

## Improving Evaluations of the Medical Home

### Key Points

By following these recommendations, future studies of the PCMH can generate high-quality, reliable evidence about the effectiveness of medical homes.

- **Focus evaluations on quality, cost, and experience.**
- **Include comparison practices.**
- **Recognize that the PCMH is a practice-level intervention and account for clustering.**
- **Include as many intervention practices as possible.**
- **Be strategic in identifying the right samples of patients to answer each evaluation question.**
- **Rethink the number of patients from whom data are collected to answer key evaluation questions.**

Primary care clinicians, health care systems, insurers, State governments, families, and communities are turning to the primary care patient-centered medical home (PCMH) as a solution to many of the troubles of the fragmented U.S. health care system. The PCMH model is a way of organizing and delivering primary health care that is patient- and family-centered, comprehensive, coordinated, accessible, and structured to continuously improve quality and safety.

No one questions these goals. Who could argue that our health care system should strive to deliver poor quality, uncoordinated care that ignores patients' values and preferences? What we don't know is whether current models of the medical home achieve these goals and, if so, how to finance them. Strong evaluations are critical in determining whether the PCMH model works, and for finding ways to refine, improve, customize, and disseminate the model if it does.

This brief highlights six key recommendations for evaluators and those considering commissioning evaluations of the PCMH model.<sup>1</sup>

### Focus evaluations on quality, cost, and experience.

**Rationale:** Quality, cost, and experience reflect our national health care goals of better care, affordable care, and improved experience of care.

**In practice:** Although not all prior evaluations have done so, future evaluations of the medical home should measure all three outcomes:

1. **Quality**, which incorporates the delivery of safe and effective care as well as patient outcomes.
2. **Cost**, which includes total cost and also can include measures of utilization that drive cost (especially hospitalizations and emergency department visits).
3. **Experience**, which encompasses not only patients' experiences, but the experiences of families, caregivers, and providers as well.

<sup>1</sup> More detailed information can be found in the longer paper from which the brief is drawn: Peikes D, Dale S, Lundquist E, Genevro J, Meyers D. Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? AHRQ Publication No. 11-0090-EF. Rockville, MD: Agency for Healthcare Research and Quality. September 2011.



**Include comparison practices. Evaluations with comparisons are more valuable than those without.**

**Rationale:** Gathering data from comparison practices (practices that do not receive the intervention) makes it possible for evaluations to demonstrate that changes in outcomes are the result of the intervention. Without comparison practices, it is not possible to identify and control for changes that would have occurred in the absence of the intervention.

**In practice:** Evaluation designs typically reflect compromises between research rigor and the limitations imposed by practical considerations (such as the availability of resources or interested participants). Investing in a high-quality study design makes it more likely that findings of the evaluation will be valid, reliable, and useful.

Study designs are ranked according to the quality of evidence they can produce:

- **Excellent: Randomized-control studies.** Randomly assigning interested practices to an intervention or comparison group is among the most powerful and definitive of evaluation designs. When studies of this type are well-implemented, changes in outcomes can be attributed to the intervention itself.
- **Very good: Matched comparison studies.** Selecting the comparison group using statistical matching can provide very good evidence of the effects of the intervention, if the practices and patients in the intervention and comparison groups are similar before the intervention begins. Intervention and comparison practices should be similar in terms of such characteristics as number and specialty of providers, use of health information technology, patient demographics, and pre-intervention values of the outcome measures of interest.
- **Poor: Pre-post evaluation.** This type of study design, which compares values of outcome measures before and after the intervention, does not include a comparison group. This makes it difficult to conclude that changes observed are due to the intervention.

**Recognize that the PCMH is a practice-level intervention.**

**Rationale:** In evaluations of the PCMH, the practice (rather than the individual patient) is the unit of intervention. This is because implementing the PCMH changes the entire practice, rather than changing care for just one patient or even one clinician's patients. Being available for late-night hours or email consultations, building a patient registry, and working in clinical teams are expected to improve care for all patients. It is not feasible, perhaps not even possible, to deliver most components of PCMH-type care to only some patients within a primary care practice.

Because the intervention affects all patients in a practice, evaluations need to account for *clustering*. Clustering occurs when outcomes for patients within a practice (that is, a cluster) are more similar to each other than to outcomes for patients in other practices, because of systematic differences between the practices. If clustering is ignored, the likelihood of concluding that an intervention works *when it does not* can be very large.

**In practice:** Hire good statisticians and involve them early in the evaluation planning process, not just in analyzing data.

Clustering affects calculations of statistical power, which are part of the evaluation design process. Evaluation designs must take clustering into account when estimating the number of practices and patients that are needed to ensure that the evaluation has adequate power to detect changes in outcomes. Analyses must also take clustering into account to accurately estimate whether findings are real or due to chance.

**Include as many PCMH practices as possible.**

**Rationale:** For evaluations of the medical home, the number of practices rather than the number of patients included in the evaluation determines the statistical power of the study.

Statistical power is the ability of an evaluation to detect a given level of change in outcomes and demonstrate with

confidence that the changes are real. The more statistical power a study has, the smaller the effect it will be able to detect.

Many factors contribute to the statistical power of an evaluation, but sample size is crucial. When an intervention changes the way a whole practice operates, the number of practices, rather than the number of patients, determines the sample size. Therefore, increasing the number of practices in the evaluation increases the statistical power of the evaluation.

Evaluators are responsible for proposing evaluation designs that have the statistical power to detect *plausible effects* of the intervention. Plausible effects are those that can reasonably be expected to occur in key outcome variables as the result of the intervention. Estimates of plausible effects are typically based on previous research findings and the experience of the intervention designers, and are determined by the type of intervention and its intensity. For example, an evaluator might be asked to evaluate a PCMH intervention that was anticipated to have a plausible effect of a 40 percent increase in the delivery of a selected preventive service or a 5 percent reduction in emergency department utilization. Evaluators should also consider the size of effect that would be meaningful to policymakers. Evaluations should be designed to have sufficient statistical power (including having enough practices) to detect changes that are both plausible and large enough to be meaningful to policymakers.

**In practice:** Increasing the number of practices in a study greatly increases the likelihood that the study will be able to detect the effects of an intervention.

The following can be used as a general rule, although the actual numbers will vary for each market, outcome, and patient population:

***Evaluations with fewer than 20 intervention practices*** typically will lack the statistical power to be able to detect plausible effect sizes for many key outcome measures, although well-designed evaluations may have the statistical power to detect effects on some well-chosen measures of quality and patient experience. Smaller PCMH interventions should still be evaluated, however. Evaluations

that do not have adequate statistical power can be considered “exploratory” (hypothesis-generating) studies—that is, studies that suggest questions for future evaluation but do not provide definitive evidence. They can also be valuable if grouped together with other studies for analysis.

***Evaluations with 20-100 intervention practices*** may have the statistical power to detect plausible effects on cost and service use outcomes among patients with multiple chronic conditions (see recommendation #5 for more on this issue). They also are likely to be able to detect effects among all patients for measures of quality and patient experience.

***Only evaluations with well over 100 intervention practices*** are likely to be able to demonstrate effects on cost and service use outcomes across all patients. These evaluations will also have the statistical power to detect effects on the full range of quality and experience outcomes.

### **Be strategic in identifying the right samples of patients to answer each evaluation question.**

**Rationale:** Transforming a practice into a medical home is expected to have beneficial effects for most or all patients, but there are likely to be bigger changes in some of the most critical outcome measures for a subset of patients who already use the most services. In addition, there is less variation in cost and service use among this subset of patients than across all patients in a practice. Lower variation improves the statistical power of the evaluation. Increased power makes it possible to detect smaller changes and more likely that a change of any size will be demonstrated to be significant. Thus, it is acceptable, even advantageous, to measure some outcomes in subgroups of patients within the medical home.

**In practice:** Treat all patients, measure costs across the entire intervention, but look for statistically significant changes in cost and service use among patients with chronic illnesses.

Data for different outcomes can be tracked using different samples of patients within a practice. The ability to detect changes in cost and utilization outcomes among people with chronic diseases is much greater than it is among the general patient population.

The ranges of values for other outcomes, such as quality and patient experience, are likely to be smaller in the general primary care population. This reflects the type of variables used to represent these outcomes, which often take on a limited number of values (for example, whether a patient received a specific service or whether a patient reported high, moderate, or low satisfaction with care). Moreover, a practice has the opportunity to affect the experience and quality of care for all of its patients, even for those who are relatively healthy. As a result, experience and quality-of-care variables, which typically take on few possible values, can be analyzed for all patients.

**Rethink the number of patients from whom data are collected to answer key evaluation questions.**

**Rationale:** Depending on the degree of clustering, a study should be able to detect plausible effects with only 20 to 100 patients per practice. In practice-level interventions, gathering data from larger numbers of patients per practice only slightly improves the minimum effect that can be detected. Collecting data on more patients per practice might not be worth the additional research costs involved.

**In practice:** Evaluations can save money by collecting survey and chart review data on a sample of the patients in a practice.

Evaluators can calculate how many patients are needed, which typically ranges between 20 and 100 patients per practice. The sample can be randomly selected, with oversampling of key populations of interest.

Evaluators also can conserve study resources by gathering data from a sample of patients when assessing outcomes that take on a limited number of values (for example, quality and patient experience). Resources can then be focused on increasing response rates among the sampled patients. Such considerations generally don't apply to claims data, in which the cost of acquiring the data tends to be the same regardless of the number of patients included.

---

This brief was prepared by David Meyers (Agency for Healthcare Research and Quality); Deborah Peikes, Stacy Dale, and Eric Lundquist (Mathematica Policy Research, Inc.); and Janice Genevro (Agency for Healthcare Research and Quality). A companion white paper commissioned by AHRQ contains additional details: Peikes D, Dale S, Lundquist E, Genevro J, Meyers D. Building the Evidence Base for the Medical Home: What Sample and Sample Size Do Studies Need? (Prepared by Mathematica Policy Research under Contract No. HHSA290200900019I TO 2.) AHRQ Publication No. 11-0090-EF. Rockville, MD: Agency for Healthcare Research and Quality. September 2011.

---

Suggested Citation: Meyers D, Peikes D, Dale S, Lundquist E, Genevro J. Improving Evaluations of the Medical Home. AHRQ Publication No. 11-0091. Rockville, MD: Agency for Healthcare Research and Quality. September 2011.

Access to these publications is available on the AHRQ Web site at <http://pcmh.ahrq.gov>.

---

