

# Methods for Evaluating Practice Change Toward a Patient-Centered Medical Home

Carlos Roberto Jaén, MD, PhD

Benjamin F. Crabtree, PhD

Raymond F. Palmer, PhD

Robert L. Ferrer, MD, MPH

Paul A. Nutting, MD, MSPH

William L. Miller, MD, MA

Elizabeth E. Stewart, PhD

Robert Wood, DrPH

Marivel Davila, MPH

Kurt C. Stange, MD, PhD



*Conflicts of interest: The authors' funding partially supports their time devoted to the evaluation, but they have no financial stake in the outcome. The authors' agreement with the funders gives them complete independence in conducting the evaluation and allows them to publish the findings without prior review by the funders. The authors have full access to and control of study data. The funders had no role in writing or submitting the manuscript.*

## CORRESPONDING AUTHOR

Carlos Roberto Jaén, MD, PhD  
Department of Family & Community  
Medicine  
University of Texas Health Science Center  
at San Antonio  
Mail Code 7794 – 7703 Floyd Curl Drive  
San Antonio, TX 78229  
jaen@uthscsa.edu

## ABSTRACT

**PURPOSE** Understanding the transformation of primary care practices to patient-centered medical homes (PCMHs) requires making sense of the change process, multilevel outcomes, and context. We describe the methods used to evaluate the country's first national demonstration project of the PCMH concept, with an emphasis on the quantitative measures and lessons for multimethod evaluation approaches.

**METHODS** The National Demonstration Project (NDP) was a group-randomized clinical trial of facilitated and self-directed implementation strategies for the PCMH. An independent evaluation team developed an integrated package of quantitative and qualitative methods to evaluate the process and outcomes of the NDP for practices and patients. Data were collected by an ethnographic analyst and a research nurse who visited each practice, and from multiple data sources including a medical record audit, patient and staff surveys, direct observation, interviews, and text review. Analyses aimed to provide real-time feedback to the NDP implementation team and lessons that would be transferable to the larger practice, policy, education, and research communities.

**RESULTS** Real-time analyses and feedback appeared to be helpful to the facilitators. Medical record audits provided data on process-of-care outcomes. Patient surveys contributed important information about patient-rated primary care attributes and patient-centered outcomes. Clinician and staff surveys provided important practice experience and organizational data. Ethnographic observations supplied insights about the process of practice development. Most practices were not able to provide detailed financial information.

**CONCLUSIONS** A multimethod approach is challenging, but feasible and vital to understanding the process and outcome of a practice development process. Additional longitudinal follow-up of NDP practices and their patients is needed.

*Ann Fam Med* 2010;8(Suppl 1):s9-s20. doi:10.1370/afm.1108.

## INTRODUCTION

The 2004 Future of Family Medicine report documented the current crisis in the US health care system and made the case for a "New Model" of practice.<sup>1,2</sup> This model has evolved to be consistent with the emerging consensus principles of the patient-centered medical home (PCMH).<sup>3</sup> The PCMH model of primary care incorporates current best practices in terms of access to care, prevention, chronic disease management, care coordination, and responsiveness to patients.<sup>4-14</sup> This model also acknowledges the trend toward health care consumerism and seeks to leverage information technology to improve outcomes and communication.<sup>15</sup>

In June 2006, the American Academy of Family Physicians (AAFP) began a trial to implement the PCMH model in 36 volunteer practices over the course of 2 years. The AAFP contracted with the Center for Research in Family Medicine and Primary Care to conduct an independent evaluation. This article describes the key methodologic strategies used for the evaluation and includes a comprehensive list of the data col-

lection tools. It also summarizes methodologic lessons learned from the evaluation over the course of 3 years. The article by Stange et al<sup>16</sup> in this supplement summarizes the context for the trial, and the article by Stewart et al<sup>17</sup> describes the conduct and evolution of the intervention.

Evaluating the NDP project required an evaluation plan having sufficient breadth and depth to capture the complex structures, processes, and outcomes likely to be affected by these efforts to bring about change.<sup>18-21</sup> The complex nature of the intervention required a combination of quantitative and qualitative strategies.<sup>22-24</sup>

The evaluation team had expertise in primary care (C.R.J., P.A.N., W.L.M., K.C.S., R.L.F.), ethnographic data collection (B.F.C., W.L.M., E.E.S.), epidemiology (C.R.J., K.C.S., R.L.F., P.A.N.), biostatistics (R.F.P., R.W., M.D.), and multimethod research (C.R.J., B.F.C., P.A.N., W.L.M., K.C.S.). The facilitators were not part of the evaluation team.

As an overall guide to the evaluation, we selected an initial practice change model, based on previous work of the evaluation team,<sup>25-27</sup> that is sensitive to both internal and external events.<sup>28,29</sup> The key elements in this change model include motivation and relationships among key stakeholders, practice resources for change, and external motivators and opportunities for change. We assessed other practice-level constructs, including staff satisfaction and organization of care according to the evolving PCMH model of TransforMED, the group implementing the intervention.<sup>17</sup>

The evaluation team selected measures of patient experience and outcomes according to both feasibility of use and a desire to represent diverse domains including critical aspects of primary care (eg, comprehensiveness of care, degree of shared knowledge between patient and clinician, quality of interpersonal communication, coordination of care, patient advocacy and trust, providing care in a family and community context, continuity, longitudinality, cultural responsiveness, accessibility, and strength of patients' preference for seeing their clinician). We also assessed medical condition-specific quality of care in the domains of acute and chronic illness, mental health, and delivery of preventive services, and patient outcomes including self-reported health status, enablement, and satisfaction.

The goals of the evaluation were (1) to describe the process of practice transformation and (2) to evaluate and compare the effects of 2 implementation approaches (ie, facilitated vs self-directed) on practice and patient outcomes. New knowledge generated from this evaluation is likely to benefit patients, primary care clinicians, researchers, evaluators, policy makers,

health care administrators, educators, and organizations advocating for better health care.

## METHODS

The AAFP recruited practices for the NDP among active academy members and graduating family medicine residents in 2006. The trial had a group-randomized design with multiple cross-sectional assessments of outcomes. A total of 36 volunteer practices were assigned to a facilitated or a self-directed intervention group. A companion article in this supplement provides a detailed description of the content of the intervention.<sup>17</sup> In short, the facilitated group received extensive assistance from 1 of the 3 facilitators during the 2 years of the study (June 2006-May 2008) in implementing the evolving model, whereas the self-directed group was left alone to implement the model.<sup>17</sup> Participating practices attempted to implement all aspects of the model. TransforMED, a wholly owned subsidiary of the AAFP, implemented the intervention.

To guide the evaluation, we created a matrix of critical areas for collecting data, shown in Table 1. This table describes the structures, process, and outcomes that we considered. For example, the relevant outcomes included patient experience, practice staff and clinician experiences, and quality of care in various areas (preventive services delivery, chronic disease care, acute illness care, and care for mental disorders). Using quantitative data strategies, we collected cross-sectional data through consecutive sampling at 3 points in time that were disclosed post hoc to the practices: baseline (July 3, 2006), 9 months (April 1, 2007), and 26 months (August 1, 2008). The evaluation team used qualitative data strategies to inform and modify the intervention throughout the study.

The AAFP Institutional Review Board (IRB) reviewed and approved the protocols for this study for primary data collection, and the IRBs of each coauthor's institution also approved secondary data analysis. Most practices did not have IRBs that represented them. In some cases, the systems accepted the AAFP IRB's approval. In 1 case, the system IRB did not approve participation of the practice in the study; this practice withdrew and all data from that practice were expunged.

### Quantitative Data Collection Strategies

Because we sought to understand the mechanisms as well as the results of practice change toward a PCMH, we collected quantitative data in 5 key domains—1 capturing baseline practice structure, 2 capturing intermediate process measures (staff perceptions about their organization; practice financial performance), and 2 capturing patient outcomes (patient ratings of their

experience with the practice; measures of care quality). This broad focus required 5 distinct sets of quantitative data collected with various tools—a baseline practice survey, medical record audits, a patient outcomes survey, a clinician staff questionnaire, and a practice financial survey—each described below.

### Baseline Practice Survey

The purpose of the baseline practice survey (BPS) was to initially determine a practice's eligibility for participation in the study, but it also served to gather baseline demographic and structural information. The BPS was an online application designed in collaboration with TransforMED (Supplemental Appendix 1, available online at [http://annfam.org/cgi/content/full/8/suppl\\_1/s9/DC1](http://annfam.org/cgi/content/full/8/suppl_1/s9/DC1)). The BPS outlined the criteria by which applicants would be evaluated and collected background information on the practice structure,

existing health information technology, team function, use of evidence in practice, attributes of the larger community and system, and characteristics of patients seen in the practice.

### Medical Record Audit

The purpose of the medical record audits was to gather information about the quality of care as measured by delivery of recommended clinical services, including selected preventive, acute, chronic, and mental health care (Supplemental Appendix 2, available online at [http://annfam.org/cgi/content/full/8/suppl\\_1/s9/DC1](http://annfam.org/cgi/content/full/8/suppl_1/s9/DC1)). We drew indicators from the Ambulatory Care Quality Alliance (ACQA) Starter Set of the Agency for Healthcare Research and Quality.<sup>30</sup> Of the 26 measures recommended, we included 16: all 7 prevention indicators, 2 coronary artery disease indicators, all 6 diabetes indicators, and 1 measure

**Table 1. Overview of Measures Used For the National Demonstration Project**

Measure (Time Interval)	Change Model	Change Process	Practice Structure		Outcomes		
			Practice Characteristics	Model Content	Patient Experience	Practice Experience	Quality of Care <sup>a</sup>
<b>Quantitative</b>							
Baseline practice survey (BPS) <sup>b,c</sup> (before 1/06)	X	–	X	X	–	X	–
Medical record audit <sup>b,c</sup> (baseline: 7/06; 9 months: 4/07; 26 months: 8/08)	–	–	–	–	–	–	X
Patient outcomes survey (POS) <sup>b,c</sup> (baseline, 9 months, 26 months)	–	–	–	X	X	–	X
Clinician staff questionnaire (CSQ) <sup>b,c</sup> (baseline, 9 months, 26 months)	X	–	X	–	–	X	–
Financial survey <sup>b,c</sup> (4/08)	X	–	X	–	–	–	–
<b>Qualitative</b>							
Facilitator-generated data							
Observation field notes <sup>b</sup> (7/06-12/07)	X	X	X	–	–	X	–
Stakeholders interviews <sup>b</sup> (7/06-8/06)	X	X	X	–	–	X	–
Practice environment checklist (PEC) <sup>b</sup> (7/06-12/07)	X	X	X	–	–	X	–
Evaluation team-generated data							
Regular monthly conference calls <sup>b</sup> (7/06-5/08)	X	X	X	–	–	–	–
E-mail streams <sup>b</sup> (7/06-5/08)	X	X	X	–	–	–	–
Learning sessions (observational field notes/interviews) <sup>b</sup> (6/06, 10/06, 9/07, 4/08 <sup>c</sup> )	X	X	X	X	–	–	–
Evaluation team conference calls <sup>b,c</sup> (7/06-12/07)	–	X	X	–	–	–	–
Multiday site visits (observations/interviews) <sup>b,c</sup> (7/07-12/07; 6/08-10/08)	X	X	X	X	–	–	–
Artifact data							
Web sites, practice documents <sup>b,c</sup>	X	X	X	X	–	X	–
<b>Multimethod</b>							
TransforMED implementation index <sup>b,c</sup> (7/06-3/08)	–	–	–	X	–	–	–

<sup>a</sup> Preventive care, chronic illness care, acute illness care, depression care.

<sup>b</sup> Facilitated practices.

<sup>c</sup> Self-directed practices.

of appropriate treatment of upper respiratory tract infection in children. We did not include measures of heart failure, asthma, prenatal care, and pharyngitis testing because of concerns about the low numbers of patients expected to have these conditions among the 60 patients whose medical records could feasibly be reviewed in each practice.

The evaluation team assessed delivery of clinical preventive services by measuring patients' receipt of services recommended by the US Preventive Services Task Force in July 2006 using sex- and age-specific recommendations.<sup>31</sup> We evaluated the quality of chronic disease care by measuring recommended quality measures for coronary artery disease (3), hypertension (2), diabetes (8), and hyperlipidemia (4). We evaluated the quality of acute care for upper respiratory infections by using the principles for judicious use of antibiotics for adults and children.<sup>32,33</sup> Finally, we evaluated the quality of depression care in the acute, continuation, and chronic care phases using a measure adapted for this study as a representative condition for mental health care (Supplemental Appendix 3, available online at [http://annfammed.org/cgi/content/full/8/3/suppl\\_1/s9/DC1](http://annfammed.org/cgi/content/full/8/3/suppl_1/s9/DC1)). Data from the medical record audit were used to generate scores for ACQA measures, preventive care, and chronic disease care.

A research nurse, employed by TransforMED but supervised by the independent evaluation team, audited 60 consecutive medical records per practice at baseline and again at 9 and 26 months. The research nurse audited the records on site or using remote access granted under a business associate agreement between TransforMED and the participating practice.

### Patient Outcomes Survey

The purpose of the patient outcomes survey (POS) was to measure patient experiences using data collection and analysis tools and techniques that have been developed by the evaluation team and others. To assess these dimensions, we included Flocke's Components of Primary Care Index (CPCI) subscales for comprehensive care, patients' shared knowledge with their clinician, interpersonal communication, personal physician preference, coordination of care, and community context.<sup>34-37</sup> We also included Safran's scales from the Ambulatory Care Experience Survey (ACES): organizational access, health promotion counseling, clinical team care, whole-person care, and patients' perception of time with the doctor.<sup>38-40</sup> We developed an all-or-none composite quality score based on the Institute of Medicine criteria: global practice experience.<sup>41-43</sup> Finally, the POS contained Howie's measure of Patient Enablement (PE) and the Consultation and Relational Empathy (CARE) measure developed by Mercer.<sup>44-47</sup>

These validated instruments have been found to be associated with patient satisfaction, preventive service delivery, chronic illness care, and health system features.<sup>35-37,46,48</sup>

TransforMED obtained a list of 120 consecutive patients visiting each practice, starting on each of the 3 dates for cross-sectional samples (baseline, 9 months, and 26 months) under a business agreement with each practice. TransforMED mailed an initial postcard followed by a POS to each patient on the list. Letters with informed consent elements were sent to the patients if they were aged 18 years or older, to both patients and parents if the patient was between ages 13 and 17 years, and to the parents of patients younger than 13 years of age. The POS included more than 100 questions (82 items), most of which used a 5-point Likert-type scale. The instructions encouraged patients to respond to items that best described their experience with their regular doctor or the practice. The items and measures are available from the specific authors who developed them and were used with permission for this study.

### Clinician Staff Questionnaire

The purpose of the clinician staff questionnaire (CSQ) was to measure and track changes over the course of the NDP in how clinicians and office staff perceived key practice attributes, such as modes of communication, leadership styles, learning culture, psychological safety, and approach to cultural diversity (Supplemental Appendix 4, available online at [http://annfammed.org/cgi/content/full/8/3/suppl\\_1/s9/DC1](http://annfammed.org/cgi/content/full/8/3/suppl_1/s9/DC1)). We selected these attributes because literature and the team's previous experience identified them as key mechanisms for successful organizational change and patient care improvement.<sup>25,49-54</sup> The CSQ was distributed to all clinical and nonclinical practice staff at each practice and collected in 3 cross-sectional waves. Staff who agreed to participate returned the questionnaire by mail directly to the study center. To comply with the IRB protocol, the CSQ did not require an individual identifier, so the 3 waves of the survey represent repeated cross-sections of the staff at each practice; thus, we analyzed organizational characteristics only at the aggregate practice level.

### Financial Survey

The purpose of the financial survey was to assess the financial status of all practices participating in the study near the end of the intervention phase of the NDP (April 2008). This survey collected information about a practice's financial status, including practice profitability, difficulty covering practice operational or capital expenses, routine financial monitoring systems available to the practice, revenue estimates, and

average salaries (Supplemental Appendix 5, available online at [http://annfammed.org/cgi/content/full/8/suppl\\_1/s9/DC1](http://annfammed.org/cgi/content/full/8/suppl_1/s9/DC1)). We mailed key stakeholders with access to financial information in each participating practice a self-administered survey. A separate, more detailed financial analysis conducted by TransforMED, although useful for practices that were able to complete it, proved infeasible to use for evaluation, since most practices were not able to provide the needed financial information on accounts receivables, accounts payable, breakdown of monthly expenses, or breakdown of net revenue by physician if they belonged to a larger system and were salaried. Some of the independent physicians were able to gather the information easily, but others lacked the time, billing support, or ability to separate personal from business finances.

### Qualitative Data Collection Strategies

In designing the qualitative data collection, we considered types of data that would be natural products of the intervention, such as e-mail streams, Web pages, and minutes from conference calls, and how to collect such data. Because the evaluation team was able to spend time with the facilitators shortly after they were hired, it was possible to integrate the collection of some qualitative observational field notes and depth interviews into the facilitators' initial assessment protocols. These data were available only for the facilitated practices, so additional strategies needed to be created for the self-directed practices. Also, although we could make a strong case that the facilitators needed to collect baseline data to guide their individualized intervention strategies for each practice, the same could not be said for the collection of follow-up data. As we reviewed the critical data collection areas (Table 1), we therefore conceived of 3 sets of data from various sources—facilitator-generated data, evaluation team-generated data, and artifact data that could be captured as natural products of the NDP—each of which is described below.

### Facilitator-Generated Data

During the first 2 to 3 months of the NDP, the facilitators made an initial site visit to each practice in their panel, took baseline observational field notes, and conducted depth interviews with key stakeholders. These visits generally lasted 2 to 3 days and gave the facilitators an opportunity to record their initial impressions and assess the baseline strengths and weaknesses of each practice. During these visits, the facilitators generated written summaries of the physical location of the practice and its staffing, and described key practice functions. If possible, they also followed 1 or more patients through the practice using a patient path strat-

egy.<sup>55</sup> Recognizing that the facilitators had limited time for extensive note taking, the evaluation team initiated conference calls with each facilitator during which the evaluation team could ask questions about the facilitator's observations. Although all facilitators later made additional visits to their practices, they were not asked to provide extensive field notes from these follow-up visits. Instead, the evaluation team received updates by conference calls and kept brief notes from these calls.

The evaluation team created a practice environment checklist (PEC) that included ratings of key concepts from the practice change and development model,<sup>29</sup> relationship systems,<sup>25,56</sup> and work relationships (Supplemental Appendix 6, available online at [http://annfammed.org/cgi/content/full/8/suppl\\_1/s9/DC1](http://annfammed.org/cgi/content/full/8/suppl_1/s9/DC1)). The form also provided space to make brief summary descriptive notes. Although the checklist had Likert scales, it was a qualitative tool that helped the facilitators focus on specific organizational characteristics in their practices and that they filled out based on their impressions of the practice. The facilitators reported difficulty filling out the PEC, in part due to their inability to assess the organizational features of a practice because they were not staff members.

During the initial site visits, the facilitators used an open-ended interview guide to conduct individual depth interviews with key practice stakeholders (Supplemental Appendix 7, available online at [http://annfammed.org/cgi/content/full/8/suppl\\_1/s9/DC1](http://annfammed.org/cgi/content/full/8/suppl_1/s9/DC1)). These questions particularly focused on motivation of key stakeholders, outside motivators, and attention to the local community and health system landscape, each of which are key components of the practice change and development model<sup>29</sup> that would not be readily available from other data sources. A second interview guide focused on stories of change and key stakeholders' recall of critical or memorable events in the practice history (Supplemental Appendix 8, available online at [http://annfammed.org/cgi/content/full/8/suppl\\_1/s9/DC1](http://annfammed.org/cgi/content/full/8/suppl_1/s9/DC1)).

Data collected by the NDP facilitators are potentially biased by their focus on and desire for practice change and by the specific obstacles and successes in their facilitation efforts. The majority of the qualitative data used for the NDP project were not directly collected by the facilitators, however, and data triangulation helped to understand and manage the potential bias. Additionally, in March 2006, before the actual initiation of the NDP, the 3 facilitators were given training in participant observation and depth interviewing, with an emphasis on taking low-inference field notes. The study's ethnographic analyst (E.E.S.) added independent observations toward the end of the study to check on our interpretation.

### Evaluation Team–Generated Data

As part of the intervention, each facilitator conducted regular monthly conference calls with their panel of 6 practices. These calls were used as opportunities to motivate stakeholders and for practices to learn from each other. To capture the content of these calls, the ethnographic analyst listened and took notes. As practices shared their experiences, these conversations helped the evaluation team to understand the change process and which NDP components were the most challenging and why.

Records of the constant e-mail communication between the 3 facilitators and their panel of practices proved to be extremely rich and enlightening. Because the e-mails included not only their message, but also detailed header information, including date, time, and recipients, they were extremely helpful in recreating the day-to-day exchange of information between the facilitators and their practices. Since the facilitators also used e-mail to connect external consultants to the practices, the e-mail stream also helped to recreate the history of different NDP components. Often, particularly as practices struggled with the demands of constant change, the e-mail streams helped to identify critical deficits in practices' adaptive reserve (capacity for change) (see the articles by Miller et al<sup>57</sup> and Nutting et al<sup>58</sup> in this supplement). The e-mail correspondences often provided more immediate notification of the ability to adapt to the relentless change required by the NDP in practice physicians or staff, since these changes were often discussed with the facilitators.

A major part of the NDP intervention was the use of 4 learning sessions, each lasting 1½ days. These sessions were designed to give the NDP intervention team an opportunity to provide educational content to the facilitated practices, while at the same time offering practice participants an opportunity to share experiences and learn from each other. To capture the complex dynamics at the learning sessions for the facilitated practices, members of the evaluation team were each assigned roles and strategic locations for taking observational field notes. They also took notes from informal interviews with session participants. Evaluation team members usually worked in pairs, with one concentrating on recording conversations as closely as possible, and the other focusing on any nonverbal cues. A final learning session in April 2008 included both the facilitated and self-directed practices.<sup>17</sup> Although the self-directed practices were not included in the initial 3 learning sessions, midway through the NDP, they self-organized a retreat (their own equivalent of a learning session). It was not possible for the whole evaluation team to observe this meeting, but we were able to arrange for our ethnographic analyst to attend to take

notes and interview participants. This attendance was especially important because it constituted the evaluation team's first contact with the self-directed practices outside of research nurse visits.

As part of real-time data analysis, and as a strategy to enhance the completeness of the data, the evaluation team convened conference calls (with note taking) throughout the NDP. Although many of these biweekly calls included only the evaluation team, other calls also allowed the team to member-check with NDP facilitators and receive updates. These conference calls typically focused on a single practice, so the evaluation team would often invite individual facilitators to join in parts of the discussion.

To conduct postintervention observations and key informant interviews comparable to those collected at baseline, near the end of the 2-year NDP, a member of the evaluation team made a 2-day visit to each facilitated practice. These visits generated 25 to 50 single-spaced pages of notes for each practice, including a description of the practice's reflection about their experience in the NDP. These visits were guided by the Site Visit Guide for the evaluation team (Supplemental Appendix 9, available online at [http://annfammed.org/cgi/content/full/8/suppl\\_1/s9/DC1](http://annfammed.org/cgi/content/full/8/suppl_1/s9/DC1)).

Because we had limited baseline data for the self-directed practices and only preliminary contact at their retreat, a member of the evaluation team conducted 2- to 3-day site visits to each self-directed practice. During these visits, interviewers asked practice staff to reflect on their experiences and to offer any critical insights about the change model and change process.

As a final determination of which model components each practice had in place and when they had been implemented, the evaluation team interviewed all facilitated and self-directed NDP practices by telephone. The calls were scheduled in advance with one of each practice's NDP champions and followed a template that included space for recording open-ended responses for each of the 8 domains and 39 measured components included in the TransformMED NDP Model (Supplemental Appendix 10, available online at [http://annfammed.org/cgi/content/full/8/suppl\\_1/s9/DC1](http://annfammed.org/cgi/content/full/8/suppl_1/s9/DC1) and accompanying articles in this supplement<sup>41,58</sup>). These responses were used to determine if and when each NDP component had been implemented. These conversations supplemented other data on the implementation process.

### Artifact Data

We used numerous artifact data to generate insights into practice characteristics, the implementation process, and outcomes. For example, practices' Web pages yielded data to confirm practice attributes such as phy-

sician staffing and presence of NDP components. The TransforMED Web page provided continuous updates on the NDP model being implemented and information about the consultants (national experts) who were available to assist facilitated practices. Handouts and Microsoft PowerPoint presentations from the learning sessions revealed which NDP components were being emphasized and when.

### Sample Size Calculations

Using methods for clustered data,<sup>59,60</sup> our initial estimates based on our anticipated sample sizes suggested that we had at least 90% power to detect moderate effect sizes (ie, approximately 0.8 of a standard deviation change in the facilitated group compared with no change in the self-directed group)<sup>61</sup> for each of the outcome variables. Because multiple outcome variables were to be evaluated, we used an  $\alpha$  level of .01 in these power calculations. For these statistical power calculations, we computed the intraclass correlation coefficients (ICCs) for the preventive service delivery score based on data from a prior study conducted by the investigative team.<sup>31</sup> For the other variables, we estimated the ICCs from the literature, or if no estimate was available, used the rule of thumb of .05 for patient outcome variables and .03 for practice process variables.<sup>62</sup> Although our completed analysis of the medical record audit data obtained the level of power desired, our patient-level data had substantially lower power (65%-80% for these analyses) because of lower response rates on the POS. Particularly with interactions, the practice-level analysis, which involved 31 practices, had even lower power (approximately 30%-40% depending on the outcome).

### Data Management Strategies

Respondents submitted their POS or CSQ directly to the research office using self-addressed prepaid envelopes. Using Snap survey software (Snap Surveys, Portsmouth, New Hampshire), we formatted these documents for rapid scanning. Data entry personnel individually reviewed each survey/questionnaire after data entry for accuracy. The research nurse recorded the medical record audit forms and mailed them in batches to the evaluation research office, and we scanned these forms using Snap survey software. After data entry, the data analysis team evaluated all data for reliability, looking for illogical patterns and eliminating duplicate entries.

We coded data to conform to previously developed subscales. Responses were scaled from 0 to 1.0 to allow ease of interpretation of the scales studied. As discussed below, we conducted factor analyses for those measures included in the CSQ to reduce data pertain-

ing to organizational factors. All qualitative data were catalogued and stored in a password-protected, secure server available only to coinvestigators for qualitative data analyses.

## Analytic Methods

### Factor Analysis

We submitted 82 items from the CSQ to a principal components factor analysis in 3 separate validation samples: sample 1 ( $n = 392$ ), sample 2 ( $n = 291$ ), and sample 3 ( $n = 292$ ). Following the procedures of Gorsuch,<sup>63</sup> we extracted 5 factors based on inspection of scree plots, eigenvalues, and simple structure (eg, items loading  $>0.60$  on only 1 factor). Items chosen for inclusion into the respective factor scales had to meet the same criteria across all 3 validation samples.

Analysis was conducted using PROC Factor in SAS software (SAS Institute Inc, Cary, North Carolina).<sup>64</sup> Because variables were nested within practices and had significant ICCs, we used multilevel software (MLwiN, University of Bristol, Bristol, United Kingdom) to export the within-subject covariance matrix as input into the factor analysis, thereby removing the portion of variance due to between-practice differences (eg, we corrected for biased standard errors due to potential effects of nested data).

The 5 factors explained approximately 53% of the total variance among the 82 items. We labeled the first factor adaptive reserve; the second, community awareness; the third, health information technology integration; the fourth, cultural sensitivity; and the fifth, patient safety culture. This approach balanced comprehensiveness vs usability.

### Other Analyses

Descriptions of specific analyses are included in the articles included in this supplement. In general, we used multilevel regression models for patient-level variables, accounting for the clustered nature of the data in practices.<sup>41</sup> For analyses of CSQ data, we used a least square means analysis of variance (ANOVA) model, weighting by the number of respondents in each practice.<sup>54</sup> The articles in this supplement that have extensive qualitative analyses detail how those analyses were conducted.<sup>57,65</sup>

## RESULTS

Articles in this supplement and elsewhere describe NDP patient outcomes<sup>41</sup> and practice outcomes,<sup>58</sup> the qualitative experiences of participating practices,<sup>65</sup> and a primary care practice development approach that emerged from the NDP and our collective experience.<sup>28,57</sup>

Overall, 31 of the original 36 practices completed the study, whereas 5 practices (2 facilitated and 3 self-

directed) withdrew. One facilitated practice withdrew because the larger system IRB could not approve participation; the other facilitated practice closed during the NDP because of financial pressures. One self-directed practice felt that the NDP data collection requirements were too burdensome in the context of other practice priorities, and 2 other self-directed practices closed during the NDP (one when the rural hospital across the street closed and the other when the larger health system closed the practice because of health system priorities beyond the practice).

We completed 98.9% of the medical record audits overall; 1 practice lost its clinical data when they switched to a different electronic medical record at baseline. The research nurse was able to collect information by remote access on 40 of 92 cross-sectional samples. All but 3 practices had electronic medical records in place by the end of the study. Despite nearly complete data collection from medical record audits, a sample size of 60 consecutive patients proved to be too small to identify sufficient numbers of patients with depression and with upper respiratory tract infections for measures of mental health and acute care to be calculated as outcome variables. For the POS, a total of 1,137, 882, and 760 were received from the baseline, 9-month, and 26-month surveys, for response rates of 29%, 24%, and 21%, respectively. Corresponding response rates for the CSQ were 60%, 48%, and 52%.

Table 2 details the items that comprise the adaptive reserve factor and the other factors that emerged from the CSQ analyses. The adaptive reserve scale was used extensively in analyses reported in the article by Nutting et al<sup>58</sup> describing the implementation of the NDP project and is discussed in terms of its conceptual fit in the article by Miller et al,<sup>57</sup> both in this supplement. This scale appears to offer a reliable way of measuring a practice's adaptive reserve.<sup>15,57,65</sup>

## DISCUSSION

This evaluation was designed to study practice change using a wide lens and multiple perspectives to understand both the details and the overall success of the transformative change process. The articles published in this supplement and elsewhere highlight the benefits of a holistic view of the practice change process and outcomes.<sup>15-17,41,57,58,65</sup>

At this early point in the evolution of the PCMH, it is important that implementation projects have a substantial evaluation component (preferably independent when there are commercial interests) to analyze complex data. This evaluation should include both numbers and narratives—quantitative data for outcomes that can be measured, and qualitative data for

emergent constructs or for areas in which the sample size or available measures are insufficient for statistical analysis. Narrative data also are particularly important for understanding meaning and context.<sup>23,24,27,28,57,66-69</sup>

We believe that the currently available disease-specific quality of care measures do not capture the higher-order primary care functions that include integrating, personalizing, and prioritizing care, and fostering healing,<sup>70</sup> and that are responsible for much of the added value of primary care.<sup>71,72</sup> In addition, because of the logistics of sampling patients with the diverse diseases seen in primary care,<sup>73,74</sup> we were unable to obtain a large enough sample size to assess disease-specific indicators of care quality for many diseases. The components of the medical record audit that were most helpful included the ACQA measure, the prevention score, and the chronic disease care score. We found the 4 pillars of primary care (easy access to first-contact care,<sup>40</sup> comprehensive care,<sup>36</sup> coordination of care,<sup>36</sup> and personal relationship over time<sup>36</sup>) as well as global practice experience as assessed with the POS to be useful as predictors of outcomes.<sup>41</sup> The patient enablement measure,<sup>46,47</sup> consultation and relational empathy measure,<sup>44,45</sup> and self-rated health status provided important patient-rated outcomes. The most useful parts of the CSQ were the items measuring adaptive reserve. The PEC and financial survey were not particularly useful.

It is important not to underestimate the logistic burden of collecting practice data in a geographically dispersed sample of diverse practices. One key issue is the tension between the need to protect participants and the need to collect valid and reliable practice data. For example, a limitation of our approach was that we were not able to follow up individual staff members with repeated measures of their opinions about their practice environments and operations. This restriction was especially important in small practices, where identifying even the respondents' roles in a practice may compromise their confidentiality and potentially jeopardize their employment. Also, because of the burden involved in obtaining informed consent from large numbers of patients, we faced substantial barriers to accessing patients' personal health information. We ultimately decided to trade off the ability to follow specific individuals in favor of the feasibility afforded by deidentified data, which prevented us from sending out multiple survey mailings to reduce patient nonresponse.

In addition, the practices themselves bore a burden of having to produce lists of consecutive patients after a particular date with the appropriate address and contact information. The practices' contact person also had the responsibility of distributing the CSQ at 3 points in the study. Facilitated practices were addition-

ally burdened with the need to respond to sustained contact with facilitators.

Other articles in this supplement describe specific limitations and potential sources of bias in the NDP. This evaluation, costing more than \$1.5 million, has 6

important overall limitations. First, participating practices were not representative of the universe of family practices in the United States. The level of motivation exhibited by these practices was much higher than that of most practices that we have studied during the last

**Table 2. Results of the Factor Analysis of the Clinician Staff Questionnaire**

Factors and Items	Attribute Assessed
Adaptive reserve (items loading >.40 across all samples; Cronbach $\alpha$ = .97)	
1. We regularly take time to consider ways to improve how we do things	Mindfulness
2. People in our practice actively seek new ways to improve how we do things	Mindfulness
3. People at all levels of this office openly talk about what is and isn't working	Mindfulness
4. People are aware of how their actions affect others in this practice	Heedful interactions
5. Most people in this practice are willing to change how they do things in response to feedback from others	Respectful interaction
6. This practice encourages everyone (front office staff, clinical staff, nurses, and clinicians) to share ideas	Cognitive diversity
7. I can rely on the other people in this practice to do their jobs well	Trust
8. Difficult problems are solved through face-to-face discussions in this practice	Communication
9. We regularly take time to reflect on how we do things	Reflection
10. After trying something new, we take time to think about how it worked	Reflection
11. The practice leadership makes sure that we have the time and space necessary to discuss changes to improve care	Leadership
12. Leadership in this practice creates an environment where things can be accomplished	Leadership
13. Practice leadership promotes an environment that is an enjoyable place to work	Leadership
14. Leadership strongly supports practice change efforts	Leadership
15. This practice learns from its mistakes	Learning culture
16. It is hard to get things to change in our practice (reverse score)	Learning culture
17. Mistakes have led to positive changes here	Learning culture
18. People in this practice have the information that they need to do their jobs well	Sensemaking
19. When we experience a problem in the practice, we make a serious effort to figure out what's really going on	Sensemaking
20. I have many opportunities to grow in my work	Teamwork (emergent property)
21. People in this practice operate as a real team	Teamwork (emergent property)
22. Most of the people who work in our practice seem to enjoy their work	Work environment (emergent property)
23. This practice is a place of joy and hope	Work environment (emergent property)
Community knowledge (items loading >.60 across all samples; Cronbach $\alpha$ = .82)	
1. This practice works effectively together as a team with community organizations	-
2. This practice utilizes community resources to meet the health care needs of patients	-
3. This practice is aware of community resources that are accessible to patients	-
4. People in this practice are connected with community organizations that serve patients	-
Health information technology integration (items loading >.60 across all samples; Cronbach $\alpha$ = .73)	
1. The electronic medical record is well integrated into the practice's daily work flow	-
2. The use of the electronic medical record during patient visits interferes with the doctor-patient relationship	-
3. I have what I need to do my work well	-
4. The practice can easily identify patients with a particular disease or medication	-
Cultural sensitivity (items loading >.60 across all samples; Cronbach $\alpha$ = .68)	
1. I am aware of my racial/ethnic/cultural stereotypes	-
2. I believe cultural issues are important in my interaction with health professional colleagues	-
3. I believe cultural issues are important in my interactions with patients	-
Patient safety culture (items loading >.60 across all samples; Cronbach $\alpha$ = .81)	
1. Staff feel like their mistakes are held against them	-
2. It is just by chance that more serious mistakes don't happen in our practice	-
3. Staff worry that mistakes they make are kept in their personal file	-
4. Staff are afraid to ask questions when something does not seem right	-
5. We tend to let setbacks and problems stop our change efforts	-

15 years.<sup>28</sup> The national spotlight on the NDP further boosted their initial motivation. Second, although we tried to maximize diversity in terms of region of the country, age of the practice, and practice size, the final selection of practices had few that served predominantly minority and poor populations. Third, although only 5 practices dropped out, the small number of practices enrolled limited the power to detect small differences in outcomes. It should also be emphasized that the study lacked a true control group, as the self-directed group received a low level of support. Fourth, the POS had a relatively low response rate compared with those seen in population surveys, in part due to the lack of monetary incentives and the limited ability to send reminders. Because of restrictions imposed by the IRB and Health Insurance Portability and Accountability Act (HIPAA) regulations, the evaluation team did not have access to patient health information and could not compare respondents with nonrespondents. Our response rates (21%-29%) are not much lower than the 29% response rate reported by Safran et al,<sup>40</sup> however. In addition, any selection bias introduced by the low response rate is likely to have been similar across the 3 cross-sectional samples. Fifth, the qualitative data collected were relatively thin, derived from only a single brief observation, but were abundant in terms of the process data collected over the 2 years of the study. Sixth, the NDP did not include payment reform. Practices needed to make all the changes within the context of the current fee-for-service structure. Lack of additional payments for participation and the practices' difficulties providing financial data make comparisons with other demonstration projects challenging.

Other logistic issues were related to the sheer scope and expense of such an evaluation. But as the descriptions and findings of the qualitative methods suggest,<sup>58,65</sup> we believe that the depth of information is critical to understanding the change processes associated with success and failure in moving toward a PCMH. Other investigators evaluating quality improvement efforts are reaching a similar conclusion, that fuller understanding of what it takes to improve organizational performance requires specific insights that are not routinely sought: it is important to know not only that a particular strategy improves outcomes, but also how and in what contexts it does so.<sup>21,75</sup> The need for information on all of these aspects is key because organizational success with complex interventions depends to a large degree on features of the local context, including leadership styles, ability to adapt practice, psychological safety, team interdependence, and others.<sup>57,76</sup> An encouraging finding from this study is that easily captured sources of electronic data, such as e-mail streams and practice Web sites, can serve as

rich repositories of data about practice history, evolution, and relationships.

In sum, our experience in conducting a multi-method evaluation of the NDP suggests that such collection and analysis of data on the process and outcome of a complex practice change process is challenging but feasible. We hope that the articles in this supplement and elsewhere<sup>15</sup> show the added value of a multimethod evaluation by an independent team in telling a more complete version of the complex, context-dependent story that a transformative practice change process involves.

**To read or post commentaries in response to this article, see it online at [http://www.annfammed.org/cgi/content/full/8/suppl\\_1/s9](http://www.annfammed.org/cgi/content/full/8/suppl_1/s9).**

**Key words:** Primary health care; family practice; professional practice; health care delivery; organizational change; mixed methods; qualitative methods; quantitative methods; National Demonstration Project; patient-centered medical home; practice-based research

Submitted July 23, 2009; submitted, revised, December 16, 2009; accepted January 19, 2010.

**Author affiliations:** Departments of Family & Community Medicine, and of Epidemiology & Biostatistics, University of Texas Health Science Center at San Antonio, San Antonio, Texas (Jaén); Department of Family Medicine, Robert Wood Johnson Medical School, University of Medicine & Dentistry of New Jersey, New Brunswick; Department of Epidemiology, School of Public Health, University of Medicine & Dentistry of New Jersey, Piscataway; and Cancer Institute of New Jersey, New Brunswick, New Jersey (Crabtree); Department of Family & Community Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas (Palmer); Department of Family & Community Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas (Ferrer); Center for Research Strategies, Denver, Colorado, and Department of Family Medicine, University of Colorado Health Sciences Center, Aurora, Colorado (Nutting); Department of Family Medicine, Lehigh Valley Health Network, Allentown; and Department of Family and Community Medicine, Pennsylvania State College of Medicine, Hershey, Pennsylvania (Miller); National Research Network, American Academy of Family Physicians, Leawood, Kansas (at the time of the project, Department of Family and Community Medicine, University of Texas Health Science Center at San Antonio) (Stewart); Department of Family & Community Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas (Wood); Department of Family & Community Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas (Davila); Departments of Family Medicine, Epidemiology & Biostatistics, and Sociology, and The Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio (Stange).

**Disclaimer:** Drs Stange, Nutting, and Ferrer, who are editors of the *Annals*, were not involved in the editorial evaluation of or decision to publish this article.

**Funding support:** The independent evaluation of the National Demonstration Project (NDP) practices was supported by the American Academy of Family Physicians (AAFP) and The Commonwealth Fund. The Commonwealth Fund is a national, private foundation based in New York City that supports independent research on health care issues and makes grants to improve health care practice and policy.

Publication of the journal supplement is supported by the American Academy of Family Physicians Foundation, the Society of Teachers of Family Medicine Foundation, the American Board of Family Medicine Foundation, and The Commonwealth Fund.

Dr Stange's time was supported in part by a Clinical Research Professorship from the American Cancer Society.

**Disclaimer:** The views presented here are those of the authors and not necessarily those of The Commonwealth Fund, its directors, officers, or staff.

**Acknowledgments:** The NDP was designed and implemented by TransformMED, LLC, a wholly-owned subsidiary of the AAFP. We are indebted to the participants in the NDP and to TransformMED for their tireless work.

## References

- Martin JC, Avant RF, Bowman MA, et al. The future of family medicine: a collaborative project of the family medicine community. *Ann Fam Med* 2004;2(Suppl 1):S3-S32.
- Green LA, Graham R, Bagley B, et al. Task Force 1. Report of the task force on patient expectations, core values, reintegration, and the new model of family medicine. *Ann Fam Med*. 2004;2(Suppl 1):S33-S50.
- American Academy of Family Physicians (AAFP), American Academy of Pediatrics (AAP), American College of Physicians (ACP), American Osteopathic Association (AOA). Joint principles of the patient-centered medical home. February 2007. <http://www.aafp.org/pcmh/principles.pdf>. Accessed Jan 19, 2010.
- Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness: the chronic care model, part 2. *JAMA*. 2002;288(15):1909-1914.
- Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. *JAMA*. 2002;288(19):2469-2475.
- Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness. *JAMA*. 2002;288(14):1775-1779.
- Ferrer RL, Hambidge SJ, Maly RC. The essential role of generalists in health care systems. *Ann Intern Med*. 2005;142(8):691-699.
- Grumbach K, Bodenheimer T. A primary care home for Americans: putting the house in order. *JAMA*. 2002;288(7):889-893.
- Grumbach K, Bodenheimer T. Can health care teams improve primary care practice? *JAMA*. 2004;291(10):1246-1251.
- Rittenhouse DR, Shortell SM. The patient-centered medical home: will it stand the test of health reform? *JAMA*. 2009;301(19):2038-2040.
- Rosenthal TC. The medical home: growing evidence to support a new approach to primary care. *J Am Board Fam Med*. 2008;21(5):427-440.
- Starfield B, Shi L. The medical home, access to care, and insurance: a review of evidence. *Pediatrics*. 2004;113(5 Suppl):1493-1498.
- The Robert Graham Center. The patient-centered medical home: history, seven core features, evidence and transformational change. 2007. <http://www.graham-center.org/online/graham/home/publications/monographs-books/2007/rgcmo-medical-home.html>. Accessed Jun 20, 2009.
- World Health Organization. *The World Health Report 2008—Primary Care (Now More Than Ever)*. <http://www.who.int/whr/2008/en/index.html>. Accessed Jun 19, 2009.
- Nutting PA, Miller WL, Crabtree BF, Jaen CR, Stewart EE, Stange KC. Initial lessons from the first national demonstration project on practice transformation to a patient-centered medical home. *Ann Fam Med*. 2009;7(3):254-260.
- Stange KC, Miller WL, Nutting PA, Crabtree BF, Stewart EE, Jaen CR. Context for understanding the National Demonstration Project and the patient-centered medical home. *Ann Fam Med*. 2010;8(Suppl 1):s2-s8.
- Stewart EE, Nutting PA, Crabtree BF, Stange KC, Miller WL, Jaen CR. Implementing the patient-centered medical home: observation and description of the National Demonstration Project. *Ann Fam Med*. 2010;8(Suppl 1):s21-s32.
- Donabedian A. Evaluating the quality of medical care. 1966. *Milbank Q*. 2005;83(4):691-729.
- Davidoff F, Batalden P, Stevens D, Ogrinc G, Mooney S; SQUIRE Development Group. Publication guidelines for quality improvement in health care: evolution of the SQUIRE project. *Qual Saf Health Care*. 2008;17(Suppl 1):i3-i9.
- Berwick DM. The science of improvement. *JAMA*. 2008;299(10):1182-1184.
- Pawson R, Tilley N. *Realistic Evaluation*. London, England: Sage Publications Ltd; 1991.
- Borkan JM. Mixed methods studies: a foundation for primary care research. *Ann Fam Med*. 2004;2(1):4-6.
- Stange KC, Miller WL, Crabtree BF, O'Connor PJ, Zyzanski SJ. Multi-method research: approaches for integrating qualitative and quantitative methods. *J Gen Intern Med*. 1994;9(5):278-282.
- Campbell M, Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ*. 2000;321(7262):694-696.
- Tallia AF, Lanham HJ, McDaniel RR Jr, Crabtree BF. 7 characteristics of successful work relationships. *Fam Pract Manag*. 2006;13(1):47-50.
- Crabtree BF, Miller WL, McDaniel RR, Stange KC, Nutting PA, Jaen CR. A survivor's guide for primary care physicians. *J Fam Pract*. 2009;58(8):E1-E7.
- Crabtree BF, Miller WL, Stange KC. Understanding practice from the ground up. *J Fam Pract*. 2001;50(10):881-887.
- Crabtree BF, Nutting PA, Jaen CR, et al. Primary care practice transformation is hard work: insights from a 15 year program of research. *Med Care*. In press.
- Cohen D, McDaniel RR Jr, Crabtree BF, et al. A practice change model for quality improvement in primary care practice. *J Healthc Manag*. 2004;49(3):155-168, discussion 169-170.
- Agency for Healthcare Research and Quality. Recommended starter set. Ambulatory Care Quality Alliance. Clinical performance measures for ambulatory care. 2005. <http://www.ahrq.gov/qual/aqa-start.htm>. Accessed Jun 11, 2009.
- Goodwin MA, Zyzanski SJ, Zronek S, et al; The Study to Enhance Prevention by Understanding Practice (STEP-UP). A clinical trial of tailored office systems for preventive service delivery. *Am J Prev Med*. 2001;21(1):20-28.
- Dowell SF, Marcy SM, Phillips WR, Gerber MA, Schwartz B. Principles of judicious use of antimicrobial agents for pediatric upper respiratory tract infections. *Pediatrics*. 1998;101(Suppl 1):163-165.
- Gonzales R, Bartlett JG, Besser RE, et al. Principles of appropriate antibiotic use for treatment of acute respiratory tract infections in adults: background, specific aims, and methods. *Ann Intern Med*. 2001;134(6):479-486.
- Flocke S. Primary care instrument [letter to editor]. *J Fam Pract*. 1998;46:12.
- Flocke SA. Measuring attributes of primary care: development of a new instrument. *J Fam Pract*. 1997;45(1):64-74.
- Flocke SA, Stange KC, Zyzanski SJ. The association of attributes of primary care with the delivery of clinical preventive services. *Med Care*. 1998;36(8 Suppl):AS21-AS30.
- Flocke SA, Miller WL, Crabtree BF. Relationships between physician practice style, patient satisfaction, and attributes of primary care. *J Fam Pract*. 2002;51(10):835-840.

38. Safran DG, Taira DA, Rogers WH, Kosinski M, Ware JE, Tarlov AR. Linking primary care performance to outcomes of care. *J Fam Pract.* 1998;47(3):213-220.
39. Safran DG, Kosinski M, Tarlov AR, et al. The Primary Care Assessment Survey: tests of data quality and measurement performance. *Med Care.* 1998;36(5):728-739.
40. Safran DG, Karp M, Coltin K, et al. Measuring patients' experiences with individual primary care physicians. Results of a statewide demonstration project. *J Gen Intern Med.* 2006;21(1):13-21.
41. Jaén CR, Ferrer RL, Miller WL, et al. Patient outcomes at 26 months in the patient-centered medical home National Demonstration Project. *Ann Fam Med.* 2010;8(Suppl 1):s57-s67.
42. Institute of Medicine. Committee on Quality of Health Care in America. *Crossing the Quality Chasm: A New Health System For the 21st Century.* Washington, DC: National Academy Press; 2001.
43. Reeves D, Campbell SM, Adams J, Shekelle PG, Kontopantelis E, Roland MO. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care.* 2007;45(6):489-496.
44. Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. *Fam Pract.* 2004;21(6):699-705.
45. Mercer SW, McConnachie A, Maxwell M, Heaney D, Watt GC. Relevance and practical use of the Consultation and Relational Empathy (CARE) Measure in general practice. *Fam Pract.* 2005;22(3):328-334.
46. Howie JG, Heaney DJ, Maxwell M, Walker JJ. A comparison of a Patient Enablement Instrument (PEI) against two established satisfaction scales as an outcome measure of primary care consultations. *Fam Pract.* 1998;15(2):165-171.
47. Howie JG, Heaney D, Maxwell M. Quality, core values and the general practice consultation: issues of definition, measurement and delivery. *Fam Pract.* 2004;21(4):458-468.
48. Howie JG, Heaney DJ, Maxwell M, Walker JJ, Freeman GK, Rai H. Quality at general practice consultations: cross sectional survey. *BMJ.* 1999;319(7212):738-743.
49. Crabtree BF, McDaniel RR, Nutting PA, Lanham HJ, Looney AJ, Miller WL. Closing the physician-staff divide: a step toward creating the medical home. *Fam Pract Manag.* 2008;15(4):20-24.
50. Edmondson AC. Learning from failure in health care: frequent opportunities, pervasive barriers. *Qual Saf Health Care.* 2004;13(Suppl 2):ii3-ii9.
51. Ohman-Strickland PA, John Orzano A, Nutting PA, et al. Measuring organizational attributes of primary care practices: development of a new instrument. *Health Serv Res.* 2007;42(3, pt 1):1257-1273.
52. Shortell SM. Increasing value: a research agenda for addressing the managerial and organizational challenges facing health care delivery in the United States. *Med Care Res Rev.* 2004;61(3 Suppl):12S-30S.
53. Zazzali JL, Alexander JA, Shortell SM, Burns LR. Organizational culture and physician satisfaction with dimensions of group practice. *Health Serv Res.* 2007;42(3 Pt 1):1150-1176.
54. Safran DG, Miller W, Beckman H. Organizational dimensions of relationship-centered care—theory, evidence, and practice. *Gen Intern Med.* 2006;21(Suppl 1):S9-S15.
55. Pommerenke FA, Dietrich A. Improving and maintaining preventive services. Part 1: applying the patient path model. *J Fam Pract.* 1992;34(1):86-91.
56. Jordan ME, Lanham HJ, Crabtree BF, et al. The role of conversation in health care interventions: enabling sensemaking and learning. *Implement Sci.* 2009;4:15.
57. Miller WL, Crabtree BF, Nutting PA, Stange KC, Jaén CR. Primary care practice development: a relationship-centered approach. *Ann Fam Med.* 2010;8(Suppl 1):s68-s79.
58. Nutting PA, Crabtree BF, Stewart EE, et al. Effect of facilitation on practice outcomes in the National Demonstration Project model of the patient-centered medical home. *Ann Fam Med.* 2010;8(Suppl 1):s33-s44.
59. Barcikowski RS. Statistical power with group means as the unit of analysis. *J Educ Behav Stat.* 1981;6(3):267-285.
60. Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? *Ann Fam Med.* 2004;2(3):204-208.
61. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum, 1988.
62. Murray DM, Blitstein JL. Methods to reduce the impact of intraclass correlation in group-randomized trials. *Eval Rev.* 2003;27:79-103.
63. Gorsuch RL. *Factor Analysis.* Hillsdale, NJ: Lawrence Erlbaum; 1983.
64. SAS Institute Inc. *Statistical Manual.* 2nd ed. Cary, NC: SAS Institute Inc; 1998.
65. Nutting PA, Crabtree BF, Miller WL, Stewart EE, Stange KC, Jaén CR. Journey to the patient-centered medical home: a qualitative analysis of the experiences of practices in the National Demonstration Project. *Ann Fam Med.* 2010;8(Suppl 1):s45-s56.
66. Stange KC, Zyzanski SJ. Integrating qualitative and quantitative research methods. *Fam Med.* 1989;21(6):448-451.
67. Stange KC, Miller WL, McWhinney I. Developing the knowledge base of family practice. *Fam Med.* 2001;33(4):286-297.
68. Crabtree BF, Miller WL, Addison RB, Gilchrist VJ, Kuzel A, eds. *Exploring Collaborative Research in Primary Care.* Thousand Oaks, CA: Sage Publications; 1994.
69. Crabtree BF, Miller WL, eds. *Doing Qualitative Research.* 2nd ed. Newbury Park, CA: Sage Publications; 1999.
70. Stange KC. A science of connectedness. *Ann Fam Med.* 2009;7(5):387-395.
71. Stange KC, Nutting PA, Miller WL, et al. Defining and measuring the patient-centered medical home. *J Gen Intern Med.* 2010;25(6):601-612.
72. Stange KC, Ferrer RL. The paradox of primary care. *Ann Fam Med.* 2009;7(4):293-299.
73. Schneeweiss R, Cherkin DC, Hart LG, et al. Diagnosis clusters adapted for ICD-9-CM and ICHPPC-2. *J Fam Pract.* 1986;22(1):69-72.
74. Stange KC, Zyzanski SJ, Jaén CR, et al. Illuminating the 'black box'. A description of 4454 patient visits to 138 family physicians. *J Fam Pract.* 1998;46(5):377-389.
75. McDaniel RR Jr, Lanham HJ, Anderson RA. Implications of complex adaptive systems theory for the design of research on health care organizations. *Health Care Manage Rev.* 2009;34(3):191-199.
76. Edmondson AC. The local and variegated nature of learning in organizations: a group-level perspective. *Organ Sci.* 2002;13(2):128-146.