What are the Effects of Different Team-based Primary Care Structures on the Quadruple Aim of Care? A Rapid Review

Prepared for:
Department of Veterans Affairs
Veterans Health Administration
Health Services Research & Development Service
Washington, DC 20420

Prepared by:
Evidence Synthesis Program (ESP) Center
West Los Angeles VA Medical Center
Los Angeles, CA
Paul G. Shekelle, MD, PhD, Director

Authors:
Principal Investigator:
Paul G. Shekelle, MD, PhD

Research Associate:
Meron Begashaw, MPH
PREFACE

The VA Evidence Synthesis Program (ESP) was established in 2007 to provide timely and accurate syntheses of targeted healthcare topics of importance to clinicians, managers, and policymakers as they work to improve the health and healthcare of Veterans. These reports help:

- Develop clinical policies informed by evidence;
- Implement effective services to improve patient outcomes and to support VA clinical practice guidelines and performance measures; and
- Set the direction for future research to address gaps in clinical knowledge.

The program comprises three ESP Centers across the US and a Coordinating Center located in Portland, Oregon. Center Directors are VA clinicians and recognized leaders in the field of evidence synthesis with close ties to the AHRQ Evidence-based Practice Center Program and Cochrane. The Coordinating Center was created to manage program operations, ensure methodological consistency and quality of products, and interface with stakeholders. To ensure responsiveness to the needs of decision-makers, the program is governed by a Steering Committee composed of health system leadership and researchers. The program solicits nominations for review topics several times a year via the program website.

Comments on this evidence report are welcome and can be sent to Nicole Floyd, Deputy Director, ESP Coordinating Center at Nicole.Floyd@va.gov.


This report is based on research conducted by the Evidence Synthesis Program (ESP) Center located at the West Los Angeles VA Medical Center, Los Angeles, CA, funded by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development. The findings and conclusions in this document are those of the author(s) who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the Department of Veterans Affairs or the United States government. Therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs. No investigators have any affiliations or financial involvement (eg, employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in the report.
ACKNOWLEDGMENTS

This topic was developed in response to a nomination by Idamay Curtis, BA, Karin Nelson, MD, MSHS, and Greg Stewart, PhD for the purpose of identifying the structure and function of primary care team members to maximize patient and provider outcomes. The scope was further developed with input from the topic nominators (ie, Operational Partners), the ESP Coordinating Center, the review team, and the technical expert panel (TEP).

In designing the study questions and methodology at the outset of this report, the ESP consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicting opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

The authors gratefully acknowledge the following individuals for their contributions to this project:

**Operational Partners**

Operational partners are system-level stakeholders who have requested the report to inform decision-making. They recommend Technical Expert Panel (TEP) participants; assure VA relevance; help develop and approve final project scope and timeframe for completion; provide feedback on draft report; and provide consultation on strategies for dissemination of the report to field and relevant groups.

Idamay Curtis, BA  
*Co-Director of the Office of Primary Care Analytics Team*  
*VA Puget Sound Health Care System*

Karin Nelson, MD, MSHS  
*Director, Primary Care Analytic Team, VHA Office of Primary Care*  
*Professor of Medicine, University of Washington*

Greg Stewart, PhD  
*Professor of Management and Entrepreneurship*  
*University of Iowa, Tippie College of Business*

**Technical Expert Panel (TEP)**

To ensure robust, scientifically relevant work, the TEP guides topic refinement; provides input on key questions and eligibility criteria, advising on substantive issues or possibly overlooked areas of research; assures VA relevance; and provides feedback on work in progress. TEP members are listed below:

Lisa Rubenstein, MD, MS  
*Professor of Medicine and Public Health, VA Greater Los Angeles and UCLA*  
*Senior Natural Scientist, RAND*
The Coordinating Center sought input from external peer reviewers to review the draft report and provide feedback on the objectives, scope, methods used, perception of bias, and omitted evidence. Peer reviewers must disclose any relevant financial or non-financial conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The Coordinating Center and the ESP Center work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................... II

EXECUTIVE SUMMARY .......................................................................................................... 1
INTRODUCTION ......................................................................................................................... 1
METHODS .................................................................................................................................... 1
RESULTS ...................................................................................................................................... 2
DISCUSSION ................................................................................................................................ 5

EVIDENCE REPORT .................................................................................................................. 7
INTRODUCTION ......................................................................................................................... 7
METHODS .................................................................................................................................... 8
  Topic Development ...................................................................................................................... 8
  Search Strategy ............................................................................................................................ 8
  Study Selection ............................................................................................................................ 8
  Data Abstraction .......................................................................................................................... 9
  Quality Assessment ...................................................................................................................... 9
  Data Synthesis ............................................................................................................................. 10
  Rating the Body of Evidence ...................................................................................................... 10
  Peer Review ............................................................................................................................... 10
RESULT ........................................................................................................................................ 11
  Literature Flow ............................................................................................................................ 11
  Figure 1. Literature Flow Chart ................................................................................................ 12
  Description of the Evidence ....................................................................................................... 13
  Key Question: What are the effects of different primary care team structures on care? .......... 14
  Hypothesis-testing studies of 2 different primary care team structures ..................................... 14
  Studies of 2 different structures for the same team, measured at different time points .......... 17
  Adding a new team member to an established team .................................................................. 17
  Pre-post or time series studies of going from a defined “no team” structure to a defined “team-
  based” structure ......................................................................................................................... 22
SUMMARY OF FINDINGS/CERTAINTY OF EVIDENCE ..................................................... 23
  Limitations ................................................................................................................................. 26
  Publication Bias .......................................................................................................................... 26
  Study Quality ............................................................................................................................. 26
  Heterogeneity ............................................................................................................................... 26
  Applicability of Findings to the VA Population ......................................................................... 26
  Research Gaps/Future Research ................................................................................................ 26
  Conclusions ............................................................................................................................... 27
REFERENCES ............................................................................................................................... 28

TABLES
  Table 1. Description of the Evidence ......................................................................................... 13
  Table 2. Comparative Studies of Different Team Member Compositions ................................. 16
  Table 3. Comparative Studies of Adding Team Members .......................................................... 20
  Table 4. Certainty of Evidence .................................................................................................. 25
APPENDIX A. SEARCH STRATEGIES .......................................................... 30
APPENDIX B. PEER REVIEWER COMMENTS AND RESPONSES .............. 31
APPENDIX C. CITATIONS FOR EXCLUDED PUBLICATIONS .............. 40
APPENDIX D. RISK OF BIAS TOOL .......................................................... 54
EXECUTIVE SUMMARY

INTRODUCTION

Team-based primary care has become a predominant model to provide accessible, high-quality care, and meet the quadruple aims of improving patient experience, population health, the work life of the health care workforce, and reduce costs. VA re-organized primary care delivery via the Patient Aligned Care Teams, which is based on the medical home model. Within the primary care team are smaller units, what Bodenheimer and Liang term “the central subunit” of the team,¹ which has been called the teamlet. The smallest composition of the teamlet is the clinician and medical assistant. Bodenheimer and Liang proposed the teamlet consist of a clinician and 2 health coaches. Other compositions have been proposed. In VA, the teamlet has been defined as a primary care provider (either a physician, a physician’s assistant, or a nurse practitioner), a registered nurse (RN), a licensed practical nurse (LPN), and a clerk or medical support assistant. Thus, the model aims to provide 3.0 full-time equivalent (FTE) staff for each PCP FTE, and each teamlet is expected to provide primary care for approximately 1200 Veterans. As VA continually seeks to improve the quality, cost, access, and wellbeing of the health care workforce, the question arises whether other compositions of the teamlet or the larger team might produce improvements in any of these domains. Thus, the Office of Primary Care requested this Rapid Review regarding team composition and outcomes.

METHODS

This topic was developed in response to a nomination by Idamay Curtis, Co-Director of Primary Care Analytics Team, Dr. Karin Nelson, Director of the Office of Primary Care Analytics Team, and Dr. Greg Stewart, Professor of Management and Entrepreneurship. Key questions were then developed with input from the topic nominator, the ESP coordinating center, the review team, and the technical expert panel (TEP).

The revised Key Question for this rapid review was:

What are the effects of different primary care team structures on care?

DATA SOURCES AND SEARCHES

We conducted broad searches using terms relating to “patient care team” or “team based” or “primary health care.” We searched OVID Medline from inception to 5/29/20.

STUDY SELECTION

Studies were included if they:

1) Were a comparative study of 2 different primary care team structures (randomized or observational);
2) Were a pre-post or time series study of 2 different structures for the same team – in other words at time point zero a team has 1 structure and then this is changed at a later time to a different structure;

3) Were hypothesis-testing studies of adding a new team member to an established team – for example, adding a nurse practitioner or a pharmacist to an existing team;

4) Were pre-post or time series studies of going from a “no team” structure to a defined “team-based” structure; or

5) Included studies needed to report a triple aim outcome (quality, cost, patient experience) or provider-based outcome (such as burnout).

DATA ABSTRACTION AND QUALITY ASSESSMENT

Data abstracted included the study design, setting, sample size, team members added or team members studied, outcomes, and data needed for the quality assessment/risk of bias tools. Randomized trials were assessed for quality/risk of bias using the Cochrane Risk of Bias Tool. Observational studies that were longitudinal and had a control group were assessed using the Cochrane Risk of Bias in Observational Studies – Interventions (ROBINS-I). Cross-sectional and pre-post studies were not assessed for risk of bias with a tool since they are by definition at high risk of bias. Modeling studies were not assessed for quality because no standardized tool exists for that purpose.

DATA SYNTHESIS AND ANALYSIS

We grouped studies into 1 of the 4 categories described above, and within category summarized the evidence narratively. We used the criteria of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group to assess the certainty of evidence across studies.

RESULTS

RESULTS OF LITERATURE SEARCH

We identified 3,463 potentially relevant citations, of which 214 were included at the abstract screening. From these, a total of 129 abstracts were excluded. A total of 14 publications were identified at full-text review as meeting initial inclusion criteria. This included 5 studies that showed comparative study of team structure A vs structure B, 8 studies that added a provider in context of team-based care, and 1 study that was not a team to team with team roles defined.

KEY FINDINGS AND STRENGTH OF EVIDENCE

KQ: What are the effects of different primary care team structures on care?

The evidence on what matters in terms of composition of the teamlet is very sparse, consisting of a few hypothesis-testing studies that address only partial aspects of the question, and modeling studies. The most robust evidence is that adding a dedicated chronic care manager can improve some outcomes for some patients – although in the prior ESP review of nurse-managed protocols
the nurse charged with doing this required prescribing authority, which is not something VA currently has for teamlet RNs. We rated this as moderate certainty evidence based on 1 RCT and 1 longitudinal study with a control group, and augmented this with the results of the 1 high-quality ESP review on nurse-managed protocols. Additional low-certainty evidence, based on a single study each, is that adding NPs as co-managers to a physician teamlet increases access (as measured by the 3rd next available appointment), that re-training medical assistants to perform screening increases screening rates, and that differing patient populations will require differing mixes of team skill FTE in order to deliver high-quality care. See Certainty of Evidence table for details. We did not include as “findings” or rate for certainty of evidence conclusions based on results of single studies that were cross-sectional or pre-post in design; thus, only 6 studies contribute evidence to “findings”.
### Table. Certainty of Evidence

<table>
<thead>
<tr>
<th>Intervention or Outcome</th>
<th>Number of studies</th>
<th>Study limitations</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Other factors</th>
<th>Overall Certainty of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adding NP or other dedicated chronic care manager improves outcomes of some chronic conditions (most notably diabetes, but not hypertension)</td>
<td>3</td>
<td>Serious</td>
<td>No serious inconsistency</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Moderate-certainty evidence that nurse-managed protocols result in improvements in multiple outcomes for patients with chronic conditions</td>
<td>Moderate</td>
</tr>
<tr>
<td>Adding NPs as co-management providers increases access</td>
<td>1</td>
<td>Very serious</td>
<td>N/A</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>None</td>
<td>Low</td>
</tr>
<tr>
<td>Retraining medical assistants to screen patients for certain conditions can increase the proportion of patients screened</td>
<td>1</td>
<td>No serious limitation</td>
<td>N/A</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>None</td>
<td>Low</td>
</tr>
<tr>
<td>3.6 to 4.0 FTE of supporting team members are needed for each 1.0 FTE of physician to deliver high-quality comprehensive care; different mixes of skillsets and staffing needed for differing patient populations</td>
<td>1</td>
<td>Serious limitations</td>
<td>N/A</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Modeling study (Limitation: modeling study)</td>
<td>Low</td>
</tr>
</tbody>
</table>
DISCUSSION

APPLICABILITY

We found only 1 study in a VA population, and it was about single versus multiple team membership roles, and not about specific team members. We can only speculate as to the applicability of the remaining findings to VA populations. At least 1 of the interventions – nurse-managed protocols requiring prescribing authority – is not currently available within VA.

RESEARCH GAPS/FUTURE RESEARCH

VA would seem to be ideally placed to provide experimental evidence about how teamlet and team structures can be optimized. Almost every VA of sufficient size organizes their teamlets into larger units (like Red, Green, or Blue teams). This would then allow for controlled comparisons of differing team and teamlet structures, with other contextual features being internally controlled (like senior leadership, incentives, and the EHR). For example, any of the recommended team staffing levels in the model of Meyers for either their “high geriatric” or their “high social needs” models could be implemented in 1 larger team (Red/Blue/Yellow) while the others serve as control. Teamlet structure could be varied (for example, 1 RN for every 2 physician providers) or team structures could be varied (for example, adding the 1.0 FTE substance abuse counselor). Data collection could come directly from the EHR. Detailed information would need to be collected about patients’ chronic conditions and social needs, as the model by Meyers consider these important variables when determining optimal team composition. An agreed-upon metric for evaluating performance – presumably based on the triple aim – would facilitate comparisons of results across studies.

CONCLUSIONS

The evidence for an optimal teamlet or team structure is very sparse. Other than adding a dedicated chronic care manager, there is no evidence above low certainty that any team structure is optimal. Complex patients almost certainly benefit from additional skills (beyond the basic teamlet of provider, medical assistant, and nurse) in the team writ large (such as pharmacist, chronic care manager, etc).
## ABBREVIATIONS TABLE

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATBC</td>
<td>Advanced Team-Based Care</td>
</tr>
<tr>
<td>HER</td>
<td>Electronic health records</td>
</tr>
<tr>
<td>FTE</td>
<td>Full-time employee</td>
</tr>
<tr>
<td>GRADE</td>
<td>Grading of Recommendations Assessment, Development and Evaluation</td>
</tr>
<tr>
<td>HEDIS</td>
<td>Healthcare Effectiveness Data and Information Set</td>
</tr>
<tr>
<td>LPN</td>
<td>Licensed practical nurse</td>
</tr>
<tr>
<td>PA</td>
<td>Physician assistant</td>
</tr>
<tr>
<td>PACT</td>
<td>Patient Aligned Care Team</td>
</tr>
<tr>
<td>PCP</td>
<td>Primary care physician</td>
</tr>
<tr>
<td>TBPC</td>
<td>Team-based primary care</td>
</tr>
</tbody>
</table>
EVIDENCE REPORT

INTRODUCTION

Team-based primary care has become a predominant model to provide accessible, high-quality care, and to meet the quadruple aims of improving patient experience, population health, the work life of the health care workforce, and reduce costs. VA re-organized primary care delivery via the Patient Aligned Care Teams (PACT). Within the primary care team are smaller units, what Bodenheimer and Liang term “the central subunit” of the team,¹ which have been called teamlets. The smallest composition of the teamlet is the clinician and medical assistant. Bodenheimer and Liang proposed the teamlet consist of a clinician and 2 health coaches. Other compositions have been proposed. In VA, the teamlet has been defined as a primary care provider (either a physician, a physician’s assistant, or a nurse practitioner), a registered nurse (RN), a licensed practical nurse (LPN), and a clerk or medical support assistant. Thus, there are 3.0 full-time equivalent (FTE) staff for each PCP FTE, and each teamlet is expected to provide primary care for approximately 1200 Veterans. As VA continually seeks to improve the quality, cost, access, and work life of the health care workforce, the question arises whether other compositions of the teamlet or the larger team might produce improvements in any of these domains. Thus, the Office of Primary Care requested this Rapid Review regarding team composition and outcomes.
METHODS

TOPIC DEVELOPMENT

This topic was developed in response to a nomination by Idamay Curtis, Co-Director of Primary Care Analytics Team, Dr. Karin Nelson, Director of the Office of Primary Care Analytics Team, and Dr. Greg Stewart, Professor of Management and Entrepreneurship. Key questions were then developed with input from the topic nominator, the ESP coordinating center, and the review team.

The initial Key Questions were:

KQ1: Who should be a member of the care team (which occupations)?

KQ2: What roles and responsibilities are needed to provide high-quality, accessible primary care?

KQ3: What extended team members are needed (e.g., Certified Diabetes Educators, Certified Pharmacy Specialists, Social Workers)?

After discussions with the topic nominator, the Key Questions were refined by the following:

When VA launched the Patient-Aligned Care Team (PACT) initiative, the team (or teamlet) was defined (based on limited data) as: 1) a provider (physician, nurse practitioner); a full-time RN care manager; 3) a full-time LPN; 4) a full-time clerk. The question then is: is there a better team-based care structure? What roles are needed and how should primary care teams be structured?

This then resulted in the revised Key Question:

KQ: What are the effects of different primary care team structures on care?

The review was not registered in PROSPERO because it was not about an eligible “intervention”.

SEARCH STRATEGY

We conducted broad searches using terms relating to “patient care team” or “team based” or “primary health care.” We searched OVID Medline from inception to 5/29/20. See Appendix A for complete search strategy.

STUDY SELECTION

All title, abstract, and full-text articles were screened by 1 reviewer, Paul G. Shekelle (PGS). Questions about the relevance of an abstract or full-text article were discussed directly with the partner. Studies were included at either the abstract or the full-text level if they were hypothesis-testing studies, modeling studies, or systematic reviews of the following types, in descending order of the degree to which it can inform the key question:
1) A comparative study of 2 different primary care team structures (randomized or observational);

2) A pre-post or time series study of 2 different structures for the same team – in other words at time point zero a team has 1 structure and this is then changed at a later time to a different structure;

3) Hypothesis-testing studies of adding a new team member to an established team – for example, like adding a nurse practitioner or a pharmacist to an existing team;

4) Pre-post or time series studies of going from a “no team” structure to a defined “team-based” structure; or

5) Included studies needed to report a triple aim outcome (quality, cost, patient experience) or provider-based outcome (such as burnout).

We did not include studies that were about implementing “team-based care”, which typically consists of a pre-implementation “usual care” that is not-team based and/or loosely defined, and then a post-implementation assessment of team-based care, with “team-based” being defined by principles but not with the associated specifications of how each team is constituted (in terms of FTE, etc).

We did not include studies from the large literature of adding to the team a new member who deals only with a specific condition – for example, adding a case manager for patients with HIV. We only included such studies of adding a designated new team member if that person had responsibilities for at least 2 different health conditions.

We did not include qualitative studies that examined experiences or beliefs about team-based primary care.

We did not include studies about team-based care in other disciplines (mental health, for example).

“Team” was primarily defined as VA defines a teamlet – a provider, nurse, case manager, clerk, etcetera. A secondary definition of “team” was the team writ large, meaning a collection of primary care providers, nurses, clerks, etcetera, working within an integrated clinic setting.

DATA ABSTRACTION

Data extraction was completed by author PGS. Data abstracted included the study design, setting, sample size, team members added or team members studied, outcomes, and data needed for the quality assessment/risk of bias tools.

QUALITY ASSESSMENT

Randomized trials were assessed for quality/risk of bias using the Cochrane Risk of Bias Tool.\(^2\) Observational studies that were longitudinal and had a control group were assessed using the Cochrane Risk of Bias in Observational Studies – interventions (ROBINS-I).\(^3\) Cross-sectional
and pre-post studies were not assessed for risk of bias with a tool since they are by definition at high risk of bias. Modeling studies were not assessed for quality because no standardized tool exists for that purpose.

DATA SYNTHESIS

We grouped studies into 1 of the 4 categories described above, and within category summarized the evidence narratively.

RATING THE BODY OF EVIDENCE

We used the criteria of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group. GRADE assesses the certainty of the evidence based on the assessment of the following domains: risk of bias, imprecision, inconsistency, indirectness, and publication bias. This results in categories as follows:

High: We are very confident that the true effect lies close to that of the estimate of the effect.

Moderate: We are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low: Our confidence in the effect estimate is limited. The true effect may be substantially different from the estimate of the effect.

Very low/Insufficient: We have very little confidence in the effect estimate. The true effect is likely to be substantially different from the estimate of effect.

PEER REVIEW

A draft version of the report was reviewed by technical experts and clinical leadership. Peer reviewer comments and responses are included in Appendix B.
RESULTS

LITERATURE FLOW

We identified 3,464 potentially relevant citations. Inspection of the titles excluded 3,249 as being irrelevant, leaving 215 titles that were selected for abstract screening. From these, a total of 129 abstracts were excluded. Excluded abstracts were categorized as not research (n=27), not about specific team member roles in TBPC (n=24), not about TBPC (n=22), qualitative study (n=17), no relevant outcomes (n=8), no abstract (n=6), descriptive study (n=1), and study protocol (n=1). This left 86 publications for full-text review, of which 72 publications were excluded for the following reasons: not about specific team member roles in TBPC (n=32), not about TBPC (n=10), adding pharmacists (n=8), no outcome of interest (n=7), qualitative study (n=4), not available (n=4), not research (n=2), background (n=1), descriptive (n=1), not primary care (n=1), non-systematic review (n=1), and study protocol (n=1). A full list of excluded studies from the full-text review is included in Appendix C. A total of 14 publications were identified at full-text review as meeting initial inclusion criteria. This included 5 studies that showed comparative study of different team structures,5-8 8 studies that added a provider in context of team-based care,9-16 and 1 study that assessed outcomes when going from a structure that was not team-based to a structure that was.17 See Figure 1 for literature flow. Descriptions of included publications are available in Tables 2 and 3.
Figure 1. Literature Flow Chart

Total titles screened: 3,464

Excluded: 3,249

Excluded = 129 references
Not research: 27
Not about specific team member roles in TBPC: 24
Adding Pharmacists: 22
Not about TBPC: 22
Qualitative study: 17
No relevant outcomes: 8
No abstract: 6
Adding Occupational Therapists: 1
Descriptive study: 1
Study protocol: 1

Abstracts reviewed: 215

Excluded = 72 references
Not about specific team member roles in TBPC: 32
Not about TBPC: 10
Adding Pharmacists: 8
No outcome of interest: 7
Qualitative study: 4
Unavailable: 4
Not research: 2
Background: 1
Descriptive study: 1
Not primary care: 1
Non-Systematic Review: 1
Study protocol: 1

Full text: 86

Included studies: 14

Comparative Study of Team Structure A vs. Structure B: 5
Pre-post or Time Series of Team Structure: 0
Adding a Provider in Context of Team-Based Care: 8
Not a Team to Team with Team Roles Defined: 1

12
DESCRIPTION OF THE EVIDENCE

We identified 1 randomized trial.12 Of the 5 studies comparing 2 different team structures, 1 was a longitudinal observational study,8 3 were cross-sectional studies,5,6,18 and 1 was a modeling study.8 (See Table 1. Description of Evidence). Three studies were in US primary care,6,7,18 1 study was from England general practices,7 and the remaining study was in Sweden.8 The team members studies included physicians and nurses. The outcomes assessed were clinical quality in 4 studies,8,12,15,16 physician burnout in 1 study,6 and a composite of “high-quality, comprehensive care”.8 Of the 7 studies about adding a team member, 2 were systematic reviews,13,14 2 were longitudinal studies,12,15 1 was a pre-post study,10 and the last 1 was a modeling study.9 The 5 non-systematic review studies were all based in the US. Three of the studies were about adding a nurse practitioner (or physician assistant),10-12 2 were about adding nurses,13,14 1 was about adding a medical scribe or medical assistant,9 and 1 was about adding a care manager.15 Outcomes assessed clinical quality, costs, access, and panel size (which we included as being related to access/costs). All included studies were at high risk of bias in at least 1 domain (see Appendix D).

Table 1. Description of the Evidence

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Study Design</th>
<th>Location</th>
<th>Outcomes Assessed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Litaker, 200312</td>
<td>RCT</td>
<td>General internal medicine at Cleveland Clinic in Cleveland, Ohio, United States</td>
<td>Diabetes care, hypertension, preventative care, HRQOL</td>
</tr>
<tr>
<td>Dorr, 200615</td>
<td>Observational longitudinal with control</td>
<td>Intermountain Healthcare in Utah and Idaho, United States</td>
<td>Diabetes control, costs for depression care</td>
</tr>
<tr>
<td>Ovhed, 20008</td>
<td>Observational longitudinal with control</td>
<td>2 primary health care centers in Blekinge county, South Sweden</td>
<td>Diabetes care practices</td>
</tr>
<tr>
<td>Stoeckle, 201916</td>
<td>Time series</td>
<td>Family medicine clinic in Philadelphia, Pennsylvania, United States</td>
<td>Screening for depression and falls</td>
</tr>
<tr>
<td>Brown-Johnson, 201917</td>
<td>Pre-post</td>
<td>Primary care clinic at an academic medical center</td>
<td>Clinic operations, wellness and proactive care, patient engagement and trust</td>
</tr>
<tr>
<td>D’Afflitti, 201810</td>
<td>Pre-post</td>
<td>General internal medicine practice at Boston Medical Center in Boston, Massachusetts</td>
<td>Access</td>
</tr>
<tr>
<td>Bower, 20035</td>
<td>Observational Cross-sectional</td>
<td>60 primary care practices across England</td>
<td>Clinical quality for angina, asthma, diabetes</td>
</tr>
<tr>
<td>Bruhl, 20206</td>
<td>Observational Cross-sectional</td>
<td>420 family medicine clinicians practicing in 59 Midwestern communities in the United States</td>
<td>Physician burnout</td>
</tr>
<tr>
<td>Crawford, 201918</td>
<td>Observational Cross-sectional</td>
<td>849 VHA primary care units across the United States</td>
<td>“Unit performance” – Emergency Department visits</td>
</tr>
<tr>
<td>Dai, 201911</td>
<td>Cross-sectional survey</td>
<td>Family practices across the United States</td>
<td>Panel size</td>
</tr>
<tr>
<td>Author, Year</td>
<td>Study Design</td>
<td>Location</td>
<td>Outcomes Assessed</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>----------</td>
<td>------------------</td>
</tr>
<tr>
<td>Basu, 2018⁹</td>
<td>Model</td>
<td>643 primary care practices across the United States</td>
<td>Costs</td>
</tr>
<tr>
<td>Meyers, 2018⁷</td>
<td>Model</td>
<td>Multiple primary care practices across the United States</td>
<td>N/A</td>
</tr>
<tr>
<td>Shaw, 2013¹³</td>
<td>Systematic review of 29 studies</td>
<td>Adult outpatient practices in Western Europe and the United States</td>
<td>Clinical quality resource use; nursing satisfaction</td>
</tr>
<tr>
<td>Wranik, 2019¹⁴</td>
<td>Systematic review of 31 quantitative studies</td>
<td>Primary care sites in Australia, Canada, the United Kingdom, and New Zealand</td>
<td>Clinical process and outcomes</td>
</tr>
</tbody>
</table>

**KEY QUESTION: WHAT ARE THE EFFECTS OF DIFFERENT PRIMARY CARE TEAM STRUCTURES ON CARE?**

**Hypothesis-testing studies of 2 different primary care team structures**

We identified 4 hypothesis-testing studies and 1 modeling study comparing different primary care team structures (Table 2).

Three studies assessed primary care in the US. The only study that was empirical and directly assessed an outcome of interest was by Bruhl and colleagues.⁶ In this cross-sectional study of 420 family medicine clinicians in 59 practices within 1 integrated health system, 217 (52%) responded to a survey containing a single item of the emotional exhaustion domain of physician burnout. Responding clinicians were working about 0.9 FTE and on teams consisting of on average 5 clinicians. Each clinician was responsible for a panel of about 1000 patients. In a multivariate linear mixed model regression analysis, an increasing proportion of physician FTE on the team (calculated by taking the physician FTE and dividing by the total care team clinician FTE) was associated with a modest reduction in burnout (that is, less burnout) – in the context of 85% of respondents indicating they were at high risk of burnout. As this study is cross-sectional in design the temporal relationship of this association cannot be discerned. The low response rate is another limitation of this study in drawing conclusions.

A second study assessed different team structures in US primary care, but via a model. The model estimated the staffing needed to provide “high-quality, comprehensive care” (not otherwise explicitly defined) for 4 exemplar patient populations: the index model, a model of a practice with a high proportion of geriatric patients, a model of a practice with a high proportion of patients with social needs, and a rural population model. Each of the first 3 models started with the goal of providing care to 10,000 patients, and then modeled how many providers and the skill mix that would be needed to provide high-quality care. The rural model assumed care for 5000 patients. The index model had 6.0 FTE MD providers and 2.0 FTE NP/PA providers, and required 3.6 FTE of supporting team members for each clinician, whereas this rose to 4.0 FTE for the high geriatric and high social needs models. The types of supporting FTE varied for each population. For example, the high geriatric model had more clinicians, more RNs, more social workers, and more dedicated care coordination FTE than the index model. The high social needs model had fewer physician providers and more NP/PA providers, more social workers, and a 1.0 FTE substance abuse counselor, in addition to community health workers. Strengths of this study
are that it gives specific estimates for a broad array of skillsets needed for a primary care team, and it does so for different types of patient populations. Limitations are that it is a model, and what the model is optimizing  high-quality comprehensive care  is not defined in detail.

The remaining 2 studies assessed primary care team structures in England and Sweden. The Swedish study compared diabetes outcomes at 2 primary health care centers. In 1 center, local guidelines were implemented an independent role for nurses, who saw diabetic patients independently 3 times per year. The comparison clinic had a traditional Swedish arrangement where nurses were assistants to the physicians and did not work independently. In addition, no local guidelines were implemented. Over 1 year there were fewer physician visits and many more nursing visits at the clinic with independent nurses, and much more secondary prevention and testing of intermediate outcomes (HbA1c, lipids, retinal exams, etc) as well as more patient knowledge and self-efficacy. In the English study, 42 general practices were assessed for skill mix (ratio of doctors to nurses), team climate, and performance measures for several chronic diseases. In a multivariable model, skill mix was not a significant predictor variable for quality of chronic disease care.

Lastly, Crawford and colleagues used data about team composition to assess the association of multiple team membership on patient Emergency Department use. Using national VA data on 849 primary care units (in hospitals and clinics) and a VA team membership report, they determined whether staff had single or multiple team membership, and how many teams a staff person participated in. They then used this as a predictor variable in a multivariable model that adjusted for patient care complexity, rural/urban location, average team size, and staff-to-provider ratio, assessing the association with Emergency Department visits by primary care unit patients. The analysis found a statistically significant association between an increasing number of team memberships and Emergency Department visits. This association was most pronounced for patients with the highest health needs (higher care complexity). The authors conclude that there is an association between multiple team memberships and worse unit performance (as measured by increasing Emergency Department use), but acknowledge that their cross-sectional study design cannot support conclusions about causation.
## Table 2. Comparative Studies of Different Team Member Compositions

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Study Design</th>
<th>Setting</th>
<th>Team Members Studied</th>
<th>Outcomes</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bruhl, 2020</td>
<td>Observational Cross-sectional</td>
<td>420 family medicine clinicians practicing in 59 Midwestern communities</td>
<td>Proportion of physician FTE on a team</td>
<td>Physician burnout</td>
<td>Increasing physician FTE on a team was associated with modest reductions in physician burnout in a setting of high risk for burnout</td>
</tr>
<tr>
<td>Meyers, 2018</td>
<td>Model</td>
<td>Multiple US primary care sites</td>
<td>MD/DDs NPs/PAs RN LPN/LVN MA</td>
<td>High-quality comprehensive care</td>
<td>Detailed specifications for skill sets and FTE to deliver high quality care under 4 different patient population scenarios</td>
</tr>
<tr>
<td>Bower, 2003</td>
<td>Observational Cross-sectional</td>
<td>60 English general practices</td>
<td>Skill mix based on team composition: Ratio of doctors: nurses Ratio of doctors: non-medical clinical staff Ratio of clinical: administrative staff</td>
<td>Clinical quality for angina, asthma, diabetes</td>
<td>Ratio of physicians to nurses not associated with changes in chronic care quality</td>
</tr>
<tr>
<td>Ovhed, 2000</td>
<td>Observational longitudinal with control N=394</td>
<td>2 Swedish primary health care centers</td>
<td>Greater independent role for nurses in diabetes management</td>
<td>Diabetes care practices</td>
<td>Independence for nurses was associated with more secondary prevention and intermediate outcomes in patients with DM</td>
</tr>
<tr>
<td>Crawford, 2019</td>
<td>Observational Cross-sectional</td>
<td>849 Veterans Health Administration primary care units</td>
<td>All team members, and the degree to which they were members of a single team or of multiple teams</td>
<td>Emergency Department visits</td>
<td>Units where staff are members of multiple teams had statistically significantly greater patient use of the Emergency Department, particularly for patients with greater health needs</td>
</tr>
</tbody>
</table>
Studies of 2 different structures for the same team, measured at different time points

We did not identify any studies of this type.

Adding a new team member to an established team

We identified 8 hypothesis-testing studies that assessed adding team members to an established primary care team: 5 empirical studies, 1 modeling study, and 2 systematic reviews (Table 3).

The only randomized trial we identified came from the Cleveland Clinic, and it compared traditional physician-only chronic disease management with “a more collaborative, team-based approach”, the key ingredient of which was adding a Nurse Practitioner trained in chronic care management. Additional intervention features included greater use of telephonic management and standardized forms to facilitate documentation of preventive care delivery. Patients with diabetes and hypertension were randomized (N=157) to 1 or the other and followed for 12 months. The mean age of subjects was about 61 years, more than half of patients were African American, and nearly 59% were female. Baseline A1c for patients with diabetes was 8.5, and 9% of patients met blood pressure target goals. A number of prevention measures and monitoring measures were performed more frequently in the intervention group (such as receipt of influenza vaccination by 78% vs 47%, p <0.001, foot exam in 100% vs 36%, etc) and as well as much more documentation of diabetic teaching. There were greater declines in A1c level in the intervention group (-0.63 vs -0.15, p=0.02) and increases in HDL (3.0 vs 0.4, p=0.02). There were no differences between groups in a number of other measures, including most HRQOL measures, an eye exam by an ophthalmologist, total cholesterol, and control of blood pressure. Personnel costs were greater in the intervention group. In a post-hoc analysis, after discontinuation of the trial and the return of patients to the “usual care” form of chronic disease management, A1c values climbed in the former intervention patients, and by 12 months there was no longer any difference between such patients and the control patients.

A study from Intermountain Healthcare also described their changes to deal with chronic disease care management. This included a reliance on the electronic health record to support information system needs and local evidence-based guidelines, but also the addition of a generalist care manager (ie, not restricted to a single disease). Each of the 7 primary care clinics had a dedicated care manager, who saw patients by referral from the primary care provider. The 3 most common reasons for referral to the care manager were diabetes, mental health, and social/organizational needs. Compared to patients that did not receive care management, those with care management (N=2,356) had a greater decrease in HbA1c (a reduction of 0.55% vs 0.18%, statistical testing not performed), increased primary care productivity (an increase of 8% vs 5.5%, statistical testing not performed), and decreased cost of care for patients with depression (8% decrease vs 19% increase, statistical testing not performed).

In a pre-post study from general internal medicine practice at Boston Medical Center, the “NP-Anchor” model of primary care was implemented. Prior to this, 8 clinic NPs functioned as independent primary care providers with their own patient panels. Due to high rates of job dissatisfaction, job stress, and burnout, after a successful pilot study the clinic staffing was re-configured such that 1 NP worked with 3-4 physicians (not all of whom were full time; the ratio of FTE was 1.0 NP to 1.5 physician) to co-manage patients. NPs no longer had their own patient
panels. The target panel size for physicians was 1350; this did not change in the new configuration. NPs saw patients 6 half-days a week, and 2 additional half-days were devoted to chronic care management via telephone calls, test results follow-up, care coordination with specialists, and the like. Following implementation, the average time to the third next available appointment decreased nearly 20 days, from 26.2 days to 6.6 days (statistical testing not done). This was judged to be due to the NP adding capacity to the care team.

A cross-sectional survey of registrants for the American Board of Family Medicine Certification Examination (N=27,836; 100% response rate for the main survey, but only 42.5% or about 11,800 answered the panel size question), panel size was statistically significantly higher for practices that had an NP or a PA (or both), being 410 patients greater in practices with a PA, 259 greater in practices with an NP, and 245 greater in practices with both an NP and a PA. Overall, panel sizes were between 1900 and 2500 patients, depending on the size of the practice (solo practice to large practices, with smaller panels sizes the larger the practice). The low response rate to the primary outcome measure and the cross-sectional study design are both limitations of this study.

One study modeled the financial effect of adding a medical scribe or of having medical assistants complete the history, documentation, counseling, and order entry. The latter has been called the advanced team-based care (ATBC) strategy. The context was a fee-for-service primary care practice. The tradeoff was less physician time spent doing documentation, counseling, etcetera, which then opened up more time to see additional revenue-generating patients. In order to break even, use of the medical scribe would need to save 3.5 minutes per encounter, thus adding 317 visit slots per year, whereas the ATBC strategy would require saving 7.4 minutes per encounter, to enable the addition of 720 extra visit slots per year.

Although not technically a study of adding a team member, we include here a time-series study of re-defining the responsibilities of a medical assistant, to deal with screening. We judged that conceptually this addition of a skillset is closer to adding a team member than it is to any of our other categories. In this study, from the Thomas Jefferson family medicine clinic, medical assistants were trained and given responsibility for screening patients for depression and for falls. Over a span of 22 months, prior to implementing this change the screening rate for these conditions was essentially zero for depression, and 20% or less for falls. Immediately after implementation, the screening rate for depression increased to 30%, and continued to increase to 60% over the next 6 months. For falls, the screening rate jumped from about 23% to 45%, and then continued to increase over the next 6 months to 75%. Although no statistical testing was done, and there is no concurrent control group, the dramatic rises in screening temporally associated with implementation of medical assistant screening is strong evidence of a causal relationship.

Lastly, there were 2 systematic reviews with data relevant to this question. The first is a VA Evidence Synthesis Project report by the Durham VA, who assessed the evidence for nurse-managed protocols in outpatient management of adults with chronic conditions. Although not required to be part of team-based care, this review was nevertheless considered to be relevant for examining the potential use of nurses in team-based care to manage chronic conditions. The review searched computerized databases from 1980 through December 2012 and identified 29 unique studies, 26 of which were RCTs, and all of which compared the nurse-managed protocol
to usual care. Most studies were about patients with elevated cardiovascular risk. Only 1 study described interactions of the nurse with the team physician, which would be a *sine qua non* for team-based care. Thus the assumption that most of what is being assessed is not in the context of team-based care. All 29 studies required the nurse to have autonomy to titrate medications. Twenty studies reported that the nurse was allowed to independently initiate a new medication. Meta-analysis of studies (anywhere from 5 to 12 depending on outcome) showed low- to moderate-certainty evidence that nurse-managed protocols compared to usual care resulted in improvements in measures such as A1c, blood pressure, total and LDL cholesterol, as well as outcomes like total hospitalizations, heart failure hospitalizations, and even mortality.

The other systematic review assessed the associations of interprofessional primary care team characteristics on use of health services and patient outcomes. The review was published in 2019, but the end date of the computerized database search was not included in the article. Thirty-one quantitative studies were included in the narrative synthesis (along with 38 qualitative studies and 8 mixed-methods studies). The authors’ synthesis states that “the addition of specific professions to teams was generally positively associated with care processes”, and goes on to state that “for example, the addition of nurses or the expansion of their role improved care coordination”, but only cited references, and not abstracted data, is given in support of this statement. Reference mining the cited studies in this review yielded 34 articles that were examined, but none of these met the inclusion criteria for this review; thus, we refrain from accepting at face value the conclusion of this review.
## Table 3. Comparative Studies of Adding Team Members

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Study Design</th>
<th>Setting</th>
<th>Team Members Added</th>
<th>Outcomes</th>
<th>Results</th>
</tr>
</thead>
</table>
| Basu, 2018* | Model        | US internal medicine and family practices seeing fee-for-service patients | 1) Medical scribe  
2) Medical assistants | Costs | In order to be cost-neutral, use of the medical scribe would need to save 3.5 minutes per visit, enabling the physician to see an additional 317 visits per year; the corresponding numbers for the medical assistant strategy is a saving of 7.4 minutes per visit with an increase of 720 visits per year. |
<p>| D’Afflitti, 2018* | Pre-post | General internal medicine practice at Boston Medical Center | NP-physician care team | Access | Decrease in average time to third next available appointment from 26.2 to 6.6 days (statistical testing not done) |
| Dai, 2019* | Cross-sectional survey, N=11,800 | US family practices | NP/PA | Panel size | Panel size is 245 to 410 patients higher in practices that have an NP or PA or both. |
| Dorr, 2006* | Observational longitudinal with control, N=2,356 treated by care manager team | Intermountain Healthcare | Chronic care manager added to PC team | Diabetes control, costs for depression care | Compared to patients treated without care management, intervention patients had greater decreases in A1c (-0.55% vs -0.18%), lower costs for depression (85 decrease vs 19% increase), and increased primary care productivity (8% vs 5.5%) Statistical testing not performed. |
| Litaker, 2003* | RCT, N=157 | General internal medicine at Cleveland Clinic | NP trained in chronic disease management added to physician team, vs physician-only care | Diabetes care, hypertension, preventative care, HRQOL | Compared to physician only care, patients treated in the teams with the NP had greater decreases in A1c (0.63 vs 0.15) but no better control of blood pressure. Personnel costs were greater in the NP group. |
| Stoeckle, 2019* | Time series, N = 12,462 | Family medicine clinic at Thomas Jefferson | Medical assistants who were already team members were trained to provide screening | Screening for depression and falls | Dramatic increases in screening coincident with implementation, from 0% to 64% for depression and from 23% to 75% for falls. |</p>
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Study Design</th>
<th>Setting</th>
<th>Team Members Added</th>
<th>Outcomes</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shaw, 2013 13</td>
<td>Systematic review of 29 studies, search ended in 2012</td>
<td>Adult outpatient practice</td>
<td>Nurse, although not required to be team-based</td>
<td>Clinical quality resource use; nursing satisfaction</td>
<td>Compared to usual care, nurse-managed protocols resulted in numerous improvements in intermediate outcomes such as A1c control, blood pressure, and lipids, as well as reductions in hospitalization and mortality.</td>
</tr>
<tr>
<td>Wranik, 2019 14</td>
<td>Systematic review of 31 quantitative studies</td>
<td>Primary care in Western publicly-funded health care systems focusing on general care, diabetes, asthma, ischemic heart disease, hypertension, and multimorbidity care</td>
<td>Nurses, pharmacists, non-clinical staff</td>
<td>Clinical process and outcomes</td>
<td>“The addition of specific professions to teams was generally positively associated with care processes”</td>
</tr>
</tbody>
</table>
Pre-post or time series studies of going from a defined “no team” structure to a defined “team-based” structure

We identified 1 hypothesis-testing study in this category. This was a description of how Stanford developed and implemented a new model of primary care in 1 of their clinics. The authors describe a careful process that included site visits to other systems noted for high-performing primary care practices, input from patient stakeholders, re-design of the physical space, implementation and evaluation of the new model of care, and more. The new care model had care teams composed of a physician and an advanced practice provider along with 4 medical assistants. Three of these teams were grouped together to care for 10,000 patients, and each of these larger care teams had access to additional onsite expertise, including a clinical pharmacist, dietician, behavioral health specialist, nurse, and physical therapist. In the pilot clinic where this was implemented, the authors report there were “positive trends” over time, including an increase in the Press-Ganey likelihood-to-recommend proportion from 81.0% to 83.6%, a HEDIS composite measure from 65.8% to 72.3%, and a measure of provider well-being increased from 49 to 64.7. However, no sample sizes or other data are provided to support these numbers, there is no statistical testing to be able to assess whether the 2 numbers are in fact statistically different, and there is no non-intervention control group to assess whether these changes are more or less than temporal trends in satisfaction, HEDIS scores, or provider well-being.
SUMMARY OF FINDINGS/CERTAINTY OF EVIDENCE

The evidence on what matters in terms of composition of the teamlet is very sparse, consisting of a few hypothesis-testing studies that address only partial aspects of the question, and modeling studies. The most robust evidence is that adding a dedicated chronic care manager can improve some outcomes for some patients – although in the prior ESP review of nurse-managed protocols study the nurse charged with doing this required prescribing authority, which is not something VA teamlet RNs currently have. We rated this as moderate-certainty evidence based on 1 RCT that found better outcomes for patients with diabetes but no difference in hypertension outcomes and 1 longitudinal study with a control group that found better outcomes for diabetes and depression care, and augmented this with the results of the 1 high-quality ESP review on nurse-managed protocols (which assessed effects on care for a number of conditions, such as diabetes, lipid control, and hypertension). Our rating of moderate certainty of evidence is supported as follows: there are 3 studies relevant to this question, 1 RCT, 1 observational study with a concurrent control group, and 1 ESP systematic review on a related topic. The RCT was judged as being at unknown risk of bias on 2 important domains, and thus for the domain of “study limitations” this body of evidence was rated as having a serious limitation. Both of the 2 original research studies reached similar (favorable) conclusions regarding the addition of a dedicated chronic care manager, and thus we judged the domain of inconsistency as having no serious inconsistency. We did not judge either study as having a serious limitation in the directness of the evidence, as they measured well-accepted outcomes for chronic conditions and were (by definition) interventions that added a chronic care manager to an existing team structure. We also did not judge these studies as having a serious limitation with respect to imprecision, not because we believed that the added benefit of the chronic care manager was precisely estimated and similar in both studies, but rather that the reported benefit in the studies exceeded commonly used thresholds for other interventions being deemed worth doing (such as the approximately 0.5 point decrease in HgbA1c, an effect size commonly seen with addition of certain drugs). Thus, both studies exceeded the “decision threshold”, and were sufficiently precise to conclude that the intervention was “worth doing”. The column headed “Other Factors” is where we considered the existing ESP review on nurse-managed protocols, which concluded that there was moderate-certainty evidence that their use resulted in beneficial outcomes for a host of chronic conditions. The synthesis of all these factors led us to conclude that the certainty of evidence is moderate for adding a dedicated chronic care manager to an existing primary care team. Similar reasoning led us to conclude that there is low-certainty evidence, based on a single study each, that adding NPs as co-managers to a physician teamlet increases access (as measured by the 3nd next available appointment), that re-training medical assistants to perform screening increases screening rates, and that differing patient populations will require differing mixes of team skill FTE in order to deliver high-quality care. See Table 4 for details. We did not include as “findings” or rate for certainty of evidence conclusions based on results of single studies that were cross-sectional or pre-post in design. Thus only 6 studies contribute evidence to “findings”.

Not included as evidence per se, but relevant to these findings, are the results of in-depth case studies of 30 primary care practices viewed as innovators in team-based care, as part of the LEAP (Learning from Effective Ambulatory Practices) program from the Robert Wood Johnson Foundation. Key findings of these case studies are that practices have expanded the role of existing staff (as seen in the medical assistants-trained-to-screen study and the Swedish study
expanding the role of RNs), and added new personnel (such as the chronic care manager studies) to provide capacity and skills needed to provide care consistent with the patient-centered medical home model. For the care of complex patients most teams have behavioral health specialists, RN care managers, and pharmacists (also noted in the modeling study described above).
## Table 4. Certainty of Evidence

<table>
<thead>
<tr>
<th>Intervention or Outcome</th>
<th>Number of studies</th>
<th>Study limitations</th>
<th>Consistency</th>
<th>Directness</th>
<th>Precision</th>
<th>Other factors</th>
<th>Overall Certainty of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adding NP or other dedicated chronic care manager improving outcomes of some chronic conditions (most notably diabetes, but not hypertension)</td>
<td>3</td>
<td>Serious</td>
<td>No serious inconsistency</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Moderate certainty evidence that nurse-managed protocols result in improvements in multiple outcomes for patients with chronic conditions</td>
<td>Moderate</td>
</tr>
<tr>
<td>Adding NPs as co-management providers increases access</td>
<td>1</td>
<td>Very serious</td>
<td>N/A</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>None</td>
<td>Low</td>
</tr>
<tr>
<td>Retraining medical assistants to screen patients for certain conditions can increase the proportion of patients screened</td>
<td>1</td>
<td>No serious limitation</td>
<td>N/A</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>None</td>
<td>Low</td>
</tr>
<tr>
<td>3.6 to 4.0 FTE of supporting team members are needed for each 1.0 FTE of physician to deliver high-quality comprehensive care; different mixes of skillsets and staffing needed for differing patient populations</td>
<td>1</td>
<td>Serious limitations</td>
<td>N/A</td>
<td>No serious indirectness</td>
<td>No serious imprecision</td>
<td>Modeling study (Limitation: modeling study)</td>
<td>Low</td>
</tr>
</tbody>
</table>
LIMITATIONS

Publication Bias

The primary limitation to this review is the paucity of hypothesis-testing studies on the subject. Only 5 studies had a longitudinal component to their data collection. Cross-sectional studies such as those identified in this review have only a very limited ability to support causal conclusions. A second limitation is the possibility of publication bias. Certainly there must have been more implementations and potentially evaluations of differing team compositions than the published studies we identified. How this publication bias might influence our conclusions is unknown.

Study Quality

As noted, study quality is a major concern for this topic. While some of the studies used a longitudinal design and might be able to support causal relationships, most did not.

Heterogeneity

Heterogeneity is a major concern for this topic. Studies’ interventions most often included multiple components, and these were all idiosyncratic—no study tested the same intervention, in all its components, as any other study. We attempted to group study interventions into categories of interventions that shared some similarities, but nevertheless within each category there is still substantial heterogeneity in interventions.

Applicability of Findings to the VA Population

We found only 1 study in a VA population, and it was about single versus multiple team membership roles, and not about specific team members. We can only speculate as to the applicability of the remaining findings to VA populations. At least 1 of the interventions – nurse-managed protocols that require prescribing authority – is not currently available within VA.

RESEARCH GAPS/FUTURE RESEARCH

VA would seem to be ideally placed to provide experimental evidence about how teamlet and team structures can be optimized. Almost every VA of sufficient size organizes their teamlets into larger units (like Red, Blue, or Yellow teams). This would then allow for controlled comparisons of differing team and teamlet structures, with other contextual features being internally controlled (like senior leadership, incentives, and the EHR). For example, any of the recommended team staffing levels in the model of Meyers for either their “high geriatric” or their “high social needs” models could be implemented in 1 larger team (Red, Green, Blue) while the others serve as control. Teamlet structure could be varied (for example, 1 RN for every 2 physician providers) or team structures could be varied (for example, adding the 1.0 substance abuse counselor). Data collection could come directly from the EHR. Detailed information would need to be collected about patients’ chronic conditions and social needs as the model by Meyers consider these important variables when determining optimal team composition. An agreed-upon metric for evaluating performance – presumably based on the triple aim – would facilitate comparisons of results across studies.
CONCLUSIONS

The evidence for an optimal teamlet or team structure is very sparse. Other than adding a dedicated chronic care manager, there is no evidence above low certainty that any team structure is optimal. Complex patients almost certainly benefit from additional skills (beyond the basic teamlet of provider, medical assistant, and nurse) in the team writ large (such as pharmacist, chronic care manager, etc).
REFERENCES


APPENDIX A. SEARCH STRATEGIES

DATABASE SEARCHED & TIME PERIOD COVERED:
OVID Medline - From inception to 5/29/2020

LANGUAGE:
English

SEARCH STRATEGY:

((patient care team/ or "patient care team*".ti,ab OR "team based".ti,ab OR "team culture".ti,ab OR "team dynamic".ti,ab OR "team function*".ti,ab OR teamwork.ti,ab OR "team work".ti,ab OR "multidisciplinary team*".ti,ab OR "interdisciplinary team*".ti,ab OR "integrated care".ti,ab OR "shared care".ti,ab) AND ("primary care" OR "primary health care").ti,ab OR primary care/) AND (role.ti,ab OR roles.ti,ab OR responsibilit*.ti,ab OR model.ti,ab OR models.ti,ab OR organization.ti,ab OR organization.ti,ab OR arrangement.ti,ab OR structure.ti,ab OR professional roles/))
### APPENDIX B. PEER REVIEWER COMMENTS AND RESPONSES

<table>
<thead>
<tr>
<th>Question</th>
<th>Reviewer comment</th>
<th>Authors Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Are there any published or unpublished studies that we may have overlooked?</td>
<td>Yes - Crawford, E. R., Reeves, C. J., Stewart, G. L., &amp; Astrove, S. L. (2019). To link or not to link? Multiple team membership and unit performance. Journal of Applied Psychology, 104(3), 341–356. <a href="https://doi.org/10.1037/apl0000381">https://doi.org/10.1037/apl0000381</a></td>
<td>We retrieved and evaluated this article and now include it in the review.</td>
</tr>
<tr>
<td>Additional suggestions or comments can be provided below. If applicable, please indicate the page and line numbers from the draft report.</td>
<td>I would classify this work as a very helpful negative study. It is disappointing that the literature did not provide more solid lessons, but the search was comprehensive and the inclusion and exclusion criteria were logical. The comments in the discussion on future research needed are compelling. I am attaching comments within the draft that do not address any major issues but include some suggestions and identification of a few places in which I found the writing difficult to follow.</td>
<td>No response needed</td>
</tr>
<tr>
<td>Page 5, line 23: nice, distilled and clear</td>
<td></td>
<td>No response needed</td>
</tr>
<tr>
<td>Page 5, line 56: All of these studies no matter what the design?</td>
<td></td>
<td>Included studies needed to be hypothesis-testing, but could be randomized or non-randomized. We added “hypothesis-testing” to the text.</td>
</tr>
<tr>
<td>Page 6, line 3: Why not put this as part of the list above? Why not make the language for the list consistent across all numbered inclusion criteria? Ok, I see below where you used the 1-4. I think it would work if you just put the outcome inclusion first, and numbered the others below</td>
<td>In the Executive Summary the full text about includes and excludes has been shortened. This text about the outcomes needed for includes is taken from a longer list in the main report on pages 11-12, but we added this as point 5 to this list</td>
<td></td>
</tr>
<tr>
<td>Page 6, line 9: were</td>
<td></td>
<td>This typo was corrected</td>
</tr>
<tr>
<td>Page 6, line 44: not something the VA RN's on teamlets have? Sentence not clear, certainly VA has RNNPs with prescribing authority</td>
<td></td>
<td>We have clarified this to state it is not currently something that teamlet RNs have authority to do.</td>
</tr>
<tr>
<td>Page 6, line 51: I expected to see the last point (differing patient populations,...) in the table below.</td>
<td></td>
<td>This was added to the table</td>
</tr>
<tr>
<td>Page 8, line 25: It could be helpful to extrapolate on the patient population comment (see my prev comment) if possible. I think that is where we may need to go in thinking about program design, in addition to understanding all the things you said about research gaps.</td>
<td></td>
<td>This additional information was added to the Future Research section</td>
</tr>
<tr>
<td>Page 11, line 51: where is the outcome inclusion? OK, I see it below after excludes. To me, it seems important enough to put above.</td>
<td>We moved this text up.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>Page 12, line 9: Given that there is so little in the literature on this, would it be worthwhile to include more of the literature referenced here for development of a conceptual framework for future research studies?</td>
<td>Although we did not assess this literature in detail, there appears to be a rich qualitative literature describing attributes for good team functioning, and VA may want to consider making this the focus of a new ESP review.</td>
<td></td>
</tr>
<tr>
<td>Page 12, line 17: Same as above, is there anything there that would be useful for framework development?</td>
<td>Same as above.</td>
<td></td>
</tr>
<tr>
<td>Page 12, line 39: were</td>
<td>This typo was corrected</td>
<td></td>
</tr>
<tr>
<td>Page 14, line 14: ok so applied on full text review...</td>
<td>No response needed</td>
<td></td>
</tr>
<tr>
<td>Page 14, line 17: Another possibly interesting article batch, given that a pharmacist is such a major part of VA teams (supposed to be per 6 teamlets)</td>
<td>There is a large body of literature, and existing reviews of portions of that literature, on adding pharmacists to teams. Not all of it appeared to be primary care, and some of it was condition-specific, such as adding a pharmacist to do diabetes medication adjustments. But this might be a fruitful area for another ESP review.</td>
<td></td>
</tr>
<tr>
<td>Page 16, line 39: Meaning that more physicians/team members resulted in less burnout??</td>
<td>That is correct, less burnout. We added this clarification to the text.</td>
<td></td>
</tr>
<tr>
<td>Page 16, line 50: I don't get this--what does assuming care mean? Should it be 1000 patients? Or is it a multi-provider practice? Sounds very interesting.</td>
<td>The model started with the idea that care was going to be delivered to 10,000 patients, and then determined how many MDs, how many RNs, etc. would be needed to deliver high quality care. We have made this clarification to the text.</td>
<td></td>
</tr>
<tr>
<td>Page 18, line 17: Whose burnout? MD's, RNs, clerks?</td>
<td>It is MD burnout. We have clarified this in the table.</td>
<td></td>
</tr>
<tr>
<td>Page 18, line 41: What was the level of variation across these practices? In team composition, or in patient population?</td>
<td>That information is not presented in the article.</td>
<td></td>
</tr>
<tr>
<td>Page 19, line 37: So the initial follow-up was very short. And, fyi, they didn't appear to have attended to depression, which was likely to be high in their demographic and is usually a state upon which improvement in all others depends</td>
<td>This observation is noted, but we did not make any changes to the text. The main point seemed to us to be that once the intervention was taken away, care regressed, making a stronger case that the improvements seen during the intervention were causally linked.</td>
<td></td>
</tr>
<tr>
<td>Page 20, line 12: very interesting.</td>
<td>No response needed</td>
<td></td>
</tr>
<tr>
<td>Page 20, line 17: I'm assuming here in addition to a physician? Does it say what the practice size or number of physicians per NP/PA was?</td>
<td>Yes, this is in addition to having a physician, as the survey was sent to Family Physicians. The panel size varied from about 1900 to 2500, with solo practitioners having larger panels. This was added to the text. The other information is not available.</td>
<td></td>
</tr>
<tr>
<td>Page 20, line 49: very interesting</td>
<td>No response needed</td>
<td></td>
</tr>
<tr>
<td>Page 21, line 7: I think you had a previous comment about RNs prescribing. If this data is what that comment was based on, adjustment of doses is so very different than prescribing ability. Maybe the titrate wording can also be added to the previous comment on prescribing.</td>
<td>Twenty of the 29 studies reported that the nurse was allowed to independently initiate a new medication, (as stated in the text), so 2/3 of the data describe a situation that currently cannot exist in VA. So we left the text as is (other than to qualify that this is currently the case, leaving open the possibility that it could change.)</td>
<td></td>
</tr>
<tr>
<td>Page 21, line 27: Makes me curious about what criteria they failed on...</td>
<td>Those articles are included in the flow and in Appendix C</td>
<td></td>
</tr>
<tr>
<td>Page 25, line 9: very interesting</td>
<td>No response needed</td>
<td></td>
</tr>
<tr>
<td>Page 25, line 58: I wonder if more work on these article sets could be the basis for developing a better model for use in future studies like the ones you suggest.</td>
<td>A good suggestion for possible future ESP work.</td>
<td></td>
</tr>
</tbody>
</table>

This is a typically competent review from the ESP group from Greater Los Angeles that addresses several aspects of an important question: “what is the ideal structure for personnel in primary care?” It is puzzling and to some extent, disheartening, that this question remains unanswered more than 50 years after the “modern” primary care model was introduced. Given that this country employs tens of thousands of health professional in primary care, spends billions of dollars, and achieves less than satisfactory outcomes, it would seem that considerable effort would be devoted to this issue. Unfortunately, as this review delineates, there is still a dearth of high-quality data on the optimal strategy for deploying workforce in the primary care setting.

Consistent with studies conducted in other contexts, this review supports the notion of a nurse care manager as a member of a primary care team. Apart from this, the available studies do not provide much needed, specific information about the ideal composition of the team. Missing

As the reviewer guessed, this information is, in general, not reported in the original research studies.
from the review, and perhaps the studies included, was meaningful description of the clinical settings, including other personnel, characteristics of the clinic populations, IT resources, etc. These are key in understanding how the nurse care managers would function. VA, for example, includes nurse care managers on every team, but in many locations, they function more as ancillary providers, dealing with phone calls and urgent care, rather than true care managers who help to proactively shepherd a panel of patients.

Another key “finding” of the review is that there is no consistency or uniformity in the literature with regard to key outcomes. Each study seems to have defined an idiosyncratic endpoint ranging from screening for specific conditions, to access, to comprehensive primary care. This makes it impossible to compare strategies.

A few studies cited appear to address the critically important cost of increment cost/gain. Adding personnel invariably enhances some output but the decision that faces every manager is whether doing so results in outcomes that could be achieved in a simpler or less expensive manner. Again, however, without an appreciation for the baseline circumstances and resources, it is difficult to generalize these findings.

As outlined by the authors, this review fits into a larger landscape of other literature that addresses more specific questions about team structures. Even so, however, the sum total remains unsatisfactory. There exists no consensus about how best to structure primary care in terms of cost and health outcomes. The authors are wholly correct that VA is in a strong position to address this issue through interventional trials. As the VA budget begins its inevitable cyclical contraction, the question about how to provide excellent care more efficiently would seem a high priority.

<p>| Page 6 line 37: Not clear why the word &quot;Occupations&quot; is added at the end of the question | This was a typo! We have removed it. |
| No response needed | No response needed, but these are very cogent policy points |</p>
<table>
<thead>
<tr>
<th>Page 6 line 42: typo: should be &quot;dedicated&quot; not &quot;dedicate&quot;</th>
<th>This typo has been corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page 7: not clear why there aren't 13 studies in the Certainty of Evidence Table</td>
<td>Some low quality studies don’t contribute to certainty of evidence findings. As noted in the text, “We did not include as “findings” or rate for certainty of evidence conclusions based on results of single studies that were cross-sectional or pre-post in design.”</td>
</tr>
<tr>
<td>Page 8: line 11: Advanced Practice Registered Nurses (APRN) are Licensed Independent Practitioner and have prescribing authority in the VA</td>
<td>The text was changed to indicate that this applies to RNs, not APNs or NPs</td>
</tr>
<tr>
<td>Page 12 Line 32: abbreviation PGS is not spelled out anywhere</td>
<td>Text changed to indicate that this is 1 of the authors</td>
</tr>
<tr>
<td>Page 15: flow chart: how were the initial 3,249 excluded?</td>
<td>This is based on titles that are obviously irrelevant</td>
</tr>
<tr>
<td>Page 17 Line 4: Typo: should be &quot;implemented an&quot; not &quot;implemented and an&quot;</td>
<td>This typo was corrected</td>
</tr>
<tr>
<td>Page 19 line 45: Typo:the word &quot;time&quot; should be removed</td>
<td>This typo was corrected</td>
</tr>
<tr>
<td>Page 24 line 40: should be &quot;dedicated&quot;</td>
<td>This typo was corrected</td>
</tr>
<tr>
<td>Page 24 line 43: see above re prescribing authority for nurses</td>
<td>This clarification was made (to indicate we are talking about teamlet RNs)</td>
</tr>
<tr>
<td>General question: could the Key Question have been revised to be broader in scope in order to obtain a greater number of articles and more information. Addition of grey literature? change in exclusion criteria?</td>
<td>This is a potentially good question for a future ESP review. There is a great deal of qualitative literature, and some quantitative literature, about team functioning that is not about team structure – it could be fruitful to evaluate this in depth.</td>
</tr>
<tr>
<td>Please update my affiliations: Karin Nelson, MD MSHS Director, Primary Care Analytic Team, VHA Office of Primary Care Professor of Medicine, University of Washington Would specify in introduction that PACT is a medical home model. Consider changing &quot;work life of the health care workforce&quot; to &quot;wellbeing...&quot;</td>
<td>This change was made.</td>
</tr>
<tr>
<td></td>
<td>This was added.</td>
</tr>
<tr>
<td></td>
<td>This change was made.</td>
</tr>
<tr>
<td>Thanks!</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Major: Throughout the document there is commentary that nurse prescribing authority is not available in VA. My understanding is that granting RNs this authority is not impossible, just not currently done. This authority may depend on state/regional laws, union positions, etc. Would not use absolute wording implying this is not something VA could consider pursuing.</td>
<td>This issue was also brought up by other reviewers, and changed have been made to indicate that this text is about teamlet RNs, not about APNs or NPs, and that it only describes the current situation, as it might change at some later date.</td>
</tr>
<tr>
<td>Avoid wording stating providers are “dealing with” patients or health problems, it sounds pejorative.</td>
<td>We re-worded this to avoid any perception of pejorative language</td>
</tr>
<tr>
<td>Page 16, line 38–39 – unclear what is meant by “proportion of physician FTE on the team”, this seems important as it is the team structure being tested (Bruhl, 2020)</td>
<td>We’ve added how this was defined by the authors of the article by Bruhl and colleagues</td>
</tr>
<tr>
<td>Page 16, line 46 – it seems important to try to define what this study meant by ‘high-quality, comprehensive care’. Even if the study did not directly define the phrase, can some insight be gained based on what modeling assumptions they used – was there a quantified amount and type of care their models were calculating how to achieve (Meyers, 2018)</td>
<td>We looked hard for a definition and could not find it. We believe this was done by doing site visits at exemplar institutions with a reputation for delivering “high quality, comprehensive care”, and seeing what their staffing ratios are. But the articles never is explicit on this.</td>
</tr>
<tr>
<td>Page 20, lines 23-32 – its not clear what the difference is between ’medical scribe’ strategy and ’ATBC’ strategy, and why the ATBC strategy is twice as ’expensive’. Is ATBC done by a higher level professional? Or does it just require more time on the part of the scribe?</td>
<td>The original article is not more explicit than how these are described in the report, but we believe the difference is that the medical scribe is essentially taking dictation from the attending physician and entering it into the electronic health record. The ATBC strategy has medical assistants meeting with the patient, without the attending doctor being present, and completing a history and doing counseling and entering that information into the electronic health record, where it is briefly reviewed by the attending physician.</td>
</tr>
<tr>
<td>Page 6, Line 17-18: “Thus, there are 3.0 full-time equivalent (FTE) staff for each PCP FTE” This ratio is often not achieved. More accurate to say “The model aims to provide 3.0 full-time equivalent…”</td>
<td>This clarification was made.</td>
</tr>
<tr>
<td>Page 5 Line 17-18: “Thus, there are 3.0 full-time equivalent (FTE) staff for each PCP FTE” This ratio is often not achieved. More accurate to say “The model aims to provide 3.0 full-time equivalent…”</td>
<td>This clarification was made.</td>
</tr>
<tr>
<td>Page 6, Line 37 – Suggest rewording this shorter version of KQ1 to better explain why it says “Occupations?”</td>
<td>The “Occupations” was a typo and has been removed.</td>
</tr>
<tr>
<td>Page 6, line 53 – States pre-post studies were excluded but studies examining practices pre-post new team structures are specified elsewhere as included.</td>
<td>Pre-post studies were included as evidence, but we refrained from drawing conclusions if the only evidence was a single study that was pre-post in design.</td>
</tr>
<tr>
<td>Page 8 – Appreciate examples of larger units for teamlets, suggest using colors such as red/blue/yellow that do not imply a hierarchy (in the current example, gold may be ‘best’, silver ‘second best’, etc)</td>
<td>This change was made.</td>
</tr>
<tr>
<td>Page 14, lines 11 – Clarify what is meant by ‘measurement’ abstract.</td>
<td>We re-classified this study to have as its exclusion criterion that it was not about specific team member roles. It was a study reporting the development of a survey.</td>
</tr>
</tbody>
</table>

Thank you for the opportunity to review this much needed review. Below are some minor suggestions, mostly clarifications, that would strengthen the quality of the reporting:

1. p. 7, Certainty of evidence (COE) table — a total of 13 publications were included in the review, but findings seem to be based only on the 6 studies in the COE table. What about the other 7? If the other 7 were “single studies that were cross-sectional or pre-post in design”, then perhaps an additional row in the table summarizing these 7 is needed. The “missing” studies from the certainty of evidence table are the studies for which we did not deem it appropriate to create rows about their conclusions, as they would have to be rated as “Very Low” in their certainty of evidence, which is tantamount to saying we don’t know whether or not a conclusion is even directionally correct. We don’t think they can be summarized in 1 additional row and don’t believe the deserve a row each. Thus we continue to leave them out of the Certainty of Evidence table, but we added some text to.
<table>
<thead>
<tr>
<th><strong>warranted. Otherwise it is confusing to read about a k of 13 and only see 6 studies discussed in the table. Re: the 1 modeling study, you may also want to clarify that the fact that it is a modeling study decreases certainty (can't really tell from the table).</strong></th>
<th><strong>alert readers as to why the discrepancy between n=13 and n = 6. The modeling study is identified as a modeling study in the current table; we have added that this is a limitation.</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2. Data abstraction (p. 12 line 32): Data extraction was completed by the PI, alone. Granted, this is not unusual in rapid reviews. Nonetheless, what measures were taken to ensure data accuracy given that only 1 person extracted the data (e.g., 10-20% check, structured database dropdowns to ensure consistency)?</strong></td>
<td><strong>The person who was supposed to be the second reviewer for this had to drop out due to COVID related family issues, and thus the review was completed by a single reviewer. There was no ability to do a data check.</strong></td>
</tr>
<tr>
<td><strong>3. GRADE criteria (p. 12 line 52): It's not clear how the domains of risk of bias, imprecision, inconsistency, indirectness, and publication bias are used to arrive at the ratings of high-very low confidence in the effect estimate (quite frankly, the original BMJ papers that present the GRADE criteria aren't much more help). Perhaps walking the reader through an example would help connect the dots for the reader. Along related lines, please also provide more detail around the domains of risk of bias (are you equating this to &quot;study limitations&quot; as stated in the BMJ paper?) imprecision (of what? the effect estimate?) inconsistency (of results), indirectness (of evidence) and reporting bias. You may want to consider adding a little bit of similar detail for the risk of bias criteria discussed in lines 37-45 of page 12.</strong></td>
<td><strong>We agree that applying GRADE to situations other than tightly defined interventions summarized with meta-analytic methods can be challenging. We have added to the text (main body only, not in the Executive Summary) how we worked through the 1 finding that had something about Low certainty evidence, where we also explain how we operationalized each domain.</strong></td>
</tr>
<tr>
<td><strong>4. Figure 1. I assume the &quot;studies of adding other members to the team&quot; were excluded? It might be more obvious to the reader to simply add &quot;(excluded)&quot; at the end of the title of that box that's just off on a corner by itself.</strong></td>
<td><strong>These were all excluded at the abstract phase, so we have added these numbers to the box in the flow about reasons for abstract rejection.</strong></td>
</tr>
<tr>
<td><strong>5. Description of the Evidence (p. 16, lines 4-20). I understand that paragraph form is the traditional way of reporting this information. But I've never found it very useful, because you're describing each dimension separately. So you don't know, for example, if the 1 England study was also the 1 RCT. I think this descriptive information will come across far better in a simple table. With only 13 studies, you can just give the entire dataset (sorted by the most</strong></td>
<td><strong>This table has been added.</strong></td>
</tr>
<tr>
<td><strong>Team-Based Primary Care Evidence Synthesis Program</strong></td>
<td><strong>Evidence Synthesis Program</strong></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td><strong>meaningful variable rather than alphabetically by author).</strong></td>
<td><strong>When we are assessing risk of bias we don't as a rule &quot;grade on a curve&quot; – a low response rate is a treated as an absolute, rather than how it fares relative to other similar studies. In this particular case, Bruhl and colleagues used an online survey instrument sent to 420 physician or NP/PAs assigned to Family Medicine care teams in 1 large, multi-state health system. Only just over half of providers responded. While the authors did present some basic demographics on the non-responders (somewhat more males, smaller teams, more physicians), in our view this is not nearly enough information on the non-responders to conclude that their burnout scores would be about the same as the responders, and thus non-response bias is a legitimate concern about the findings in this study, We did not make any change to the response rate critique.</strong></td>
</tr>
<tr>
<td><strong>6. Response rate critique (p. 16 line 41). The Bruhl and colleagues study is critiqued for having a &quot;low response rate&quot;, and this is cited as a limitation. 52% is considerably above average. Anseel et al. (2010) did a review and meta-analysis of survey study response rates and found that for surveys of employees in organizations (both web-based and paper-and-pencil) the median response rate is 41% (average 43%), and that rate decreases considerably the busier the type of respondent is (e.g., median response rate for top leadership is 31%). Further, if the sample is representative, the response rate is not nearly as much of an issue. Suggest reconsidering the impact of the response rate on the overall quality rating of the study.</strong></td>
<td><strong>Are the findings presented in a way that is helpful for decision-making? Do you have any recommendations on how this report can be revised to more directly address or assist with implementation?</strong></td>
</tr>
<tr>
<td><strong>7. Appendix C. Only 63 references are listed in the excluded studies list. Yet Figure 1 (Flow diagram) identifies 201 studies that were excluded (129 at the abstract level, 72 at the full text level). Suggest listing all 201 in the appendix, organized by reason for exclusion at each stage (i.e., just like in Figure 1).</strong></td>
<td><strong>I wonder if the helpfulness of the findings could have been enhanced by broadening the question or changing the exclusion criteria in order to increase the return of included publications. This might provide more evidence about the structure of teams and any outcomes related to different structures.</strong></td>
</tr>
<tr>
<td><strong>The abstract rejects have now been added to this.</strong></td>
<td><strong>Nice presentation</strong></td>
</tr>
<tr>
<td><strong>We can’t change the key question at this stage, but exploring the qualitative and quantitative literature for information about what makes teams effective teams might be a promising topic for a new ESP review.</strong></td>
<td><strong>No response needed.</strong></td>
</tr>
</tbody>
</table>
APPENDIX C. CITATIONS FOR EXCLUDED PUBLICATIONS

Abstract Excludes

Not Research (n=27)


Not about Specific Team Member Roles in TBPC (n=24)

7. Hall, T.L., et al., Understanding adaptations to patient-centered medical home activities:


Adding Pharmacists (n=22)


Not about Team-based Primary Care (n=22)

11. Kilpatrick, K., et al., Implementing primary healthcare nurse practitioners in long-term


Qualitative Study (n=17)


No Relevant Outcomes (n=8)
5. Quinlan, E. and S. Robertson, The communicative power of nurse practitioners in


No Abstract (n=6)


Adding Occupational Therapists (n=1)


Descriptive Study (n=1)


Study Protocol (n=1)


Full Text Excludes

Not about Specific Team Member Roles in TBPC (n=32)

Not about Team-based Primary Care (n=10)

Adding Pharmacists (n=8)

2. Dennis, S., et al., What evidence is there to support skill mix changes between GPs, pharmacists and practice nurses in the care of elderly people living in the community? Australia & New Zealand Health Policy, 2009. 6: p. 23.
5. Santschi, V., et al., Improving blood pressure control through pharmacist interventions: a


No Outcome of Interest (n=7)


Qualitative Study (n=4)


3. Laing, B.Y., et al., Introducing the "teamlet": initiating a primary care innovation at san

Unavailable (n=4)


Not Research (n=2)


Background (n=1)


Descriptive Study (n=1)


Not Primary Care (n=1)


Non-Systematic Review (n=1)

Study Protocol (n=1)

## APPENDIX D. RISK OF BIAS TOOL

### Risk of Bias Assessment for Included RCT

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Random sequence generation</th>
<th>Allocation concealment</th>
<th>Blinding of participants and personnel</th>
<th>Blinding of outcome assessment</th>
<th>Incomplete outcome data</th>
<th>Selective reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Litaker, 2003</td>
<td>Unknown</td>
<td>Unknown</td>
<td>High risk</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
</tbody>
</table>

### Risk of Bias for Included Observation Studies with Control

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Confounding bias</th>
<th>Selection bias</th>
<th>Bias in measurement classification of interventions</th>
<th>Bias due to deviations from intended interventions</th>
<th>Bias due to missing data</th>
<th>Bias in measurement of outcomes</th>
<th>Bias in selection of reported results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ohved, 2000</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
<tr>
<td>Dorr, 2006</td>
<td>High risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
<td>Low risk</td>
</tr>
</tbody>
</table>