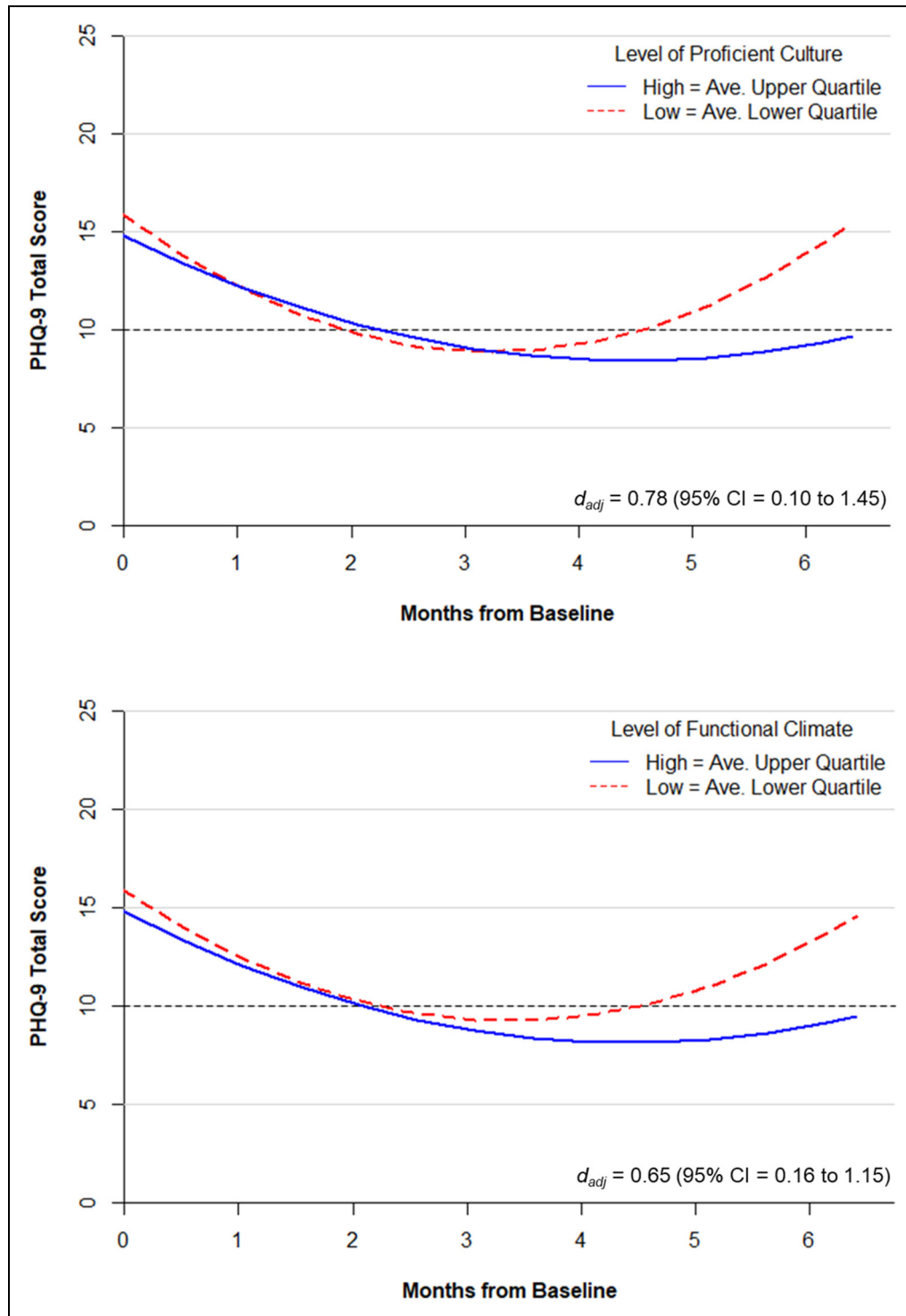
* $p < .05$.

baseline to 6.5-month post-baseline even after controlling for factors known to be associated with outcomes, including level of implementation support (Unützer et al., 2020). Specifically, patients who received CoCM in health centers with highly proficient cultures and highly functional climates had significantly greater reductions in depression

symptoms at 6.5-month post-baseline, compared to other patients, with medium to large adjusted effect sizes of $d = 0.78$ and $d = 0.65$, respectively. These results offer some of the first evidence linking organizational culture and climate to variation in the clinical outcomes of evidence-based practices implemented in primary care.

Figure 1

Relations Between Proficient Organizational Culture and Functional Climate With Change in Patient Depression



Note. $N = 1,455$ observations, $N = 468$ patients, $K = 10$ FQHCs. Low = average value of culture/climate dimension among centers in the lower quartile. High = average value of the culture/climate dimension among centers in the upper quartile. Lines show the covariate-adjusted, average change in depression symptoms for patients treated by health centers with high vs. low values of proficient culture or functional climate from baseline to 6.5-month post-baseline. Effects sizes (d_{adj}) represent the standardized mean difference between groups in mean change in PHQ-9 total score from baseline to 6.5-month post-baseline (i.e., Cohen's d ; Feingold, 2019).

Given the prominence of culture and climate within implementation frameworks, these findings represent an important advance for implementation theory and practice.

Importantly, patients served by centers with more proficient cultures and more functional climates only began exhibiting superior clinical outcomes at 4-month post-baseline. Prior to 4 months, patients in all centers improved, on average, to depression levels below the clinical cutoff; however, beginning at 4-month post-baseline, patients served by centers with less proficient cultures and less functional climates began to worsen significantly, resulting in a large difference in depression outcomes at 6.5-month post-baseline. Further work is needed to replicate this finding and to identify the mechanism for this difference; however, one possibility is that culture and climate influenced clinical outcomes by shaping fidelity to CoCM, particularly during the maintenance phase of treatment. The CoCM relies on systematic, proactive, and sustained follow-up of patients to ensure steady improvement; it is plausible that this varied by health center culture and climate. For example, health centers with proficient cultures may have had stronger procedures in place to ensure that patient follow-up assessments were completed and may have offered higher informal and formal incentives for team members to take responsive actions based on the assessments, compared to centers with less proficient cultures. Similarly, coordination and teamwork may have been better within multidisciplinary teams with functional climates, resulting in higher completion of follow-up assessments and actions. Future research should examine the mechanisms through which organizational culture and climate relate to CoCM clinical outcomes.

Most studies of variation in clinical outcomes of CoCM have focused on patient-level characteristics or policy interventions to incentivize care delivery as predictors (Bauer et al., 2012; Huang et al., 2012). The health centers included in our study varied in the rates of clinically significant patient improvement by approximately three-fold, from 17% to 57%, showing the need to look at health center-level factors influencing outcomes for this complex intervention for chronic disease. Given the difficulty of implementing CoCM, including creating new clinical relationships and workflows across distinct providers and staff, it is not surprising that patient improvement was associated with health center-level measures of culture and climate even after adjusting for a range of patient and health center characteristics. A high level of organizational support is needed to carry out the practice changes required to implement and sustain CoCM.

Our findings are consistent with other research showing that proficient culture and functional climate explain variation in implementation (Beidas et al., 2015; Beidas et al., 2019; Williams & Glisson, 2020) and clinical (Williams & Glisson, 2014) outcomes across a range of medical and social service settings. However, the linkage of only two of six culture and climate dimensions with

CoCM outcomes raises the question, “Why?” One possible explanation is conceptual proximity. Of the six dimensions studied, proficient culture is the only one that directly addresses clinicians’ perceptions of the extent to which they are expected to prioritize patient well-being (i.e., clinical effectiveness) and their own competence in effective treatment models (i.e., implementation of CoCM) *over and above* other organizational priorities. The fact that experimentally-induced improvement in proficient culture has been shown to explain improvement in clinicians’ adoption of evidence-based practices lends credence to this hypothesis (Williams et al., 2017). The nature of the CoCM intervention functionality was an important dimension of climate. Clinicians in functional climates receive the support and cooperation they need from colleagues to effectively do their jobs and have a clear understanding of how their role fits within the organization. These conditions are essential to the effective implementation of the team-based CoCM model.

The results from this study suggest that the implementation of evidence-based practices such as CoCM for depression may be improved by implementation strategies that generate proficient cultures and functional climates. Proficient cultures and functional climates can be developed in social service and specialty mental health organizations through organizational interventions that engage treatment staff in collaborative monitoring of quality and outcomes to identify and remove service barriers (Glisson et al., 2013; Glisson, Hemmelgarn, et al., 2012). Prior research indicates that proficient organizational culture can be improved within 2 years (Williams et al., 2017) and functional climate within 1 year (Glisson et al., 2012).

While the current study supports the linkage between health center culture and climate and variation in patient outcomes of CoCM, there are a number of limitations. First, this is an observational study that does not provide a basis for causal inference. Clinical outcomes of CoCM are undoubtedly influenced by many factors, and it was not possible to control for all of these variables in the present study. Additional research that replicates these findings and rules out alternative confounds—perhaps by including organizational culture and climate as a treatment moderator in an implementation-effectiveness trial of CoCM for depression—is needed. Second, the OSC measure was assessed at one time in the middle of the study period, which raises issues related to temporality. Although the measurement of culture/climate did not precede all measurements of patients’ clinical outcomes in time, there is evidence that culture and climate can be stable in the absence of intervention; consequently, we believe an argument can be made that the cultures and climates of these clinics were in place prior to the actual measurement period and remained in place during the months following measurement. Furthermore, the use of a single measurement may actually increase the robustness of the results as remote measurement of culture/climate

relative to clinical outcomes should reduce (not increase) the likelihood of statistically significant relations between these variables. Third, while the results of this study must be limited to the patient population and health centers included in the sample, we believe that these findings are likely to generalize to patients with common mental disorders and other chronic diseases. The evidence-based principles of effective chronic care of diseases are reflected in the CoCM and likely apply to other disorders. Fourth, the study included a small health center sample ($N = 10$) and a relatively small patient sample ($N = 468$); larger studies are needed to confirm this work. Fifth, our use of a clinical dataset with unstandardized assessment periods for each patient makes it impossible to rule out potential confounds associated with missing data patterns even though we included covariates in the models to address this concern. Studies are needed to replicate these findings using structured research designs with standardized follow-up periods.

Conclusion

This study tested whether variation in the clinical outcomes of CoCM implemented in primary care health centers was associated with variation in the centers' cultures and climates. We found that parenting women with depression were more likely to experience clinical improvement and maintain this improvement if they were treated in health centers with higher levels of proficient culture and functional climate even after adjusting for other center and patient characteristics associated with outcomes. The results of this study highlight the potential value of implementation strategies that target proficient organizational culture and functional climate, improving depression outcomes in health centers.

List of abbreviations

CoCM	collaborative care model
ICC	intraclass correlation coefficient
OSC	organizational social context
PHQ-9	patient health questionnaire-9



Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Washington University Implementation Research Institute, Eitel Foundation.

ORCID iDs

Melinda Vredevoogd  <https://orcid.org/0000-0001-6278-8524>
Ian M. Bennett  <https://orcid.org/0000-0002-7139-9456>

Supplemental material

Supplemental material for this article is available online.

References

- Bauer, A. M., Azzone, V., Alexander, L., Goldman, H. H., Unützer, J., & Frank, R. G. (2012). Are patient characteristics associated with quality of depression care and outcomes in collaborative care programs for depression? *General Hospital Psychiatry, 34*(1), 1–8. <https://doi.org/10.1016/j.genhosppsych.2011.08.019>
- Beidas, R. S., Marcus, S., Aarons, G. A., Hoagwood, K. E., Schoenwald, S., Evans, A. C., Hurford, M. O., Hadley, T., Barg, F. K., Walsh, L. M., Adams, D. R., & Mandell, D. S. (2015). Predictors of community therapists' use of therapy techniques in a large public mental health system. *JAMA Pediatrics, 169*(4), 374–382. <https://doi.org/10.1001/jamapediatrics.2014.3736>
- Beidas, R. S., Williams, N. J., Becker-Haimes, E. M., Aarons, G. A., Barg, F. K., Evans, A. C., Jackson, K., Jones, D., Hadley, T., Hoagwood, K., Marcus, S. C., Neimark, G., Rubin, R. M., Schoenwald, S. K., Adams, D. R., Walsh, L. M., Zentgraf, K., & Mandell, D. S. (2019). A repeated cross-sectional study of clinicians' use of psychotherapy techniques during 5 years of a system-wide effort to implement evidence-based practices in Philadelphia. *Implementation Science, 14*(1), 67. <https://doi.org/10.1186/s13012-019-0912-4>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bennett, I. M., Marcus, S. C., Palmer, S. C., & Coyne, J. C. (2010). Pregnancy-related discontinuation of antidepressants and depression care visits among Medicaid recipients. *Psychiatric Services, 61*(4), 386–391. <https://doi.org/10.1176/ps.2010.61.4.386>
- Bennett, I. M., Palmer, S., Marcus, S., Nicholson, J. M., Hantsoo, L., Bellamy, S., Rinaldi, J., & Coyne, J. C. (2009). "One end has nothing to do with the other:" Patient attitudes regarding help seeking intention for depression in gynecologic and obstetric settings. *Archives of Women's Mental Health, 12*(5), 301–308. <https://doi.org/10.1007/s00737-009-0103-4>
- Bennett, I. M., Schott, W., Krutikova, S., & Behrman, J. R. (2016). Maternal mental health, and child growth and development, in four low-income and middle-income countries. *Journal of Epidemiology and Community Health, 70*(2), 168–173. <https://doi.org/10.1136/jech-2014-205311>
- Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 349–381). Jossey-Bass.
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of Applied Psychology, 83*(2), 234–246. <https://doi.org/10.1037/0021-9010.83.2.234>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. <https://doi.org/10.4324/9780203771587>

- Croghan, T. W., Schoenbaum, M., Sherbourne, C. D., & Koegel, P. (2006). A framework to improve the quality of treatment for depression in primary care. *Psychiatric Services, 57*(5), 623–630. <https://doi.org/10.1176/appi.ps.57.5.623>
- Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*(9), 917–928. <https://doi.org/10.1093/jpepsy/jsp004>
- Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science, 1*(1), 1. <https://doi.org/10.1186/1748-5908-1-1>
- Feingold, A. (2019). Time-varying effect sizes for quadratic growth models in multilevel and latent growth modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(3), 418–429. <https://doi.org/10.1080/10705511.2018.1547110>
- Ferrari, A. J., Norman, R. E., Freedman, G., Baxter, A. J., Pirkis, J. E., Harris, M. G., Page, A., Carnahan, E., Degenhardt, L., Vos, T., & Whiteford, H. A. (2014). The burden attributable to mental and substance use disorders as risk factors for suicide: Findings from the Global Burden of Disease Study 2010. *PloS One, 9*(4), e91936. <https://doi.org/10.1371/journal.pone.0091936>
- Gaynes, B. N., Gavin, N., Meltzer-Brody, S., Lohr, K. N., Swinson, T., Gartlehner, G., Brody, S., & Miller, W. C. (2005). Perinatal depression: Prevalence, screening accuracy, and screening outcomes. *Evidence Report/Technology Assessment (Summary), 119*, 1–8. <https://doi.org/10.1037/e439372005-001>
- Gilbody, S., Bower, P., Fletcher, J., Richards, D., & Sutton, A. J. (2006). Collaborative care for depression: A cumulative meta-analysis and review of longer-term outcomes. *Archives of Internal Medicine, 166*(21), 2314–2321. <https://doi.org/10.1001/archinte.166.21.2314>
- Gilbody, S., Bower, P., & Whitty, P. (2006). Costs and consequences of enhanced primary care for depression: Systematic review of randomised economic evaluations. *British Journal of Psychiatry, 189*, 297–308. <https://doi.org/10.1192/bjp.bp.105.016006>
- Gjerdingen, D., Crow, S., McGovern, P., Miner, M., & Center, B. (2009). Postpartum depression screening at well-child visits: Validity of a 2-question screen and the PHQ-9. *Annals of Family Medicine, 7*(1), 63–70. <https://doi.org/10.1370/afm.933>
- Glisson, C., & Green, P. (2011). Organizational climate, services, and outcomes in child welfare systems. *Child Abuse & Neglect, 35*(8), 582–591. <https://doi.org/10.1016/j.chiabu.2011.04.009>
- Glisson, C., Green, P., & Williams, N. J. (2012). Assessing the Organizational Social Context (OSC) of child welfare systems: Implications for research and practice. *Child Abuse & Neglect, 36*(9), 621–632. <https://doi.org/10.1016/j.chiabu.2012.06.002>
- Glisson, C., Hemmelgarn, A., Green, P., Dukes, D., Atkinson, S., & Williams, N. J. (2012). Randomized trial of the Availability, Responsiveness, and Continuity (ARC) organizational intervention with community-based mental health programs and clinicians serving youth. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(8), 780–787. <https://doi.org/10.1016/j.jaac.2012.05.010>
- Glisson, C., Hemmelgarn, A., Green, P., & Williams, N. J. (2013). Randomized trial of the Availability, Responsiveness and Continuity (ARC) organizational intervention for improving youth outcomes in community mental health programs. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(5), 493–500. <https://doi.org/10.1016/j.jaac.2013.02.005>
- Glisson, C., Landsverk, J., Schoenwald, S., Kelleher, K., Hoagwood, K. E., Mayberg, S., & Green, P., & Research Network on Youth Mental, H. (2008). Assessing the organizational social context (OSC) of mental health services: Implications for research and practice. *Administration and Policy in Mental Health and Mental Health Services Research, 35*(1–2), 98–113. <https://doi.org/10.1007/s10488-007-0148-5>
- Glisson, C., Schoenwald, S. K., Hemmelgarn, A., Green, P., Dukes, D., Armstrong, K. S., & Chapman, J. E. (2010). Randomized trial of MST and ARC in a two-level EBT implementation strategy. *Journal of Consulting and Clinical Psychology, 78*(4), 537–550. <https://doi.org/10.1037/a0019160>
- Glisson, C., & Williams, N. J. (2015). Assessing and changing organizational social contexts for effective mental health services. *Annual Review of Public Health, 36*, 507–523. <https://doi.org/10.1146/annurev-publhealth-031914-122435>
- Glisson, C., Williams, N. J., Hemmelgarn, A., Proctor, E., & Green, P. (2016a). Aligning organizational priorities with ARC to improve youth mental health service outcomes. *Journal of Consulting and Clinical Psychology, 84*(8), 713–725. <https://doi.org/10.1037/ccp0000107>
- Glisson, C., Williams, N. J., Hemmelgarn, A., Proctor, E., & Green, P. (2016b). Increasing clinicians' EBT exploration and preparation behavior in youth mental health services by changing organizational culture with ARC. *Behaviour Research and Therapy, 76*, 40–46. <https://doi.org/10.1016/j.brat.2015.11.008>
- Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T., & Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *The Journal of Clinical Psychiatry, 76*(2), 155–162. <https://doi.org/10.4088/JCP.14m09298>
- Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. John Wiley & Sons.
- Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24*(5), 623–641. [https://doi.org/10.1016/S0149-2063\(99\)80077-4](https://doi.org/10.1016/S0149-2063(99)80077-4)
- Huang, H., Chan, Y. F., Katon, W., Tabb, K., Sieu, N., Bauer, A. M., Wasse, J. K., & Unützer, J. (2012). Variations in depression care and outcomes among high-risk mothers from different racial/ethnic groups. *Family Practice, 29*(4), 394–400. <https://doi.org/10.1093/fampra/cmr108>
- Huang, H., Russo, J., Bauer, A. M., Chan, Y. F., Katon, W., Hogan, D., & Unützer, J. (2013). Depression care and treatment in a chronically ill Medicare population. *General Hospital Psychiatry, 35*(4), 382–386. <https://doi.org/10.1016/j.genhosppsych.2013.02.006>
- James, L. R., Demaree, R. G., & Wolf, G. (1993). Rwg: An assessment of within-group interrater agreement. *Journal of Applied Psychology, 78*(2), 306–309. <https://doi.org/10.1037/0021-9010.78.2.306>
- Katon, W., & Unützer, J. (2006). Collaborative care models for depression: Time to move from evidence to practice.

- Archives of Internal Medicine*, 166(21), 2304–2306. <https://doi.org/10.1001/archinte.166.21.2304>
- Katon, W., Unützer, J., Wells, K., & Jones, L. (2010). Collaborative depression care: History, evolution and ways to enhance dissemination and sustainability. *General Hospital Psychiatry*, 32(5), 456–464. <https://doi.org/10.1016/j.genhosppsych.2010.04.001>
- Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, 19(2), 195–229. <https://doi.org/10.5465/amr.1994.9410210745>
- Kramer, T. L., Drummond, K. L., Curran, G. M., & Fortney, J. C. (2017). Assessing culture and climate of federally qualified health centers: A plan for implementing behavioral health interventions. *Journal of Health Care for the Poor and Underserved*, 28(3), 973–987. <https://doi.org/10.1353/hpu.2017.0091>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9), 606–613. <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815–852. <https://doi.org/10.1177/1094428106296642>
- Li, S.-A., Jeffs, L., Barwick, M., & Stevens, B. (2018). Organizational contextual features that influence the implementation of evidence-based practices across healthcare settings: A systematic integrative review. *Systematic Reviews*, 7(1), 72. <https://doi.org/10.1186/s13643-018-0734-5>
- Liao, J. M., Navathe, A. S., & Press, M. J. (2019). Medicare's approach to paying for services that promote coordinated care. *JAMA*, 321(2), 147–148. <https://doi.org/10.1001/jama.2018.19315>
- McMillan, D., Gilbody, S., & Richards, D. (2010). Defining successful treatment outcome in depression using the PHQ-9: A comparison of methods. *Journal of Affective Disorders*, 127(1–3), 122–129. <https://doi.org/10.1016/j.jad.2010.04.030>
- Melville, J. L., Gavin, A., Guo, Y., Fan, M.-Y., & Katon, W. J. (2010). Depressive disorders during pregnancy: Prevalence and risk factors in a large urban sample. *Obstetrics and Gynecology*, 116(5), 1064–1070. <https://doi.org/10.1097/AOG.0b013e3181f60b0a>
- Melville, J. L., Reed, S. D., Russo, J., Croicu, C. A., Ludman, E., LaRocco-Cockburn, A., & Katon, W. (2014). Improving care for depression in obstetrics and gynecology: A randomized controlled trial. *Obstetrics and Gynecology*, 123(6), 1237–1246. <https://doi.org/10.1097/AOG.0000000000000231>
- National Committee for Quality Assurance. (n.d.). *HEDIS depression measures for electronic clinical data*. <https://www.ncqa.org/hedis/the-future-of-hedis/hedis-depression-measures-for-electronic-clinical-data/>
- Olin, S. S., Williams, N., Pollock, M., Armusewicz, K., Kutash, K., Glisson, C., & Hoagwood, K. E. (2014). Quality indicators for family support services and their relationship to organizational social context. *Administration and Policy in Mental Health and Mental Health Services Research*, 41(1), 43–54. <https://doi.org/10.1007/s10488-013-0499-z>
- Pilowsky, D. J., Wickramaratne, P. J., Rush, A. J., Hughes, C. W., Garber, J., Malloy, E., King, C. A., Cerda, G., Sood, A. B., Alpert, J. E., Wisniewski, S. R., Trivedi, M. H., Talati, A., Carlson, M. M., Liu, H. H., Fava, M., & Weissman, M. M. (2006). Children of currently depressed mothers: A STAR*D ancillary study. *The Journal of Clinical Psychiatry*, 67(1), 126–136. <https://doi.org/10.4088/jcp.v67n0119>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunker, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65–76. <https://doi.org/10.1007/s10488-010-0319-7>
- Ratzliff, A. D., Ni, K., Chan, Y. F., Park, M., & Unützer, J. (2013). A collaborative care approach to depression treatment for Asian Americans. *Psychiatric Services*, 64(5), 487–490. <https://doi.org/10.1176/appi.ps.001742012>
- Raudenbush, S. W. (2004). *HLM 6: Hierarchical linear and non-linear modeling*. Scientific Software International.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.
- Regier, D. A., Narrow, W. E., Rae, D. S., Manderscheid, R. W., Locke, B. Z., & Goodwin, F. K. (1993). The de facto US mental and addictive disorders service system. Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. *Archives of General Psychiatry*, 50(2), 85–94. <https://doi.org/10.1001/archpsyc.1993.01820140007001>
- Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior*, 7, 1–37.
- Smith, S. N., Almirall, D., Prenovost, K., Goodrich, D. E., Abraham, K. M., Liebrecht, C., & Kilbourne, A. M. (2018). Organizational culture and climate as moderators of enhanced outreach for persons with serious mental illness: Results from a cluster-randomized trial of adaptive implementation strategies. *Implementation Science*, 13(1), 93. <https://doi.org/10.1186/s13012-018-0787-9>
- Solberg, L. I., Crain, A. L., Jaeckels, N., Ohnsorg, K. A., Margolis, K. L., Beck, A., Whitebird, R. R., Rossom, R. C., Crabtree, B. F., & Van de Ven, A. H. (2013). The DIAMOND initiative: Implementing collaborative care for depression in 75 primary care clinics. *Implementation Science*, 8, 135. <https://doi.org/10.1186/1748-5908-8-135>
- Spitzer, R. L., Williams, J. B., Kroenke, K., Hornyak, R., & McMurray, J. (2000). Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: The PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *American Journal of Obstetrics and Gynecology*, 183(3), 759–769. <https://doi.org/10.1067/mob.2000.106580>
- Stein, A., Pearson, R. M., Goodman, S. H., Rapa, E., Rahman, A., McCallum, M., Howard, L. M., & Pariante, C. M. (2014). Effects of perinatal mental disorders on the fetus and child. *The Lancet*, 384(9956), 1800–1819. [https://doi.org/10.1016/S0140-6736\(14\)61277-0](https://doi.org/10.1016/S0140-6736(14)61277-0)
- Unützer, J. (2008). Evidence-based treatments for anxiety and depression: Lost in translation? *Depression and Anxiety*, 25(9), 726–729. <https://doi.org/10.1002/da.20529>
- Unützer, J., Carlo, A. C., Arao, R., Vredevoogd, M., Fortney, J., Powers, D., & Russo, J. (2020). Variation in the effectiveness of collaborative care for depression: Does it matter where you

- get your care? *Health Affairs*, 39(11), 1943–1950. <https://doi.org/10.1377/hlthaff.2019.01714>
- Unützer, J., Choi, Y., Cook, I. A., & Oishi, S. (2002). Clinical computing: A web-based data management system to improve care for depression in a multicenter clinical trial. *Psychiatric Services*, 53(6), 671–678. <https://doi.org/10.1176/ps.53.6.671>
- Unützer, J., Katon, W. J., Fan, M.-Y., Schoenbaum, M. C., Lin, E. H. B., Della Penna, R. D., & Powers, D. (2008). Long-term cost effects of collaborative care for late-life depression. *The American Journal of Managed Care*, 14(2), 95–100.
- Verbeke, G., & Molenberghs, G. (2009). *Linear mixed models for longitudinal data*. Springer Science & Business Media.
- Vesga-López, O., Blanco, C., Keyes, K., Olfson, M., Grant, B. F., & Hasin, D. S. (2008). Psychiatric disorders in pregnant and postpartum women in the United States. *Archives of General Psychiatry*, 65(7), 805–815. <https://doi.org/10.1001/archpsyc.65.7.805>
- Wells, K. B., Miranda, J., & Gonzalez, J. J. (2002). Overcoming barriers and creating opportunities to reduce burden of affective disorders: A new research agenda: Introduction. *Mental Health Services Research*, 4(4), 175–178. <https://doi.org/10.1023/A:1020947829820>
- Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., Charlson, F. J., Norman, R. E., Flaxman, A. D., Johns, N., Burstein, R., Murray, C. J., & Vos, T. (2013). Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, 382(9904), 1575–1586. [https://doi.org/10.1016/S0140-6736\(13\)61611-6](https://doi.org/10.1016/S0140-6736(13)61611-6)
- Williams, N. J., & Beidas, R. S. (2018). Navigating the storm: How proficient organizational culture promotes clinician retention in the shift to evidence-based practice. *PLoS ONE*, 13(12). <https://doi.org/10.1371/journal.pone.0209745>
- Williams, N. J., Ehrhart, M. G., Aarons, G. A., Marcus, S. C., & Beidas, R. S. (2018). Linking molar organizational climate and strategic implementation climate to clinicians' use of evidence-based psychotherapy techniques: Cross-sectional and lagged analyses from a 2-year observational study. *Implementation Science*, 13(1), 85. <https://doi.org/10.1186/s13012-018-0781-2>
- Williams, N. J., Frank, H. E., Frederick, L., Beidas, R. S., Mandell, D. S., Aarons, G. A., Green, P., & Locke, J. (2019). Organizational culture and climate profiles: Relationships with fidelity to three evidence-based practices for autism in elementary schools. *Implementation Science*, 14(1), 15. <https://doi.org/10.1186/s13012-019-0863-9>
- Williams, N. J., & Glisson, C. (2013). Reducing turnover is not enough: The need for proficient organizational cultures to support positive youth outcomes in child welfare. *Children and Youth Services Review*, 35(11), 1871–1877. <https://doi.org/10.1016/j.childyouth.2013.09.002>
- Williams, N. J., & Glisson, C. (2014). Testing a theory of organizational culture, climate and youth outcomes in child welfare systems: A United States national study. *Child Abuse & Neglect*, 38(4), 757–767. <https://doi.org/10.1016/j.chiabu.2013.09.003>
- Williams, N. J., & Glisson, C. (2020). Changing organizational social context to support evidence-based practice implementation: A conceptual and empirical review. In *Implementation science 3.0* (pp. 145–172). Springer. https://doi.org/10.1007/978-3-030-03874-8_6
- Williams, N. J., Glisson, C., Hemmelgarn, A., & Green, P. (2017). Mechanisms of change in the ARC organizational strategy: Increasing mental health clinicians' EBP adoption through improved organizational culture and capacity. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(2), 269–283. <https://doi.org/10.1007/s10488-016-0742-5>
- Woolston, J. L. (2005). Implementing evidence-based treatments in organizations. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44(12), 1313–1316. <https://doi.org/10.1097/01.chi.0000181041.21584.7a>
- Yu, H., Jiang, S., & Land, K. C. (2015). Multicollinearity in hierarchical linear models. *Social Science Research*, 53, 118–136. <https://doi.org/10.1016/j.ssresearch.2015.04.008>