



The Medical Home: What Do We Know, What Do We Need to Know?

A Review of the Earliest Evidence
on the Effectiveness of the Patient-Centered
Medical Home Model



Agency for Healthcare Research and Quality

Advancing Excellence in Health Care • www.ahrq.gov



IMPROVING
PRIMARY CARE

The Medical Home: What Do We Know, What Do We Need to Know? A Review of the Earliest Evidence on the Effectiveness of the Patient-Centered Medical Home Model

Prepared for:

Agency for Healthcare Research and Policy
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

**Contract Nos. HHS290200900019I/HHS29032002T,
HHS290200900019I/HHS29032005T**

Prepared by:

Mathematica Policy Research, Princeton, NJ

Aparajita Zutshi, Ph.D., Mathematica Policy Research
Deborah Peikes, Ph.D., M.P.A., Mathematica Policy Research
Kimberly Smith, Ph.D., M.P.A., Mathematica Policy Research
Janice Genevro, Ph.D., Agency for Healthcare Research and Quality
Melissa Azur, Ph.D., Mathematica Policy Research
Michael Parchman, M.D., Agency for Healthcare Research and Quality
David Meyers, M.D., Agency for Healthcare Research and Quality

AHRQ Pub. No. 12(14)-0020-1-EF

Acknowledgments

We would like to thank a number of people for their assistance with this paper. The authors of many of the evaluations patiently answered questions about the interventions and their evaluation methods. Kristin Geonnotti at Mathematica Policy Research provided in-depth analyses of some of the papers included in this review. Michael Barr at the American College of Physicians; Robert Reid at Group Health Research Institute; and Randall Brown, Christopher Trenholm, Silvie Colman, Brian Goesling, and Tim Novak at Mathematica provided helpful comments and guidance.

Abstract

Purpose: The patient-centered medical home (PCMH, or medical home) aims to reinvigorate primary care and achieve the triple aim of better quality, improved experience, and lower costs. This study systematically reviews the early evidence on the effectiveness of the PCMH.

Methods: Of 498 articles on U.S.-based interventions, published or disseminated from January 2000 to September 2010, 14 evaluations of 12 interventions met our inclusion criteria: (1) tested a primary-care, practice-based intervention with three or more of five key PCMH components; and (2) conducted a quantitative evaluation of either (a) a triple aim outcome (quality of care, costs [or hospital use or emergency department use, two major cost drivers], and patient or caregiver experience), or (b) health care professional experience. We describe the interventions, their target populations, and implementation settings, and provide a broad overview of the research approaches used to evaluate these interventions. We developed and applied a formal rating system to identify interventions that have been evaluated using rigorous methods, and synthesized the evidence of effectiveness on each outcome generated by rigorous evaluations. Using these findings, we provide guidance to inform current efforts and structure future evaluations to maximize learning.

Results: The joint principles that first defined the PCMH were released in 2007, and we reviewed evidence through September 2010. Reflecting the time required to evaluate and publish findings on the model, the 12 interventions reviewed here—many of which are often cited in support of the medical home—are best viewed as precursors to the medical home. While these early interventions varied, most essentially tested the addition of a care manager operating from within the primary care practice, rather than a fundamentally transformed practice. Most interventions were evaluated in practices that were part of larger delivery systems and targeted patients who were older and sicker than average. Six of the 12 interventions evaluated at least one outcome using rigorous methods. This rigorous evidence indicates mostly inconclusive results (because of insufficient sample sizes to detect effects that might exist or uncertain statistical significance of results because analyses did not account for clustering of patients within practices); however, we found some favorable effects on quality of care, hospital and emergency department use, and patient or caregiver experience, and a few unfavorable effects on costs. Our review of these early interventions indicates that we need more evaluations of the medical home to assess its effectiveness.

Conclusions: Improving primary care is the key to achieving the triple aim outcomes. Although the PCMH is a promising innovation, rigorous quantitative evaluations and comprehensive implementation analyses are needed to assess effectiveness and refine the model to meet stakeholders' needs. Findings from future evaluations will help guide the substantial efforts practices and payers need to adopt the PCMH with the goal of achieving the triple aim outcomes.

Contents

- 1. Why Evidence on the Effectiveness of the Medical Home Is Important..... 1
 - What Is the PCMH, and How Might It Improve Outcomes?..... 3
 - Road Map..... 5
- 2. Key Questions and Approach..... 7
 - Key Questions 7
 - Selection of Evaluations for Key Questions..... 8
 - Methods for Key Questions 1-4..... 10
- 3. The Medical Home Landscape: Which Interventions Have Been Tested, for Which Patients, and in Which Implementation Settings? 16
 - Which Interventions Have Been Tested? 16
 - Where Were the Interventions Tested?..... 20
 - Whom Did the Interventions Serve?..... 24
- 4. How Do Studies Evaluate the Interventions? 25
 - Evaluation Designs..... 25
 - Outcome Measures 26
- 5. Which Evaluations Provide Rigorous Evidence of Effects on Each Outcome?..... 29
 - Ratings by Intervention..... 29
 - Ratings by Outcome..... 33
- 6. Evidence Synthesis 35
 - Approach..... 35
 - Findings From Rigorous Evaluations..... 35
- 7. Conclusion and Next Steps 47
 - Summary of Findings 48
 - Placing the Findings in Context..... 50
 - Limitations 51

Guidance to Improve the Future Evidence Base	52
Looking Forward	56
References and Included Studies	57

Figures

1. Conceptual Framework for the Effectiveness of the Medical Home	4
2. Selection of Evaluations for Key Questions 1–4	15
3. Rating Criteria for Randomized Controlled Trials.....	72
4. Rating Criteria for Comparison Group Evaluations	73
5. What Works Clearinghouse Liberal Attrition Standards	74

Tables

Table 1. Overview of the 12 interventions reviewed	17
Table 2. Target populations and implementation settings.....	21
Table 3. Evaluation designs and outcomes of the 12 interventions	27
Table 4.1. Evaluations with ratings of high or moderate on at least one outcome.....	31
Table 4.2. Evaluations with ratings of low or excluded on all outcomes.....	33
Table 5. Number of evaluations that assessed each triple aim outcome and health care professional experience	34
Table 6. Summary of evidence on processes of care	36
Table 7. Summary of evidence on health outcomes	37
Table 8. Summary of evidence on mortality	38
Table 9. Summary of evidence on cost	39
Table 10. Summary of evidence on hospital use	40
Table 11. Summary of evidence on emergency department use.....	42
Table 12. Summary of evidence on patient experience	43
Table 13. Summary of evidence on caregiver experience	44

Table 14. Summary of evidence on professional experience.....	45
Table 15. Snapshot of findings.....	49
Table 16. Definition of ratings.....	69
Table 17. Descriptions of the interventions, by AHRQ PCMH principles and facilitators	77

Appendices

Appendix A AHRQ’S Definition of the Patient-Centered Medical Home.....	63
Appendix B Methods for Reviewing the Evidence on the Patient-Centered Medical Home	65
Evaluation Selection.....	65
Methods to Assess the Rigor of the Evaluations	66
Appendix C Supplemental Table on Descriptions of the Interventions, by AHRQ PCMH Principles and Facilitators.....	75

Chapter 1. Why Evidence on the Effectiveness of the Medical Home Is Important

- The patient-centered medical home (PCMH) is a promising approach to improving primary care delivery.
- The PCMH aims to improve quality, reduce cost, and improve the experience of patients, caregivers, and health care professionals.
- We systematically review the quantitative evidence generated by early evaluations of the PCMH.
- We also distill lessons for future evaluations, to build a better evidence base.

Reinventing primary care is a task that is “far too important to fail” (Meyers and Clancy, 2009) and central to reforming health care delivery. While patient-centered primary care was once the backbone of our health care system, over time the system has become more specialized and technologically sophisticated (Bodenheimer and Pham, 2010), and fewer medical residents are choosing to become primary care physicians (Bodenheimer, 2006). The current health care system, with its incentives for volume over value, produces fragmented care that lacks coordination, patient-centeredness, and proactive population health management (Berenson and Rich, 2010b; Bodenheimer and Pham, 2010; Dentzer, 2010; Rittenhouse, Shortell, and Fisher, 2009; Howell, 2010). Although 93 percent of Americans want one place or doctor that provides primary care and coordinates care with specialists, only half report having such an experience (Schoen, Osborn, Doty, et al., 2007; Stremikis, Schoen, and Fryer, 2011). The PCMH is a promising model that aims to reinvent primary care so that it is “accessible, continuous, comprehensive, and coordinated and delivered in the context of family and community” (American Academy of Family Physicians, American Academy of Pediatrics, American College of Physicians, et al., 2007), and, in so doing, improve the triple aim outcomes of quality, affordability, and patient and caregiver experience.

The medical home concept arose in the 1960s as a way of improving care for children with special needs, and policy interest outside pediatrics grew over time (Kilo and Wasson, 2010). In 2007, primary care physician societies endorsed the “Joint Principles of the Patient-Centered Medical Home” (AAFP, AAP, ACP, et al., 2007). Intrigued by the potential of the PCMH model, major employers, private insurers and State Medicaid agencies across the Nation are currently rolling out pilots and demonstrations of the concept. The Centers for Medicare & Medicaid Services, the Department of Veterans Affairs, and other Federal agencies are also testing the model (http://pcmh.ahrq.gov/portal/server.pt/community/pcmh_home/1483/pcmh_federal_pcmh_activities_v2). It will likely be many years before results of current evaluations become available. Transforming care will require recognizing and addressing many barriers to change using lessons from these evaluations (Landon, Gill, Antonelli, et al., 2010).

Against this backdrop, decisionmakers must consider whether the current evidence on the model is already strong enough to proceed with widespread adoption, or whether gathering additional evidence is warranted. To contribute to this discussion, researchers at the Agency for Healthcare Research and Quality (AHRQ) and Mathematica Policy Research undertook a systematic review of quantitative evaluations of the medical home model to summarize the evidence on medical home effectiveness as well as identify lessons for future evaluations to generate a solid evidence base to guide health system reform. In addition to the triple aim

outcomes, we review the model's effectiveness on the experience of health care professionals (hereafter referred to as *professional experience*), since the success of primary care transformation and improvements in care are contingent on the satisfaction and ongoing engagement of health care personnel. This paper provides a more detailed version of papers published earlier by AHRQ and the American Journal of Managed Care (Peikes, Zutshi, Genevro, et al.; 2012a; Peikes, Zutshi, Genevro, et al., 2012b).

Given the relative newness of the PCMH model, and the time it takes to design, implement, and evaluate an intervention and disseminate findings, we were not surprised to find that many of the interventions evaluated to date, and currently cited as evidence in favor of the model, are precursors to the model. Many of these pioneering efforts were undertaken before the recent interest in the medical home and contained many, but not all, of its components.¹ To emphasize the difference between the interventions reviewed here and those being tested now, we refer to the ones we review as "PCMH precursors."

The review limits the synthesis of findings to interventions evaluated using rigorous methods. While much can be learned from rapid-cycle evaluations of small pilots and from evaluations of narrowly targeted interventions, this review intends to fulfill stakeholders' need for rigorous quantitative evidence on broad medical home-like interventions that test multiple components and examine effects on key outcomes.² Qualitative evaluations of PCMH implementation can also offer valuable insights into the implementation of these interventions and provide context for generalizing findings; they were excluded from this review, however, because our focus is on outcomes and because existing evaluations rarely documented their implementation experiences in published reports.

Some readers may not consider an evidence review of the PCMH to be necessary because they believe that the evaluations conducted to date, combined with the vast cross-sectional literature on the positive relationship between more primary care and better outcomes, provide sufficient evidence to proceed with widespread adoption of the model. Others may feel that the model is being held to a higher standard than many clinical interventions currently being used without strong evidentiary support. However, we believe that, given the significant investments required to transform practices and revitalize our primary health care system, many decisionmakers are, appropriately, going to demand rigorous evidence of effectiveness.

Historically, rigorous evaluations of a number of promising health care interventions have shown the interventions to be ineffective in achieving their goals. For example, telephonic disease management seemed to address obvious problems in coordination and patient self-management, but a number of randomized trials showed that many programs were ineffective and pointed the way to refining the model to offer better integration with providers, more in-person contact, and careful focusing of efforts to those most likely to benefit (Peikes, Peterson, Brown, et al., 2012c; Brown, Peikes, Peterson, et al., 2012; McCall and Cromwell, 2011; Peikes,

¹ These interventions are not static; although most were implemented before the joint principles were released, many were subsequently adapted to look more like the medical home and continue to evolve today.

² For example, a practice interested in decreasing the time between the receipt of laboratory results and patient notification need not wait for the results of a rigorous, controlled evaluation. It could convene the practice team members to redesign their workflow and measure changes in outcomes of interest (such as percentage of results delivered within two days) before and after implementation of the redesigned process. This approach provides quick answers to a low-cost initiative. While decisionmakers may require solid evidence on outcomes to justify large, transformative investments in primary care, for smaller initiatives, overreliance on rigorous evaluations carries the risk of delaying beneficial changes (Gold, Helms, and Guterman, 2011).

Chen, Schore, et al., 2009; Peikes, Peterson, Brown, et al., 2010). Similarly, rigorous evidence regarding the effectiveness of the PCMH model and how best to refine it is critical if it can be used for transforming primary care, especially given the substantial investments the model requires.

This review makes two important methodological contributions. First, we limited it to multi-component interventions with at least three of the five components of the PCMH model. Earlier reviews typically included results from interventions with as few as one feature, largely as a result of the infancy of the model. Homer, Klatka, Romm, et al. (2008) found that only 1 of the 33 studies they reviewed was of an intervention modeled after the medical home, while the others tested selected components. Rosenthal (2008), the Robert Graham Center (2007), and DePalma (2007) each reviewed the literature on individual components, such as team-based care, rather than multi-component interventions that more closely resemble the PCMH model.

Second, we limit the synthesis of the evidence to that generated by rigorous evaluations, which we assess using a systematic review process. Three previous reviews did not consider the rigor of the evidence (Grumbach and Grundy, 2010; Fields, Leshen, and Patel, 2010; DePalma, 2007). Two conducted a limited assessment by restricting the summary of the evidence to comparison group (Homer, Klatka, Romm, et al., 2008) or peer-reviewed (Friedberg, Lai, Hussey, et al., 2009) studies, with neither assessing the strength of the analytical methods used by the studies or excluding studies that did not use rigorous methods from their summary of the evidence.

What Is the PCMH, and How Might It Improve Outcomes?

There are several definitions of the medical home model, and most contain similar elements. For this review, we use the AHRQ definition, which like many others views the medical home as a way to organize primary care to deliver patient-centered, continuous, comprehensive, and coordinated care with a system-based focus on safety and quality.³ The definition focuses on five principles of primary care delivery⁴:

1. **A patient-centered orientation** toward each patient's unique needs, culture, values, and preferences; support of the patient's self-care efforts; and involvement of the patient in care plans.
2. **Comprehensive care** that meets the large majority of each patient's physical and mental health care needs, including prevention and wellness, acute care, and chronic care, and is provided by a cohesive team.
3. **Care that is coordinated** across all elements of the complex health care system and connects patients to both medical and social resources in the community.

³ The AHRQ definition builds on the traditional definition of primary care established by the Institute of Medicine and Barbara Starfield (Donaldson, Yordy, Lohr, et al., 1996; Starfield, 1992, 2008) and incorporates aspects of the expanded care model (Barr, Robinson, and Marin-Link, 2003; Glasgow, Orleans, Wagner, et al., 2001). It is similar to the definition of the medical home provided in the joint principles but places a greater emphasis on team-based care.

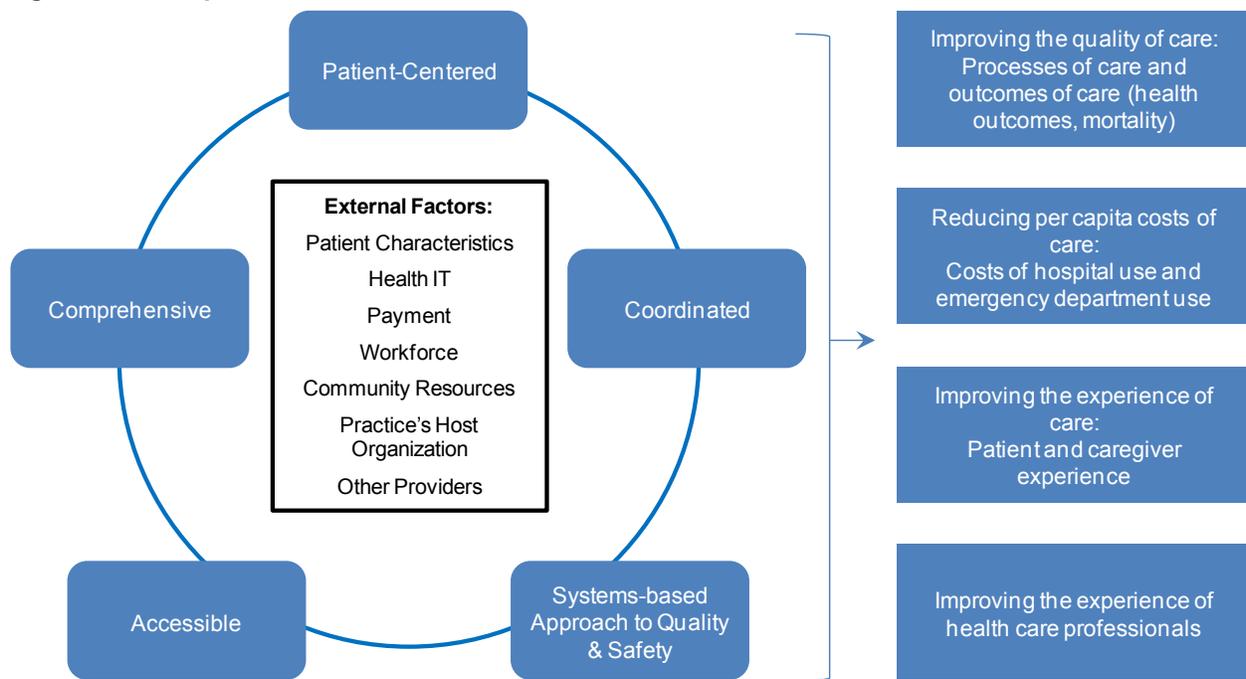
⁴ See Appendix A for a detailed version of the AHRQ PCMH definition. Also available at: http://pcmh.ahrq.gov/portal/server.pt/community/pcmh_home/1483/PCMH_Defining%20the%20PCMH_v2. Accessed August 23, 2012.

4. **Superb access to care** that meets patients’ needs and preferences, including care provided after hours and by email and telephone.
5. **A systems-based approach to quality and safety** that includes gathering and responding to patient experience data, having a commitment to ongoing quality improvement, and practicing population health management.

In the AHRQ definition, health IT, workforce development, and enhanced payment are considered to be important facilitators of change that support the medical home.⁵

A conceptual framework guides this review (Figure 1). The framework illustrates the medical home and the outcomes it is hypothesized to improve. We categorize the outcomes using a triple aim framework inspired by the frameworks of Don Berwick and the National Quality Strategy (Berwick et al., 2008; Department of Health and Human Services, March 2011). Specifically, implementing the medical home model is expected to improve quality of care (including processes of care and health outcomes), reduce costs (including use of hospital and emergency department [ED] services—two key drivers of cost), and enhance the experience of care (for patients and caregivers, who are the users of the health care system). We also expect the model to improve the experience of health care professionals.

Figure 1. Conceptual Framework for the Effectiveness of the Medical Home



As the figure shows, factors external to the medical home will also influence the outcomes of interest. These include patient factors (such as health risk, motivation, behaviors, and socioeconomic status), as well as system-wide factors (such as adoption of health IT, payment policies affecting primary care and other providers, workforce development, community resources, the organizational context of the primary care setting, and the nature and

⁵ We note that pilots and demonstrations are testing different variants of the model. The variants reflect different ways of operationalizing the principles that we refer to collectively as the PCMH model.

cooperativeness of other providers in the medical neighborhood). These factors, in addition to the medical home components, will collectively determine whether an intervention improves key outcomes, the magnitude of these improvements, and the applicability of the findings to other implementation settings and populations. It is important to interpret the findings of evaluations in the context of these multiple external factors to distill lessons to guide the transformation of primary care.

Road Map

This review is organized as follows. In Chapter 2, we present five key questions and the review methods. Chapter 3 describes the interventions included in this review along with their target populations and implementation settings. Chapter 4 provides a broad overview of the research designs used to evaluate these interventions. In Chapter 5, we assess the rigor of the evaluations of these interventions. Chapter 6 synthesizes the evidence on each outcome using only findings from rigorous evaluations. Chapter 7 summarizes the evidence, discusses how to apply the evidence to current medical home initiatives, and describes limitations of this review. Finally, we share lessons on how to improve the rigor, comprehensiveness, and generalizability of the evidence in the future.

Chapter 2. Key Questions and Approach

- This review of the evidence on the effectiveness of the medical home addresses five key questions:
 1. Which interventions have been evaluated using quantitative methods, for which patients, and in which implementation settings?
 2. Which evaluation designs and outcome measures were used to evaluate these interventions?
 3. Which interventions were evaluated using rigorous methods?
 4. What are the effects of the rigorously evaluated interventions on key outcomes?
 5. What lessons does the current evidence teach us?
- Of 498 studies released between January 2000 and September 2010 on U.S.-based interventions, 14 evaluations of 12 interventions tested a multi-component, practice-based intervention and quantitatively examined effects on quality, cost, patient or caregiver experience of care, or professional experience.
- We draw on the U.S. Preventive Services Task Force methods and well-regarded systematic reviews from other fields to rate the internal validity of each analysis undertaken by an evaluation; we do not factor generalizability into the rating.
- We select only rigorous analyses when we summarize the evidence of effectiveness on each outcome.
- We view results that are not statistically significant as inconclusive rather than as evidence of no effects, because we suspect that many of the evaluations lacked sufficient statistical power to detect effects if they existed.
- For findings reported as statistically significant from practice-level evaluations that did not account for clustering of patients within practices, we used published estimates of clustering to adjust their statistical significance; when such estimates were not available, we view the findings as having uncertain statistical significance and classified them as inconclusive, because it is possible that a clustering adjustment could render the findings not statistically significant.

This review examines five key questions. Our goals are to understand what evidence supports the medical home and what lessons can be drawn at this time to ensure that current pilots and demonstrations can generate a solid evidence base in the future. Below we describe the key questions and the sample of evaluations and methods used to answer each question.

Key Questions

1. **Which interventions have been evaluated using quantitative methods, for which patients, and in which implementation settings?** Specifically, we ask how close to the PCMH, as defined by AHRQ, these early interventions are, and whether they have been tested among the intended users of the PCMH model: typical patients and practices. Our goal is to establish how relevant the current evidence base is for understanding the effectiveness of the PCMH.
2. **Which research designs and outcome measures were used to evaluate these interventions?** Ideally, the evidence base would consist of evaluations with strong research designs and analyses of the full range of outcomes that stakeholders need to guide their decisions. We assess whether the existing body of evidence meets this goal.
3. **Which interventions were evaluated using rigorous methods?** We develop and apply a systematic assessment approach to identify interventions that have been

rigorously evaluated (and thus can provide rigorous evidence) as well as those that require further evaluation.

4. **What is the evidence from rigorous evaluations regarding the effects of early PCMH interventions on key outcomes?** We use findings from rigorously evaluated interventions to summarize the evidence on effectiveness of the medical home. We exclude findings from weak evaluations because we have less confidence in their findings about the intervention's effectiveness.
5. **What lessons can be learned from the early evaluations?** We provide insights on how to strengthen the evidence base to meet the high demand for reliable information on the effectiveness of the medical home. These insights emerged from reviewing the interventions that have been tested so far, and identifying factors that have limited the rigor and scope of the current evidence. This information is intended for model developers, funders, and evaluators of medical homes, and has relevance to evaluations of other practice-level interventions.

Selection of Evaluations for Key Questions

We used two samples of evaluations to answer the key questions. From a literature search, we identified 14 evaluations of 12 interventions that met our inclusion criteria (described below). We used this sample to describe the interventions and their evaluations, and to assess the rigor of the evaluations (key questions 1-3). We found that 6 of the 12 interventions were evaluated using rigorous methods for at least one outcome, and we used this sample of evaluations to synthesize the evidence on each outcome (key question 4). We show this process in a flowchart in Figure 2 and describe it in detail below. Finally, we draw on the sample of 14 evaluations and our analyses of key questions 1-4 to draw lessons on how to improve the evidence base on the medical home (key question 5).

Sample Selection for Key Questions 1-3: Description of the Interventions, Their Evaluations, and Assessment of the Rigor of the Evaluations

The review team conducted a broad search to identify English-language studies in the published and grey literature on the PCMH in the United States. To capture published studies, we used Ovid and EBSCO search engines to search within multiple social science, health services, and medical databases for articles from January 2000 to September 2010 containing the words “medical home” or “primary care transformation.” We also conducted targeted searches to identify studies of initiatives for which no start dates were specified, but that are widely cited as being part of the evidence base on the medical home. We identified additional studies by reviewing content on 100 relevant Web sites, examining bibliographies in existing review articles, and gathering input from experts in the field. This search process yielded 498 potentially relevant citations. As with all evidence reviews, because of publication bias, the evaluations summarized here may be more likely to include favorable effects and less likely to include no effects or unfavorable effects.

Of the 498 citations, we found 14 evaluations of 12 interventions⁶ that met our two criteria for inclusion in the review:

⁶ One intervention, Community Care of North Carolina [CCNC], was evaluated by three distinct studies.

1. The evaluation tested a primary-care, practice-based intervention with three or more of the five medical home components defined by AHRQ (delivering care that is patient-centered, comprehensive and team-based, coordinated, accessible, and systems-based in its approach to quality and safety). We excluded evaluations of care coordination and disease management interventions that met these criteria but were not provided from within, or in close partnership with, the practice (for example, interventions delivered by off-site care managers via telephone).
2. The evaluation used quantitative methods to examine effects on either (a) a triple aim outcome: quality of care, costs⁷ (or hospital use or emergency department use, two major cost drivers), and patient or caregiver experience; or (b) professional experience.

Because most of the interventions target different subgroups of the U.S. primary care population, our inclusion criteria did not consider the population served. We also did not require that the intervention include health IT or provide enhanced payment. For most evaluations, findings were published in multiple articles.

Two evaluations of medical home interventions—the American Academy of Family Practice’s National Demonstration Project (NDP), which is often cited in the medical home literature, and the Illinois Medical Home Project (IMHP)—are not included in the sample. Because these evaluations tested the effect of *facilitation* as an intervention for practice redesign efforts, they did not meet the first criterion of testing the effects of a medical home intervention. In other words, they tested the effect of helping practices redesign themselves to become medical homes relative to the effect of practices becoming medical homes on their own. We also did not include studies of the Physician Practice Connections-Primary Care Medical Home Assessment (PPC-PC) (Holmboe, Arnold, Weng, et al., 2010) and the Physician Practice Connections Readiness Survey Assessment (PPC-RS) (Solberg, Asche, Pawlson, et al., 2008). These studies did not test a specific intervention; instead, they surveyed physicians about the presence of medical home capabilities to examine whether physicians who reported such capabilities delivered better quality of care.

Sample Selection for Key Question 4: Synthesis of Evidence from Rigorous Evaluations

We included only findings from rigorous evaluations in the synthesis of the evidence. Overall, 6 of the 14 evaluations were rated as providing rigorous evidence on at least one outcome: 4 evaluations on different aspects of quality of care (3 on processes of care, 3 on health outcomes, and 2 on mortality); 6 on cost or utilization (4 on total costs, 5 on hospital use, and 3 on ED use); 3 on experience of care (3 on patient experience, 2 on caregiver experience); and 1 on professional experience.

⁷ None of the studies reported effects on out-of-pocket patient costs or practice revenues.

Methods for Key Questions 1-4

We turn now to the methods we used to answer key questions 1-4.

Key Question 1: Description of Interventions

We describe each of the 12 interventions in terms of AHRQ's five core principles, as well as two of the facilitators of the model: health IT and payment (none of the interventions undertook workforce development). We also describe the patient populations each intervention targeted (for example, Medicare patients, patients with chronic physical or mental illnesses), as well as the type of insurance coverage patients had (both fee-for-service and managed care). Finally, we describe the implementation setting in which the intervention was implemented (in larger care delivery settings, such as integrated delivery systems, versus in independent practices) and whether the practice had access to electronic health records (EHRs). We draw particular attention to the number of practices in which the intervention was implemented, as this has important implications for the ability of the evaluation to identify effects of the intervention when they occur. Chapter 3 presents these findings.

Key Question 2: Description of Evaluations

As a precursor to the formal assessment of the rigor of the evaluations (undertaken to answer key question 3), we present a broad overview of the rigor and scope of the current research landscape on the medical home by describing the evaluation designs and outcome measures used in the 14 evaluations. Selecting a strong evaluation design is a first step that evaluators can take to ensure that the evaluation produces unbiased estimates of the causal effects of an intervention. We therefore classify the designs into those that employed a control or comparison group (randomized, controlled trials [RCTs] and nonexperimental comparison group designs)⁸ and those that did not (pre-post and cross-sectional designs). Designs without a control or comparison group often make it difficult to assess what the sample's outcomes would have been absent the intervention. (The purpose of a control/comparison group is to establish that counterfactual—a necessary condition for obtaining an unbiased impact estimate.) We also describe which of the key outcomes the evaluations examined. We present these findings in Chapter 4.

Key Question 3: Assessment of the Rigor of the Evaluations

We developed a systematic approach to assess the rigor of the 14 evaluations. We drew broadly from the published methods of the U.S. Preventive Services Task Force (USPSTF) and supplemented them with specific criteria from well-regarded evidence reviews.⁹

⁸ The term *control group* is used exclusively when the group was assigned using an RCT. The term *comparison group* indicates that the group was selected using nonexperimental comparison group methods.

⁹ In addition to the USPSTF review methods (see Harris, Helfand, Woolf, et al., 2001), we drew specific operational criteria from the What Works Clearinghouse (WWC) review of educational interventions (which also typically employ clustered designs, like the many practice-level interventions reviewed here (see http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_1_standards_handbook.pdf) and from an evidence review of home visiting programs for families with pregnant women and children (see <http://www.mathematica-mpr.com/EarlyChildhood/homvee.asp>).

Rather than give a global rating to each evaluation, we individually rated the internal validity of each analysis conducted by an evaluation as *high*, *moderate*, *low*, or *excluded*. We rated individual analyses because evaluations often used different designs, samples, and methods (and sometimes different subgroups of patients) to analyze different outcomes over varying followup periods. Therefore, to allow for the possibility that the evaluation of a single intervention could provide more rigorous evidence on some outcomes than others, we separately assessed the analysis of each outcome measure at each followup period and, if applicable, for each subgroup of patients. We include only analyses rated as high or moderate in our synthesis of the evidence.

Our rating of each analysis is based solely on an assessment of its internal validity. We do not factor generalizability (or external validity) into the rating because most interventions included in this review targeted a specific subset of primary care patients, were implemented in unique settings, and either purposely selected practices or relied on them to volunteer; therefore, nearly all of them have limited generalizability. We summarize the characteristics of patients and implementation settings used in the rigorous evaluations to alert decisionmakers to the possibility that findings may differ if interventions are implemented in other populations and settings.

We rated each analysis using a sequence of criteria, starting with the most general (evaluation design) and ending with the most specific (such as whether the analysis controlled for outcome values prior to the start of the intervention (at baseline)). We rated an analysis “excluded” if the evaluation design and methods were not described in enough detail to permit assessment of the internal validity of the results. We always rated analyses “low” if they did not employ a control or comparison group (and instead used pre-post or cross-sectional evaluations). If they did employ a control or comparison group, such as analyses from RCTs and nonexperimental comparison group evaluations, we assessed the strength of the methods to identify causal effects and produce unbiased estimates of the interventions’ effects, and accordingly rated them high, moderate, or low.

Analyses from RCTs were given a high rating if they had:

- No systematic confounders
- No endogenous subgroups¹⁰
- Low attrition
- Adjustment for any statistically significant baseline differences in the outcome between the intervention and control groups

Analyses from comparison group evaluations, and from RCTs with high attrition or with endogenous subgroups, were given a moderate rating if they had:

- No systematic confounders
- Equivalence of the outcome between the intervention and comparison group samples at baseline

¹⁰ A subgroup is considered endogenously formed, and estimates of effects for this subgroup considered biased, if the subgroup is defined on the basis of a followup (or post-randomization) value of an outcome that the intervention can potentially affect. The extent of this bias may be small if the intervention and control arms of the subgroup are comparable at baseline, or if the intervention had no effect on the outcome that defines the subgroup.

- Adjustment for baseline values of the outcome

Analyses from RCTs and comparison group evaluations were given a low rating if they did not meet the criteria for a high or moderate rating. In many cases, because of the limited information that study authors can include in a journal article, we sought additional details from authors to be able to determine the rating. Appendix B describes the assessment approach in detail. Chapter 5 presents the ratings.

Key Question 4: Synthesis of Rigorous Evidence

Chapter 6 synthesizes evidence on four broad outcomes: (1) quality of care (including processes of care, health outcomes, and mortality); (2) costs, hospital use, and ED use; (3) patient and caregiver experience of care; and (4) professional experience. The synthesis of evidence on any given outcome includes only findings from analyses found to be rigorous, that is, those rated high or moderate. We did not synthesize results from analyses rated low, because we believe that if these interventions were evaluated using better methods, the results could differ substantially. For example, results could change from suggesting an intervention did not work to indicating it did, or vice versa. We also discuss the generalizability of the findings to the U.S. primary care population, using the target populations and implementation settings described in key question 1.

We categorized the rigorous findings as (1) statistically significant and favorable, (2) statistically significant and unfavorable, (3) inconclusive (that is, they fail to indicate whether or not the intervention worked) because they were not statistically significant, or (4) inconclusive because their statistical significance was uncertain due to lack of adjustment for clustering of patients within practices. We recognize that decisionmakers, from payers to providers, may be frustrated with inconclusive findings, but we believe this label accurately reflects the lack of certainty about whether or not certain interventions had an effect, as we explain below.

Statistical significance thresholds. To determine whether the effect on an outcome measure is statistically significant, we use a significance level of 5 percent for all measures except costs. If an intervention-control difference is statistically significant at the 5 percent level, we can be 95 percent confident that there is a true effect. Because of the large variation in cost outcomes, we use a more liberal significance level of 10 percent.

Results that are not statistically significant are considered inconclusive. When an intervention-control difference is not statistically significant, we treat the result as inconclusive, rather than as evidence that the intervention is not effective. Interpretation of tests of statistical significance depends on the statistical power of the test—the chance of detecting an effect that is truly there. A test is conventionally considered to be adequately powered if there is an 80 percent chance (in other words, reasonable certainty) that if a true effect of a prespecified magnitude exists, it will show up as a statistically significant intervention-control difference. When a test is not adequately powered, the effect can be difficult to detect with reasonable certainty. In other words, testing is likely to show that the difference is not statistically significant, even though it is real.

We suspect that most evaluations reviewed here had inadequately powered tests of statistical significance because of small sample sizes: none of the rigorous evaluations of practice-level interventions had more than 11 practices in the intervention group. Peikes, Dale, Lundquist, et al.

(2011) estimate that, assuming a moderate amount of clustering, a model that is tested in 20 intervention practices (with 20 control practices) and targets all patients might need to reduce costs by 45 percent or more (a very large effect) to have an 80 percent chance of detecting the reduction. If cost and service use were measured among the chronically ill, as many of these evaluations do, the intervention might still need to reduce costs by 20 percent or more for the test to have an 80 percent chance of detecting it. These are large effects for an intervention to achieve, and an evaluation would need even larger sample sizes (and therefore greater power) to detect smaller, more plausible effects. In such cases, therefore, a difference that is not statistically significant does not imply that the intervention is not effective. Instead, it has not been *proven* to be effective; well-powered evaluations in the future would be required to assess the intervention's effectiveness.

Statistically significant results that do not account for clustering are considered inconclusive. In some evaluations, and for some outcomes, statistical testing of practice-level interventions did not account for the clustering of patients within practices when determining statistical significance. Because clustering reduces the effective sample size, a test that ignores clustering overstates the statistical significance of a finding and might show a difference to be statistically significant when it was not (Peikes, Dale, Lundquist, et al., 2011). As a result, what at first appears to be a statistically significant difference may no longer remain so after accounting for clustering. For differences reported to be statistically significant without accounting for clustering, we adjusted the statistical significance by using published estimates of clustering.¹¹ Differences that were no longer statistically significant post-adjustment were reclassified as not statistically significant. However, in two cases, we classified differences as “inconclusive due to uncertain statistical significance”: (1) when statistical significance was sensitive to the range of clustering estimates used; and (2) for outcomes other than cost and service use, for which we could not find published estimates of clustering.¹² In both cases, because we were unable to determine the correct statistical significance, we consider significance as uncertain and the evidence inconclusive. We note that this issue is irrelevant for differences that were reported as not statistically significant because they would continue to remain so after a clustering adjustment.

Reporting the magnitude of effects. Our approach to reporting the magnitude of effects varies by outcome and category of evidence. For statistically significant effects for cost, hospital use, and ED use, we report the magnitude of effects. Ideally, we would indicate the uncertainty about the estimated effect by reporting the confidence interval to convey the possible range of the true effect. However, most evaluations did not report this information. For other outcomes, because of the diverse measures that were examined across different evaluations, we do not report the magnitude of effects. We also do not report the magnitude of results that are not statistically significant or have uncertain statistical significance. This is because the confidence

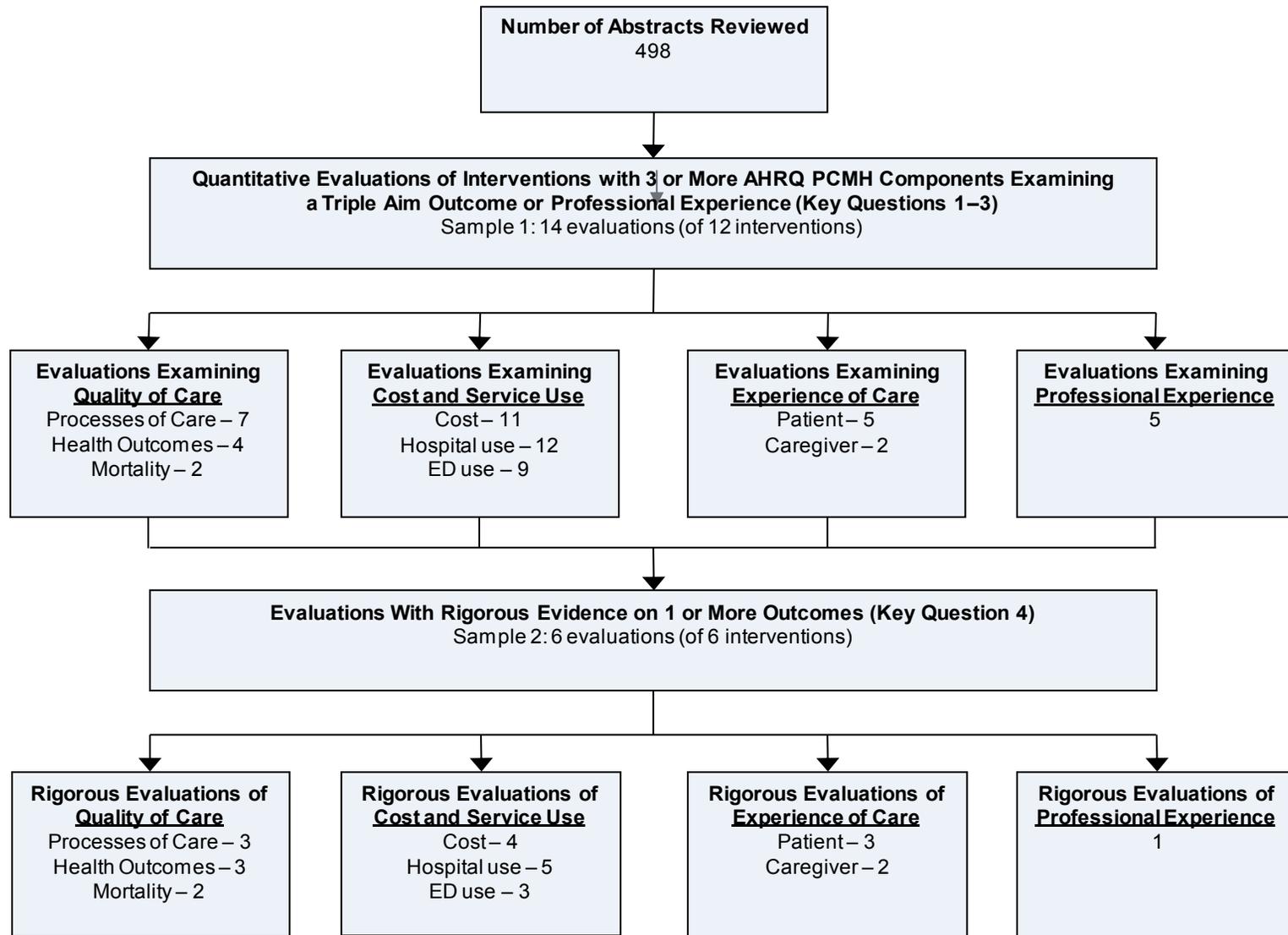
¹¹ To make the correction for clustering, we used the methodology of the What Works Clearinghouse (WWC), which can be found in Appendix C of the WWC Procedures and Standards Handbook (<http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&toCId=9>).

¹² See Appendix B for details on this approach. Peikes, Dale, Lundquist, et al. (2011) discuss the importance of adjusting for clustering and how to do so in more detail.

interval around such results includes both favorable and unfavorable intervention-control differences.¹³

¹³ For example, Guided Care found that the costs for the intervention group were \$75,000 lower than those for the control group, but this difference was not statistically different from zero. In this case, the researchers reported the 95 percent confidence interval, which conveys that, if the intervention were tested 100 times, 95 times out of 100, the difference could randomly fluctuate anywhere from a *savings* of \$244,000 to a *cost increase* of \$151,000.

Figure 2. Selection of Evaluations for Key Questions 1–4



Chapter 3. The Medical Home Landscape: Which Interventions Have Been Tested, for Which Patients, and in Which Implementation Settings?

- Most of the 12 interventions were developed before the recent interest in the medical home.
- Most interventions added a care manager to operate within the primary care practice. Care management is an important component of the medical home, but does not fundamentally transform a practice in the way becoming a medical home does.
- These precursor interventions differed considerably from one another.
- Few evaluations documented how well the interventions were implemented, so our descriptions may be closer to what was planned than to what was delivered.
- Half the interventions provided financial support to the practices and among those that did not, all but two provided a care manager.
- Most interventions were designed to serve patients who were older or sicker than those in the average primary care patient panel.
- Most interventions were tested in practices that were part of larger delivery systems, and are not typical of primary care practices in the United States.

In this chapter, we describe the 12 interventions that met our criteria for inclusion in the review, interventions often cited as evidence in favor of the medical home model. We describe how similar the interventions are to the PCMH model, the patient populations they targeted, and the settings in which they were implemented. To enable comparison to the PCMH, we categorize the features of each intervention into AHRQ's five core principles and two of the facilitators (health IT and payment). We summarize the interventions as the study authors and implementers described them. Publications varied in the degree of detail they provided about how the intervention operationalized and implemented different functions and processes. We also note that *few evaluations documented how well the models were implemented*, so our descriptions may be closer to what was planned than what was delivered when there was a difference.

Which Interventions Have Been Tested?

Table 1 provides an overview of the interventions, and Table 17 in Appendix C categorizes them using the framework of the AHRQ PCMH definition.

Because many of the interventions were developed before the recent interest in the medical home, *most of them embedded a care manager within the primary care practice, and did not fundamentally transform the practice, as would be required to become a medical home*. The care manager was typically responsible for developing an individualized care plan for high-risk patients and for coordinating their care. By providing care management from within the practice, these early interventions tested a core component of the medical home. Most of these early interventions also included each of the five AHRQ PCMH principles, but they did so in a less integrated and comprehensive manner than current demonstrations and typically for a subset of patients. A few introduced more fundamental changes to the way primary care practices operate (such as daily care team huddles, 24/7 access, and longer patient appointments) and could be considered early prototypes of the medical home model. The field is evolving rapidly, and many

of these diverse interventions have been adapted to look more like the medical home model since the joint principles were released. However, at the time they were studied, they represent precursors to the model being tested today. This serves as a reminder that the evidence commonly cited in support of the PCMH is actually on precursors, and needs to be interpreted in that context.

Table 1. Overview of the 12 interventions reviewed

Intervention	Overview	Sources Cited^a
Aetna's Embedded Case Managers	Nurse case managers are embedded in primary care practices to help manage care for Medicare Advantage members and collaborate with the clinical team.	Hostetter, 2010
Care Management Plus	Nurse care managers supported by specialized health IT tools are embedded within primary care clinics to orchestrate care for chronically ill elderly patients.	Agency for Healthcare Research and Quality, 2011; Dorr, Wilcox, Brunner, et al., 2008.
Community Care of North Carolina	Community-based care management provided through networks of primary care physicians (PCPs), a hospital, the Department of Social Services, and the health department. Case managers from a nonprofit work with PCPs in the network to coordinate care and undertake population health management.	Domino, Humble, Lawrence, et al., 2009; Lodh, 2005; Ricketts, Greene, Silberman, et al., 2004; Steiner, Denham, Ashkin, et al., 2008; Wilhide and Henderson, 2006.
Geisinger Health System Proven-Health Navigator	Geisinger Health Plan embedded a nurse case manager for every 900 Medicare Advantage patients in primary care practices to identify high-risk patients, design patient-centered care plans, provide care coordination and care transition support, and monitor patients using patient-accessible electronic health records.	Gilfillan, Tomcavage, Rosenthal, 2010; Graff, 2009; Paulus, Davis, and Steele, 2008; Steele, Haynes, Davis, et al., 2010.
Geriatric Resources for Assessment and Care of Elders (GRACE)	An advanced practice nurse and social worker assess low-income seniors in the home, and develop and implement a care plan with a geriatrics interdisciplinary team, in collaboration with the patient's PCP.	Bielaszka-DuVernay, 2011; Counsell, Callahan, Tu, et al., 2009; Counsell, Callahan, Clark, et al., 2007; Counsell, Callahan, Buttar, et al., 2006.
Group Health Cooperative Medical Home	Group Health redesigned a clinic to be a PCMH by changing staffing, scheduling, point-of-care, patient outreach, health IT, and management; reducing caseloads; increasing visit times; using team huddles; and introducing rapid process improvements.	Group Health News, 2010; Reid, Coleman, Johnson, et al., 2010; Reid, Fishman, Yu, et al., 2009.
Guided Care	Guided Care nurses are embedded in the primary care practice to provide assessments, care plans, monthly monitoring, and transitional care to the highest-risk Medicare patients.	Boult, Reider, Leff, et al., 2011; Boyd, Reider, Frey et al., 2010; Guided Care Web site, 2010; Leff, Reider, Frick et al., 2009; Marsteller, Hsu, Reider, et al., 2010; Wolff, Rand-Giovanetti, Palmer, et al., 2009; Wolff, Rand-Giovanetti, Boyd, et al., 2010.
Improving Mood-Promoting Access to Collaborative Treatment for Late-Life Depression (IMPACT)	A depression clinical specialist care manager (a nurse or psychologist) is embedded in the primary care practice to provide depression care for elderly depressed patients in coordination with the PCP, a consulting PCP, and a psychiatrist.	Hunkeler, Katon, Tang, et al., 2006; IMPACT Implementation Center Web site, 2010; Levine, Unützer, Yip, et al., 2005; Unützer, Katon, Williams, et al., 2001; Unützer, Katon, Callahan, et al., 2002; Unützer, Katon, Fan, et al., 2008.
Merit Health System and Blue Cross Blue Shield (BCBS) of North Dakota Chronic Disease Management Pilot	BCBS embedded a chronic disease management nurse in the clinic for patients with diabetes. The nurse assesses the patients' knowledge of diabetes, sets goals for disease self-management, establishes the need for in-person or telephone followup, and refers patients to services.	Fields, Leshen, and Patel, 2010; McCarthy, Nuzum, Mika, et al., 2008.

Intervention	Overview	Sources Cited ^a
Pediatric Alliance for Coordinated Care	A pediatric nurse practitioner from each practice allocates 8 hours per week to coordinate the care of children with special health care needs and make expedited referrals to specialists and hospitals; a local parent of a child with special health care needs provides consultations to the practice.	Palfrey, Sofis, Davidson, et al., 2004; Silvia, Sofis, and Palfrey, 2000.
Pennsylvania Chronic Care Initiative	Integrates the chronic care model and the medical home model for patients with diabetes and pediatric patients with asthma and includes the following key components: patient-centered care, teaching self-management of chronic conditions, forming partnerships with community organizations, financial incentives for providers, and making data-driven decisions.	AcademyHealth State Health Research and Policy Interest Group, 2009; Chronic Care Management, Reimbursement and Cost Reduction Commission, 2008; Houy, 2008; Torregrossa, 2010.
Veterans Affairs Team-Managed Home-Based Primary Care	Comprehensive and longitudinal primary care provided by an interdisciplinary team that includes a home-based primary care (HBPC) nurse in the homes of veterans with complex, chronic, terminal, or disabling diseases.	Department of Veterans Affairs, 2007; Hughes, Weaver, Giobbie-Hurder, et al., 2000.

^aThe findings in this paper are based on information from the sources cited here.

As Table 17 in Appendix C shows, nearly all 12 interventions incorporated each of the core PCMH principles, but varied in the degree to which they emphasized them.

- **Patient-Centered.** Most (9 of 12) of the interventions emphasized the development of an individualized care plan by a nurse care manager. Four of these interventions reported that the plan was developed in collaboration with the patient and/or caregiver, and it is difficult to determine from the available information which interventions shared the plan with patients. Several interventions also included specific protocols for developing plans and skills for patient self-management. Systematic patient outreach efforts (such as GRACE’s monthly contacts with patients and caregivers and Aetna’s option for the family to sit in on patient office visits) were less common across the interventions. One intervention (CCNC) was described as focusing on culturally sensitive and whole-person care.
- **Comprehensive Care.** All 12 interventions adopted a team-based approach to patient assessment and care planning. Although the size and composition of the teams varied across interventions, all teams included a PCP and a care manager. In 10 of the 12 interventions, this care manager was a nurse who may have worked with other clinical and nonclinical staff. The other two interventions used a nurse in some practices, and a psychologist, social worker, or other clinician in others. Several interventions also included team members from various disciplines, such as specialists, physical therapists, social workers, and pharmacists, depending on the needs of the patient. All 12 interventions were tested in practices led by physicians; none were tested in practices led by nurse practitioners or physician assistants.
- **Coordinated Care.** The mechanisms and processes for ensuring effective coordination of patient care by the care manager were rarely described. Half the interventions reported coordinating care with community-based services.

- **Access to Care.** Many of these interventions implemented mechanisms to enhance access to care and facilitate scheduling. These included dedicated telephone lines to a nurse care manager or care team and, among selected interventions, use of email messaging with care providers, telephone visits, and self-scheduling using electronic systems.
- **Systems Approach to Quality and Safety.** In 8 of the 12 interventions, there appeared to be some emphasis on systematizing the use of evidence-based care protocols. Most practices with EHRs embedded such protocols and care reminders in the EHR system. Eight interventions implemented quality assurance mechanisms, including regular training for care managers, random or periodic chart reviews, and analysis of quality and clinical outcomes (among practices with EHR capabilities).

Interventions also varied in the extent to which they emphasized payment and health IT.

- **Payment: Financial and In-Kind Support.** Practices and/or physicians in 6 of the 12 interventions received financial support or incentive payments to facilitate and incentivize practice changes. Of those six interventions, three provided practices with funding for startup, infrastructure (for example, health IT), and/or practice transformation costs. Three of the six provided stipends to physicians or other providers, and two provided a payment to the practice according to the number of patients served or enrolled in the program. Five interventions implemented shared savings or performance incentive programs. Five of the six interventions providing financial support to practices also provided a care manager.¹⁴ Among the six that did not provide practices with financial support, four provided a care manager.¹⁵
- **Health Information Technology.** Half of the 12 interventions used EHRs to help primary care teams better manage their patients and systematize the application of evidence-based care protocols. Most EHR systems were used to track patients, generate reminders, and calculate patient statistics to better target their disease management and care coordination efforts to high-risk patients. A few EHR systems were accessible to patients via a Web-based interface.

¹⁴ Aetna's Embedded Case Managers, CCNC, Group Health Cooperative Medical Home, GHS ProvenHealth Navigator, Merit Health System and BCBS of North Dakota Chronic Disease Management Pilot, and Pennsylvania Chronic Care Initiative provided financial support or incentive payments. Of these, GHS ProvenHealth Navigator, Merit Health System and BCBS of North Dakota Chronic Disease Management Pilot, and Pennsylvania Chronic Care Initiative provided funding for startup, infrastructure, and/or practice transformation costs. GHS ProvenHealth Navigator, Group Health Cooperative Medical Home, and Pennsylvania Chronic Care Initiative provided stipends to providers while Aetna's Embedded Case Managers and CCNC provided payments for patients served or enrolled in the program. Finally, all but CCNC implemented shared savings or performance incentive programs and all but Pennsylvania Chronic Care Initiative provided care managers.

¹⁵ The four included IMPACT, Guided Care, GRACE, and Care Management Plus.

Where Were the Interventions Tested?

Most interventions were tested in unique settings not typical of primary care practices in the United States. Table 2 summarizes the implementation settings and target populations for the 12 interventions.

Only 4 of the 12 interventions were not tested in practices in larger delivery systems such as an integrated delivery system (IDS) or comparable organizational context (Table 2). No estimates of the proportion of primary care practices that operate nationwide within larger delivery systems are available. However, an earlier study estimated that there were only about 100 IDSs in the country, most commonly in the West and Upper Midwest (Enthoven, 2009), so these are not typical of the primary care landscape. The evaluations tested interventions implemented in rural communities, in cities, across multiple counties, across entire States, and across multiple States.

Most interventions were tested in a relatively small number of intervention practices. Two interventions were tested in a single practice, four in 5 to 10 practices, three in 11 to 20, and only three were tested in more than 20 practices. As discussed in another AHRQ white paper (Peikes, Dale, Lundquist, et al., 2011), a small number of intervention practices limits the statistical power of evaluations to detect effects of interventions delivered at the practice level. From an implementation perspective, a trial with 10 or more intervention practices is large and challenging. However, from an evaluation perspective, given the need to account for clustering, a practice-level intervention with fewer than 20 intervention practices is generally too small to be able to detect effects on some outcomes, such as costs. An evaluation might require well over 500 intervention practices/sites to be likely to detect changes in costs across all patients, and at least 20 intervention practices to detect changes in costs only among high-risk patients.

Table 2. Target populations and implementation settings

Intervention Name	Target Population					Implementation Setting			
	Target Population (and Consent Rate When Reported)	Includes All Patients	Limited to Medicare Patients	Limited to Patients With Chronic Physical or Mental Illness	Includes Patients With Both Fee-for-Service and Managed Insurance Coverage	Practice Setting	Limited to Larger Delivery Systems	Number of Practices	Use Electronic Health Records
Aetna's Embedded Case Managers	Medicare beneficiaries enrolled in Medicare Advantage		Yes			Practices across the United States		36	
Care Management Plus	Medicare fee-for-service patients aged 65 or older with complex chronic care needs identified by the primary care physician		Yes	Yes		Moderate-sized primary care clinics (4 family medicine and 3 internal medicine practices) in a large Integrated Delivery System (IDS) in Utah	Yes	7	Yes
Community Care of North Carolina	Medicaid patients in North Carolina			Yes	Yes	12 PCMH networks in North Carolina (exact number of primary care practices in networks not reported)		Many	
Geisinger Health System ProvenHealth Navigator	Enrollees in Geisinger Health Plan's Medicare Advantage (MA) plan		Yes			GHS practices in rural central Pennsylvania in a large IDS	Yes	11	Yes
Geriatric Resources for Assessment and Care of Elders	Medicare patients with incomes less than 200% of the Federal poverty level; 43% consent rate		Yes		Yes	Primary care physicians in community-based health centers in urban area of Indiana in an IDS	Yes	6	Yes
Group Health Cooperative Medical Home	All adult patients in the Seattle clinic	Yes			Yes	Moderate-size primary care clinic in Seattle operating within an IDS.	Yes	1	Yes

Table 2 (continued)

Intervention Name	Target Population					Implementation Setting			
	Target Population (and Consent Rate When Reported)	Includes All Patients	Limited to Medicare Patients	Limited to Patients With Chronic Physical or Mental Illness	Includes Patients With Both Fee-for-Service and Managed Insurance Coverage	Practice Setting	Limited to Larger Delivery Systems	Number of Practices	Use Electronic Health Records
Guided Care	Roughly 25% of a practice's sickest aged Medicare patients; 38% consent rate		Yes	Yes	Yes	Primary care teams ("pods" of 2 to 5 physicians), including 18 physicians from practices in 3 large delivery systems in Baltimore/Washington metropolitan areas	Yes	8	Yes
Improving Mood-Promoting Access to Collaborative Treatment for Late-Life Depression	Socioeconomically diverse sample of elderly (≥60 years) patients with major depression and/or dysthymia			Yes	Yes	450 primary care providers in primary care clinics operating in IDSs (mostly academic medical centers) within 8 health care organizations in 5 States	Yes	18	
Merit Health System and Blue Cross Blue Shield (BCBS) of North Dakota Chronic Disease Management Pilot	Patients with diabetes			Yes	Yes	A MeritCare internal medicine clinic in an IDS in North Dakota	Yes	1	Yes
PA Chronic Care Initiative	Adult patients with diabetes and pediatric patients with asthma			Yes	Yes	Practices in southeastern Pennsylvania		32	
Pediatric Alliance for Coordinating Care (PACC)	Children with special health care needs			Yes	Yes	Pediatric primary care practices in the Boston area (4 private practices; 2 community health centers)		6	

Table 2 (continued)

Intervention Name	Target Population					Implementation Setting			
	Target Population (and Consent Rate When Reported)	Includes All Patients	Limited to Medicare Patients	Limited to Patients With Chronic Physical or Mental Illness	Includes Patients With Both Fee-for-Service and Managed Insurance Coverage	Practice Setting	Limited to Larger Delivery Systems	Number of Practices	Use Electronic Health Records
Veterans Affairs Team-Managed Home-Based Primary Care	Veterans with limitations in 2 or more activities of daily living (ADLs) or a prognosis of terminal illness or homebound with congestive heart failure or chronic obstructive pulmonary disease; 89% consent rate			Yes	Yes	VA medical centers with HBPC programs	Yes	16	

Note: “Practices” also includes clinics and health centers.

Whom Did the Interventions Serve?

Most of the interventions targeted patients who were older or sicker than the average patient panel. Of the 12 interventions, the Group Health Cooperative Medical Home intervention was the only one designed to serve all patients in a practice (see Table 2). The other interventions targeted either patients with chronic illnesses or Medicare patients, who tend to be sicker than the average patient in a primary care practice. Most evaluations tested the interventions on adult patient populations; however, PACC was tested solely on pediatric patients. Five interventions served Medicare patients exclusively, and one served Medicaid patients exclusively. Nine of the interventions were tested on patients with both fee-for-service and managed care insurance coverage.

These interventions have limited generalizability to the current PCMH model being tested. They were tested largely among patients sicker and older than average and in practices that were part of larger delivery systems that are not representative of most practices in the United States. Most new PCMH demonstrations and initiatives are designed to serve all patients in the primary care practice, and many encompass diverse practice types, including small and medium-sized independent practices, as well as federally qualified health centers.¹⁶

¹⁶ As described in Chapter II, NDP and IMHP are excluded because they tested the effect of facilitated versus non-facilitated implementation of the medical home model, rather than the effect of the model itself. These interventions made an important contribution to the field, so we briefly describe them along with their implementation settings and target populations:

The NDP was comprehensive. It included elements of each of the five principles in the medical home definition and emphasized the use of health IT. Practices did not receive any funding or additional practice staff. Eighteen practices received practice facilitation to become a medical home; the other 18 sought to become medical homes on their own. The 36 practices are diverse, and include small, independent practices in 25 States. The practices served all patients (Jaén, Ferrer, Miller, et al., 2010; Nutting, Crabtree, Stewart, et al., 2010). In contrast to most studies, the NDP provided rich insights about the implementation experience.

The IMHP model was tested in pediatric primary care practices and included elements of four of the five core principles; the model included no changes in access to care. Five practices received practice facilitation to become a medical home; the other five received technical assistance, but no facilitation, to become a medical home. These 10 practices served children and youth with special health care needs in urban, suburban, and rural regions of Illinois. All practices received small mini-grant stipends (up to \$5,000 per year for 2 years); training for quality improvement teams, which included family members; and medical home resources, including the Center for Medical Home Improvement tool kit, training materials, and other resources (American Academy of Pediatrics Illinois Chapter, 2009).

Chapter 4. How Do Studies Evaluate the Interventions?

- Among the 14 evaluations, study designs varied:
 - 6 used comparison group designs.
 - 4 used randomized, controlled trials.
 - 2 used pre-post studies.
 - 2 did not report the study design they used.
- Among the 14 evaluations, the outcomes varied:
 - 5 examined all triple aim outcomes.
 - 8 examined process of care or health outcome measures.
 - 8 examined total costs (including intervention costs).
 - 5 examined patient or caregiver experience of care.
 - 5 examined health care professional experience.

To understand the scope of current research for evaluating the PCMH model, the review team collected information on evaluation designs and outcomes for the 12 interventions included in the review. Because one intervention—CCNC—was evaluated by three distinct evaluations, we examine 14 evaluations here.

Evaluation Designs

The evaluations used a range of designs to assess the effects of these interventions (Table 3). These designs include patient-level RCTs, cluster RCTs (where practices, physicians, or groups of physicians, rather than patients, are randomized), comparison group designs, and pre-post designs. A comparison group design, considered to be fairly rigorous, was the most commonly used (by evaluations of CMP, GHS ProvenHealth Navigator, Group Health, and Merit Health System and Blue Cross Blue Shield (BCBS) of North Dakota Chronic Disease Management Pilot, and two evaluations of CCNC). An RCT (including cluster RCT), considered the strongest possible research design, was used by four evaluations: two RCTs (IMPACT and VA TM/HBPC), and two cluster RCTs (GRACE and Guided Care).¹⁷ Weaker designs, such as pre-post designs, were used by two evaluations (Pennsylvania Chronic Care Initiative and PACC) to examine all outcomes. The remaining two evaluations (Aetna’s Embedded Case Managers and one of the three evaluations of CCNC) did not report the design they used.

Overall, strong evaluation designs were used to evaluate many of the interventions. Because strong designs might not yield unbiased estimates of the effects of an intervention if the evaluation suffers from problems (such as high attrition of participants) or if analytic methods lack rigor, we also examined these additional features when rating the rigor of each evaluation.

¹⁷ Two of these studies used cross-sectional or pre-post designs to examine some outcomes, as noted in Table 3.

Outcome Measures

Among these early evaluations, fewer than half (5 of 14) looked at all three triple aim outcomes (quality, cost, and patient or caregiver experience) (Table 3). Seven evaluations examined process-of-care measures; four investigated health outcome measures (such as functioning, self-rated health status, and quality of life); and two examined mortality. Eleven evaluations examined cost, but only eight of these examined total costs including the intervention. Without information on the costs of providing the intervention, it is difficult to assess whether the intervention increased costs, generated savings, or was cost neutral from the payer's perspective. Twelve evaluations examined hospital use, and nine examined ED use. Although a major focus of the *patient-centered* medical home is on patient experience, only five of these evaluations examined measures related to patient experience, and only two examined caregiver experience. The dearth of studies examining patient and caregiver experience may reflect the relatively high cost of collecting survey data, or the fact that these models predated the current interest in the PCMH, which emphasizes patient-centeredness. Five of the 14 evaluations examined effects on professional experience.

Given the nascent stage of research on the medical home and the different goals of stakeholders, it is not surprising that evaluations varied in the measures and the length of followup periods for each outcome. This variation makes it difficult to synthesize results across evaluations. An example illustrates how measures within each outcome varied across the evaluations: process-of-care measures ranged from Prevention Quality Indicators (PQIs, used in CMP), to HEDIS measures (Group Health Cooperative Medical Home), to Assessing Care of Vulnerable Elders (ACOVE) general and geriatric care measures (GRACE). Similarly, measures of health outcomes included the SF-36, Barthel Index, quality of life, and self-rated health status. Additionally, outcomes were evaluated for different durations of followup, ranging from 8 months to 3 years (see Chapter 5 for more details). The Commonwealth Fund recognized the need to harmonize measures, data sources, and followup periods and has recently released core measures for evaluating the PCMH (Rosenthal, Abrams, Bitton, et al., 2012).¹⁸

¹⁸ See <http://www.commonwealthfund.org/Publications/Data-Briefs/2012/May/Measures-Medical-Home.aspx>.

Table 3. Evaluation designs and outcomes of the 12 interventions

Intervention Name	Evaluation Design	Cost and Service Use										
		Quality of Care			Cost		Service Use		Experience			
		Outcomes	Health	Mortality	Processes of Care	Without Intervention	With Intervention	Hospital Use	ED Use	Patient	Caregiver	Professional
Aetna's Embedded Case Managers	Unknown due to limited information				Yes			Yes				Yes
Care Management Plus	Comparison group design			Yes	Yes			Yes	Yes			
Community Care of North Carolina (Evaluation 1, Domino, Humble, Lawrence, et al., 2009)	Comparison group design					Yes	Yes	Yes	Yes			
Community Care of North Carolina (Evaluation 2, Ricketts, Greene, Silberman, et al., 2004)	Comparison group design					Yes	Yes	Yes	Yes			
Community Care of North Carolina (Evaluation 3, Lodh, 2005)	Unknown due to limited information					Yes	Yes					
Geisinger Health System ProvenHealth Navigator	Comparison group design					Yes		Yes				
Geriatric Resources for Assessment and Care of Elders	Cluster RCT ^a	Yes	Yes	Yes		Yes	Yes	Yes	Yes			Yes
Group Health Cooperative Medical Home	Comparison group design				Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Guided Care	Cluster RCT					Yes	Yes	Yes	Yes	Yes	Yes	Yes
Improving Mood—Promoting Access to Collaborative Treatment for Late-Life Depression	RCT ^b	Yes			Yes	Yes	Yes			Yes		Yes
Merit Health System and Blue Cross Blue Shield (BCBS) of North Dakota Chronic Disease Management Pilot	Comparison group design				Yes	Yes		Yes	Yes			

Table 3 (continued)

Intervention Name	Evaluation Design	Cost and Service Use								
		Quality of Care			Cost		Service Use		Experience	
		Outcomes	Mortality	Processes of Care	Without Intervention	With Intervention	Hospital Use	ED Use	Patient	Caregiver
Pediatric Alliance for Coordinated Care	Pre-post design						Yes	Yes	Yes	
Pennsylvania Chronic Care Initiative	Pre-post design	Yes		Yes	Yes		Yes	Yes		
Veterans Affairs Team-Managed Home-Based Primary Care	RCT	Yes			Yes	Yes	Yes		Yes	Yes

Notes: Table 1 contains the references for each intervention.

Health outcomes include measures of functioning, quality of life, and self-rated health status.

Processes of care include measures such as proportion of patients with controlled blood pressure, or with preventable hospitalizations.

Professional experience includes the experience of staff and other health care professionals.

^aThe main study design was a cluster randomized, controlled trial, but professional experience was examined only for intervention group providers using a cross-sectional study.

^bThe main study design was a randomized, controlled trial, but professional experience was examined only for intervention group providers using a pre-post study.

Chapter 5. Which Evaluations Provide Rigorous Evidence of Effects on Each Outcome?

- Six of the 14 evaluations received a high or moderate rating for analysis of at least one outcome.
- When the evidence on an outcome is rated low or excluded, the intervention itself could be effective (or not); however, limitations in the evaluation methods prevent us from reliably drawing such a conclusion.

Having described the 12 interventions and the research designs and outcomes used to evaluate them, we now report the findings of a formal assessment of the rigor of the evidence on effectiveness of these interventions. Our goal was to identify analyses based on strong designs and methods that generated internally valid estimates of the effects of the intervention. We rated the rigor of each analysis conducted by an evaluation as high, moderate, low, or excluded. Tables 4.1 and 4.2 show these ratings along with the primary rationale for the rating.

We note here that the ratings developed in this chapter are of the evidence generated by the evaluations. When the evidence on an outcome is rated low or excluded, the intervention itself could be effective (or not), but limitations in the evaluation design or methods prevent us from reliably drawing such a conclusion. Such evaluations represent important efforts to build the evidence base and may provide important insights about how best to refine a specific intervention and guidance on how to design future evaluations; however, their usefulness in determining the quantitative effectiveness of the intervention is limited. To learn about the potential of these interventions to improve outcomes, evaluations using stronger methods are needed.

Ratings by Intervention

We began the sequential assessment of each analysis conducted as part of an evaluation by assessing the strength of the evaluation design. As noted in Chapter 4, 12 of the 14 evaluations provided information on design, while 2 (Aetna and 1 of the CCNC evaluations) did not provide sufficient information to make it possible to assess the strength of their designs. Because knowledge of the evaluation design is crucial in assessing whether the evaluation produced internally valid estimates of effects, we gave these two evaluations an “excluded” rating to indicate that we do not have sufficient information to make this assessment.

Among the 12 evaluations that provided information on design, 10 used strong designs of an RCT or a nonexperimental comparison group for most of their outcomes, and 2 (Pennsylvania Chronic Care Initiative and PACC) used pre-post designs. We gave a low rating to all analyses conducted by the two pre-post evaluations, because it is difficult to attribute changes in the outcomes to the intervention alone rather than to other factors (such as secular trends) that might have influenced the outcome—factors that could be controlled with the use of a comparison group.

The 10 evaluations that used RCTs and nonexperimental comparison group designs were next assessed for the strength of their analytic methods. Of these, 6 received a high or moderate rating for at least one outcome, while 4 received a low rating for all outcomes examined.¹⁹

Of the six evaluations that received a high or moderate rating for at least one outcome (GRACE, IMPACT, VA TM/HBPC, Guided Care, CMP, and GHS ProvenHealth Navigator), ratings varied considerably within the evaluation—an indication that the evaluation produced more rigorous evidence on some outcomes than others. Table 4.1 shows the ratings for these six evaluations. The lack of an appropriate comparison group was the most common reason for a low rating. Well-matched comparison groups (that are similar to the intervention group in terms of baseline values of key patient outcome measures, as well as practice variables such as the mix of patients, number of providers, and key infrastructure such as EHRs) are important because they help gauge the extent to which change over the course of the evaluation is due to the intervention rather than to other concurrent changes.²⁰ In general, an evaluation that compares patients in pioneering, high-performing practices that chose to participate in an intervention, with patients in practices that had average performance before the intervention and who did not choose to change may artificially make the intervention look more effective than it truly is. Next we describe how we arrived at the ratings for the six evaluations that received a high or moderate rating for at least one outcome:

- **GRACE:** We gave a high rating to analyses of most outcomes examined by this cluster RCT evaluation. We rated one process-of-care measure low because it was examined for an endogenous subgroup—patients who had one or more hospital stay in the first year—and the evaluation did not report whether the intervention and control patients in this subgroup were comparable at baseline. We also rated the analysis of professional experience low, because the evaluation examined effects on this outcome without using a comparison group.
- **IMPACT:** We gave a high rating to analyses of most outcomes examined by this RCT evaluation. The analysis of patient experience at two of four followups received a low rating because the outcome was examined for an endogenous subgroup, and the evaluation did not report whether the intervention and control group subgroups were comparable at baseline. We also rated the analysis of professional experience low because the evaluation examined effects on this outcome without using a comparison group.

¹⁹ The primary rationale for a low rating varied across the four evaluations as follows:

- **Community Care of North Carolina.** Two evaluations of CCNC used comparison group designs, but their analyses did not establish whether the intervention and comparison groups had comparable values of the outcome at baseline.
- **Group Health Cooperative Medical Home and Merit Health System and Blue Cross Blue Shield (BCBS) of North Dakota Chronic Disease Management Pilot.** Analyses from both evaluations received a low rating because these interventions were implemented in a single intervention practice. While such a design can represent an important opportunity to pilot a new intervention and break ground toward a larger evaluation, it cannot distinguish the effects of the intervention from other characteristics of the particular practice that implemented it, thereby undermining the ability of attributing an observed effect to the intervention.

²⁰ Because most studies do not report all this information, our formal rating criterion was more liberal: we assessed only whether the intervention and comparison groups had comparable baseline values of the outcome being examined.

- **VA TM/HBPC:** The analyses of costs and service use for patients both with and without a terminal illness, and the analyses of health outcomes and experience of care for patients and caregivers among patients without a terminal illness, were rated high. However, during the course of the evaluation, there was high attrition among the subgroup of patients with a terminal illness, and the evaluation did not ascertain whether the intervention and control groups who were alive at followup had similar baseline values of the outcomes being examined. Analyses of all outcomes for this subgroup received a low rating because we do not know whether the intervention rather than pre-intervention differences between the surviving members of intervention and control groups are driving the results.
- **Guided Care:** Analyses of all outcomes, except professional experience, received a high rating. The sample of providers experienced high attrition during the evaluation. However, because the evaluation showed that intervention and control group providers had similar baseline values of the professional experience measures, and also controlled for these values in the analyses, the analyses of professional experience received a moderate rating.
- **GHS ProvenHealth Navigator:** This comparison group evaluation received a moderate rating (the highest rating a comparison group evaluation can receive) for its analysis of hospital use because the intervention and comparison groups had comparable hospital use before the intervention began, and the evaluation controlled for baseline hospital use in the analyses. Analysis of costs, however, received a low rating because the intervention and comparison groups did not have similar values of costs at baseline.
- **CMP:** This comparison group design evaluation received a moderate rating for all analyses it conducted.

Table 4.1. Evaluations with ratings of high or moderate on at least one outcome

Intervention	Evaluation Design	Evidence Rating: Outcome	Primary Rationale
Care Management Plus	Comparison group design	Moderate: Hospital use and emergency department (ED) use, process of care measures, and mortality	Intervention and comparison groups had equivalent outcomes at baseline and the study controlled for baseline values of the outcome
Geisinger Health System ProvenHealth Navigator	Comparison group design	Moderate: Hospital Use	Intervention and comparison groups had equivalent outcomes at baseline and the study controlled for baseline values of the outcome
		Low: Costs	Intervention and comparison groups did not have equivalent values of the outcome at baseline

Table 4.1 (continued)

Intervention	Evaluation Design	Evidence Rating: Outcome	Primary Rationale
Geriatric Resources for Assessment and Care of Elders	Cluster RCT ^a	High: Health outcomes; mortality; costs, hospital use, and ED use; 18 of 19 process of care measures	RCT with low attrition at followup and no reported statistically significant baseline differences in the outcomes
		Low: 1 of 19 process of care measures (followup primary care visit occurred within 6 weeks of hospital discharge)	Outcome based on an endogenous subgroup (those with a hospitalization in the first year of the intervention)
		Low: Health care professional experience	Evaluation design prevents attribution of changes in the outcome to the intervention rather than other factors
Guided Care	Cluster RCT	High: Costs, hospital use, and ED use; patient experience of care outcomes; caregiver experience of care outcomes	RCT with low attrition and no reported statistically significant baseline differences in the outcomes
		Moderate: Health care professional experience	RCT with high attrition but intervention and control group samples at followup had equivalent values of the outcome at baseline and study controlled for baseline values of the outcome
Improving Mood—Promoting Access to Collaborative Treatment for Late-Life Depression	RCT ^b	High: Health outcomes; process of care outcomes; patient experience of care at 3 and 12 months; costs	RCT with low attrition and no reported statistically significant baseline differences in the outcomes
		Low: Patient experience of care at 18 and 24 months	Outcome based on endogenous subgroup (those reporting depression care in the past 6 months)
		Low: Health care professional experience	Evaluation design prevents attribution of changes in the outcome to the intervention rather than other factors
Veterans Affairs Team-Managed Home-Based Primary Care	RCT	High: Hospital use; costs; for subgroup of nonterminally ill patients—health outcomes and patient and caregiver experience of care	RCT with low attrition and no reported statistically significant baseline differences in the outcomes
		Low: For subgroup of terminally ill patients—patient and caregiver health outcomes; patient and caregiver experience of care	RCT with high attrition among terminally ill patients; study does not report if the intervention and control group samples at followup had equivalent outcomes at baseline

Note: The attrition criterion for RCTs accounts for both overall attrition and differential attrition between the intervention and control groups.

^aThe main evaluation design was a cluster randomized, controlled trial, but health care professional experience was examined only for intervention group providers using a cross-sectional study.

^bThe main evaluation design was a randomized, controlled trial, but health care professional experience was examined only for intervention group providers using a pre-post study.

Table 4.2. Evaluations with ratings of low or excluded on all outcomes

Intervention	Evaluation Design	Evidence Rating: Outcome	Primary Rationale
Aetna's Embedded Case Managers	Unknown	Excluded: Hospital use; process of care; health care professional experience	Limited information on design and analysis reported
Community Care of North Carolina (Evaluation 1)	Comparison group design	Low: Costs, hospital use, and ED use	Intervention and comparison groups did not have equivalent outcomes at baseline
Community Care of North Carolina (Evaluation 2)	Comparison group design	Low: Costs, hospital use, and, ED use	Study did not report if the intervention and comparison groups had equivalent outcomes at baseline
Community Care of North Carolina (Evaluation 3)	Unknown	Excluded: Costs	Limited information on design and analysis reported
Group Health Cooperative Medical Home	Comparison group design	Low: Costs, hospital use, and ED use; process of care; patient experience of care; health care professional experience	Systematic confounding due to implementation in only one clinic
Merit Health System and Blue Cross Blue Shield (BCBS) of North Dakota Chronic Disease Management Pilot	Comparison group design	Low: Costs, hospital use, and ED use; process of care	Systematic confounding due to implementation in only one clinic
Pediatric Alliance for Coordinated Care	Pre-post design	Low: Hospital use and ED use; patient experience of care	Evaluation design prevents attribution of changes in the outcome to the intervention rather than other factors
Pennsylvania Chronic Care Initiative	Pre-post design	Low: Costs, hospital use, and ED use; health outcomes; process of care	Evaluation design prevents attribution of changes in the outcome to the intervention rather than other factors

Ratings by Outcome

Not all evaluations examined all outcomes, and among those that did, not all used rigorous methods. Table 5 shows the number of evaluations receiving a high or moderate rating for an outcome among those that examined that outcome (which is also shown in Figure 2). This ranges from one to five studies, with three studies for most outcomes.

Table 5. Number of evaluations that assessed each triple aim outcome and health care professional experience

Outcome	Number of Evaluations That Assessed the Outcome	
	Using Any Method	Using Rigorous Methods
Quality of Care		
Processes of Care	7	3
Health Outcomes	4	3
Mortality	2	2
Cost and Service Use		
Costs (with or without the Intervention)	11	4
Hospital Use	12	5
Emergency Department Use	9	3
Experience of Care		
Patient	5	3
Caregiver	2	2
Health Care Professional Experience		
Health Care Professional Experience	5	1

Chapter 6. Evidence Synthesis

- The rigorous evidence on PCMH effectiveness indicates some favorable effects across quality of care, hospital and ED use, and patient and caregiver experience, a few unfavorable effects on costs, and mostly inconclusive results (because of insufficient sample sizes or inadequate methods to account for clustering in the data).
- The interventions were tested in atypical practices and among older and sicker-than-average patients.

This chapter synthesizes the evidence on each key outcome from analyses rated as high or moderate.

Approach

As described in Chapter 2, we synthesize findings on four broad outcomes of interest: (1) quality of care (including processes of care, health outcomes, and mortality); (2) costs, hospital use, and ED use; (3) patient and caregiver experience; and (4) health care professional experience. Within these outcomes, we classify the findings on each measure into one of four categories: (1) statistically significant with a favorable effect, (2) statistically significant with an unfavorable effect, (3) inconclusive because of lack of statistical significance, and (4) inconclusive because of uncertain statistical significance. Inconclusive findings cannot indicate whether the intervention worked or not. As described in Chapter 2, we consider findings that are not statistically significant as inconclusive rather than as evidence that the intervention had no effects because we suspect that the evaluations lacked adequate power to detect effects that might have existed. We also view as inconclusive findings from analyses that did not correctly account for clustering of patients within practices, because their reported statistical significance is likely overstated. We report the magnitude of effects for statistically significant effects on cost and service use, but not for other outcomes, and note that the true effect on cost and service use could fall in a wide range around the reported estimate.

Findings From Rigorous Evaluations

Improving the Quality of Care: Processes of Care, Health Outcomes, and Mortality

The evidence on processes of care and health outcomes indicates some improvements with the majority of evidence being inconclusive.

Processes of Care. Of the three rigorous evaluations of processes of care (IMPACT, GRACE, and CMP), only IMPACT had favorable effects (Table 6). It increased rates of medication use throughout its 2 years and increased use of psychotherapy or specialty mental health care in the first year but not the second. However, for GRACE and CMP, findings on processes of care were inconclusive, because some results were not statistically significant and because statistical significance for other results had been determined without accounting for the clustered nature of the data.

Table 6. Summary of evidence on processes of care

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Care Management Plus			1 year: Preventive Quality Indicator (PQI) hospitalizations among all patients, patients with diabetes, and patients without diabetes 2 years: PQI hospitalizations among all patients	2 years: PQI hospitalizations among patients with and without diabetes
Geriatric Resources for Assessment and Care of Elders Improving Mood—Promoting Access to Collaborative Treatment	3 months; 6 months: Increased rates of antidepressant use, psychotherapy 1 year: Increased rates of antidepressant use, psychotherapy 1.5 years; 2 years: Increased rates of antidepressant use		1 Year: 2 process of care measures 1.5 years; 2 years: Rates of psychotherapy	1 Year: 16 process of care measures

Note: A result is considered statistically significant if $p < 0.05$. Statistical significance is considered uncertain if the study did not account for clustering of patients within practices when calculating statistical significance.

Health Outcomes. Evaluations of GRACE, IMPACT, and VA TM/HBPC provided rigorous evidence on measures of health outcomes. IMPACT improved nearly all the health measures that were examined, GRACE improved some of the measures, and VA TM/HBPC did not improve any (Table 7). IMPACT reduced depression symptoms and improved overall quality of life and the SF-12 physical component score throughout the first and second years of the program, and reduced overall functional impairment through the first 1.5 years. General health status, which was first measured at 1 year, showed an improvement that continued through the second year. GRACE improved half the SF-36 scales; however, for the other scales, as well as for activities of daily living, instrumental activities of daily living, and days in bed, the findings were inconclusive because the results were not statistically significant. The findings from the

evaluation of VA TM/HBPC were also inconclusive because the results for functional status and quality of life of patients with a nonterminal illness were not statistically significant.²¹

Table 7. Summary of evidence on health outcomes

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Geriatric Resources for Assessment and Care of Elders	2 years: Improved 4 of 8 Short Form (SF)-36 scales		2 years: 4 of 8 SF-36 scales, Activities of Daily Living (ADLs), Instrumental Activities of Daily Living, and days in bed	
Improving Mood—Promoting Access to Collaborative Treatment	3 months; 6 months: Reduced depression symptoms, overall impairment; improved overall quality of life 1 year: Reduced depression symptoms, overall impairment; improved SF-12 physical component score, quality of life, and general health 1.5 years: Reduced depression symptoms, overall impairment; improved SF-12 physical component score, quality of life, and general health 2 years: Reduced depression symptoms, improved SF-12 physical component score, quality of life, and general health		2 years: Overall impairment	
Veterans Affairs Team-Managed Home-Based Primary Care		1 year: Worsened 1 of 8 SF-36 scales for nonterminal patients	1 year: Barthel index and 7 of 8 SF-36 scales for nonterminal patients	

Note: For all outcomes, a result is considered statistically significant if $p < 0.05$. Statistical significance is considered uncertain if the study did not account for clustering of patients within practices when calculating statistical significance.

²¹ As the table indicates, the intervention group had a worse score than the control group for one of the eight SF-36 scales; however, it is unlikely that the intervention could have reduced functioning. Because we expect 1 of every 20 intervention-control differences to be statistically significant by chance, this statistically significant finding may have arisen simply from testing multiple differences rather than as an effect of the intervention.

Mortality. Two evaluations (of GRACE and CMP) examined effects on mortality using rigorous methods. Although effects among all patients would not be expected in the short followup periods employed, they are possible among the high-risk Medicare patients these interventions served. The evidence was inconclusive for both interventions (Table 8). The evaluation of CMP did not adjust statistical significance for clustering, which rendered the results they reported as statistically significant inconclusive. Results from its remaining analyses, and results from GRACE, were not statistically significant. Without knowing the power, it is difficult to determine whether these indicate no effects.

Table 8. Summary of evidence on mortality

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Care Management Plus			<p>2 years: All patients</p>	<p>1 year: All patients and patients with diabetes</p> <p>2 years: Patients with diabetes</p>
Geriatric Resources for Assessment and Care of Elders			<p>2 years: All patients</p>	

Note: A result is considered statistically significant if $p < 0.05$. Statistical significance is considered uncertain if the study did not account for clustering of patients within practices when calculating statistical significance.

Reducing Costs of Care and Service Use

The evidence on costs is mixed, with some increases, one reduction, and some inconclusive results. The evidence on hospital use and ED use, two important drivers of costs, is mostly inconclusive but shows some reductions.

Cost. Four evaluations examined the effects on total costs (including intervention costs) using rigorous methods (GRACE, IMPACT, VA TM/HBPC, and Guided Care; Table 9). Two of these evaluations reported statistically significant cost *increases* and the other two were inconclusive because the cost differences were not statistically significant.

The VA TM/HBPC *increased* costs by 12 percent over a 1-year period among its target population of chronically and terminally ill veterans, and GRACE *increased* costs by 28 and 14 percent among Medicare patients in the first and second years of the intervention.²² The cost increases for GRACE were driven by increases of 46 and 30 percent in each year among the low-risk subgroup of Medicare patients. Although GRACE increased costs in both years, it is the only evaluation to report any savings for some enrollees. These savings of 23 percent for high-risk Medicare patients (those with a probability of repeated admission [PRA] score of 0.4 or above, which represented one-fourth of the enrollees) emerged a year after the 2-year

²² While we present point estimates of the statistically significant effects, we note that the true effect could lie anywhere in a confidence interval around the point estimate.

intervention ended (when there were no intervention costs).²³ These reductions were enough to offset the 19 percent increase in costs of low-risk patients. As a result, there were no statistically significant differences between intervention and control group costs for all patients in the third year.

The evaluations of IMPACT and Guided Care reported no statistically significant differences in costs. Without sufficient information on statistical power, it is not possible to determine whether this represents lack of effects, and therefore cost neutrality, or insufficient sample size to detect effects. As a result, the evidence on costs from these evaluations is inconclusive.

Table 9. Summary of evidence on cost

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Geriatric Resources for Assessment and Care of Elders ^a	Year 3: Reduced 23% among high-risk patients	Year 1: Increased 28% among all patients and 46% among low-risk patients Year 2: Increased 14% among all patients and 30% among low-risk patients Year 3: Increased 19% among low-risk patients	Year 1: High-risk patients Year 2: High-risk patients Year 3: All patients	
Guided Care			8 months: All patients	
Improving Mood-Promoting Access to Collaborative Treatment for Late-Life Depression			4 years: All patients	
VA Team-Managed Home-Based Primary Care		Months 1–12: Increased 12%		

Notes: A result is considered statistically significant if $p < 0.1$. The effects reported here are point estimates; the true effects can lie anywhere within the confidence intervals around these estimates, and may well be different from the point estimate. We do not include the confidence intervals because not all studies reported them.

^aWe adjusted the p-values for clustering for findings on costs for the GRACE evaluation using a range of estimates of the intracluster correlation (ICC) from the literature. The findings remained statistically significant after the adjustment.

²³ The evaluation of GRACE was the only one to measure outcomes after the intervention had ended.

Hospital Use. Evaluations of five interventions (GRACE, VA TM/HBPC, Guided Care, CMP, and GHS ProvenHealth Navigator) provide rigorous evidence on hospital use. Some evidence indicates reductions in use, but the remaining evidence is inconclusive (Table 10). GHS ProvenHealth Navigator, the only intervention that reduced hospitalizations for all patients (Medicare Advantage enrollees), decreased hospitalizations by an estimated 18 percent and readmissions by 36 percent.

Two other interventions (GRACE and VA TM/HBPC) also reduced the number of hospitalizations (readmissions in the case of VA TM/HBPC), but these were limited to certain high-risk subgroups served by these interventions. GRACE did not have a statistically significant effect on hospitalizations in the first year for a high-risk subgroup of patients (defined by a high probability of repeated hospital admission), but it did reduce hospitalizations by 44 percent in the second year and 40 percent in the third (postintervention) year for this subgroup. VA TM/HBPC reduced the number of readmissions among the subgroup of severely disabled VA patients by 22 percent in the first 6 months of the yearlong intervention, but did not have a statistically significant effect over the first year when the next 6 months were included, which suggests that the estimate for the first 6 months was not sustainable. Turning to all enrollees, the evidence from GRACE and VA TM/HBPC evaluations is inconclusive.

Evidence from the remaining two interventions is also inconclusive. Guided Care’s evaluation found that intervention-control differences in the number of hospitalizations among its targeted population of high-risk Medicare patients were not statistically significant. In the case of CMP, intervention-comparison differences in the odds of hospitalization among all patients and among the subgroup without diabetes were not significant, and among the subgroup with diabetes had uncertain significance, rendering all these findings inconclusive.

Table 10. Summary of evidence on hospital use

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Care Management Plus ^a			Year 1; Year 2: Odds of hospitalization among all patients and patients without diabetes	Year 1; Year 2: Odds of hospitalization among patients with diabetes
Geisinger Health System ProvenHealth Navigator	4 years: Reduced number of stays by 18% Reduced number of readmissions by 36%			

Table 10 (continued)

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Geriatric Resources for Assessment and Care of Elders ^b	<p>Year 2: Reduced number of stays by 44% among high-risk (high PRA score) patients</p> <p>Year 3: Reduced number of stays by 40% among high-risk (high PRA score) patients</p>		<p>Year 1: Number of stays among all patients and high-risk (high PRA score) patients</p> <p>Year 2: Number of stays among all patients</p>	
Guided Care			<p>8 Months; 20 Months: Number of stays</p>	
Veterans Affairs Team-Managed Home-Based Primary Care	<p>Months 1-6: Reduced number of readmissions by 22% among severely disabled patients</p>		<p>Months 1-6; Months 1-12: Proportion readmitted among all patients and severely disabled patients</p> <p>Months 1-6: Number of readmissions among all patients</p> <p>Months 1-12: Number of readmissions among all patients and severely disabled patients</p>	

Notes: A result is considered statistically significant if $p < 0.05$. The effects reported here are point estimates; the true effects can lie anywhere within the confidence intervals around these estimates, and may well be different from the point estimate. We do not include the confidence intervals because not all studies reported them.

^aBecause the statistical significance of findings on the odds of hospital use in the CMP evaluation did not take clustering into account, we made this adjustment by using ICC estimates from the literature. However, we found that the statistical significance of these findings was sensitive to the estimate of ICC used; the findings remained significant when some estimates of ICC were used but lost their significance when other estimates were used. Therefore, we consider the statistical significance of these findings to be uncertain.

^bThe p-values for findings on hospital stays for the GRACE evaluation were adjusted for clustering using a range of estimates of the intracluster correlation (ICC) from the literature. The findings remained statistically significant after the adjustment.

Emergency Department Use. Evaluations of the GRACE, Guided Care, and CMP interventions used rigorous methods to examine effects on ED use (Table 11). Only one of these evaluations found favorable effects: intervention-control differences in the first year of the GRACE intervention were not statistically significant, but in the second year, GRACE reduced the number of ED visits by 24 percent for all patients and by 35 percent for its high-risk patients. The other two evaluations—Guided Care and CMP—provided inconclusive evidence on ED use. The evaluation of Guided Care did not find statistically significant differences in either the first 8

or 20 months of the intervention. Similarly, the evaluation of CMP did not find statistically significant differences in the odds of an ED visit in the first year among all patients, or among patients with or without diabetes. In the second year, results for all patients continued to be not significant, while among the subgroups with and without diabetes, the results had uncertain significance because the analyses did not account for clustering.

Table 11. Summary of evidence on emergency department use

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Care Management Plus ^a			Year 1: Odds of an ED visit among all patients, patients without diabetes, and patients with diabetes Year 2: Odds of an ED visit among patients with diabetes	Year 2: Odds of an ED visit among all patients and patients without diabetes
Geriatric Resources for Assessment and Care of Elders ^b	Year 2: Reduced number of visits by 24% among all patients and by 35% among high-risk patients		Year 1: Number of visits among all patients and high-risk patients	
Guided Care			8 Months; 20 Months: Number of visits	

Notes: A result is considered statistically significant if $p < 0.05$. The effects reported here are point estimates; the true effects can lie anywhere within the confidence intervals around these estimates, and may well be different from the point estimate. We do not include the confidence intervals because not all studies reported them.

^aBecause the statistical significance of findings on the odds of ED use in the CMP study did not take clustering into account, we made this adjustment ourselves by using ICC estimates from the literature. However, we found that the statistical significance of these findings was sensitive to the estimate of ICC used; the finding remained significant when some estimates of ICC were used but lost their significance when other estimates were used. Therefore, we consider the statistical significance of these findings to be uncertain.

^bWe adjusted the p-values for findings for GRACE on ED use for clustering using a range of estimates of the intraclass correlation (ICC) from the literature. The findings remained statistically significant after the adjustment.

Improving the Experience of Care

Patients and Caregivers

Some evidence on patient and caregiver experience of care is favorable and the rest is inconclusive.

Patient Experience of Care. Of the three rigorous evaluations of patient experience (IMPACT, VA TM/HBPC, and Guided Care), two found evidence of some improvements, while the third was inconclusive (Table 12). IMPACT improved patients’ satisfaction with depression care at 3 months and 1 year. Patients with a nonterminal illness in the VA TM/HBPC reported improvements in all but one measure: patients in the intervention group had better access to care, interpersonal experience, technical quality, communication, and self-reported outcomes, but did not have statistically significant differences in satisfaction with care, compared to patients in the control group. The findings from Guided Care were inconclusive, in large part because the analyses did not account for clustering when calculating statistical significance and because for one measure—decision support—the intervention-control difference was not statistically significant.

Table 12. Summary of evidence on patient experience

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Guided Care			1.5 years: Decision support	1.5 years: Care coordination, overall quality of care, goal setting, problem solving, patient activation
Improving Mood—Promoting Access to Collaborative Treatment	3 months; 12 months: Improved satisfaction with care			
Veterans Affairs Team-Managed Home-Based Primary Care	Year 1: Improved access to care, interpersonal experience, technical quality, communication, and self-reported outcomes among nonterminal patients		Year 1: Satisfaction with care among nonterminal patients	

Notes: A result is considered statistically significant if $p < 0.05$.

Caregiver Experience of Care. Of the two rigorous evaluations of caregiver experience (VA TM/HBPC and Guided Care), the evaluation of VA TM/HBPC found mostly favorable effects, while the evaluation of Guided Care was inconclusive (Table 13). The VA TM/HBPC improved caregiver ratings of the quality of care received by patients with nonterminal illness. The intervention also improved most measures of caregiver functional status as measured by the SF-36 and reduced one of two measures of burden among caregivers of patients with nonterminal illness. For Guided Care, intervention-control differences in caregiver experience with the quality of care provided to patients, caregiver burden, and caregiver productivity at 18 months were not statistically significant; the statistical significance of differences in the remaining measures was uncertain because the analyses did not account for clustering.

Table 13. Summary of evidence on caregiver experience

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Guided Care ^a			<p>6 months: Caregiver burden among all caregivers</p> <p>18 months: 1 of 6 measures of caregiver experience with quality of care provided to patients among all caregivers, 4 of 6 measures among high-intensity caregivers, and 3 of 6 measures among low-intensity caregivers</p> <p>18 months: Caregiver burden among all, high-intensity, and low-intensity caregivers</p> <p>18 months: Caregiver productivity among all, high-intensity, and low-intensity caregivers</p>	<p>6 months: Caregiver burden among high-intensity caregivers</p> <p>18 months: 5 of 6 measures of caregiver experience with quality of care provided to patients among all caregivers, 2 of 6 measures among high-intensity caregivers, 3 of 6 measures among low-intensity caregivers</p>
Veterans Affairs Team-Managed Home-Based Primary Care	<p>12 months: Improved caregiver experience with quality of care provided to patients among nonterminal patients</p> <p>12 months: Reduced 1 of 2 measures of caregiver burden among nonterminal patients</p> <p>12 months: Improved 6 of 8 SF-36 scales among nonterminal patients</p>		<p>12 months: 1 of 2 measures of caregiver burden among nonterminal patients</p> <p>12 months: 2 of 8 SF-36 scales among nonterminal patients</p>	

Notes: A result is considered statistically significant if $p < 0.05$.

^aHigh-intensity caregivers are those providing more than 14 hours per week of assistance, and low-intensity caregivers are those providing less than 14 hours per week of assistance.

Health Care Professional Experience

The evaluation of Guided Care is the only rigorous evaluation of professional experience. Its findings were inconclusive; the intervention-control differences were either not statistically significant or had uncertain statistical significance because the analyses did not account for clustering (Table 14).

Table 14. Summary of evidence on professional experience

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Guided Care			12 months: Satisfaction with care management, time spent on chronic care, knowledge of patients' personal circumstances, and coordination of care	12 months: Satisfaction with communication and knowledge of patients' clinical circumstance

Note: A result is considered statistically significant if $p < 0.05$. Statistical significance is considered uncertain if the study did not account for clustering of patients within practices when calculating statistical significance.

Effects Over Time

We expect the effects of the medical home on key outcomes to follow different time paths. Transformation to a medical home could change outcomes in multiple, often nonlinear, ways, and improvements in outcomes could occur at different times. For example, improving access to care could lead to identification and management of ambulatory-care-sensitive conditions, which could reduce ambulatory-care-sensitive hospitalizations and ED visits. However, other changes could lead to worse outcomes in the short term. For example, more systematic assessment of patient care needs could uncover unmet needs, which could increase cost and service use in the short run but eventually reduce them through avoidance of acute exacerbations in the long run. Similarly, the introduction of health IT and new team members could create significant new organizational challenges for a practice, which could initially worsen patient and professional experience; however, these outcomes could improve over time through, for example, lower medication errors and quicker record-keeping.

In addition, for a given outcome, we expect the time path of effects to vary across different subgroups of patients. For example, costs for patients who had been receiving unnecessary care might fall, while those for patients with unmet needs might rise. For example, increased use of preventive care could increase short term costs but decrease them over longer periods. In the case of hospital use, patients who are at high risk of immediate hospitalization, such as those with severe congestive heart failure, might benefit from improved access to care and experience reductions in hospitalizations.

Unfortunately, summarizing the evidence on effects over time is currently difficult, for several reasons. First, the interventions, implementation settings, and target populations differ, which makes it impossible to ascertain whether differential effects over time reflect a true time path rather than these differences. Second, evaluations often used different followup periods, which make it difficult to compare time paths of effects on a given outcome across evaluations. Third, with the exception of costs, evaluations differed in the measures they used to examine a given outcome, which again makes it difficult to summarize effects over time across evaluations. Fourth, within a study, formal statistical testing would be needed to determine whether an effect that appears to grow over time is really statistically different across the followup periods; none of the evaluations performed such testing.

For example, in the case of costs, only one of the four rigorous evaluations (GRACE) examined effects over time. It found that it took 3 years for favorable effects on costs for high-

risk patients to emerge; however, differences over time were not formally tested. In the case of hospital use, four of the five evaluations providing rigorous evidence on different measures of hospital use examined effects over time. Of these, two provided conclusive evidence but found different time paths. For GRACE, it took time for favorable effects on the number of hospital stays for high-risk patients to emerge. In contrast, VA TM/HBPC's reduction in readmissions over the first 6 months disappeared with an additional 6 months of followup.

Current evidence on time paths is limited for the reasons cited above. Future evaluations could make a valuable contribution to the field by explicitly incorporating logic models based on theories of change and clinical experience to guide examination of key outcomes over time.

Generalizability of This Evidence Is Limited

Because the interventions were tested on particular types of patients and practices, the evidence from these six evaluations may not be applicable to the typical patient panel in the average primary care setting in the United States. All interventions targeted patients who were older and sicker (and thus at higher risk of poor health outcomes) than the general population in a practice. GRACE served low-income Medicare beneficiaries; IMPACT, patients aged 60 and over with depression; VA TM/HBPC, veterans with terminal or very complex chronic diseases; Guided Care, 25 percent of the sickest Medicare beneficiaries in a practice; CMP, chronically ill Medicare beneficiaries; and GHS ProvenHealth Navigator, Medicare Advantage patients. All were tested in practices that were part of a larger delivery system. Decisionmakers must use caution when trying to extrapolate from lessons learned from these early evaluations.

Chapter 7. Conclusion and Next Steps

- The PCMH model is a promising approach to improving health care delivery, but stronger evaluations of the current model are needed to determine how it alters each of the key outcomes and how it could be refined to enhance effectiveness. Information about which patients benefit and under what circumstances will be key to interpreting this evidence.
- There are many opportunities to improve the future evidence base. Some of these are specific to evaluations of the PCMH, and others are general best practices for conducting high-quality health service evaluations:
 - Use strong evaluation designs and methods.
 - Conduct comprehensive implementation studies.
 - Test the model in an adequate number of practices and measure different outcomes for different subgroups of patients.
 - Follow outcomes for longer periods of time (despite pressure for rapid research results).
 - Improve reporting and documentation of evaluation methods.
 - Independently evaluate the models to ensure transparency and objectivity.
 - Test the model in typical practices and among typical patients.
 - Develop a core set of outcome measures and standardized measures of PCMH components.
 - Measure effects on all triple aim outcomes and professional experience.
 - Explore novel approaches to evaluating PCMH interventions.

Our review of the PCMH reveals that most of the current evidence on the PCMH has limited direct applicability to the PCMH model being tested today, for two reasons. First, most interventions are best viewed as precursors to the PCMH. They include some, but not all, aspects of the model. This reflects the recent rise in interest in the model: the joint principles on the PCMH were released in 2007, and we reviewed evidence through September of 2010. Second, they were tested on generally older and sicker-than-average patients, and in unique implementation settings. More research is needed to determine whether the PCMH model currently being tested will improve outcomes for all patients in typical primary care practices around the Nation.

We found that six interventions were evaluated using rigorous methods for at least one outcome. These interventions had a range of effects on key outcomes: some favorable effects on quality, and patient and caregiver experience of care; a few unfavorable effects on costs; and many inconclusive results on all outcomes. Inconclusive results are not evidence for or against the effectiveness of an intervention. Well-designed evaluations with adequate power and rigorous methods are needed to determine effectiveness.

We now summarize the findings and how they fit into the literature, cite considerations for decisionmakers applying the evidence to current medical home initiatives, discuss limitations of this review, and suggest ways to build a better evidence base.

Summary of Findings

With the exception of some favorable effects on quality of care, hospital and ED use, and patient and caregiver experience of care, and a few unfavorable effects on costs, the findings on the effectiveness of precursors to the PCMH are largely inconclusive, either because the sample size was insufficient to detect effects if they exist or because the statistical significance of the effects was potentially overstated owing to lack of adjustment for clustering of patients within practices. More evidence from ongoing evaluations of current PCMH initiatives is needed to determine whether the PCMH as it is currently being implemented improves the quality, affordability, and experience of care; under what circumstances it does so; and how refinements to the model can increase its effectiveness. Table 15 provides a high-level summary of the rigorous evidence by outcome (described in detail in Chapter 6).

- **Quality of Care.** There were favorable effects on quality of care among the rigorous evaluations: one of three evaluations reported improvements in processes of care, and two of three reported improvements in some measures of health outcomes. However, the remaining evaluations examining these outcomes, as well as both evaluations examining mortality, produced inconclusive evidence.
- **Cost and Service Use.** The rigorous evidence on cost and service use shows limited favorable effects, some unfavorable effects on cost, and many inconclusive results. The evaluation of GRACE was the only one of four evaluations to find any evidence of savings, and these were limited to the high-risk subgroup of Medicare patients in the third (or post-intervention) year. However, both GRACE and VA TM/HBPC increased total costs during the intervention, while evidence from the other two interventions—Guided Care and IMPACT—was inconclusive. Similarly, GHS ProvenHealth Navigator was the only one of five rigorous evaluations to report reductions in hospital use for its full sample of patients; two other evaluations (of GRACE and VA TM/HBPC) reported reductions only for their high-risk subgroups in some followup periods. Evidence on hospital use from the remaining evaluations (of Guided Care and CMP) was inconclusive. Similarly, one of three evaluations (of GRACE) found reductions in ED use in one of the two followup periods, but evidence from the remaining evaluations (of Guided Care and CMP) was inconclusive.
- **Experience of Care.** The rigorous evidence on patient and caregiver experience shows some favorable effects, while the rest of the evidence is inconclusive.
- **Professional Experience.** The lone evaluation with rigorous evidence on professional experience is inconclusive.

Table 15. Snapshot of findings

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Processes of Care				
Geriatric Resources for Assessment and Care of Elders			Yes	Yes
Care Management Plus			Yes	Yes
Improving Mood—Promoting Access to Collaborative Treatment for Late-Life Depression	Yes		Yes	
Health Outcomes				
Geriatric Resources for Assessment and Care of Elders	Yes		Yes	
Improving Mood—Promoting Access to Collaborative Treatment for Late-Life Depression	Yes		Yes	
Veterans Affairs Team-Managed Home-Based Primary Care		Yes	Yes	
Mortality				
Care Management Plus			Yes	Yes
Geriatric Resources for Assessment and Care of Elders			Yes	
Cost				
Geriatric Resources for Assessment and Care of Elders	Yes	Yes	Yes	
Guided Care			Yes	
Improving Mood—Promoting Access to Collaborative Treatment for Late-Life Depression			Yes	
Veterans Affairs Team-Managed Home-Based Primary Care		Yes		
Hospital Use				
Care Management Plus			Yes	Yes
Geisinger Health System ProvenHealth Navigator	Yes			
Geriatric Resources for Assessment and Care of Elders	Yes		Yes	
Guided Care			Yes	
Veterans Affairs Team-Managed Home-Based Primary Care	Yes		Yes	
Emergency Department Use				
Care Management Plus			Yes	Yes
Geriatric Resources for Assessment and Care of Elders	Yes		Yes	
Guided Care			Yes	

Table 15 (continued)

Intervention	Statistically Significant		Inconclusive	
	Favorable	Unfavorable	Not Statistically Significant	Uncertain Statistical Significance
Patient Experience				
Guided Care			Yes	Yes
Improving Mood–Promoting Access to Collaborative Treatment for Late-Life Depression	Yes			
Veterans Affairs Team-Managed Home-Based Primary Care	Yes		Yes	
Caregiver Experience				
Guided Care			Yes	Yes
Veterans Affairs Team-Managed Home-Based Primary Care	Yes		Yes	
Health Care Professional Experience				
Guided Care			Yes	Yes

Placing the Findings in Context

The findings are less conclusive than those from most prior reviews. We found some promising results across quality of care, hospital and ED use, and patient and caregiver experience; however, the majority of findings were inconclusive. The conclusions we draw are consistent with those of Friedberg, Lai, Hussey, et al. (2009), who described the evidence in favor of the medical home as “scant.” Our conclusions are more tentative than those of Homer, Klatka, Romm, et al. (2008); Fields, Leshen, and Patel (2010); and Grumbach and Grundy (2010), who claimed overwhelming evidence in support of the medical home. We conclude that more work, including additional evaluations that are well-designed, implemented, and analyzed, is needed to guide decisionmakers on this young, rapidly evolving model.

Findings from the rigorous evaluations reflect unique contexts and patients. All were tested in practices that were part of larger delivery systems and targeted patients who were older and sicker than the average in U.S. primary care practices (see Table 5). As a caveat, we expect it will be harder to generate cost and service use effects of the same size among healthier patients, who do not use many services.

The improvements in cost and service use may have been concentrated among the sickest patients. Two of the six rigorous evaluations examined outcomes for different high-risk subgroups of patients among their target population of older or sicker patients.²⁴ The evaluation of GRACE reported that even among its low-income, elderly patients, improvements were concentrated among the sickest. The evaluation of the VA TM/HBPC intervention found favorable effects among severely disabled veterans but not among other high-risk veterans; it is unclear whether this reflects lack of power to detect effects (a result of small samples), lack of followup periods long enough for effects to emerge, or a true lack of effects.

These results, while limited, raise the question of whether conducting additional analyses on sicker patients could be a useful approach for future evaluations. The highest-risk patients present providers with more opportunities to take action to reduce service use and costs in the relatively short followup periods observed because a medical home intervention is likely to reduce hospitalizations more for patients who are frequently hospitalized. In addition, there is better power to detect effects among the highest-risk patients than among all patients, reducing the likelihood of missing important beneficial effects (Peikes, Dale, Lundquist, et al., 2011). This does not imply that the PCMH should be targeted only to patients with complex medical needs. The PCMH is a whole-practice-level intervention and is expected to improve care for all. It is critical not to confuse the goal of the intervention with suggestions for refining evaluations.

Findings from multi-component medical home interventions in other settings will likely differ. Decisionmakers should consider context when interpreting findings: the findings on effectiveness will differ if the full medical home model is implemented, and with different practices, markets, and patients. For example, implementing the PCMH model in certain markets or delivery settings where there is overuse of care could produce results different from those in areas where care is underused. Similarly, modifications of the interventions might alter outcomes. For example, it is possible that adding certain components of the medical home, such as health IT and stronger financial incentives to practices, could improve outcomes. In addition, program designers may be able to identify areas in which efficiency can be increased to achieve cost neutrality or generate savings. For example, although this information was not provided in reports of these evaluations, a careful review of which team members are best suited to provide which parts of interventions could lower the costs of care.

Limitations

This evidence review has several limitations, many of which stem from the fact that the PCMH, and therefore research on the PCMH, are in their nascent stages. First, the summary we present here is based on the limited number of well-evaluated interventions currently available, and on interventions that are often related, but not identical, to the PCMH model. Therefore, these findings have limited applicability to the medical home as it is currently being rolled out throughout the country. Second, we did not have resources to survey the model implementers to fully characterize the models tested and how faithful the actual implementation was to what was planned. A survey or an in-depth interview would be needed for this purpose, because most articles in the literature gave brief descriptions of the interventions, using language and categories that were not standardized, and few described the fidelity of the implementation to its original design. Third, most evaluations did not report the statistical power to detect effects. If an

²⁴ CMP did so, too, but results are inconclusive owing to the lack of adjustment for clustering.

evaluation has an insufficient number of practices and is underpowered, it is unlikely to detect plausible effects even when they exist. Therefore, there is a chance that the number of programs that truly had effects (favorable or unfavorable) is understated in our synthesis, because we were forced to classify so many of them as inconclusive. Finally, publication bias might lead us to overstate favorable effects because journals are less likely to publish studies that find no effects or unfavorable effects.

Guidance to Improve the Future Evidence Base

This review highlights opportunities to improve the evidence base on the PCMH going forward. There is a large risk that research currently under way on PCMH interventions (not reviewed here) will fail to support decisionmakers' information needs. A survey of 26 current medical home pilots in 18 States concluded that only 40 percent had well-developed evaluation plans. Among those with plans, only about 40 percent planned to use a comparison group design, and the others planned to use pre-post designs (Bitton, Martin, and Landon, 2010), which typically provide weak evidence.

The challenges to conducting rigorous evaluations are not unique to the PCMH. In 2011, the GAO criticized evaluations of 127 diverse health care interventions for weak evaluation designs, limited generalizability, and failure to report the outcomes of interest (in their case, quality and cost) (U.S. Government Accountability Office, 2011). Below we describe a number of steps that can be taken to improve the evidence base. Some of these are specific to evaluations of the PCMH, and others are general best practices for conducting rigorous health service evaluations:

- **Use strong evaluation designs and methods.** Weak designs and analytical methods severely limit the potential of a strong intervention to produce rigorous evidence for decisionmakers. Current and future evaluators of PCMH interventions have an opportunity to fill knowledge gaps and contribute to ongoing learning on PCMH effectiveness. One challenge for a good evaluation of the medical home is to make sure that practices and patients in the intervention and comparison groups are comparable at baseline. Otherwise, it is difficult to distinguish effects that are due to the intervention from preexisting differences between the intervention and comparison practices and patients. Evaluations should also use rigorous analytical methods, including adjusting analyses for clustering of patients within practices (see Peikes, Dale, Lundquist, et al., 2011).
- **Conduct comprehensive implementation studies.** We found that most evaluations did not report how the intervention was implemented. While undertaking an implementation study requires additional expertise and resources, it adds tremendous value in identifying barriers and facilitators to improving outcomes, how findings might generalize to other contexts, and ways to refine the model. Implementation evaluations can provide powerful insights on their own, as well as when combined with quantitative outcome studies (a mixed-methods approach).²⁵

²⁵ Creswell, Klassen, Clark, et al., (2011) provide useful guidance on mixed methods. Crabtree, Chase, Wise, et al., (2011) emphasize the necessity of a mixed-methods approach when evaluating the PCMH model.

- **Test the model in an adequate number of practices and measure different outcomes for different subgroups of patients.** Because the PCMH is a practice-level intervention, it must be tested in a large number of practices, or the evaluation is likely to lack the statistical power to detect effects even when they exist. As discussed in Chapter 2, measuring costs and service use among sicker patients permits detection of smaller effects than among all patients. In contrast, measures for quality of care and patient experience typically take on a small number of values and result in less variation; therefore, effects on these outcomes can be examined and more easily detected among all patients.
- **Follow outcomes for longer periods of time.** Evaluations examined outcomes for 1 to 3 years, with most following patients for 2 years. While most decisionmakers are eager to obtain results, given the dramatic changes many practices must undergo to become medical homes, a short followup period might provide an overly pessimistic view of the medical home by capturing the negative effects of disruptive transformation. Consistent with this possibility, GRACE substantially *increased* costs by 28 percent early in the intervention, but became cost neutral a year after the intervention ended. However, the VA TM/HBPC evaluation found that short-term favorable effects on readmissions dissipated over time. Evaluation designs should also explicitly consider the periods of time needed to observe the effects of PCMH interventions on health care processes and subsequently on various health outcomes; information from early evaluations might be useful in modeling time paths of effects on different outcomes.
- **Improve reporting and documentation.** Many evaluations were not documented well enough to assess the strength of their methods. To allow objective assessment of the evidence, evaluation results—even preliminary results or results from pilot studies—should be accompanied by a detailed description of the methods used.
- **Independently evaluate the interventions to ensure objectivity.** Many evaluations were conducted by intervention developers. While developers have deep knowledge of their initiatives and a commitment to learning about them, independent evaluations might provide more credible evidence. At a minimum, peer review of evaluations conducted by developers would build a better evidence base.
- **Test the model in typical practices and among typical patients.** All six interventions with rigorous evidence were tested exclusively in practices in larger delivery systems that had some degree of integration across providers. Therefore, these results may not apply to independent practices. Ideally, future research would test the PCMH model in practices that are representative of the primary care landscape. In terms of patients, all six interventions were tested on those that were older or sicker than average. Also, while testing effects on specific patient populations is appropriate for evaluating specific research questions, as a practice-level intervention, the PCMH must be implemented for more diverse populations as decisionmakers still require evidence of effectiveness for the general patient population.

- **Examine a core set of outcome measures and develop standardized measures of PCMH components.** Estimating effects on a core set of standardized outcome measures would enable a meta-analysis of findings across different interventions, while still enabling researchers to accommodate the unique needs of their stakeholders. A meta-analysis can dramatically improve the power to detect effects compared to individual evaluations, which are often underpowered. The body of evidence would also be improved if researchers used detailed, standardized measures of PCMH components and processes. Such measures would enable meta-analyses to discern which intervention features are most effective in which settings. The Commonwealth Fund (2011) has convened a collaborative for medical home evaluators to support the creation of a uniform research infrastructure that can guide PCMH evaluations.
- **Measure effects on all triple aim outcomes and professional experience.** The PCMH model grew out of the need to improve quality and experience while reducing costs. It is critical that evaluations examine all these outcomes if they are to provide comprehensive information to decisionmakers. Improving one type of outcome may not warrant model adoption if it comes at the expense of deterioration in other outcomes. Examining the full range of outcomes might require addressing a number of barriers, including payer concerns about confidentiality of cost data, limited resources to collect and analyze multiple data sources, and lack of validated tools to measure certain outcomes.²⁶
- **Explore novel approaches to evaluating PCMH interventions.** A number of studies over the past decade have shown that health care interventions can be viewed as complex interventions within a complex adaptive system (CAS), similar to processes in ecology, computer science, and organizational science. Complexity science views the multiple components of complex interventions such as the PCMH as dependent on each other, as well as on the primary care practice and health care setting (Plsek and Greenhalgh, 2001). For example, quality of care delivered by a practice can be viewed as a system-level property that arises over time from the interactions among the members of the practice (Lanham, McDaniel, Crabtree, et al., 2009). As a result, in addition to individual processes or components, the relationships among practice team members become key levers for improving outcomes. Furthermore, the framework's emphasis on the importance of the external environment underscores the influence of the medical neighborhood on key outcomes. Some evidence indicates that interventions designed and implemented using CAS principles are more effective at improving clinical outcomes (Leykum, Parchman, Pugh, et al., 2010; Leykum, Pugh, Lawrence, et al., 2007).

²⁶ The 2011 release from AHRQ of PCMH-Consumer Assessment of Healthcare Providers and Systems (PCMH-CAHPS) may address one such barrier and enable future evaluators to measure patient experience more easily. Built on the existing, well-validated Clinician and Group survey, the PCMH-CAHPS covers topics such as provider-patient communication, coordination of care, and shared decisionmaking, and is available in adult and child versions, and in English and Spanish (<https://www.cahps.ahrq.gov/Surveys-Guidance/CG/PCMH.aspx>).

Principles of complexity science might be used to create better approaches to evaluate PCMH interventions, including designing more insightful implementation analyses (Litaker, Tomolo, Liberatore, et al., 2006; Campbell, Fitzpatrick, Haines, et al., 2000; Craig, Dieppe, Macintyre, et al., 2008; Stetler, Damschroder, Helfrich, et al., 2011; Damschroder, Aron, Keith, et al., 2009; Nutting, Crabtree, Stewart, et al. 2010; May, Mair, Dowrick, et al., 2007; Cohen, McDaniel, Crabtree, et al., 2004). Measures of the internal and external environment might be useful in both (1) selecting comparison practices that closely resemble the intervention practices, and (2) helping explain why an intervention was more successful in certain contexts than in others. More work is needed to develop such measures.²⁷ In addition, from a complexity framework, attempts to isolate the relative contributions of individual components of the medical home are ill-advised and are likely to result in misleading findings, because these components are dependent on each other to achieve the desired outcomes of medical home implementation.

By moving away from a mechanistic and reductionist perspective, complexity frameworks and other new approaches may help us evaluate PCMH interventions in more insightful ways.

²⁷ For example, measures of the external environment within which a PCMH operates could build on Parchman, Scoglio, and Schumm's (2011) modeling of health care delivery across a network of providers.

Looking Forward

Although the PCMH is a promising innovation, rigorous quantitative evaluations and comprehensive implementation analyses are needed to assess effectiveness and refine the model to meet stakeholders' needs. Findings from future evaluations can guide the substantial efforts of practices and payers to adopt the PCMH and ensure that the revitalized primary care system achieves the triple aim outcomes in a sustainable manner.

References and Included Studies

- AcademyHealth State Health Research and Policy Interest Group. The Pennsylvania Chronic Care Initiative. Presentation at the Annual AcademyHealth Research Meeting, 2009. http://www.academyhealth.org/files/interestgroups/shrp/SHRP_Breakfast_2009_Magistro.pdf.
- Agency for Healthcare Research and Quality. AHRQ's medical home definition. http://pcmh.ahrq.gov/portal/server.pt/community/pcmh_home/1483/PCMH_Defining%20the%20PCMH_v2.
- Agency for Healthcare Research and Quality. Primary care managers supported by information technology systems improve outcomes, reduce costs for patients with complex conditions. AHRQ Health Care Innovations Exchange Profile. Last updated: July 20, 2011. <http://www.innovations.ahrq.gov/content.aspx?id=264&tab=1>.
- American Academy of Family Physicians, American Academy of Pediatrics, American College of Physicians, American Osteopathic Association. Joint principles of a patient-centered medical home. March 2007. http://www.aafp.org/online/etc/medialib/aafp_org/documents/policy/fed/jointprinciplespcmh0207.Par.0001.File.dat/022107medicalhome.pdf.
- American Academy of Family Physicians. TransforMED Patient-Centered Medical Home model continues to evolve in part based on findings from National Demonstration Project. AAFP Press release, June 07, 2010. <http://www.aafp.org/online/en/home/media/releases/2010b/transformed-nationaldemonstrationproject.html>.
- American Academy of Pediatrics Illinois Chapter. Illinois Medical Home Project final report. June 29, 2009. <http://illinoisAAP.org/2010/01/illinois-medical-home-project-final-report/>.
- Backer LA. The patient centered medical home: still a work in progress. *Fam Pract Manag* 2010 Sep-Oct;17(5):30-5.
- Barr VJ, Robinson S, Marin-Link B, et al. The expanded Chronic Care Model: an integration of concepts and strategies from population health promotion and the Chronic Care Model. *Hosp Q*. 2003;7(1):73-82.
- Berenson RA, Devers KJ, Burton RA; for the Urban Institute. Will the patient-centered medical home transform the delivery of health care? August 1, 2011. <http://www.urban.org/publications/412373.html>.
- Berenson RA, Rich EC. How to buy a medical home? Policy options and practical questions. *J Gen Intern Med* 2010a Jun;25(6):619-24.
- Berenson RA, Rich EC. U.S. approaches to physician payment: The deconstruction of primary care. *J Gen Intern Med* 2010b Jun;25(6):613-8.
- Berwick DM. The science of improvement. *JAMA* 2008 Mar 12;299(10):1182-4.
- Berwick DM, Nolan TW, Whittington J. The triple aim: Care, health, and cost. *Health Aff (Millwood)* 2008 May-Jun;27(3):3759-69.
- Bielaszka-DuVernay C. The 'GRACE' model: In-home assessments lead to better care for dual eligibles. *Health Aff (Millwood)* 2011 Mar;30(3):431-4.
- Bitton A, Martin C, Landon BE. A nationwide survey of patient centered medical home demonstration projects. *J Gen Intern Med* 2010 Jun;25(6):584-92.
- Bodenheimer T, Pham HH. Primary care: Current problems and proposed solutions. *Health Aff (Millwood)* 2010 May;29(5):799-805.
- Bodenheimer T, Grumbach K, Berenson RA. A lifeline for primary care. *N Engl J Med* 2009 Jun 25;360(26):2693-6.
- Bodenheimer T, Berenson RA, Rudolph P. The primary care-specialty income gap: Why it matters. *Ann Intern Med* 2007 Feb;146(4):301-7.
- Bodenheimer T. Primary care—will it survive? *N Engl J Med* 2006 Aug 31;355(9):861-4.
- Boyd CM, Reider L, Frey K, et al. The effects of Guided Care on the perceived quality of health care for multi-morbid older persons: 18-month outcomes from a cluster-randomized controlled trial. *J Gen Intern Med* 2010 Mar;25(3):235-42.
- Boult C, Reider L, Leff B, et al. The effect of Guided Care teams on the use of health services: Results from a cluster-randomized controlled trial. *Arch Intern Med* 2011 Mar 14;171(5):460-6.
- Brown R, Peikes D, Peterson G, Schore J, and Razafindrakoto C. "Six Features of Medicare Coordinated Care Demonstration Programs That Cut Hospital Admissions of High-Risk Patients." *Health Affairs*, vol. 31, no. 6, June 2012, pp. 1156-1166.
- Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;321:694-6.
- Campbell MK, Elbourne DR, Altman DG. CONSORT statement: Extension to cluster randomised trials. *BMJ* 2004 Mar 20;328(7441):702-8.

- Chronic Care Management, Reimbursement and Cost Reduction Commission. Prescription for Pennsylvania. Right state, right plan, right now. Strategic plan. 2008. Available at <http://www.rxforpa.com/assets/pdfs/ChronicCareCommissionReport.pdf>. Accessed January 31, 2011.
- Cohen D, McDaniel RR Jr, Crabtree BF, Ruhe MC, Weyer SM, Tallia A, Miller WL, Goodwin MA, Nutting P, Solberg LI, Zyzanski SJ, Jaén CR, Gilchrist V, Stange KC. A practice change model for quality improvement in primary care practice. *J Healthc Manag* 2004 May-Jun;49(3):155-68.
- Commonwealth Fund. *The Patient-Centered Medical Home Evaluators' Collaborative*. <http://www.commonwealthfund.org/Content/Newsletters/The-Commonwealth-Fund-Connection/2011/Mar/March-18-2011/Announcements/Patient-Centered-Medical.aspx>.
- Counsell SR, Callahan CM, Buttar AB, et al. Geriatric Resources for Assessment and Care of Elders (GRACE): A new model of primary care for low-income seniors. *J Am Geriatr Soc* 2006;54(7):1136-41.
- Counsell SR, Callahan CM, Clark DO, et al. Geriatric care management for low-income seniors: A randomized control trial. *JAMA* 2007 Dec 12;298(22):2623-33.
- Counsell SR, Callahan CM, Tu W, et al. Cost analysis of the Geriatric Resources for Assessment and Care of Elders care management intervention. *J Am Ger Soc* 2009 Aug;57(8):1420-6.
- Crabtree BF, Chase SM, Wise CG, et al. Evaluation of patient centered medical home practice transformation initiatives. *Med Care*. 2011 Jan;49(1):10-6.
- Crabtree BF, Nutting PA, Miller WL, et al. Summary of the National Demonstration Project and recommendations for the patient-centered medical home. *Ann Fam Med* 2010;8(Suppl 1):S80-90.
- Craig P, Dieppe P, Macintyre S, Michie S, et al. Developing and evaluating complex interventions: new guidance. London, UK: Medical Research Council; 2008. <http://www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC004871>.
- Creswell JW, Klassen AC, Plano Clark VL, Smith KC; for the Office of Behavioral and Social Sciences Research. Best practices for mixed methods research in the health sciences. National Institutes of Health. http://obssr.od.nih.gov/mixed_methods_research. Published August 2011. Accessed December 1, 2011.
- Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci*. 2009;4:50.
- Dentzer S. Reinventing primary care: A task that is far 'too important to fail'. *Health Aff (Millwood)* 2010 May;29(5):757.
- DePalma JA. Evidence to support medical home concept for children with special health care needs. *HHCMP* 2007 Oct;19(6):473-5.
- Domino ME, Humble C, Lawrence WW Jr, et al. Enhancing the medical homes models for children with asthma. *Med Care* 2009 Nov;47(11):1113-20.
- Donaldson MS, Yordy KD, Lohr KN, Vanselow NA, eds; Committee on the Future of Primary Care, Division of Health Care Services, Institute of Medicine. *Primary Care: America's Health in a New Era*. Washington, DC: National Academies Press; 1996.
- Dorr DA, Wilcox AB, Brunner CP, et al. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. *J Am Geriatr Soc* 2008 Dec;56(12):2195-202.
- Duane D, Tomcavage J. ProvenHealth Navigator: A patient-centered primary care model. www.allhealth.org/briefingmaterials/davispresentation-1302.ppt.
- Enthoven AC. Integrated delivery systems: The cure for fragmentation. *Am J Manag Care* 2009 Dec;15(10 Suppl):S284-90.
- Fields D, Leshen E, Patel K. Driving quality gains and cost savings through adoption of medical homes. *Health Aff (Millwood)* 2010 May;29(5):819-26.
- Friedberg MW, Coltin KL, Safran DG, et al. Medical home capabilities of primary care practices that serve sociodemographically vulnerable neighborhoods. *Arch Intern Med* 2010 Jun 14;170(11):938-44.
- Friedberg MW, Lai DJ, Hussey PS, et al. A guide to the medical home as a practice-level intervention. *Am J Manag Care* 2009;15(10 Suppl):S291-9.
- Gilfillan RJ, Tomcavage J, Rosenthal MB, et al. Value and the medical home: Effects of transformed primary care. *Am J Manag Care* 2010 Aug;16(8):607-14.
- Glasgow RE, Orleans CT, Wagner EH. Does the chronic care model serve also as a template for improving prevention? *Milbank Q*. 2001;79(4):579-612.
- Gold M, Helms D, Guterman S. Identifying, monitoring, and assessing promising innovations: using evaluation to support rapid-cycle change, The Commonwealth Fund, June 2011.
- Graff T. ProvenHealth Navigator ICSI update. Geisinger Health System. 2009 May. http://www.icsi.org/icsi_colloquium_-_2009_/proven_health_navigator_.html.

- Group Health News. Medical home pays off with improved care and reduced costs. 2010 May 11. <http://www.ghc.org/news/news.jhtml?repositid=%2fcommon%2fnews%2fnews%2f20100504-medicalHome.html>.
- Grumbach K, Grundy P. Outcomes of implementing patient centered medical home interventions: a review of the evidence from prospective evaluation studies in the United States. Patient Centered Primary Care Collaborative 2010. http://www.pcpc.net/files/evidence_outcomes_in_pcmh.pdf. Accessed February 3, 2011.
- Guided Care Web site. 2010. <http://www.guidedcare.org/index.asp>.
- Harris RP, Helfand M, Woolf SH, et al. Current methods of the U.S. Preventive Services Task Force: A review of the process. *Am J Prev Med* 2001 Apr;20(3 Suppl):21-35. <http://www.uspreventiveservicestaskforce.org/uspstf08/methods/procedure.htm>.
- Hauer KE, Durning SJ, Kernan WN, et al. Factors associated with medical students' career choices regarding internal medicine. *JAMA* 2008 Sep 10;300(10):1154-64.
- Holmboe ES, Arnold GK, Weng W, et al. Current yardsticks may be inadequate for measuring quality improvements from the medical home. *Health Aff (Millwood)* 2010 May;29(5):859-66.
- Homer CJ, Klatka K, Romm D, et al. A review of the evidence for the medical home for children with special health care needs. *Pediatrics* 2008 Oct;122(4):e922-37.
- Hostetter M. Case study: Aetna's embedded case managers seek to strengthen primary care. Commonwealth Fund. August/September 2010. <http://www.commonwealthfund.org/Content/Newsletters/Quality-Matters/2010/August-September-2010/Case-Study.aspx>.
- Houy M. Bailit Health Purchasing. Pennsylvania Chronic Care Initiative. Presentation prepared for Maryland Health Quality and Cost Council. 2008. http://dhmh.maryland.gov/mhqcc/materials/council/121208/PA_Houy_M_121208.pdf.
- Houy M. Pennsylvania Chronic Care Initiative. Appendix. Presentation prepared for Maryland Health Quality and Cost Council. 2008. http://dhmh.maryland.gov/mhqcc/materials/council/121208/PA_appendix_121208.pdf.
- Howell JD. Reflections on the past and future of primary care. *Health Aff (Millwood)*. 2010;29(5):760-765.
- Hughes SL, Weaver FM, Giobbie-Hurder A, et al. Effectiveness of team-managed home-based primary care: A randomized multicenter trial. *JAMA* 2000 Dec 13;284(22):2877-85.
- Hunkeler EM, Katon W, Tang L, et al. Long term outcomes from the IMPACT randomised trial for depressed elderly patients in primary care. *BMJ* 2006 Feb 4;332(7536):259-63.
- IMPACT Implementation Center Web site. <http://impact-uw.org/>.
- Jaén CR, Crabtree BF, Palmer RF, et al. Methods for evaluating practice change toward a patient-centered medical home. *Ann Fam Med* 2010;8(Suppl 1):S9-20.
- 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):S57-67.
- Johns Hopkins Bloomberg School of Public Health. Guided Care: Better care for older people with chronic conditions. <http://www.guidedcare.org/pdf/Guided%20Care%20summary%20and%20results.pdf>.
- Keckley PH, Underwood HR. The Medical Home: Disruptive innovation for a new primary care model. Deloitte Center for Health Solutions, 2008. <http://www.dhcs.ca.gov/provgovpart/Documents/Deloitte%20-%20Financial%20Model%20for%20Medical%20Home.pdf>
- Kilo CM, Wasson JH. Practice redesign and the patient-centered medical home: history, promises, and challenges. *Health Aff (Millwood)*. 2010;29(5):773-778.
- Landon BE, Gill JM, Antonelli RC, et al. Prospects for rebuilding primary care using the patient-centered medical home. *Health Aff (Millwood)* 2010 May;29(5):827-34.
- Lanham HJ, McDaniel RR Jr, Crabtree BF, et al. How improving practice relationships among clinicians and nonclinicians can improve quality in primary care. *Jt Comm J Qual Patient Saf*. 2009;35(9): 457-466.
- Leff B, Reider L, Frick K, et al. Guided Care and the cost of complex health care: A preliminary report. *Am J Manag Care* 2009 Aug;15(8):555-9.
- Levine S, Unützer J, Yip J, et al. Physicians' satisfaction with a collaborative disease management program for late-life depression in primary care. *Gen Hosp Psychiatr*. 2005 Nov-Dec;27(6):383-91.
- Leykum LK, Pugh J, Lawrence V, Parchman M, Noël PH, Cornell J, McDaniel RR Jr. Organizational interventions employing principles of complexity science have improved outcomes for patients with Type II diabetes. *Implement Sci* 2007 Aug 28;2:28.

- Leykum LK, Parchman M, Pugh J, Lawrence V, Noël PH, and McDaniel RR Jr. The importance of organizational characteristics for improving outcomes in patients with chronic disease: a systematic review of congestive heart failure. *Implement Sci* 2010;5:66.
- Litaker D, Tomolo A, Liberatore V, Stange KC, Aron D. Using complexity theory to build interventions that improve health care delivery in primary care. *J Gen Intern Med* 2006 Feb;21 Suppl 2:S30-4.
- Lodh M. Mercer Government Human Services Consulting. ACCESS cost savings—state fiscal year 2004 analysis. Letter to Jeffrey Simms, State of North Carolina, Office of Managed Care, Division of Medical Assistance. 2005.
- McCall N, Cromwell C. Results of the Medicare Health Support disease-management pilot program. *N Engl J Med*. 2011;365(18): 1704-1712.
- Marsteller JA, Hsu YJ, Reider L, et al. Physician satisfaction with chronic care processes: A cluster-randomized trial of guided care. *Ann Fam Med* 2010 Jul-Aug;8(4):308-15.
- Mathematica Policy Research. How effective is home visiting? <http://www.mathematica-mpr.com/EarlyChildhood/homvee.asp>.
- Maxwell S, Zuckerman S, Berenson R. Use of physicians' services under Medicare's resource-based payments. *N Engl J Med* 2007 May 3;356(18):1853-61.
- May CR, Mair FS, Dowrick CF, Finch TL. Process evaluation for complex interventions in primary care: understanding trials using the normalization process model. *BMC Fam Prac* 2007;8:42.
- McCarthy D, Nuzum R, Mika S, et al. The North Dakota experience: Achieving high-performance health care through rural innovation and cooperation. Commonwealth Fund Commission on a High Performance Health System. 2008. <http://www.commonwealthfund.org/Content/Publications/Fund-Reports/2008/May/The-North-Dakota-Experience--Achieving-High-Performance-Health-Care-Through-Rural-Innovation-and-Coo.aspx>.
- Meyers DS, Clancy CM. Primary care: too important to fail. *Ann Intern Med*. 2009;150(4):272-273.
- National Committee for Quality Assurance. Patient-centered medical home survey tool. 2008. <http://www.ncqa.org/tabid/629/Default.aspx>.
- 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):S33-44.
- ORH PL110-329. Funding execution—summary of selected program office projects. <http://www.ruralhealth.va.gov/RURALHEALTH/docs/popjectsummariefy09.pdf>.
- Palfrey JS, Sofis LA, Davidson EJ, et al. The Pediatric Alliance for Coordinated Care: Evaluation of a medical home model. *Pediatrics* 2004 May;113(5 suppl):1507-16.
- Parchman ML, Scoglio CM, Schumm P. Understanding the implementation of evidence-based care: a structural network approach. *Implement Sci* 2011;6:14-23.
- Paulus RA, Davis K, Steele GD. Continuous innovation in health care: Implications of the Geisinger experience. *Health Aff (Millwood)* 2008 Sep-Oct;27(5):1235-45.
- Peikes D, Zutshi A, Genevro JL, Smith K, Parchman ML, Meyers DS. Early Evidence on the Patient-Centered Medical Home. Final paper submitted to the Agency for Healthcare Research and Quality. Princeton, NJ: Mathematica Policy Research; 2012a.
- Peikes D, Zutshi A, Genevro JL, Parchman ML, Meyers DS. Early evaluations of the medical home: building on a promising start. *Am J Manag Care* 2012b;18(2):105-116.
- Peikes D, Peterson G, Brown RS, Graff S, Lynch JP. "How Changes in Washington University's Medicare Coordinated Care Demonstration Pilot Ultimately Achieved Savings." *Health Affairs*, vol. 31, no. 6, June 2012c, pp. 1216-1226.
- Peikes D, Dale S, Lundquist E, et al. Choosing the sample and sample size for medical home evaluations: How to ensure that studies can answer the key research questions. Paper presented at Third National Medical Home Summit, Philadelphia, March 14, 2011. http://pcmh.ahrq.gov/portal/server.pt/gateway/PTARG_S_0_11787_955919_0_0_18/Building%20Evidence%20Base%20PCMH%20White%20Paper.pdf
- Peikes D, Chen A, Schore J, Brown R. Effects of care coordination on hospitalization, quality of care, and health care expenditures among Medicare beneficiaries: 15 randomized trials. *JAMA*. 2009;301(6): 603-618.
- Peikes D, Peterson G, Brown R, Schore J, Razafindrakoto C. Results from a radical makeover of a care coordination program show how program design affects success in reducing hospitalizations and costs: evidence from a randomized controlled trial before and after key changes in program design. Paper presented at: AcademyHealth Annual Research Meeting; 2010 June 27; Boston, MA.
- Pennsylvania Chronic Care Initiative. Project status. <http://www.pcpcc.net/content/pennsylvania-chronic-care-initiative>.
- Plsek PE, Greenhalgh T. Complexity science: the challenge of complexity in health care. *BMJ* 2001;323:625-628.

- Rankin KM, Cooper A, Sanabria K, et al. Illinois Medical Home Project: Pilot intervention and evaluation. *Am J Med Qual* 2009 Jul-Aug;24(4):302-9.
- Reid RJ, Coleman K, Johnson EA, et al. The group health medical home at year two: Cost savings, higher patient satisfaction, and less burnout for providers. *Health Aff (Millwood)* 2010 May;29(5):835-43.
- Reid RJ, Fishman PA, Yu O, et al. Patient-centered medical home demonstration: A prospective, quasi-experimental, before and after evaluation. *Am J Manag Care* 2009 Sep 1;15(9):e71-87.
- Ricketts TC III, Greene S, Silberman P, et al. *Evaluation of Community Care of North Carolina Asthma and Diabetes Initiative: January 2000-December 2002*. Report submitted under University of North Carolina Contract No. 5-35174a with the Foundation for Advanced Health Programs, 2004. <http://www.communitycarenc.com/PDFDocs/Sheps%20Eval.pdf>.
- Rittenhouse DR, Shortell SM, Fisher ES. Primary care and accountable care—two essential elements of delivery-system reform. *N Engl J Med* 2009 Dec 10;361(24):2301-3.
- Robert Graham Center. *The Patient Centered Medical Home History, Seven Core Features, Evidence and Transformational Change, 2007*. http://www.aafp.org/online/etc/medialib/aafp_org/documents/about/pcmh.Par.0001.File.dat/PCMH.pdf.
- Rosenthal MB, Abrams MK, Bitton A, and the Patient-Centered Medical Home Evaluators' Collaborative. Recommended Core Measures for Evaluating the Patient-Centered Medical Home: Cost, Utilization, and Clinical Quality, The Commonwealth Fund, May 2012.
- Rosenthal TC. The medical home: Growing evidence to support a new approach to primary care. *J Am Board Fam Med* 2008 Sept-Oct;21(5):427-40.
- Sanabria K. Illinois Chapter of the American Academy of Pediatrics. State Implementation Grant (D70) for Improving Systems of Services for Children and Youth with Special Health Care Needs (CYSHCN) abstract. http://www.medicalhomeinfo.org/state_pages/illinois.aspx#projects.
- Schoen C, Osborn R, Doty MM, et al. Toward higher-performance health systems: Adults' health care experiences in seven countries, 2007. *Health Aff (Millwood)* 2007 Nov-Dec;26(6):w717-34.
- Silvia TJ, Sofis LA, Palfrey JS. *Practicing Comprehensive Care: A Physician's Manual for Implementing a Medical Home for Children With Special Health Care Needs*. Boston: Institute for Community Inclusion, Children's Hospital; 2000. http://www.communityinclusion.org/article.php?article_id=193&type=topic&id=2.
- Solberg LI, Glasgow RE, Unützer J, et al. Partnership research: A practical trial design for evaluation of a natural experiment to improve depression care. *Med Care* 2010 July;48(7):576-82.
- Solberg LI, Asche SE, Pawlson LG, et al. Practice systems are associated with high-quality care for diabetes. *Am J Manag Care* 2008 Feb;14(2):85-92.
- Starfield B. *Is the Patient-Centered Medical Home the Same as "Primary Care"?* *Measurement Issues*. <http://ncvhs.hhs.gov/080520p01.pdf>. Published May 20, 2008.
- Starfield B. *Primary Care: Concept, Evaluation, and Policy*. New York: Oxford University Press; 1992.
- Steele GD, Haynes JA, Davis DE, et al. How Geisinger's advanced medical home model argues the case for rapid-cycle innovation. *Health Aff (Millwood)* 2010 Nov;29(11):2047-53.
- Steiner BD, Denham AC, Ashkin E, et al. Community Care of North Carolina: Improving care through community health networks. *Ann Fam Med* 2008 Jul-Aug;6(4):361-7.
- Stetler CB, Damschroder LJ, Helfrich CD, Hagedorn HJ. A guide for applying a revised version of the PARIHS framework for implementation. *Implement Sci* 2011;6:99.
- Stewart EE, Nutting PA, Crabtree BF, et al. Implementing the patient-centered medical home: Observation and description of the national demonstration project. *Ann Fam Med* 2010;8(Suppl 1):S21-32.
- Stremikis K, Schoen C, Fryer A-K. A call for change: The 2011 Commonwealth Fund survey of public views of the U.S. health system. New York: Commonwealth Fund; April 2011.
- Torregrossa AS, Governor's Office of Health Care Reform, Commonwealth of Pennsylvania. Pennsylvania's efforts to transform primary care. http://www.cthealthpolicy.org/webinars/20100223_ato rregrossa_webinar.pdf.
- Unützer J, Katon WJ, Callahan CM, et al. Collaborative care management of late-life depression in the primary care setting: A randomized controlled trial. *JAMA* 2002 Dec 11;288(22):2836-45.
- Unützer J, Katon WJ, Fan MY, et al. Long-term cost effects of collaborative care for late-life depression. *Am J Manag Care* 2008 Feb;14(2):95-100.
- Unützer J, Katon WJ, Williams JW Jr, et al. Improving primary care for depression in late life: The design of a multi-center randomized trial. *Med Care* 2001 Aug;39(8):785-99.
- U.S. Congress. Patient Protection and Affordable Care Act. H.R. 3590. Washington, DC: Government Printing Office, 2010.

- U.S. Department of Education, Institute of Education Sciences. What Works Clearinghouse. <http://ies.ed.gov/ncee/wwc/>.
- U.S. Department of Health and Human Services. *Report to Congress: National Strategy for Quality Improvement in Health Care*. March 2011. <http://www.healthcare.gov/law/resources/reports/quality03212011a.html#es>.
- U.S. Department of Health and Human Services. *Overview of the Teen Pregnancy Prevention Research Evidence Review*. <http://www.hhs.gov/ash/oah/prevention/research/index.html>.
- U.S. Department of Veterans Affairs. Home-based primary care program. VHA Handbook 1141.01. Washington, DC, January 31, 2007. http://www1.va.gov/vhapublications/ViewPublication.asp?pub_ID=1534.
- U.S. Government Accountability Office. *Value in Health. Key Information for Policymakers to Assess Efforts to Improve Quality While Reducing Costs*. Report to Congressional Requesters. <http://www.gao.gov/new.items/d11445.pdf>. Published July 2011. Accessed July 26, 2011.
- Wilhide S, Henderson T. Community Care of North Carolina: A provider-led strategy for delivering cost-effective primary care to Medicaid beneficiaries. American Academy of Family Physicians, 2006. Available at http://www.aafp.org/online/etc/medialib/aafp_org/documents/policy/state/medicaid/ncfull.Par.0001.File.tmp/ncfullreport.pdf. Accessed January 31, 2011.
- Wolff JL, Rand-Giovanetti E, Palmer S, et al. Caregiving and chronic care: The Guided Care program for family and friends. *J Gerontol A Biol Med Sci* 2009 Jul;64(7):785-91.
- Wolff JL, Rand-Giovanetti E, Boyd CM, et al. Effects of Guided Care on family caregivers; *Gerontologist* 2010 Aug;50(4):459-70.

Appendix A

AHRQ'S Definition of the Patient-Centered Medical Home

The patient-centered medical home (PCMH, or *medical home*) model holds promise as a way to improve health care in America by transforming how primary care is organized and delivered. Building on the work of a large and growing community, the Agency for Healthcare Research and Quality (AHRQ) defines a PCMH not simply as a place but as a model of the organization of primary health care that delivers the core functions of primary health care (AHRQ, 2012).

The PCMH encompasses five functions and attributes:

1. **Patient-centered.** The PCMH provides relationship-based primary health care that is oriented toward the “whole person.” Partnering with patients and their families requires understanding and respecting each patient’s unique needs, culture, values, and preferences. The PCMH practice actively supports patients in learning to manage and organize their own care at the level the patient chooses. Recognizing that patients and families are core members of the care team, PCMH practices ensure that they are fully informed partners in establishing care plans.
2. **Comprehensive care.** The PCMH is accountable for meeting the bulk of each patient’s physical and mental health care needs, including prevention and wellness, acute care, and chronic care. Comprehensive care requires a team of care providers, possibly including physicians, advanced practice nurses, physician assistants, nurses, pharmacists, nutritionists, social workers, educators, and care coordinators. Although some PCMH practices may bring together large and diverse teams of care providers to meet the needs of their patients, many others, including smaller practices, will build virtual teams linking themselves and their patients to providers and services in their communities.
3. **Coordinated care.** The PCMH coordinates care across all elements of the broader health care system, including specialty care, hospitals, home health care, and community services and supports. Such coordination is particularly critical during transitions between sites of care, such as when patients are being discharged from the hospital. PCMH practices also excel at building clear and open communication among patients and families, the medical home, and members of the broader care team.
4. **Superb access to care.** The PCMH delivers accessible services with shorter waiting times for urgent needs, enhanced in-person hours, around-the-clock telephone or electronic access to a member of the care team, and alternative methods of communication, such as email and telephone care. The medical home practice is responsive to patients’ preferences regarding access.
5. **A systems-based approach to quality and safety.** The PCMH demonstrates a commitment to quality and quality improvement by ongoing engagement in activities

such as using evidence-based medicine and clinical decision-support tools to guide shared decisionmaking with patients and families, engaging in performance measurement and improvement, measuring and responding to patient experiences and patient satisfaction, and practicing population health management. Publicly sharing robust quality and safety data and improvement activities is also an important marker of a system-level commitment to quality.

AHRQ recognizes the central role of health IT in successfully operationalizing and implementing the key features of the medical home. In addition, AHRQ notes that building a primary care delivery platform that the Nation can rely on for accessible, affordable, high-quality health care will require significant workforce development and fundamental payment reform. Without these critical elements, the potential of primary care will not be achieved.

Appendix B

Methods for Reviewing the Evidence on the Patient-Centered Medical Home

This appendix describes the methods used for reviewing the evidence on the PCMH, beginning with selecting evaluations for inclusion in the review and developing and applying a formal rating system to identify rigorously evaluated interventions.

Evaluation Selection

The review team conducted a broad search to identify English-language studies in the published and grey literature on the PCMH in the United States. To capture published studies, we searched several databases using the Ovid and EBSCO search engines for articles from January 2000 to September 2010 containing the words “medical home” or “primary care transformation.” Using the Ovid search engine, we searched the following databases: Journals@Ovid, HealthSTAR, Ovid MEDLINE, and PsycINFO. We used the EBSCO search engine on the following databases: Academic Search Premier, Business Source Corporate, Cumulative Index to Nursing and Allied Health Literature, Cochrane Center Register of Controlled Trials, Cochrane Systematic Database of Reviews, Cochrane Methodology Register, Database of Abstracts of Reviews of Effects, E-Journals, Econ Lit, Health Technology Assessments, National Health Service, Economic Evaluation Database, Health Policy Reference Center Database, and Health Policy Reference Center.

We also conducted targeted searches to identify evaluations of initiatives for which no start dates were specified, but that are widely cited as part of the evidence base for the medical home. We identified additional evaluations by reviewing content on 100 relevant Web sites, examining bibliographies in existing review articles, and gathering input from experts in the field. This search process yielded 498 potentially relevant citations. As with all evidence reviews, owing to publication bias, the evaluations selected and synthesized here may be more likely to include favorable effects (versus those that show no effects or unfavorable effects) than those excluded by our search and synthesis criteria.²⁸

Of the 498 citations, we selected evaluations that met the following two criteria:

1. The evaluation tested a primary-care, practice-based intervention with three or more of the five medical home components defined by AHRQ (delivering care that is patient-centered, comprehensive and team-based, coordinated, accessible, and systems-based in its approach to quality and safety) described in detail in Appendix A. We excluded evaluations of care coordination and disease management interventions that met these criteria but were not provided from within, or in close partnership with, the practice (for example, interventions delivered by off-site care managers via telephone).

²⁸ See <http://www.cochrane-net.org/openlearning/html/mod15-2.htm> for more details on publication bias.

2. The evaluation used quantitative methods to examine effects on either (a) a triple aim outcome: quality of care, costs²⁹ (or hospital use or emergency department use, two major cost drivers), and patient or caregiver experience; or (b) professional experience.

Because most interventions targeted different subgroups of the U.S. primary care population, our inclusion criteria did not consider the population served. We also did not require that the intervention use health IT or provide enhanced payment.

Using these criteria, the review team identified 14 evaluations of 12 distinct interventions (one intervention, CCNC, was evaluated by three distinct evaluations) for inclusion in the review.³⁰ Although most of these interventions can be viewed as precursors to the medical home, they share multiple components of the medical home and are frequently cited as part of the evidence base for it.

Methods to Assess the Rigor of the Evaluations

In this section, we first provide an overview of the rating system and then describe, in detail, the individual criteria that factor into this system.

Rating System

To assess the rigor of the 14 evaluations selected for review, we developed a systematic approach by drawing broadly from the USPSTF review methods and supplementing them with specific criteria used by well-regarded evidence reviews from the fields of education and home visiting programs for families with pregnant women and children.³¹

Rather than give a global rating to each evaluation, we individually rated the internal validity of each analysis conducted as part of an evaluation as *high*, *moderate*, *low*, or *excluded*. We rated individual analyses because evaluations often used different designs, samples, and methods (and sometimes different subgroups of patients) to analyze different outcomes over varying followup periods. Therefore, to allow for the possibility that the evaluation of a single intervention could provide more rigorous evidence on some outcomes than others, we separately assessed the analysis of each outcome measure at each followup period and, if applicable, for each subgroup of patients. We consider analyses rated high or moderate as providing rigorous evidence and include such analyses only in our synthesis of the evidence.

Our rating of each analysis is based solely on an assessment of its internal validity. We do not factor generalizability (external validity) into the rating because most interventions included in this review targeted a specific subset of primary care patients, were implemented in unique settings, and either purposively selected practices or relied on them to volunteer; therefore, nearly all these evaluations have limited generalizability. We summarize the characteristics of patients and practice settings used in the rigorous evaluations to alert decisionmakers to the

²⁹ None of the studies reported effects on out-of-pocket patient costs or practice revenues.

³⁰ In general, we found that, for most of the interventions, different analyses from the same study design were published in multiple articles.

³¹ For the USPSTF guideline, see Harris et al., 2001. For education guidelines, see <http://ies.ed.gov/ncee/wwc/>. For the home visiting guidelines, see <http://www.mathematica-mpr.com/EarlyChildhood/homvee.asp>.

possibility that findings may differ if interventions are implemented in other populations and settings.

We rated each analysis using a sequence of criteria, starting with the most general (evaluation design) and ending with the most specific (such as whether the analysis controlled for outcome values prior to the start of the intervention, in other words, at baseline). As a first step in assessment, we considered only analyses conducted as part of randomized, controlled trials (RCTs) and nonexperimental comparison group evaluations for a high or moderate rating, based on the strength of the methods the evaluations used to produce unbiased estimates of the effects of the interventions. If they failed to meet criteria for either a high or a moderate rating, they received a low rating. Analyses from evaluations that did not include a control or comparison group³² (for example, pre-post or cross-sectional evaluations) always received a low rating. This is because such designs make it difficult to assess what the outcomes would have been absent the intervention (the purpose of a control/comparison group is to establish this counterfactual). Analyses were rated *excluded* if the evaluation design or methods were not described in sufficient detail to permit assessment of their internal validity. In many cases, because of the limits on what study authors can include in a journal article, we sought additional details from authors to be able to rate analyses.

We note that the rating of the internal validity of the evidence does not take into account whether an evaluation has sufficient power to detect policy-relevant effects, or whether tests of statistical significance in clustered designs (that is, ones that intervene with entire practices or sets of providers) account appropriately for clustering. However, because these are important considerations for the interpretation of findings, we do consider them when we synthesize the findings, as described in Chapter 2.

Below, we define the ratings and the criteria they are based on.

High rating. A high rating reflects high confidence that the analysis accurately estimated the effect of the intervention (where the effect might be favorable, unfavorable, or zero). A high rating was reserved for analyses from RCTs with no systematic confounding factors, no endogenous subgroups, high maintenance of the intervention and control groups at followup through low attrition rates, and use of regression analysis to control for reported statistically significant baseline differences between the intervention and control groups in the outcome. (These terms are defined in more detail below.)

Moderate rating. A moderate rating reflects moderate confidence that the analysis accurately estimated the effect of the intervention. Future research based on more rigorous evaluation designs or methods might alter the estimated effect. A moderate rating was assigned to analyses from RCTs that fulfilled all criteria for a high rating but failed to control for reported statistically significant baseline intervention-control differences in the outcome. A moderate rating was also given to analyses from comparison group designs that (1) had no systematic confounding factors, (2) were based on intervention and comparison groups with equivalent outcomes at baseline, and (3) used regression analysis to control for baseline values of the

³² The term *control group* is used exclusively when the group was assigned using an RCT. The term *comparison group* indicates that the group was selected using nonexperimental comparison group methods.

outcome.³³ Analyses from RCTs that suffered from high attrition or were based on endogenous subgroups were treated similarly to those from comparison group evaluations, and had to meet the same criteria as a comparison group evaluation.

Low rating. A low rating reflects low confidence that the analysis accurately estimated the effect of the intervention. Future research based on more rigorous evaluation designs or methods is likely to alter the estimated effect. A low rating was given to analyses from RCTs and comparison group designs that suffered from systematic confounding. It was also given to analyses from comparison group designs and from RCTs with high attrition or endogenous subgroups under two conditions: (1) the intervention and control/comparison group analysis samples did not have equivalent baseline values of the outcome; or (2) if that condition was met, the analysis did not control for baseline values of the outcome. Finally, analyses from pre-post and cross-sectional evaluations always received a low rating.

Excluded rating. Some evaluations provided insufficient information to establish whether the estimates accurately reflect the effect of the intervention. Analyses were rated excluded if the design or methods were not described in sufficient detail to enable assessment. In this case, we cannot know with certainty whether the reported effects are a result of the intervention.

³³ Comparison group evaluations and RCTs with high attrition, or endogenous subgroups that show baseline equivalence on the outcome being examined, are also required to control for the baseline values of the outcome in their analyses because this ensures that any small differences at baseline do not bias the impact estimates.

Table 16. Definition of ratings

High	<ul style="list-style-type: none">• RCTs (including cluster-RCTs) with no systematic confounding factors, no endogenous subgroups, no sample reassignment from the control to the intervention group or vice versa,³⁴ and low attrition at followup. To receive a high rating, RCTs that meet all these criteria also need to control for any reported baseline difference between the intervention and control groups on the outcome.
Moderate	<ul style="list-style-type: none">• Comparison group evaluations (including case control and cohort studies with comparison groups) with no systematic confounding factors, analysis showing the intervention and comparison groups have equivalent outcomes at baseline, and controls for baseline values of the outcome.• RCTs with:<ul style="list-style-type: none">– No systematic confounding factors, no endogenous subgroups, no sample reassignment, and low attrition at the unit of analysis, but that fail to control for reported statistically significant baseline differences in the outcome between the intervention and control groups.– No systematic confounding factors, but with (1) endogenous subgroups, (2) high attrition of the analysis sample at followup, or (3) sample reassignment. These are reviewed as comparison group evaluations and receive a moderate rating if they meet applicable criteria for a comparison group evaluation.
Low	<ul style="list-style-type: none">• RCTs that did not meet the criteria for a moderate or high rating and comparison group evaluations that did not meet the criteria for a moderate rating, as well as pre-post and cross-sectional evaluations.
Excluded	<ul style="list-style-type: none">• Analyses from evaluations for which the design and/or methods were not described in sufficient detail to enable assessment.

Description of the Individual Criteria

Here, we describe in detail the key criteria that factor into the rating system.

Evaluation design. The highest rating is reserved for analyses from evaluations that randomly assigned subjects to the evaluation’s research groups. Evaluations using random assignment can—if well implemented and analyzed—provide the strongest evidence that differences in the outcomes between the intervention and control groups can be attributed to the intervention.

Comparison group evaluations can achieve a moderate rating at best. In such evaluations, subjects are sorted into intervention and comparison groups in a nonrandom way; therefore, even if the groups have comparable observed characteristics before the intervention, they still may differ on unmeasured characteristics. We therefore cannot rule out the possibility that the findings are attributable to unmeasured differences between the intervention and comparison groups. Certain RCTs (as described in Table 16) are treated similar to comparison group evaluations and, at best, considered for a moderate rating.

³⁴ In RCTs, deviation from the original random assignment (“sample reassignment”) can bias the study’s impact estimates. This can occur if patients in control practices obtained care from intervention practices, or vice versa. Therefore, for an RCT to receive a high rating, the analysis (in addition to meeting other criteria for a high rating) should be performed on the sample as originally assigned. RCTs that somehow alter the original random assignment must establish baseline equivalence of the intervention and control group members in the analysis sample to be considered for a moderate rating. None of the studies we reviewed reported sample reassignment.

For a stepwise illustration of the rating process for RCTs and comparison group evaluations, see figures 3 and 4, respectively.

Attrition among RCTs. We assess attrition in RCTs but not in comparison group evaluations. Comparison group evaluations examine outcomes based on the final analysis samples, from which there is, by definition, no attrition.

In RCTs, loss of data on some evaluation participants can bias the evaluation’s impact estimates by creating, over time, differences in the characteristics of the intervention and control groups that had originally been comparable because of randomization. Bias can arise from *overall* attrition (the percentage of evaluation participants lost among the total evaluation sample) and *differential* attrition (the difference in attrition rates between the intervention and control groups).

We use “liberal standards” employed by the What Works Clearinghouse to assess the level of attrition for each outcome examined in a given evaluation. To determine whether attrition may be a source of bias in the impact estimates, this assessment takes into account both overall attrition and differential attrition between the intervention and control groups. Figure 5 shows the cutoffs for combinations of overall and differential attrition used to determine “low” or “high” attrition. Evaluations with a relatively high level of differential attrition can still meet standards for the “low” attrition category if they have a relatively low level of overall attrition, whereas evaluations with a relatively high level of overall attrition require a lower level of differential attrition to meet standards. For example, as Figure 5 indicates, if the rate of attrition is the same for the intervention and control groups (that is, there is zero differential attrition), the evaluation can fall in the low attrition category even with 60 percent overall attrition. However, even a small amount of differential attrition (say 10 percent) requires the overall attrition rate to be very low (in this case, less than 13 percent) to meet the standards for low attrition.³⁵

We consider attrition due to mortality as inconsequential for analysis of claims-based outcomes, such as service use and costs, and do not apply it as a criterion for assessment of the rating because we know with certainty that there are no service use and costs for people who died. For survey-based outcomes, however, we treat mortality as any other type of attrition and factor it into the rating process.

Only RCTs meeting the standard for acceptable combinations of overall and differential attrition are considered for the high rating. RCTs that do not meet these standards are considered for the moderate rating.

Baseline equivalence of the intervention and control or comparison groups. To obtain a moderate rating, RCTs with high attrition or endogenous groups and comparison group evaluations must (1) demonstrate baseline equivalence of the two research groups, and (2) control for baseline values of the outcome when estimating the effect of the intervention. We use the first criterion because the use of comparable intervention and control/comparison groups minimizes the bias in the estimated effect. We examine statistical tests of the difference in means

³⁵ More information on the attrition calculations used can be found in the *What Works Clearinghouse Procedures and Standards Handbook* at: http://ies.ed.gov/ncee/wwc/pdf/wwc_procedures_v2_standards_handbook.pdf. Future reviews of primary care interventions could consider whether there is a need to tailor these attrition standards to the primary care setting.

to show baseline equivalence.³⁶ Evaluations must establish baseline equivalence using the analysis sample at followup (as opposed to the sample at the start of the intervention). The second criterion ensures that any differences at baseline do not bias the estimated effects at followup. For example, if a comparison group evaluation examines effects on two outcomes—costs and hospitalizations—but finds baseline equivalence only on costs and not on hospitalizations, only costs will be considered for a moderate rating, while hospitalizations will receive a low rating. To actually receive a moderate rating on costs, the analysis would also need to control for baseline costs. Finally, if the outcomes of the intervention and control/comparison groups are not equivalent at baseline, then the analysis will receive a low rating even if it controlled for baseline values of the outcome. This is because controlling for baseline values of the outcome will not account for the potential differences in unobserved characteristics between the intervention and control/comparison groups that can bias the estimated effect.³⁷

Systematic confounding. A systematic confounding factor is a component of the research design or methods that undermines the credibility of attributing an observed effect to the intervention. One example of a systematic confounder is the use of one practice in the intervention or control/comparison group. Using a single practice precludes factoring in the variation in outcomes that occurs at the practice level (in addition to the variation that occurs across patients within a practice) in estimating the overall variance in the outcome and conducting tests of statistical significance to determine whether the observed intervention-control difference is due to the intervention or to chance. Another example of a confounding factor is systematic differences in data collection methods for the intervention and control/comparison groups. Because the presence of such confounding factors severely weakens the credibility of an analysis' findings, a low rating is assigned to analyses from RCTs or comparison group evaluations with such factors.

Endogenous subgroups. A subgroup is considered endogenously formed and estimated effects for this subgroup considered biased if the subgroup is based on the followup (or postrandomization) value of an outcome that could be affected by the intervention. The extent of this bias may be small if the intervention and control arms of the subgroup are comparable at baseline, or if the intervention had no effect on the outcome that defines the subgroup. For example, in an intervention aimed at improving depression care, examining satisfaction with depression care among people who reported receiving such care during the intervention constitutes, by definition, analysis of an endogenous subgroup, because the intervention may affect receipt of depression care. Analyses based on endogenous subgroups in an RCT are treated similarly to those from a comparison group evaluation and must meet criteria applicable to a comparison group evaluation to receive a moderate rating.

³⁶ This is a liberal criterion for the evaluations in this review that have small sample sizes and are likely underpowered. Such studies are more likely to find differences at baseline not to be statistically significant. Alternatively, with large samples, it is possible that even a very small difference appears as statistically significant. To address this possibility, future reviews could establish a threshold (such as 0.25 standard deviations from the pooled mean) below which even statistically significant differences would be considered as meeting the baseline equivalence criterion.

³⁷ RCTs that otherwise meet the criteria for the highest rating are not required to establish baseline equivalence, because randomization is expected to produce intervention and control groups that are equivalent, on average, on both observed and unobserved characteristics. Nevertheless, chance differences between the two groups can arise despite randomization, especially with small samples. As a result, to meet the criteria for the highest study rating, RCTs that showed evidence of statistically significant baseline differences on outcome measures are required to control for these differences in their statistical impact analyses. RCTs that do not control for statistically significant baseline differences in the outcome measure are assigned the moderate rating.

Figure 3. Rating Criteria for Randomized Controlled Trials

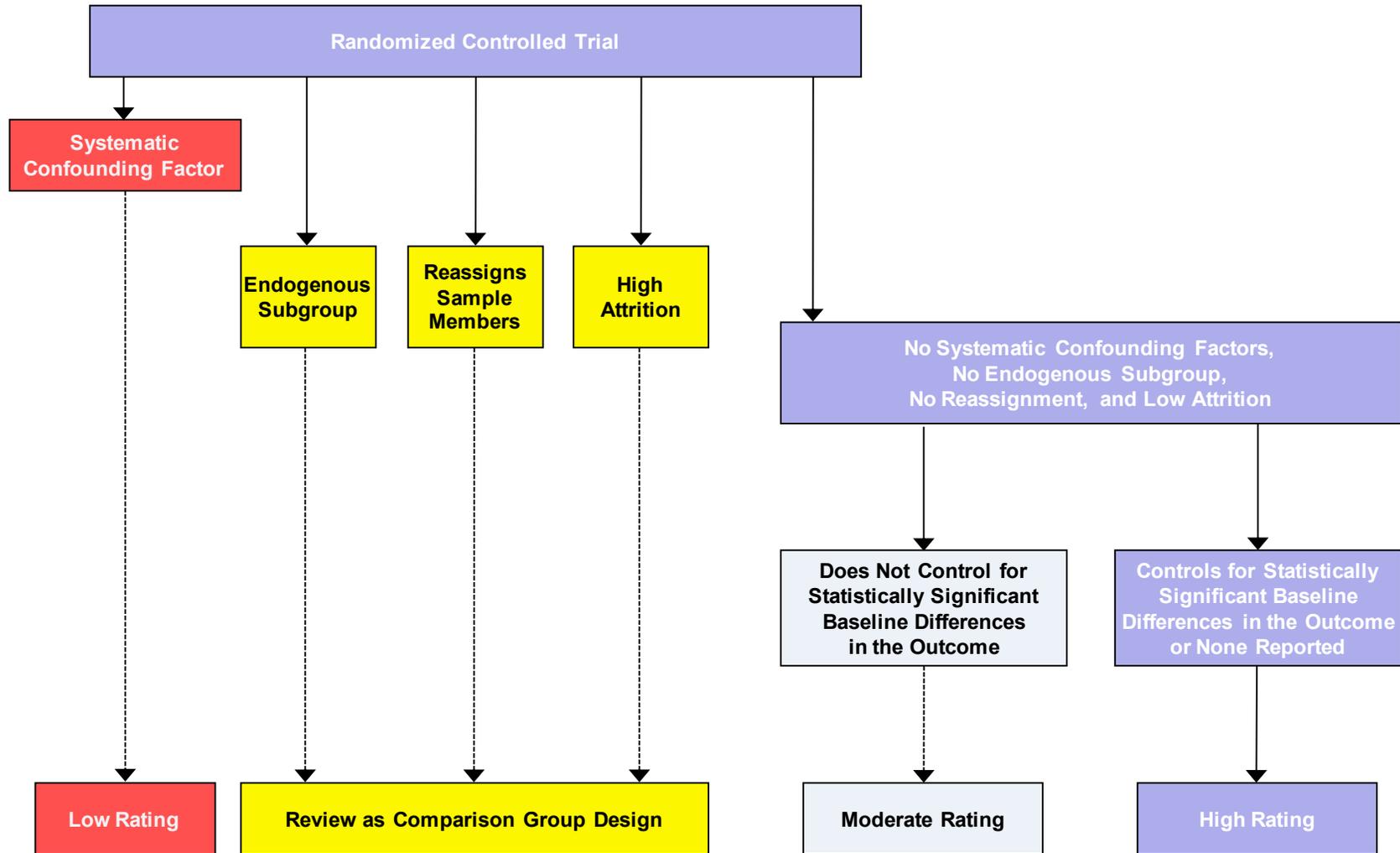


Figure 4. Rating Criteria for Comparison Group Evaluations

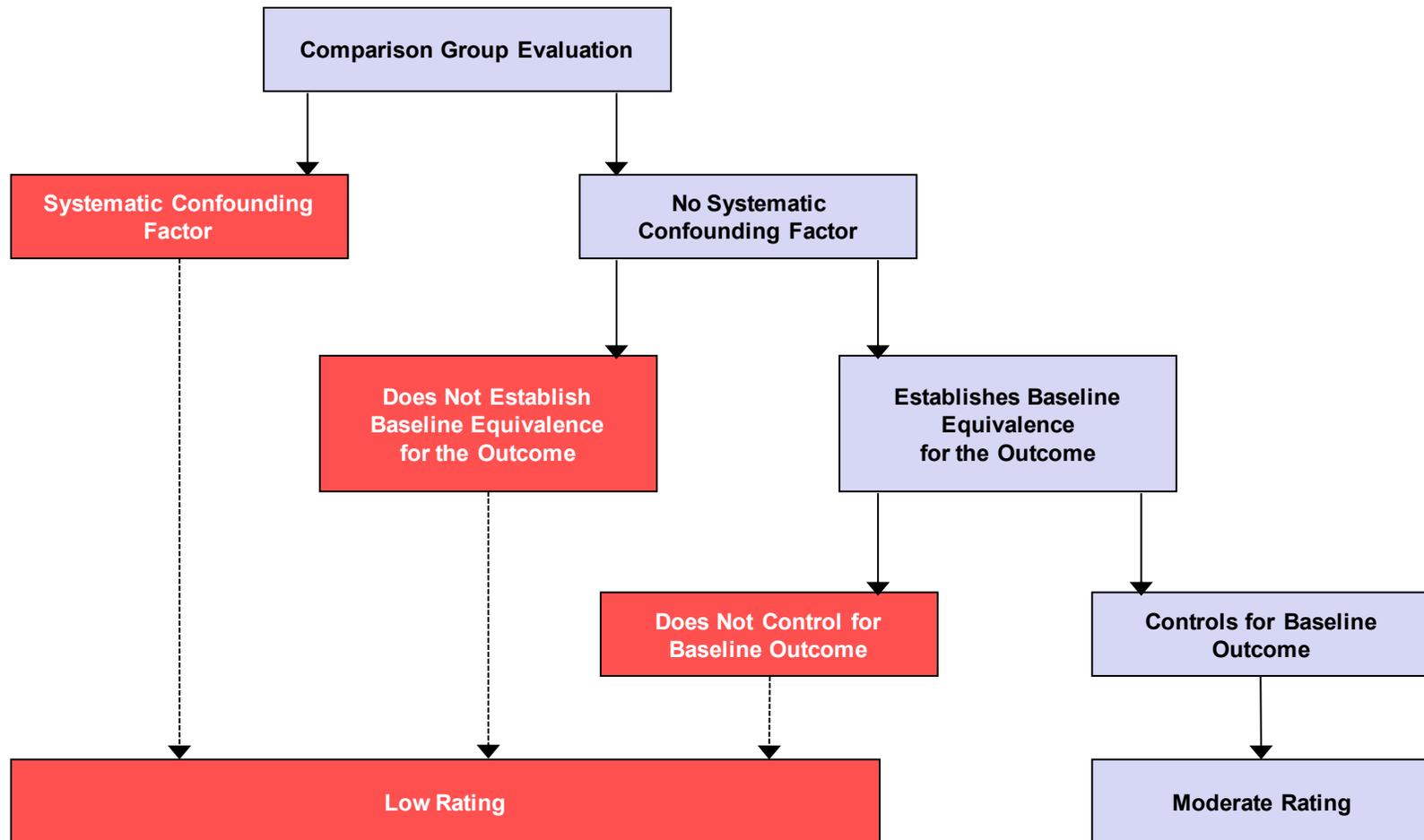
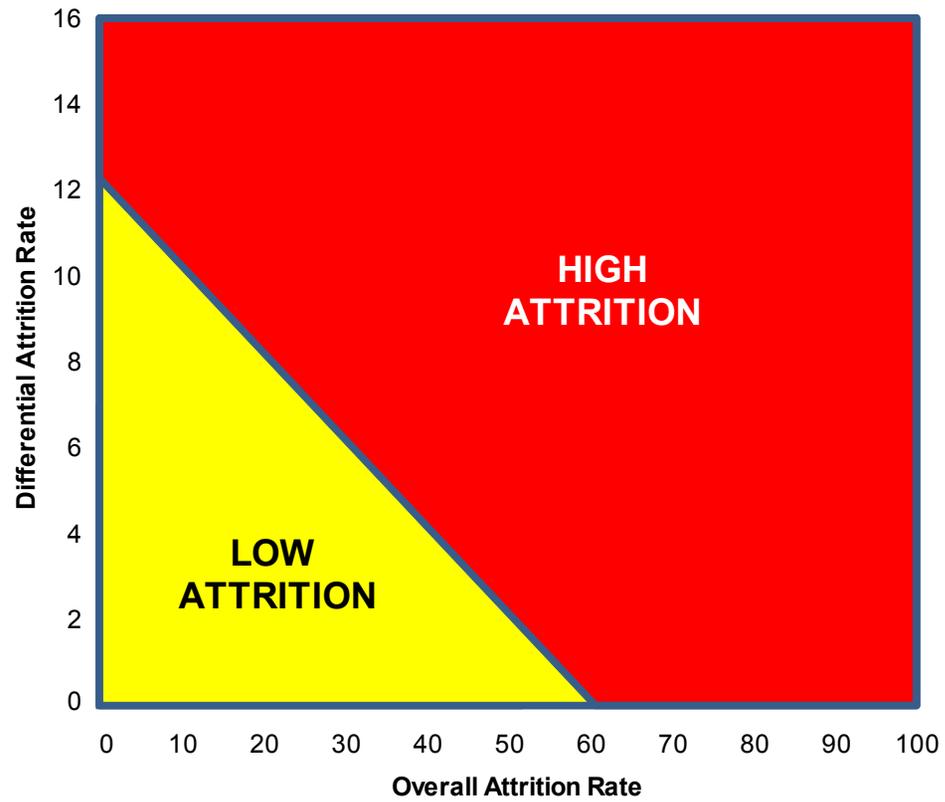


Figure 5. What Works Clearinghouse Liberal Attrition Standards



Appendix C

Supplemental Table on Descriptions of the Interventions, by AHRQ PCMH Principles and Facilitators

Table 17. Descriptions of the interventions, by AHRQ PCMH principles and facilitators

AHRQ PCMH Principles and Facilitators								
Intervention	Overview	Patient-Centered	Comprehensive Care	Coordinated Care	Access to Care	Systems Approach to Quality and Safety	Payment and Other Resources to the Primary Care Practice	Health IT
Aetna's Embedded Case Managers	Program assigns nurse case managers to primary care practices to help manage care for Medicare Advantage members and collaborate with the clinical team	Care plans; disease management coaching; family members can sit in on patient office visits	Team-based care, including the nurse case manager and clinical team, who address needs of patients with multiple chronic conditions, including dementia and depression, and provide end-of-life care	Case manager coordinates care, including hospital discharge plan, and links patients to social services	No changes in access to care	Case manager uses clinical decision support software to identify gaps in treatment; reviews data weekly with the clinical team and monthly with the medical director	Program provides nurse case managers; practice receives an extra fee for patients enrolled in program and incentives for meeting quality targets	Clinical decision support software
Care Management Plus	Nurse care managers, supported by specialized health IT tools within primary care clinics, orchestrate care for chronically ill elderly patients	Develop care plan with patients and family; teach self-management to patients	Team-based approach to patient assessment and care planning	Care manager coordinates care across providers	Patient-specific secure messaging system facilitates communication	Care management tracking (CMT) database embeds disease protocols and generates flexible, patient-specific care plans, as well as aggregate statistics	No payment component. Program provides care manager and specialized IT tools	Existing electronic health records (EHRs) and CMT to track all contacts with patients, families and providers; generate reminders, calculate patient statistics; and provide electronic protocols

Table 17 (continued)

AHRQ PCMH Principles and Facilitators								
Intervention	Overview	Patient-Centered	Comprehensive Care	Coordinated Care	Access to Care	Systems Approach to Quality and Safety	Payment and Other Resources to the Primary Care Practice	Health IT
Community Care of North Carolina	Community-based care management provided through networks of primary care providers (PCPs), a hospital, the Department of Social Services, and the health department. Case managers from a nonprofit work with PCPs to coordinate care and undertake population health management	Providers and/or case managers (a nurse, social worker, or other clinician) coach and educate patients on disease management and assess psychosocial needs	Practice team includes primary care provider and case managers who provide comprehensive case management	Case manager coordinates with providers, hospitals, health departments, and social service agencies that are part of network; web-based program used to coordinate care	24/7 on-call assistance; case managers make home visits	Random chart reviews to assess adherence to case management protocols; review of claims data and charts to assess clinical improvements	PCPs receive \$2.50 per member per month (PMPM) for medical home and population management activities and the help of the case manager; networks receive \$3 PMPM (\$5 PMPM for elderly or disabled patients)	No standardized health IT component; some participating physicians may be using EHRs
Geisinger Health System ProvenHealth Navigator	Geisinger Health Plan (GHP) provided one nurse case manager for every 900 Medicare Advantage patients in each primary care practice to identify high-risk patients, design patient-centered care plans, provide care coordination and care transition support, and monitor patients using patient-accessible EHRs	Case manager develops individualized care plans; provides self-management education to patient and family; assesses patient satisfaction.	Care teams composed of PCP, physician's assistant, nurse practitioners, nurses, administrative staff, and case manager address patient's care needs, including medication management and end-of-life planning	Case manager coordinates care across providers, including during care transitions, and conducts outreach to home health agencies and nursing homes	24/7 access, same-day appointments, self-scheduling using EHR, direct telephone lines to case managers, home interactive voice response for high-risk or postdischarge patients	EHRs provide preventive and chronic care reminders and embedded care workflows; program tracks 10 quality-of-care metrics, including chronic and preventive care, postdischarge followup, and patient satisfaction and experience; monthly meetings with primary care practices, navigators, and GHP staff to review results	Program provided case manager and funding for new services, physician and practice transformation stipends, and staff incentives, including employee stipends and quarterly performance-based payments; program also used a shared savings incentive model based on quality and efficiency performance	Existing EHR embeds care workflows, captures patient information, tracks patient care, generates reminders, and calculates patient statistics; EHR is patient-accessible via a Web-based interface; Bluetooth scales for daily monitoring of heart failure patients

Table 17 (continued)

AHRQ PCMH Principles and Facilitators								
Intervention	Overview	Patient-Centered	Comprehensive Care	Coordinated Care	Access to Care	Systems Approach to Quality and Safety	Payment and Other Resources to the Primary Care Practice	Health IT
Geriatric Resources for Assessment and Care of Elders	Advanced practice nurse and social worker (GRACE support team) assess low-income seniors in home, and develop and implement a care plan with a geriatrics interdisciplinary team, in collaboration with the patient's primary care provider	Initial and annual in-home comprehensive geriatric assessment; annual individualized care plan; minimum 1 in-home visit to review care plan and 1 face-to-face or telephone contact per month with patients and family members or caregivers	Care plan developed and implemented in collaboration with the GRACE interdisciplinary team of a pharmacist, physical therapist, community resource expert, and mental health case manager, led by a geriatrician and the patient's PCP. The care plan covers physical, mental, and social needs	The nurse practitioner-social worker team coordinates with the inpatient and nursing home teams for patients who have been hospitalized or using skilled nursing facility services; the team conducts a home visit and full review of the case after hospital and ED visits. They also coordinate specialty visits	Dedicated telephone line to GRACE support team	Care protocols for evaluation and management of 12 common geriatric conditions	No payment component. Program provides assistance of GRACE support team to primary care practice.	Integrated EHRs and web-based tracking tool support care management and coordination of care.
Group Health Cooperative Medical Home	Group Health redesigned one pilot clinic to be a PCMH by changing staffing, scheduling, point of care, patient outreach, health IT, and management; reducing caseloads; increasing visit times; using team huddles; and rapid process improvements	Individualized care plans viewable through patient EHRs	Care team composed of PCP, nurse care manager, pharmacist, medical assistant, and a Licensed Practical Nurse deliver primary care to patients, which includes pre-visit contact to discuss concerns	Nurse works with PCP to coordinate care across providers, including during transitions between care sites	24-hour telephone access to consulting nurse, same-day appointments, online services, self-scheduling using EHRs' direct telephone lines to case managers	EHR provides preventive and chronic care reminders and embeds care workflows	Physicians paid a salary and shared savings based on quality targets achieved; program provided additional staff	Existing EHR records patient information and care, generates reminders; its messaging feature is used for real-time specialist consultations. Patients can access EHRs

Table 17 (continued)

AHRQ PCMH Principles and Facilitators

Intervention	Overview	Patient-Centered	Comprehensive Care	Coordinated Care	Access to Care	Systems Approach to Quality and Safety	Payment and Other Resources to the Primary Care Practice	Health IT
Guided Care	GC nurse (GCN) joins primary care practice, provides assessments, care plans, monthly monitoring, and transitional care to highest-risk Medicare patients	Home-based assessment; individualized care plan and a patient self-care plan to promote self-management; group classes for caregivers	GCN and PCP discuss and modify individualized care guide. GCN proactively manages patients, mostly by telephone	GCN coordinates care and provides care plan to other providers; facilitates care transitions; monitors patients during hospital stays; and facilitates access to community services	Telephone access to GCN	Evidence-based guidelines, embedded in Guided Care EHRs, used to generate individualized care guides and monthly reports on GCN performance. GCN, study team, and nurse managers met monthly to review performance	No payment component. Program provides on-site registered nurse (the GCN)	EHR embeds evidence-based guidelines; generates individualized care guides based on guidelines and patient information; tracks patients; and sends reminders to GCN
Improving Mood-Promoting Access to Collaborative Treatment for Late-Life Depression	Depression care for elderly depressed patients is integrated into primary care via a depression clinical specialist (DCS) (a nurse or psychologist) who coordinates care between the PCP, consulting PCP, and psychiatrist	Patient and DCS establish individualized care plan, which includes education, care management, problem-solving treatment, support for antidepressant use, and relapse prevention	DCS, in consultation with the consulting PCP and team psychiatrist, works with patient and regular PCP to provide depression care. DCS supports antidepressant therapy and behavioral activation	DCS does not coordinate with external providers (psychiatrist and DCS become part of internal team)	Telephone and in-person contact with DCS	Evidence-based treatment algorithm used by DCS and care team. The DCS and psychiatrist review progress weekly over the year-long intervention	No payment component. Program provides DCS, consulting PCP, and psychiatrist	Internet-based system used to record patient contacts; electronic reminders to DCS if time for a contact or on ineffective treatment

Table 17 (continued)

AHRQ PCMH Principles and Facilitators

Intervention	Overview	Patient-Centered	Comprehensive Care	Coordinated Care	Access to Care	Systems Approach to Quality and Safety	Payment and Other Resources to the Primary Care Practice	Health IT
Merit Health System and Blue Cross Blue Shield of North Dakota Chronic Disease Management Pilot	BCBS embedded a chronic disease management nurse in a clinic for patients with diabetes. The nurse assesses patient knowledge of diabetes, sets goals for disease self-management, establishes the need for in-person or telephone followup, and refers to services	Nurse and patients set goals, and nurse provides self-management education	Focused on diabetes care	Nurses make referrals for services such as nutrition counseling	Nurse available by telephone (unclear whether 24/7 access is available)	EHRs allow patients and physicians to track patient outcomes and provide aggregate performance information to physicians	\$20,000 startup grant and 50% of savings generated in the first year of the pilot. Program provides a disease management nurse in the clinic. After the pilot, BCBS replaced the startup grant and in-kind nurse with a disease management fee	Existing EHR used by physicians and patients to track patient care
Pediatric Alliance for Coordinated Care	A pediatric nurse practitioner (PNP) from each practice allocates 8 hours per week to coordinate care of children with special health care needs and make expedited referrals to specialists and hospitals; a local parent of a child with special health care needs consults to the practice	Individualized health plan developed with the patient and family	Practice-based team care that includes physicians, PNP, office staff, and family consultants. Provides 8 hours per week of comprehensive case management; social support and activities	PNP makes expedited referrals and coordinates care across providers (e.g., therapists, school nurses), and education, social services, and recreation	After-hours coverage; PNP conducts home visits	PNPs and physicians receive ongoing training. Local parent provides feedback to practice	No payment to practices. Stipend to family members serving as consultants. Continuing medical education for physicians	No health IT component

Table 17 (continued)

AHRQ PCMH Principles and Facilitators								
Intervention	Overview	Patient-Centered	Comprehensive Care	Coordinated Care	Access to Care	Systems Approach to Quality and Safety	Payment and Other Resources to the Primary Care Practice	Health IT
Pennsylvania Chronic Care Initiative	Integrates the chronic care model and the medical home model for patients with diabetes and pediatric patients with asthma and includes patient-centered care, teaching self-management of chronic conditions, forming partnerships with community organizations, financial incentives for providers, and making data driven-decisions	Self-management support and coaching	Practice-based team care, which includes case managers, physicians, nurses, and office staff	Referral process to community services	Timely or same-day appointments	Use of performance measures and evidence-based guidelines to inform planning and treatment	Providers in practices that meet National Committee for Quality Assurance (NCQA) PCMH standards are eligible for supplemental payment, including an annual payment for clinicians (\$40,000 to \$95,000), infrastructure payments (starting at \$20,895), and provider performance incentives	Electronic patient registry
Veterans Affairs Team-Managed Home-Based Primary Care	Comprehensive and longitudinal primary care provided by an interdisciplinary team that includes a home-based primary care (HBPC) nurse in the homes of veterans with complex, chronic, terminal, and disabling diseases	Individualized treatment plan developed in collaboration with patient and caregiver; HBPC nurse teaches both patients and caregivers about the disease, treatment, and self-care; caregiver support provided	Patient assessment by HBPC team members from at least three different disciplines (social workers, dietitians, therapists, pharmacists, and paraprofessional aides); weekly team meetings	HBPC team coordinates patient care across all settings, and is involved in hospital discharge planning	24-hour contact for patients	Mandatory annual performance improvement plan; quarterly medical record reviews	No payment component. Physicians are salaried staff who devote a specific percentage of time to the HBPC program	HBPC information system designed to help HBPC teams manage their patients and resources, as well as to provide VA Central Office with site-specific information for all programs



AHRQ Publication No. 12(14)-0020-1-EF
February 2014