Development and Validation of the Modified Patient-Centered Medical Home Assessment for the Comprehensive Primary Care Initiative

Dmitriy Poznyak , Deborah N. Peikes, Breanna A. Wakar, Randall S. Brown, and Robert J. Reid

Objective. To describe the modified Patient-Centered Medical Home Assessment (M-PCMH-A) survey module developed to track primary care practices' care delivery approaches over time, assess whether its underlying factor structure is reliable, and produce factor scores that provide a more reliable summary measure of the practice's care delivery than would a simple average of question responses.

Data Sources/Study Setting. Survey data collected from diverse practices participating in the Comprehensive Primary Care (CPC) initiative in 2012 ($n = 497$) and 2014 ($n = 493$) and matched comparison practices in 2014 ($n = 423$).

Study Design. Confirmatory factor analysis.

Data Collection. Thirty-eight questions organized in six domains: Access and Continuity of Care, Planned Care for Chronic Conditions and Preventive Care, Risk-Stratified Care Management, Patient and Caregiver Engagement, Coordination of Care across the Medical Neighborhood, and Continuous Data-Driven Improvement.

Principal Findings. Confirmatory factor analysis suggested using seven factors (splitting one domain into two), reassigning two questions to different domain factors, and removing one question, resulting in high reliability, construct validity, and stability in all but one factor. The seven factors together formed a single higher-order factor summary measure. Factor scores guard against potential biases from equal weighting.

Conclusions. The M-PCMH-A can validly and reliably track primary care delivery across practices and over time using factors representing seven key components of care as well as an overall score. Researchers should calculate factor loadings for their specific data if possible, but average scores may be suitable if they cannot use factor analysis due to resource or sample constraints.

Key Words. Psychometrics, survey research and questionnaire design, quality of care/patient safety, quality improvement/report cards

Payers around the country are testing the patient-centered medical home (PCMH) and similar models to improve primary care delivery. Accurately measuring how primary care practices are transforming care delivery is critical to tracking progress and areas for improvement in this rapidly changing landscape. One of the largest efforts to test the effects of these models is the Comprehensive Primary Care (CPC) initiative, which the Centers for Medicare & Medicaid Services (CMS) launched in 2012 in collaboration with 39 public and private payers. CPC provides 497 participating practices with enhanced payment, learning support and technical assistance, and data feedback to help them make required improvements in five functions of primary care delivery: (1) Access and Continuity of Care, (2) Planned Care for Chronic Conditions and Preventive Care, (3) Risk-Stratified Care Management, (4) Patient and Caregiver Engagement, and (5) Coordination of Care across the Medical Neighborhood (the providers caring for the practice's patients). CMS expects practices to use continuous data-driven improvement to support these changes. CPC requires that practices meet annual Milestones to guide progress in delivering CPC's functions. (Taylor et al. 2015; Dale et al. 2016; Peikes et al. 2016 provide details of the initiative and interim evaluation results.)

Our evaluation of CPC estimates the nature and degree of participating practices' improvements in key aspects of care delivery, and how—if at all—such changes are associated with improved patient outcomes. We do this by analyzing Medicare claims data; surveys of practices, patients, clinicians, and staff; and qualitative data collected from practices, payers, patients, and stakeholders. The evaluation uses a quasi-experimental comparison group to estimate differences in care delivery and outcomes between CPC practices and otherwise similar practices. Although several existing PCMH recognition tools, survey instruments, and practice self-improvement tools measure primary care delivery, none completely covers the functions that CPC targets (Burton, Devers, and Berenson 2012). Many also use binary response categories, which do not support tracking of incremental improvements over time. This article describes the development of the modified Patient-Centered Medical Home Assessment (M-PCMH-A), a module in the survey instrument that

Address correspondence to Dmitriy Poznyak, Ph.D., Mathematica Policy Research, 600 Alexander Park Drive, Princeton, NJ 08540; e-mail: dpoznyak@mathematica-mpr.com. Deborah N. Peikes, Ph.D., M.P.A., and Randall S. Brown, Ph.D., are with Mathematica Policy Research, Princeton, NJ. Breanna A. Wakar, M.A., is with Mathematica Policy Research, Washington, DC. Robert J. Reid, M.D., Ph.D., M.P.H., is with Trillium Health Partners and Institute for Health Policy, Management and Evaluation, University of Toronto, Mississauga, ON, Canada.

we administered to CPC and comparison practices to measure changes in care delivery approaches. We then analyze the factor structure across the two groups of practices and over time among CPC practices. Finally, we measure practices' performance on the domain factors and overall using factor scores based on the factor loadings estimated on the first round of data collected from the CPC practices. These factor scores enable us to measure change over time and between the CPC and comparison practices.

METHODS

Measure Development

We developed the M-PCMH-A in 2012 to capture practices' self-reported approaches to delivering various aspects of primary care over time. We reviewed literature and existing instruments used to assess primary care delivery and solicited input from physicians, health and survey researchers, and CMS's CPC implementation team. The module in the first survey round contained 41 questions, of which only 38 were included consistently in subsequent survey rounds and therefore can be used to measure change over time. Twenty-six of the 38 questions in the M-PCMH-A were adopted verbatim or slightly edited from the PCMH-A instrument (v.1.3) developed by the MacColl Center for Health Care Innovation at the Group Health Research Institute and Qualis Health for the Safety Net Medical Home Initiative (Daniel et al. 2013; Safety Net Medical Home Initiative 2014). A study of that initiative found high agreement between 65 clinics' self-reported PCMH-A scores and those of external practice coaches. It also found positive correlations between the PCMH-A scores and structured, independent assessments of practice transformation capacity by the same coaches, the percentage of quality improvement activities underway reported by the clinic, and early achievement of NCQA PCMH recognition (Daniel et al. 2013). To more closely measure the aspects of the five CPC functions and continuous data-driven improvement, an additional focus of CPC, we changed the order and domain assignment for some of these questions. Because the PCMH-A did not cover all aspects of primary care delivery relevant to the CPC model, we added 12 new questions developed in-house or adapted from other practice assessment tools. Table 1 summarizes each question and its source. Although the M-PCMH-A domains and questions do not map exactly with the CPC Milestones or functions, these areas are consistent with them, cover

approaches that prior studies suggest are important to primary care redesign, and can be used to track progress.

Practices were asked to select the point value that best describes the level of care the practice currently provides. Following the PCMH-A's format, responses to each M-PCMH-A question are coded on a 12-point scale, divided into four ranges (1–3, 4–6, 7–9, and 10–12), with higher ratings reflecting better approaches to care delivery. To guide respondents on how to score themselves, the instrument described the behavior or activities expected for each of the four ranges for each of the questions (Mathematica Policy Research 2012, 2014). Question A17 appears below, and the full 2014 version of the module is available in Appendix S1.

Clinical care management services for high-risk patients	... are not available.			... are provided by external care managers with limited connection to the practice.			... are provided by external care managers who regularly communicate with the care team.			... are systematically provided by care managers functioning as members of the practice team.		
	1	2	3	4	5	6	7	8	9	10	11	12

CPC and Comparison Practices

CMS selected approximately 500 practices from applicants in seven regions, based largely on practices' use of health information technology. These regions include four states (Arkansas, Colorado, New Jersey, and Oregon) and portions of three states (New York's Capital District-Hudson Valley region, Ohio-Kentucky's Cincinnati-Dayton region, and Oklahoma's Greater Tulsa region). The evaluation identified nearby comparison regions and then used propensity score matching to select from those regions and from unselected applicants a set of comparison practices with similar pre-CPC practice-, market-, and patient-level characteristics (Taylor et al. 2015; Dale et al. 2016; Peikes et al. 2016). Table S1 reports the similarity of the CPC and comparison practices.

Survey Administration and Study Sample

We surveyed CPC practices from October through December 2012, at the start of CPC. Comparison practices were not selected until July 2013; thus, we

Table 1: Questions and Factors in the Factor-Analyzed M-PCMH-A, Descriptive Statistics, and Factor Loadings for CPC Practices (2012 Survey Data; N= 497); Sources for Questions

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Question Loading on a Factor (Factor Loading)	Reliability Weights Derived from Factor Analysis		Sum-Score Weights (Simple Average)		
					Question Weight in a Factor (Normalized to Sum to 1 within a Factor)	Question Weight in the Total Score (Normalized to Sum to 1 within a Factor)	Simple Average Weight for First-Order Factor (Normalized to Sum to 1 within a Factor)	Simple Average Weight in the Total Score (Normalized to Sum to 1 within a Factor)	Source
Continuity of Care	1. Patient assignment to specific provider and use of that assignment to schedule and monitor supply and demand	9.31	2.46	0.819	0.472	0.014	0.500	0.071	PCMH-A, Item 1, verbatim
	2. The extent to which patients are encouraged to, and usually see their own provider and practice team	9.96	2.16	0.918	0.528	0.016	0.500	0.071	PCMH-A, Item 5, modified
Access to Care	3. Flexibility of appointment systems for different length and same-day visits	10.25	1.64	0.402	0.288	0.038	0.333	0.048	PCMH-A, Item 20, verbatim
	4. Asynchronous communication with practice team including patients' preferred mode	4.27	2.85	0.585	0.419	0.056	0.333	0.048	Developed by Mathematica

Continued

Table 1: Continued

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Reliability Weights Derived from Factor Analysis			Sum-Score Weights (Simple Average)		
				Question Loading on a Factor (Factor Loading)	Question Weight in a Factor (Normalized to Sum to 1 within a Factor)	Question Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Simple Average Weight for First-Order Factor		
							Simple Average Weight in the Total Score (Normalized to Sum to 1 within Total Score)		
	10. The extent to which evidence-based reminders to providers are specific to the individual patient encounter	7.47	2.83	0.637	0.195	0.035	0.167	0.024	TransformMED, Item 31, modified
	11. Extent of role of non-physician practice team members in providing clinical care	8.28	2.74	0.547	0.167	0.030	0.167	0.024	PCMH-A, Item 6, verbatim
	12. Extent to which medication reconciliation is regularly done and documented in the patient's medical record	10.20	1.83	0.304	0.093	0.017	0.167	0.024	TransformMED, Item 36, modified

Continued

Table 1: Continued

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Question Loading on a Factor (Factor Loading)	Reliability Weights Derived from Factor Analysis		Sum-Score Weights (Simple Average)		
					Question Weight in a Factor (Normalized to Sum to 1 within a Factor)	Question Total Score (Normalized to Sum to 1 within Total Score)	Simple Average Weight for First-Order Factor (Normalized to Sum to 1 within a Factor)	Simple Average Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Source
Risk-Stratified Care Management	16. Degree to which a standard method or tool to stratify patients by risk level is used and guides care delivery	3.77	2.75	0.724	0.353	0.059	0.333	0.048	Developed by Mathematica
	17. The provision of clinical care management services for high-risk patients by care managers integrated into the practice team	4.76	3.52	0.703	0.342	0.057	0.333	0.048	PCMH-A, Item 33, modified
	18. The availability of registry or panel-level data to assess and manage care for practice populations	5.41	3.15	0.626	0.305	0.051	0.333	0.048	PCMH-A, Item 2, verbatim
Patient and Caregiver Engagement	19. Assessment and incorporation of patient and family preferences in planning and organizing care	6.55	2.82	0.573	0.148	0.026	0.167	0.024	PCMH-A, Item 8, verbatim

Continued

Table 1: Continued

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Reliability Weights Derived from Factor Analysis			Sum-Score Weights (Simple Average)		
				Question Loading on a Factor (Factor Loading)	Question Weight in a Factor (Normalized to Sum to 1 within a Factor)	Question Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Simple Average Weight for First-Order Factor (Normalized to Sum to 1 within a Factor)	Simple Average Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Source
	20. How systematically practice teams involve patients in decision making	6.92	2.29	0.739	0.191	0.034	0.167	0.024	PCMH-A, Item 9, verbatim
	21. Extent to which patient comprehension of written and verbal communication is assessed and accomplished	6.28	2.59	0.706	0.183	0.033	0.167	0.024	PCMH-A, Item 10, modified
	22. The type of self-management support provided by members of the practice team	5.84	2.45	0.717	0.185	0.033	0.167	0.024	PCMH-A, Item 11, verbatim
	23. How test results and care plans are communicated to patients	8.74	1.98	0.579	0.150	0.027	0.167	0.024	PCMH-A, Item 29, verbatim
	24. The use of feedback from a patient and family caregiver council to guide practice improvements	5.35	3.26	0.553	0.143	0.026	0.167	0.024	Developed by Mathematica

Continued

Table 1: Continued

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Reliability Weights Derived from Factor Analysis			Sum-Score Weights (Simple Average)		
				Question Loading on a Factor (Factor Loading)	Question Weight in a Factor (Normalized within a Factor)	Question Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Simple Average Weight for First-Order Factor (Normalized to Sum to 1 within a Factor)	Simple Average Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Source
Coordination of Care across the Medical Neighborhood	14. The extent of tracking of patient referrals to specialists	7.80	2.86	0.556	0.096	0.014	0.100	0.014	Developed by Mathematica
	15. The collaborative development of care plans with patients and families that include self-management and clinical management goals, and are used to guide care	6.47	2.80	0.615	0.106	0.016	0.100	0.014	PCMH-A, Item 32, verbatim
	26. The extent to which referral relationships with a range of specialists are formalized	7.14	2.52	0.583	0.101	0.015	0.100	0.014	TransformED, Item 36, modified
	27. Availability of behavioral health services for patients	5.80	2.56	0.568	0.098	0.015	0.100	0.014	PCMH-A, Item 25, verbatim
	28. The ease of obtaining referrals for specialty care,	8.45	2.07	0.614	0.106	0.016	0.100	0.014	PCMH-A, Item 26, verbatim

Continued

Table 1: Continued

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Reliability Weights Derived from Factor Analysis		Sum-Score Weights (Simple Average)			
				Question Weight in a Factor (Normalized to Sum to 1 within a Factor)	Question Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Simple Average Weight for First-Order Factor (Normalized to Sum to 1 within a Factor)	Simple Average Weight in the Total Score (Normalized to Sum to 1 within Total Score)		Source
	hospital care, or supportive community-based resources and exchange of relevant information with other providers before and after the patient visit	7.16	2.60	0.586	0.101	0.015	0.100	0.014	PCMH-A, Item 27, modified
	29. Practice staff follow-up with patients following emergency room (ER)/hospital visits	5.86	2.31	0.709	0.122	0.018	0.100	0.014	PCMH-A, Item 28, verbatim
	30. How practices link patients to supportive community-based resources								

Continued

Table 1: Continued

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Question Loading on a Factor (Factor Loading)	Reliability Weights Derived from Factor Analysis		Sum-Score Weights (Simple Average)		
					Question Weight in a Factor (Normalized to Sum to 1 within a Factor)	Question Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Simple Average Weight for First-Order Factor (Normalized to Sum to 1 within a Factor)		
							Simple Average Weight in the Total Score (Normalized to Sum to 1 within Total Score)		
Continuous Data-Driven Improvement	31. Transmission of patient information when this practice refers patients to hospitals, ERs, and specialists	8.67	2.22	0.537	0.093	0.014	0.100	0.014	Developed by Mathematica
	32. The timeliness of information received from hospitals and ERs following a patient's visit	6.82	2.78	0.473	0.082	0.012	0.100	0.014	Developed by Mathematica
	34. The proportion of patients for whom the practice knows the total cost to payers for medical care	2.81	2.33	0.557	0.096	0.014	0.100	0.014	Developed by Mathematica
Continuous Data-Driven Improvement	35. Practice's use of quality improvement (QI) activities that are done continuously and based on	6.71	2.71	0.774	0.154	0.025	0.143	0.020	PCMH-A, Item 17, verbatim

Continued

Table 1: Continued

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Reliability Weights Derived from Factor Analysis			Sum-Score Weights (Simple Average)		
				Question Loading on a Factor (Factor Loading)	Question Weight in a Factor (Normalized to Sum to 1 within a Factor)	Question Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Simple Average Weight for First-Order Factor (Normalized to Sum to 1 within a Factor)	Simple Average Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Source
proven improvement strategies	36. Extent to which QI activities are conducted by practice teams supported by a QI infrastructure with meaningful involvement of patients and their families	4.89	2.88	0.768	0.153	0.025	0.143	0.020	PCMH-A, Item 19, modified
	37. The availability of comprehensive performance measures to practice site and individual providers	6.77	3.32	0.696	0.139	0.023	0.143	0.020	PCMH-A, Item 18, modified
Availability of feedback reports on patient care experiences, and care processes or outcomes to practice site, individual providers, practice teams,	38. Availability of feedback reports on patient care experiences, and care processes or outcomes to practice site, individual providers, practice teams,	4.40	3.17	0.646	0.129	0.021	0.143	0.020	PCMH-A, Item 4, modified

Continued

Table 1: Continued

M-PCMH-A Questions									
Factor	Question	Mean	Standard Deviation (SD)	Question Loading on a Factor (Factor Loading)	Reliability Weights Derived from Factor Analysis		Sum-Score Weights (Simple Average)		
					Question Weight in a Factor (Normalized to Sum to 1 within a Factor)	Question Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Simple Average Weight for First-Order Factor (Normalized to Sum to 1 within a Factor)	Simple Average Weight in the Total Score (Normalized to Sum to 1 within Total Score)	Source
	patients, other teams, and external agencies	5.33	2.44	0.792	0.158	0.025	0.143	0.020	Developed by Mathematica
	39. The availability of staff, resources, and time for QI activities								
	40. The extent to which hiring and training processes focus on improving care and creating patient-centered care	5.97	2.87	0.569	0.113	0.018	0.143	0.020	PCMH-A, Item 15, verbatim
	41. The extent to which responsibility for conducting QI activities is shared by staff and is made explicit through protected time to meet and specific resources to engage in QI								
		5.69	3.10	0.779	0.155	0.025	0.143	0.020	PCMH-A, Item 16, verbatim

did not collect data from them until 2014. In 2014, both CPC and comparison practices completed the survey from April (March for the comparison practices) through July. We analyzed 2012 CPC practice data and 2014 CPC and comparison practice data.

Comprehensive Primary Care practices completed a fillable PDF in 2012 and web survey in 2014. The practices that did not respond to the 2014 survey within one month received a hard copy. Comparison practices received a web survey and those that did not respond by week 5 received a hard copy. Comparison practices and withdrawn CPC practices received \$100 for completing the survey; active CPC practices were not paid because completing the survey was a condition of CPC participation.

Both the introduction to the 2012 survey and a 2014 accompanying document addressing frequently asked questions explained that practices would not receive financial incentives for providing specific responses. However, it is possible that practices responded more favorably than their actual performance due to the social desirability of providing better care. Practices' understanding of or interpretation of questions also may vary from those of their own or other practices over time.

Practice managers were asked to complete the survey after discussing the responses with others in the practice. Respondents were asked to describe the process they used to fill out the survey (one team member or a group) and indicate who provided input for completing the survey (lead clinician, other clinicians, care manager or care coordinator, practice manager, nursing staff, billing staff, front desk staff, patients, and others).

Seventy-one percent of CPC practices in 2012 and 69 percent in 2014 used input from more than one staff member to respond to the survey. A single-staff member was more likely to respond in comparison practices (61 percent of all responses), typically a practice manager (70 percent of responses from a single-staff member; see Table S2 for more details). Responses and reliability of responses were similar whether the responses were provided by a single person or based on group input, and whether that group input included clinician participation (Duda and Poznyak 2016). However, this does not rule out the possibility of a systematic social-desirability bias across all types of respondents from the practice, which would be much less likely in an objective external assessment.

Reflecting CPC's requirements, 100 percent of the practices participating in CPC at the end of the fielding period responded to the survey (497 practices in 2012; 493 in 2014 because five of the original CPC practices left the initiative and one practice that split from an existing practice joined the

initiative). A total of 423 (48 percent) of the 881 comparison practices responded in 2014. Question-level nonresponse ranged from 0 to 0.6 percent for CPC practices in 2012 and from 0 to 1.6 percent in 2014 on all but one question (8.5 percent for question A36, in 2012, which asks about practices' quality improvement activities).

Rates of item nonresponse for comparison practices were less than 4 percent for all items except for items excluded from the short form of the survey. (To increase responses, initial nonresponding comparison practices received a shortened version of the instrument, which included only six M-PCMH-A items. Of the 423 comparison practice respondents, 69 [16 percent] completed the short form. Rather than excluding these practices from the analysis, we used a full-information maximum-likelihood method to handle the missing data.)

Validation Approach

We used confirmatory factor analysis (CFA) to test the factor structure of the 38 questions and the six domains they were designed to measure. First, we validated the M-PCMH-A using 2012 CPC practice data ($N = 497$). Second, we cross-validated the resulting factor structure using the 2014 comparison practice data ($N = 423$) to ensure the proposed factor structure generalized to other samples. Third, we examined reliability and validity of the M-PCMH-A using 2012 CPC data and 2014 comparison data. Fourth, we assessed the stability (or measurement invariance) of the M-PCMH-A's factor structure across the CPC and comparison respondents and over time within CPC practices using 2012 and 2014 ($N = 493$) data.

We hypothesized a hierarchical factor structure of the M-PCMH-A where six lower-order factors (the six M-PCMH-A domains) measured by 38 survey questions form a single higher-order summary factor that captures correlations among the lower-order factors. We also assumed that measurement errors of the questions are uncorrelated (i.e., that variables are not systematically influenced by the same unmeasured variable in addition to the latent variables they represent) (top figure in Figure 1). We iteratively tested and refined the structure using CFA to arrive at the model that best fit the data in the validation sample based on model fit statistics, reliability, and convergent and discriminant validity.

We examined model fit statistics including standardized root-mean-squared residual (SRMR) and root-mean-squared error of approximation (RMSEA) indices, which compare the hypothesized models to a saturated

Figure 1: Hypothesized (Model 1; top figure) and Revised (Model 2; bottom figure) factor structure of the M-PCMH-A

Notes: Circles represent latent factors; rectangles represent observed variables. Arrows connecting latent factors and observed variables represent a causal path; short, pointed arrows indicate error variances implied from total question variances. In Model 2, residual errors of the Continuity of Care and Access to Care factors were allowed to correlate to account for their conceptual relationship beyond the summary factor.

model that perfectly fits the data (Hu and Bentler 1999; Steiger 2007). Consistent with past research, we defined acceptable cutoff values ≤ 0.08 for SRMR and 0.06 for RMSEA (Hu and Bentler 1999).

We assessed internal-consistency reliability in the CPC and comparison data using Cronbach's alpha (α) and McDonald's omega (ω) coefficients. In contrast to α , ω calculates reliability of latent factors controlling for other factors in the model. Values ≥ 0.7 indicate acceptable internal-consistency reliability for both coefficients (Nunnally and Bernstein 1994).

We tested convergent validity by checking whether each question's loading on its assigned factor is greater than 0.40 and statistically significant (Cole 1987; Guo et al. 2008). We tested discriminant validity by setting the correlation among factors to 1.0, equivalent to a single-factor model, and testing its fit against the hypothesized model with freely estimated factor correlations (Guo et al. 2008).

To determine whether the mean differences in the factors are attributable to differences in primary care delivery rather than differences in

measurement properties of the M-PCMH-A, we tested its measurement invariance properties across the CPC and comparison groups (using 2012 data for CPC practices and 2014 data for comparison practices) and separately over time for CPC practices (using their 2012 and 2014 data). We conducted four increasingly stringent tests using the multigroup CFA framework (Steenkamp and Baumgartner 1998; Gregorich 2006). First, we examined whether the survey questions consistently map onto the same set of factors. This test is referred to as configural invariance and the tested model serves as a baseline for comparing subsequent models. If supported, this type of invariance indicates that the number and content of the factors remain the same for the two groups of practices. Second, we constrained first- and second-order factor loadings and question intercepts to be the same to test metric invariance of the model. If supported, this type of invariance suggests that the module's questions remain equally important indicators of the factors in both groups of practices. Consequently, scales of latent factors have the same metric across the two groups, so a comparison practice with the same observed questionnaire responses as a CPC practice has the same expected score on a given factor (Wang and Wang 2012). Third, we tested the invariance of the first-order factor intercepts by restricting them to zero in a reference group (2012 CPC data) and setting the mean of the summary factor to zero in both groups. This model tests the differences in the first-order factor intercepts (or the means of the first-order factors corresponding to the zero value of the summary factor) between the CPC and comparison groups. If confirmed, this type of invariance suggests the measurement scales for the first-order factors have the same origin in both groups of practices; therefore, differences between the data in the two groups can be attributed to true differences in the underlying factor rather than to measurement scaling (Fiske, Gilbert, and Lindzey 2010). Fourth, we tested the invariance of the overall summary index by constraining the first-order factor intercepts to zero in both groups and the mean of the summary factor to zero in the 2012 CPC reference group. If confirmed, this type of invariance suggests that the means of the CPC and comparison practices on the summary index can be validly compared as the latent scales in both groups have equivalent units of measurement (metric invariance) and origins (scalar invariance) (Gregorich 2006).

To test the measurement invariance of the M-PCMH-A's factor structure over time, we estimated configural and metric invariance models. A test for scalar invariance (no differences in the intercepts/means between the two time

periods) is not appropriate in this context because we expect the initiative to increase CPC practices' intercepts and means from 2012 to 2014 (Vandenberg and Lance 2000).

We followed Chen's (2007) recommendation that given $N > 300$, a change of ≥ 0.015 in RMSEA and ≥ 0.030 in SRMR between each tested model indicates noninvariance (Keith 2014). Tables S3 and S4 show the results.

For each CPC practice in 2012, we then computed factor scores, which combine practice responses to all questions related to a given factor based on the strength of the association between the question and the factor. Although calculating factor scores separately for each survey round would better reflect each round's data structure, this would not allow us to measure change over time, as practices with the same responses each round would have different scores due to a change in factor weights. Therefore, we used factor loadings estimated on the 2012 CPC data as weights to obtain summary measures for each factor and the overall score for each CPC practice in 2012 and 2014, and for comparison practices in 2014. Weights (factor loadings) of nonmissing questions were scaled to have a sum of one to ease interpretation and account for practices in which some, but not all, questions are missing in a factor. We computed a practice's score for each factor based on the statistically determined strength of association between each question and a factor it measures. We then computed a practice's overall score on the M-PCMH-A by further weighting a practice's scores on each individual factor by the correlation of that factor with the second-order (summary) factor (see Table 1 for the weights). Practices that skipped all questions corresponding to a factor did not receive a score for that factor, and second-order factor loadings of nonmissing first-order factors were scaled to sum to one and subsequently used when computing the overall score.

We conducted all analyses in *Mplus* 7.1 using a robust maximum-likelihood estimator to account for non-normal distribution of the variables and full-information maximum-likelihood for missing data.

RESULTS

M-PCMH-A Validation (2012 CPC Data)

The SRMR of 0.089 suggests that the hypothesized model did not provide a good fit to the 2012 CPC data (Table S5, Model 1). The misfit was primarily due to the Access and Continuity of Care questions, as well as questions A14 and A15, which more strongly related to the Coordination of Care than the

Planned Care factor they were hypothesized to measure. Additionally, question A13 (consistency of notification of patients of their laboratory and radiology results) was not strongly related to any factor and was dropped from other iterations of the model, leaving 37 questions in all other models.

We respecified the model (Model 2) to reflect the correlations in the data (bottom figure in Figure 1). Model 2 better fit the data (Table S5), with the SRMR decreasing to 0.066, which indicates a good fit, and with the RMSEA changing slightly to 0.072 (although improvement in fit cannot be measured formally as Model 2 is not nested within Model 1). A RMSEA of 0.072 implies that a small degree of misfit is still possible, but this misfit may be due to the non-normal distribution of the questions in the 2012 CPC data rather than model misspecification. Therefore, we accepted Model 2 as our final model. Although the Model 2 factor loading of question A12 (extent to which medication reconciliation is regularly done and documented) fell below the standard cutoff of 0.4, we retained it because we expect it to be an important determinant of outcomes, despite the lack of correlation with other questions. Table 1 presents descriptive statistics and factor loadings.

M-PCMH-A Cross-Validation (2014 Comparison Data)

To ensure the proposed factor structure generalized to other practices, we fit our final model to the data from a cross-validation sample of comparison practices. The proposed model fit the comparison practices' data even better than the CPC data on which the model was developed (Table S5). The consistency across two different groups supports the validity of the M-PCMH-A's factor structure (descriptive statistics and factor loadings available upon request.)

M-PCMH-A's Reliability and Validity (2012 CPC and 2014 Comparison Data)

The seven factors displayed high internal consistency, with reliability exceeding 0.70 in CPC and comparison samples' data, with one exception (Table 2). The Access to Care factor was the least reliable. Practice performance on the items in this scale was only weakly correlated (average correlation among the items is 0.274). However, despite low internal consistency, these items measure different aspects of Access to Care and serve as an important "checklist" of the aspects of comprehensive primary care that constitute this domain.

Large and significant factor loadings (question-factor correlations) of the questions onto a corresponding factor also demonstrate the M-PCMH-A's convergent validity (Table 1). We confirmed discriminant validity by the lack

of fit of a single-factor model ($RMSEA = 0.090$; $SRMR = 0.077$). That is, the model fit is improved by first aggregating the 37 individual items into seven domain-specific factors (which can then be combined into a single overall factor) rather than by requiring all 37 items to load on a single factor (results available upon request). Notably, the Continuity of Care factor had a low interfactor correlation, suggesting that this factor has less content overlap with the rest of the factors (Table 2).

Stability of the M-PCMH-A's Factor Structure Between 2012 CPC and 2014 Comparison Groups

The baseline configural model (Table S3, Model 1.1) and the model with the invariant first- and second-order factor loadings and question intercepts (Model 2.1) fit the data well. This suggests that the module's questions consistently map onto the same set of factors and the relative importance of the module questions remains largely the same across groups (Gregorich 2012). The model with the invariant first-order factor intercepts (Model 3.1) showed only minor reduction in fit compared to the baseline model, which was not significant according to Chen's (2007) criteria. This, in turn, indicates that differences between groups on the first-order factor means are unbiased—CPC practices are no more likely to overstate their responses to questions than the comparison practices. Model 4.1, with the invariant mean of the second-order summary factor, was supported by the data and did not demonstrate a significant decrease in model fit compared to the baseline model. This indicates that estimated differences between the CPC and comparison groups on the overall summary index (second-order factor) are not contaminated by underlying differences between the two groups in the factor structure. All of these results confirm that the measurement structure, reliability, and validity of the M-PCMH-A remain the same across groups. The relationship between module questions and factors does not differ between the CPC and comparison groups, and we can validly compare scores on the seven factors as well as the summary index between the two groups of practices.

Stability of the M-PCMH-A's Factor Structure over Time (2012 and 2014 CPC Data)

In preparation for the multigroup CFA, we fit the measurement model from 2012 CPC data to 2014 CPC data. The model provided an excellent fit to the data ($RMSEA = 0.063$; $SRMR = 0.062$) and was used for testing invariance.

Table 2: Reliability and Validity Statistics for the Seven Factors of the M-PCMH-A

Factor	2012 CPC Data (N = 497)				2014 Comparison Data (N = 423)			
	Cronbach's α	McDonald's ω	Average Inter-Factor Correlation	Question-Factor Correlation Range	Cronbach's α	McDonald's ω	Average Inter-Factor Correlation	Question-Factor Correlation Range
Continuity of Care	0.86	0.86	0.17	0.82–0.92	0.79	0.85	0.12	0.79–0.93
Access to Care	0.44	0.46	0.58	0.40–0.59	0.52	0.54	0.53	0.49–0.57
Planned Care for Chronic Conditions and Preventive Care	0.71	0.72	0.68	0.30–0.65	0.79	0.81	0.65	0.40–0.80
Risk-Stratified Care Management	0.72	0.73	0.65	0.63–0.72	0.82	0.82	0.67	0.76–0.79
Patient and Caregiver Engagement	0.79	0.81	0.69	0.55–0.74	0.85	0.86	0.67	0.61–0.76
Coordination of Care across the Medical Neighborhood	0.83	0.84	0.60	0.47–0.71	0.82	0.82	0.70	0.38–0.73
Continuous Data-Driven Improvement	0.88	0.88	0.63	0.57–0.78	0.92	0.92	0.60	0.69–0.84
Summary Factor	0.83	0.92	0.76	0.16–0.95	0.89	0.92	0.76	0.14–0.97

Notes: ω reliability was computed controlling for the effect of all other factors in the model. Question-factor correlation is the correlation of each question with the total factor score. Inter-factor correlation is the average correlation of each factor with all other factors.

The baseline configural model (Table S4, Model 1.1) and the metric invariance model (Model 2.1) fit the data well, indicating that the number and content of the factors remained the same before and after the intervention and that factor loadings remained largely the same; thus, module questions remain equally important indicators of the factors over time. These results confirm the CPC intervention did not alter conceptual and statistical relationships between the module's questions and the factors they measure.

Simple Average Versus Weighted Scores Based on Factor Loadings

We compared the weighted factor score (summing question scores multiplied by factor loadings) to a simple average of question scores. We calculated the simple average score by averaging all questions mapping to a domain to create domain scores and averaging all domain scores to create an overall score, using the same 37 questions mapped to the seven domains we obtained from the factor analysis.

In our sample, the factor-weighted average and simple average approaches generate practice scores that are very similar in the 2012 and 2014 CPC and 2014 comparison data (Table 3). The factor-weighted method shows a slightly greater *change* in CPC practice scores between 2012 and 2014 on average (2.3 points versus 2.0 points using the simple average method, see Table 3). This was true for 392 of the 493 CPC practices participating in both 2012 and 2014, although the difference between the indices in score improvement over time was greater than 1 point for only eight of these practices (see Table S6).

While we would recommend that prospective users of the M-PCMH-A weight practices' scores using factor analysis, some users may not have adequate resources or comfort with this method. The scores created using the simple average method are based on the assumption that each item carries the same weight within factors and that the factors themselves are equally strongly related to the overall score, which is not always consistent with estimates obtained from a factor analysis of the data. However, in our sample, the results were similar for the simple and factor-weighted averages. Concordance correlation between the practices' overall score of the second-order (summary) factor scores using the two methods is 0.93 and 0.96 for the CPC practices in 2012 and 2014, respectively, and 0.97 for the comparison practices in 2012. The concordance correlation between the change score in the CPC practices from 2012 to 2014 is 0.97 (Table 3).

Table 3: Mean and Correlation of Overall Practice Scores as Calculated Based on the Weighted Factor Scores and Simple Average Methods

	<i>CPC Practices 2012</i>	<i>CPC Practices 2014</i>	<i>Comparison Practices 2014</i>	<i>Change in CPC Practices 2012 to 2014</i>
Mean as calculated by weighted factor score method	6.477	8.737	7.857	2.260
Mean as calculated by simple average method	6.947	8.964	8.128	2.017
Concordance correlation between weighted factor and simple average methods	0.927	0.963	0.967	0.971

Note: This table includes the 493 CPC practices responding in both 2012 and 2014. The concordance correlation coefficient (Lin 1989) between the practices' overall score of the second-order (summary) factor scores was computed using the Concord package in Stata.

DISCUSSION

Our findings suggest that researchers can use the M-PCMH-A survey module as one way to construct and track statistically valid and reliable summary measures of the ways in which practices deliver different aspects of comprehensive primary care. With minor revisions, the validated factor structure of the M-PCMH-A corresponds to the domains of primary care delivery that the M-PCMH-A was designed to measure. The analysis confirmed the M-PCMH-A's hierarchical factor structure, with 37 of the 38 questions forming seven factors that capture the five core CPC functions (with Access and Continuity split into two factors) and a key support (continuous data-driven improvement) jointly forming one higher-order summary factor. The resulting factor structure demonstrates sufficient internal-consistency reliability and convergent and discriminant validity in measuring all but one factor (Access to Care) in CPC and comparison practices. The structure also exhibits adequate stability of the factor configurations, factor loadings, and intercepts across CPC and comparison groups and stability of the factor configurations and factor loadings within the CPC group over time.

In this application, using factor scores derived from factor loadings on the 2012 CPC group as weights yielded factor scores similar to those calculated by the traditional approach of constructing indices by simply averaging the question scores. However, the scores produced from CFA in other applications will more reliably capture practice performance on a given factor for several reasons. First, the weights account for differences in the

relative strength of the relationship between the seven factors and the summary factor, and the questions that measure these factors. Second, factor scores minimize the influence of measurement error in the questions contributing to each of the factors, which increases reliability and validity of the computed scores. Third, factor scores are generated by an objective statistical model ensuring that they are consistent with the underlying structure of the data. Fourth, the factor analysis identifies the set of questions that are appropriately grouped together in a given domain, whereas the simple sum approach provides no such discrimination. In contrast, the simple average approach uses arbitrary equal weighting of individual questions in each domain and overall. While in this data set the simple method yielded an overall index that was highly correlated with the overall index generated from the factor scores, this will not necessarily be the case in other data sets or applications. We encourage other researchers to use the M-PCMH-A as a credible way to measure differences across practices and over time in care delivery approaches, but we also recommend that when possible, factor loadings in other samples should be calculated based on the data collected, while relying on the basic approach described here.

There are several limitations to this analysis. First, practices self-reported the data, which raises questions about the validity of the responses. Although practice coaches reviewed the 26 questions from the PCMH-A for a sample of 65 safety net clinics from another initiative and determined that the questions largely agree with clinics' self-reports (Daniel et al. 2013), none of the 37 questions used in the M-PCMH-A was objectively validated in *this* sample using an external rater or additional data sources. However, it would have been difficult, time-consuming, and expensive (and beyond the scope of the study) to visit each of the 920 practices that responded to the survey to obtain external assessments of care delivery features. Survey responses had no implications for the practices' financial rewards from CMS, so practices had no incentive to overrate their care delivery features. However, some practices might overstate their self-ratings by giving what they feel to be more socially desirable responses. On the other hand, CPC may raise practices' standards for high performance, leading to lower self-ratings of the same approaches over time. Additionally, the respondent for a given practice may interpret the response categories or the practice's care delivery approaches differently from respondents for other practices or from other respondents for the practice in other survey rounds. Despite these potential limitations, our measurement invariance tests suggest the absence of differential response patterns or biases across the groups. Second, the practices selected for CPC and as comparisons are

diverse but not representative of practices nationwide. Third, future delivery reforms and market changes may alter how practices respond to survey questions. Thus, our results may not generalize to other practices, settings, or time periods. Fourth, the model fits the data well but has a few weaknesses, particularly in the 2012 CPC data: the Access to Care factor has lower-than-desired reliability, the Continuity of Care factor has a low average interfactor correlation, and question A12 has a low factor loading.

A final limitation is that this article does not examine criterion validity of the module—the extent to which the factor scores are useful for assessing the relationship between M-PCMH-A changes and external outcomes important to practices and payers. Our ongoing work will link levels and changes in the total, domain, and item scores to changes in outcomes such as cost, service use, quality, and patient experience (see Peikes et al. [2016] for our preliminary analyses of such relationships).

As payers around the country continue to move toward transformed primary care, we believe the M-PCMH-A can be valuable for researchers, practices, improvement experts, and payers. These stakeholders can use the M-PCMH-A to develop summary measures of primary care delivery (as measured by the seven factors and the overall index) and transformation, and identify and track areas for improvement. Researchers can also use the M-PCMH-A to assess linkages between changes in these specific aspects of care delivery and changes in cost, quality, and patient and provider experience over time. Future research should examine whether the number of questions in the M-PCMH-A could be reduced substantially while maintaining reliability and validity of the survey.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Supported by the Department of Health and Human Services, Centers for Medicare and Medicaid Services (CMS) under contract (HHSM-500-2010-00026I/HHSM-500-T0006) with Mathematica Policy Research (MPR). The views expressed in this article are solely those of the authors and do not necessarily represent the policy or views of the Centers for Medicare & Medicaid Services (CMS). The authors thank Erin Taylor, Ann O'Malley, Stacy Dale, Nancy Duda, Karen Bogen, and Martha Kovac of Mathematica Policy Research; Michael Prachman of Group Health Research Institute; and Timothy Day of the CMS for their input on the contents of this survey, and Frank Yoon of Truven Health Analytics for his

comments on this manuscript. We also acknowledge the many practices that responded to the survey.

Disclosures: None.

Disclaimer: None.

REFERENCES

- Burton, B. R. A., K. J. Devers, and R. A. Berenson. 2012. "Patient-Centered Medical Home Recognition Tools: A Comparison of Ten Surveys' Content and Operational Details" [accessed on October 28, 2016]. Available at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Reports/downloads/Burton_PCMHRT_Survey_Compare_March_2012.pdf
- Chen, F. F. 2007. "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling* 4: 464–504.
- Cole, D. A. 1987. "Utility of Confirmatory Factor Analysis in Test Validation Research." *Journal of Consulting and Clinical Psychology* 55: 584–94.
- Dale, S. B., A. Ghosh, D. N. Peikes, T. J. Day, F. B. Yoon, E. F. Taylor, K. Swankoski, A. S. O'Malley, P. H. Conway, R. Rajkumar, M. J. Press, L. Sessums, and R. Brown. 2016. "Two-Year Costs and Quality in the Comprehensive Primary Care Initiative." *New England Journal of Medicine* 374: 2345–56.
- Daniel, D. M., E. H. Wagner, K. Coleman, J. Schaefer, B. T. Austin, M. K. Abrams, K. E. Phillips, and J. R. Sugarman. 2013. "Assessing Progress toward Becoming a Patient-Centered Medical Home: An Assessment Tool for Practice Transformation." *Health Services Research* 48: 1879–97.
- Duda, N., and D. Poznyak. 2016. "The Effect of Individual versus Group Response on Instrument Reliability for a Medical Practice Survey." Presentation at the 71st Annual AAPOR Conference Austin, TX.
- Fiske, S. T., D. T. Gilbert, and G. Lindzey. 2010. *Handbook of Social Psychology*, vol. 1, 5th Edition. Hoboken, NJ: Wiley.
- Gregorich, S. E. 2006. "Do Self-Report Instruments Allow Meaningful Comparisons across Diverse Population Groups? Testing Measurement Invariance Using the Confirmatory Factor Analysis Framework." *Medical Care* 44 (11 Suppl. 3): S78–S94.
- Guo, B., P. Aveyard, A. Fielding, and S. Sutton. 2008. "Testing the Convergent and Discriminant Validity of the Decisional Balance Scale of the Transtheoretical Model Using the Multi-Trait Multi-Method Approach." *Psychology of Addictive Behaviors* 22: 288–94.
- Hu, L., and P. M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives." *Structural Equation Modeling* 6 (1): 1–55.
- Keith, T. K. 2014. *Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modeling*, 2nd Edition. New York: Routledge.

- Lin, L. I.-K. 1989. "A Concordance Correlation Coefficient to Evaluate Reproducibility." *Biometrics* 45 (1): 255–68.
- Mathematica Policy Research. 2012. *Evaluation of the Comprehensive Primary Care Initiative 2012 Practice Survey*. Princeton, NJ: Mathematica Policy Research.
- Mathematica Policy Research. 2014. *Evaluation of the Comprehensive Primary Care Initiative 2014 Practice Survey*. Princeton, NJ: Mathematica Policy Research.
- Nunnally, J. C., and I. H. Bernstein. 1994. *Psychometric Theory*. New York: McGraw-Hill.
- Peikes, D., E. F. Taylor, S. Dale, A. O'Malley, A. Ghosh, G. Anglin, K. Swankoski, L. Converse, A. Zutshi, and R. Brown. 2016. *Evaluation of the Comprehensive Primary Care Initiative: Second Annual Report* [accessed on October 28, 2016]. Prepared for the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services. Princeton, NJ: Mathematica Policy Research. Available at <https://innovation.cms.gov/files/reports/cpci-evalrpt2.pdf>
- Safety Net Medical Home Initiative. 2014. *The Patient-Centered Medical Home Assessment (PCMH-A), version 4.0*. Seattle, WA: The MacColl Center for Health Care Innovation at Group Health Research Institute and Qualis Health [accessed on October 28, 2016]. Available at <http://www.improvingchroniccare.org/downloads/pcmha.pdf>
- Steenkamp, J. B. E. M., and H. Baumgartner. 1998. "Assessing Measurement Invariance in Cross-National Consumer Research." *Journal of Consumer Research* 25: 78–90.
- Steiger, J. H. 2007. "Understanding the Limitations of Global Fit Assessment in Structural Equation Modeling." *Personality and Individual Differences* 42: 893–8.
- Taylor, E. F., S. Dale, D. Peikes, R. Brown, A. Ghosh, J. Crosson, G. Anglin, R. Keith, and R. Shapiro. 2015. *Evaluation of the Comprehensive Primary Care Initiative: First Annual Report. Prepared for the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services* [accessed on October 28, 2016]. Prepared for the U.S. Department of Health and Human Services, Centers for Medicare & Medicaid Services. Princeton, NJ: Mathematica Policy Research. Available at <https://innovation.cms.gov/files/reports/cpci-evalrpt1.pdf>
- TransforMED. 2012. "Medical Home Implementation Quotient." No longer in circulation.
- Vandenberg, R. J., and C. E. Lance. 2000. "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research." *Organizational Research Methods* 3 (1): 4–70.
- Wang, J., and X. Wang. 2012. *Structural Equation Modeling*. New York: Wiley.

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.

Appendix S1: Modified Patient-Centered Medical Home Assessment (M-PCMH-A) Module in the 2014 Comprehensive Primary Care Practice Survey.

Appendix S2:

Table S1: Baseline Characteristics of CPC Practices and Matched Comparison Practices.

Table S2: Respondents to the CPC Practice Survey.

Table S3: Model Fit and Comparison Statistics for the Models Testing Second-Order Measurement Invariance across 2012 CPC and 2014 Comparison Practices.

Table S4: Model Fit and Comparison Statistics for the Models Testing Second-Order Measurement Invariance across 2012 and 2014 Data of CPC Practices.

Table S5: Goodness-of-Fit Statistics for the Confirmatory Factor Analysis Models.

Table S6: Comparison of Primary Care Practices' Care Delivery Approaches Based on the Weighted Factor Scores and Simple Average Methods.