QUERI

Challenges and Opportunities for Pay-for-Performance as Veteran Care Moves into the Community

October 2017

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

Prepared by:
Evidence-based Synthesis Program (ESP)
Portland VA Health Care System
Portland, OR
Devan Kansagara, MD, MCR, Director

Investigators:
Principal Investigator:
Karli Kondo, PhD

Co-Investigators:
Jessica Wyse
Aaron Mendelson
Gabriella Beard
Allison Low
Michele Freeman
Devan Kansagara

VA HEALTH CARE Defining EXCELLENCE in the 21st Century

PREFACE

The VA Evidence-based Synthesis Program (ESP) was established in 2007 to provide timely and accurate syntheses of targeted healthcare topics of particular importance to clinicians, managers, and policymakers as they work to improve the health and healthcare of Veterans. QUERI provides funding for 4 ESP Centers, and each Center has an active University affiliation. Center Directors are recognized leaders in the field of evidence synthesis with close ties to the AHRQ Evidence-based Practice Centers. The ESP is governed by a Steering Committee comprised of participants from VHA Policy, Program, and Operations Offices, VISN leadership, field-based investigators, and others as designated appropriate by QUERI/HSR&D.

The ESP Centers generate evidence syntheses on important clinical practice topics. 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 ESP disseminates these reports throughout VA and in the published literature; some evidence syntheses have informed the clinical guidelines of large professional organizations.

The ESP Coordinating Center (ESP CC), located in Portland, Oregon, was created in 2009 to expand the capacity of QUERI/HSR&D and is charged with oversight of national ESP program operations, program development and evaluation, and dissemination efforts. The ESP CC establishes standard operating procedures for the production of evidence synthesis reports; facilitates a national topic nomination, prioritization, and selection process; manages the research portfolio of each Center; facilitates editorial review processes; ensures methodological consistency and quality of products; produces “rapid response evidence briefs” at the request of VHA senior leadership; collaborates with HSR&D Center for Information Dissemination and Education Resources (CIDER) to develop a national dissemination strategy for all ESP products; and interfaces with stakeholders to effectively engage the program.

Comments on this evidence report are welcome and can be sent to Nicole Floyd, ESP CC Program Manager, at Nicole.Floyd@va.gov.

Recommended citation: Kondo K, Wyse J, Mendelson A, Beard G, Low A, Freeman M, Kansagara D. Challenges and Opportunities for Pay-for-performance as Veteran Care Moves into the Community. VA ESP Project #05-225; 2017.

This report is based on research conducted by the Evidence-based Synthesis Program (ESP) Center located at the **Portland VA Healthcare System, Portland, Oregon**, funded by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Quality Enhancement Research Initiative. 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.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

INTRODUCTION

Pay-for-performance (P4P) is commonly used in the Veterans Health Administration (VHA) system, and is expected to be an important strategy to incentivize quality and appropriate utilization as Veteran care moves into the community. The purpose of the current project is to 1) assess the effects of pay-for-performance programs on the quality of care and health of Veterans, 2) identify potential unintended consequences of pay-for-performance programs targeting Veteran health, 3) identify performance metrics that have been incentivized in published P4P literature, 4) identify the program design features and implementation factors that might modify the effectiveness of P4P targeting Veteran populations, both in VHA settings and in the community, and 5) identify novel P4P approaches in VHA settings and Veterans Affairs (VA)-funded research examining P4P or related program features or implementation factors.

METHODS

Data Sources and Searches

We identified studies from a previous ESP review on P4P, as well as from a targeted search of known VA P4P and quality improvement researchers. In addition, we conducted an update search of PubMed, PsycINFO©, and CINAHL© (January 2014 to March 2017). We used snowball sampling to identify additional studies and novel approaches currently being tested or implemented in the VHA.

We included English-language studies of P4P programs targeting healthcare providers at the individual, group, managerial, or institutional level in VHA or Veterans Choice Program (VCP) settings. To better understand factors that might contribute to successful P4P programs for Veterans both in VHA settings and in the community, we interviewed 17 key informants (KIs). KIs had extensive P4P research or administrative experience, and knowledge of the VHA health system. Using conventional content analysis to guide protocol development, we drafted a semi-structured interview that was informed by themes identified in our previous P4P review, which also allowed for new themes and concepts to emerge. Interviews averaged 60 minutes, were led by 2 investigators, and were conducted by phone.

We qualitatively synthesized and organized the results of included studies and key informant interviews according to an implementation framework that describes the relationship between the features of P4P programs, external factors, implementation factors, and provider cognitive/affective and behavioral responses on processes of care and patient outcomes (see Figure 1 in the main report).

RESULTS

Results of Literature Search

We included 68 articles representing 62 studies, from 1,031 titles and abstracts. We identified 23 relevant VA-funded projects, programs, and initiatives.

Summary of Results for Key Questions

Key Question 1. What are the effects of pay-for-performance programs on the quality of care and health of Veterans?

We found insufficient evidence to draw firm conclusions about P4P's effectiveness in VHA settings. One RCT found that the combination of audit and feedback and physician-directed incentives resulted in a small, short-term positive effect on blood pressure control, but incentives directed at the practice or physician and practice were not associated with improved outcomes. 2 observational studies report evidence of positive effects on processes of care. However, it is possible that the findings of these studies may have been influenced by concomitant public reporting and denominator management. Table 2 in the main report provides study-level detail.

Key Question 2. In Veteran populations, what are the potential unintended consequences of pay-for-performance in healthcare?

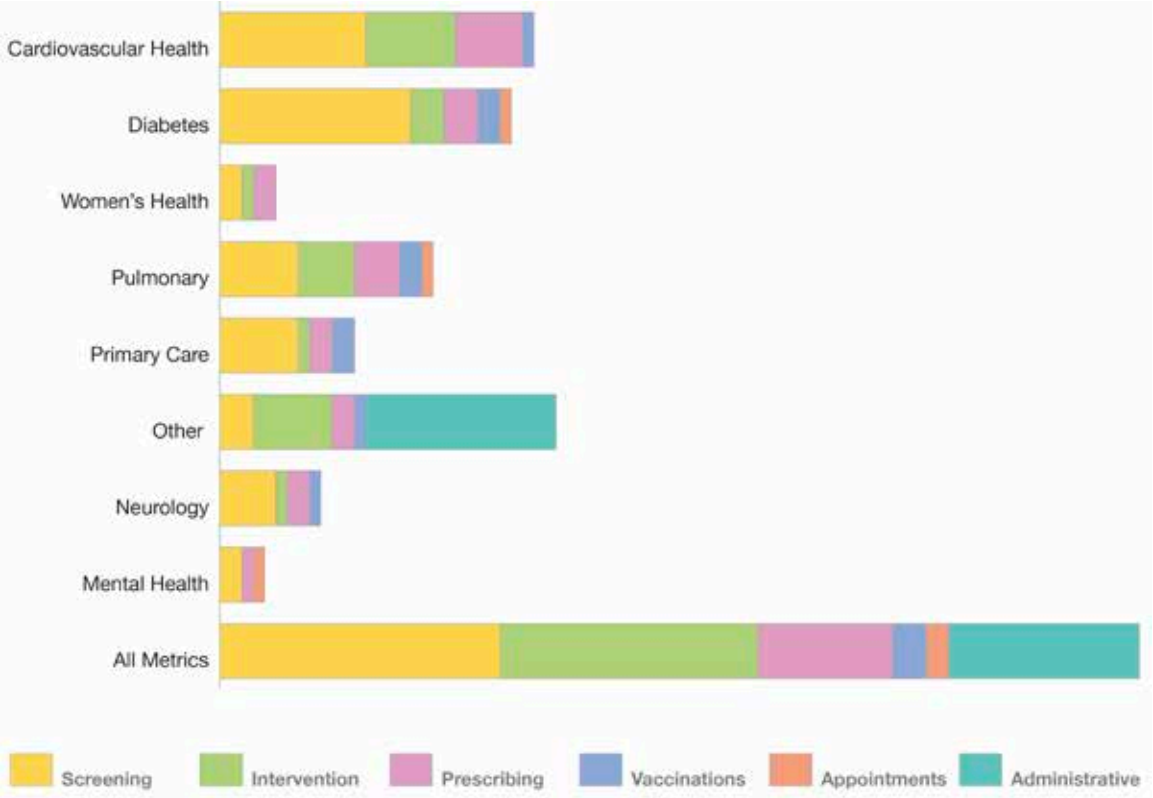
Thirteen articles from 11 studies examined potential unintended consequences associated with pay-for-performance in VHA settings. In general, studies using administrative data and qualitative studies of VHA providers and leaders support the potential for overtreatment associated with performance metrics. However, the sole RCT of P4P specifically found no association between P4P for hypertension and hypotension. Furthermore, a qualitative sub-study of the same RCT found that despite no evidence of hypotension, a number of study participants reported concern for potential overtreatment. Other studies found evidence of denominator management associated with a VISN Director-aimed incentive, and no evidence of risk selection. Qualitative studies explored provider perceptions of both negative and positive unintended consequences associated with performance metrics. Tables 3 and 4 in the main report provide study-level detail.

Key informants were concerned about potential overtreatment, as well as denominator management, risk selection/health disparities, teaching to the test/attention shift, and gaming (see Figure 3 in the main report).

Key Question 3. What metrics have been commonly incentivized in published literature examining P4P?

Across 39 studies, we identified 82 process of care or administrative metrics, and 10 patient outcome metrics (*ie*, intermediate and health outcomes). Tables 5, 6, and 7 in the main report provide a detailed tally of the measures. Metrics most commonly targeted cardiovascular health and diabetes, followed by pulmonary conditions and primary care. Screenings were the most common type of incentivized metric, followed by interventions/procedures, and prescribing. Metrics classified as "other" were predominantly administrative in nature (*eg*, trainings and EHR use). Very few patient outcome metrics were reported in published P4P research. The following figure illustrates the relative proportion of metric types examined in the P4P literature.

Figure. Incentivized Process of Care and Administrative Metrics Reported in Published Literature by Condition and Type



Key Question 4. In Veteran populations, what program features and implementation factors modify the effectiveness of pay-for-performance programs?

VHA Settings

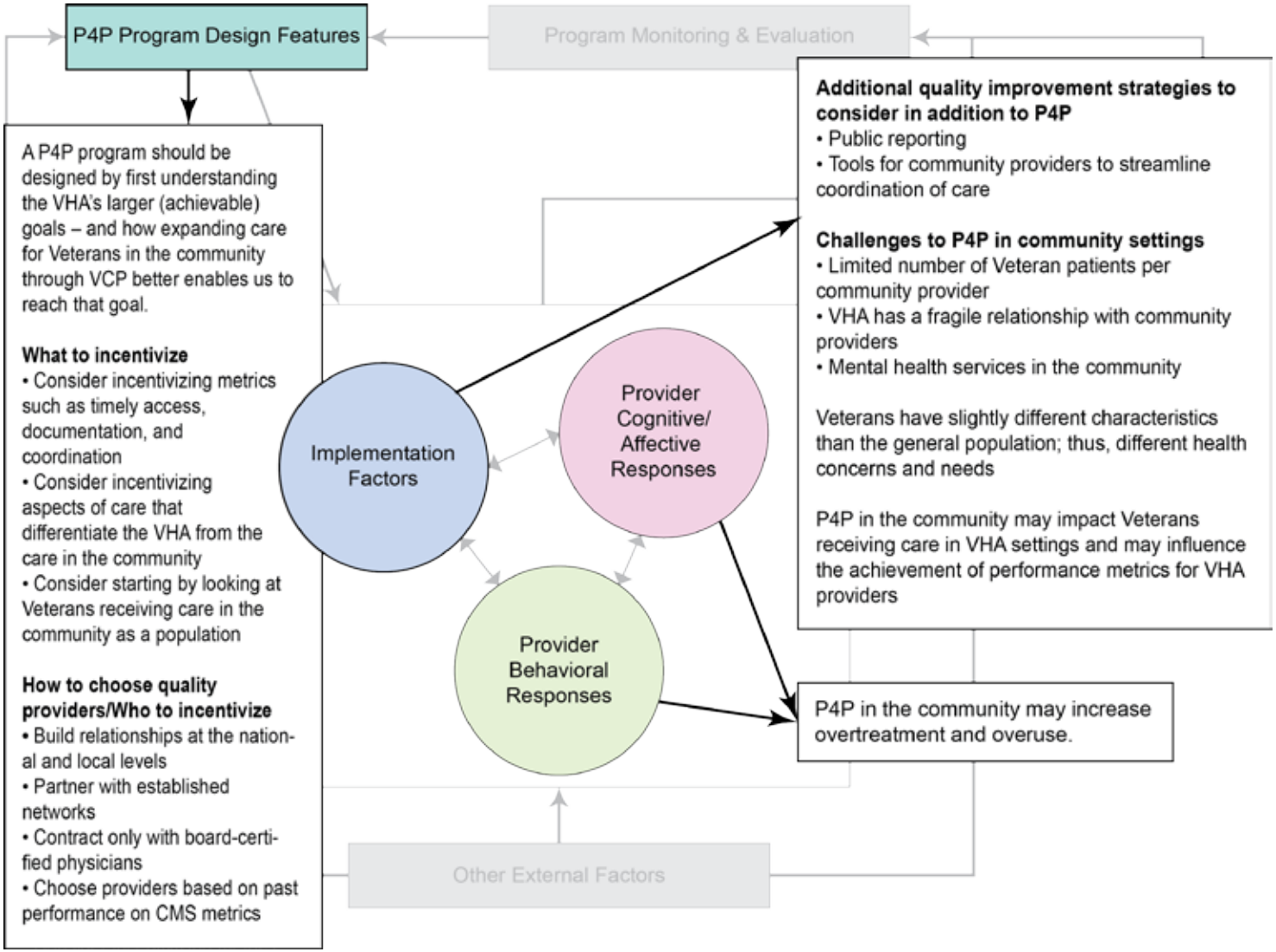
Sixteen articles from 13 studies provide data examining program factors or implementation factors of pay-for-performance programs in VHA settings. In general, studies examining program design features found physician-targeted incentives to be more effective than those targeting groups/practices, that the degree of agreement between EHR data and manual review varied by metric, that the relationship between access metrics and patient satisfaction varied by access metric and whether the patient was new or returning, and that the difficulty of achieving multi-tasked metrics was not directly related to the number of tasks involved. Studies examining implementation processes found no difference in the achievement of actively versus passively monitored metrics, provide mixed evidence related to the impact of the removal of incentives on performance, found a relationship between high-performing facilities and a timely individualized audit and feedback process, and suggest areas of improvement for implementing performance metrics at the local level. One study examined provider affective/cognitive responses, and found that P4P had no impact on goal commitment. Tables 8 and 9 in the main report provides study level detail.

Themes from KI interviews focused on incentive structure, the validity and achievability of performance metrics, and creating an organizational culture that fosters learning and quality (see Figure 5 in the main report).

In Community Settings

Five studies examined P4P or related design features or implementation factors in Veteran populations in community settings. In general, studies found that a number of survey instruments examining cross-system access and coordination exist, and that Veterans, providers, and administrators expressed concern that VCP had resulted in fragmented care for patients, poor communication and coordination amongst providers, and that it placed an additional burden on VHA providers. Other concerns included barriers to sharing medical records, and differences between providers interested in VCP and those who are not. Table 10 in the main report provides study-level detail. The figure below illustrates themes related to P4P in community care, identified through key informant interviews.

Figure. Key Informant Interviews: Themes – P4P in Community Settings



Note. Implementation Factors include implementation processes; outer setting; inner setting; and provider characteristics. Abbreviations: CMS = Centers for Medicare and Medicaid Services; P4P = pay-for-performance; VCP = Veterans Choice Program; VHA = Veterans Health Administration.

Key Question 5. What novel approaches and/or current or recently closed research projects funded by VA examine the effectiveness, implementation factors, or unintended consequences associated with pay-for-performance in Veteran populations?

We identified no novel approaches to P4P being tested in clinical settings in the VHA. However, we did identify 23 current and recently closed (2016 – present) projects, initiatives, and programs funded by VA (see Table 11 in the main report). To our knowledge, only the Partnered Evidence-Based Policy Resource Center (PEPReC) is currently engaged in work directly related to P4P. Along with the Office of Community Care, they are developing performance standards for P4P in the community, and in addition, are performing a randomized evaluation of a P4P program to improve outcomes related to opioid use in Veterans in community care settings. All identified Quality Enhancement Research Initiative (QUERI) activities and one additional project relate to community care.

SUMMARY OF FINDINGS

We examined 68 articles and conducted interviews with 17 key informants to help inform the implementation of pay-for-performance programs for Veterans in the VHA and in community settings. While we found insufficient evidence to determine whether and how much P4P affects Veteran outcomes, we did find information in the literature and through KI interviews that may help guide the implementation of P4P and maximize potential benefits while minimizing negative unintended consequences.

Several themes related to general issues with P4P in VHA emerged from key informant interviews that are consistent with the findings from published literature (see Table 12 in the main report):

- *Regardless of whether performance metrics are incentivized, they should be valid, achievable, and within a provider's control.*
- *Potential overtreatment and overuse may be an unintended consequence of performance metrics, and de-intensification metrics should be considered.*
- *Consider re-evaluation of the size (monetary), frequency, and target (provider vs team) of performance pay in the VHA.*
- *Use a transparent, bottom-up approach for selecting and implementing metrics, and secure provider and staff buy-in.*
- *Foster overall and local-level cultures that encourage learning and value quality improvement.*
- *Gaming will likely be mitigated by providing the resources support necessary for achievement.*

A number of themes related to the design and implementation of P4P in community settings also emerged (see Table 13 in the main report).

- *Initially target areas in need of improvement such as documentation and coordination (eg, receipt of records from community providers).*
- *Develop relationships with providers and health systems with records of strong performance on commonly used, well-validated, and well-established metrics.*
- *The likely small number of Veteran patients per community provider may pose a challenge, both in terms of accurately assessing quality and the potential for an incentive to influence behavior. Consider beginning with alternate approaches, such population-based incentives.*
- *Use strategies such as public reporting to complement P4P.*
- *Developing tools and resources to streamline the data-sharing and coordination necessary to inform a cross-system P4P program.*
- *Consider how funding expanded care in the community might affect funding for Veterans receiving care in VHA settings.*
- *Consider how performance by community providers might impact measured performance for VHA providers.*
- *Be vigilant for overtreatment and for differences in standards of care (eg, opioid prescriptions).*

Conclusions

The effectiveness of pay-for-performance in the VHA settings has been largely understudied, but we highlight a number of key lessons learned from the implementation of programs that may help guide future P4P program improvements in the VHA. In P4P programs targeting Veteran health in community settings, care should be taken to establish relationships with providers with track records of quality; consideration should be given to the impact of the small number of Veterans per community provider; efforts should be made to develop resources and tools to better enable coordination of care, data-sharing, and record transfer; and special attention should be paid to mitigate the potential for overtreatment and ensure quality care for all Veterans.

ABBREVIATIONS TABLE

AA	African American
ACEI	Angiotensin converting enzyme inhibitor
ACG	Adjusted Clinical Group
ACS	Acute Coronary Syndrome
AMI	Acute Myocardial Infarction
AP-EHR	Automatic processing electronic health record
ARB	Angiotensin II receptor blockers
BMI	Body mass index
BP	Blood pressure
CFIR	Consolidated Framework for Implementation Research
CKD	Chronic Kidney Disease
CMS	Centers for Medicare and Medicaid Services
CoC	Community of Care
COPD	Chronic Obstructive Pulmonary Disease
DCG	Diagnostic Cost Group
ED	Emergency Department
EHR	Electronic health record
FOBT	Fecal Occult Blood Test
FY	Fiscal year
HbA1C	Hemoglobin A1C
HDL	High-density lipoprotein
HCV	Hepatitis C Virus
HF	Heart Failure
HSR&D	Health Services Research and Development
HWR	Hospital wide readmission
KI	Key informant
KQ	Key question
LARC	Long acting reversible contraception
LDL	Low-density lipoprotein
LTC	Long term care
MACRA	Medicare Access and CHIP Reauthorization Act of 2015
MDD	Major Depressive Disorder
MIPS	Merit-Based Incentive Payment System
NA	Not applicable
NR	Not reported
NRCT	Non-randomized controlled trial
ORD	VA Office of Research Development
P4P	Pay-for-performance

PACT	Patient Aligned Care Team
PCI	Percutaneous coronary intervention
PM	Performance metric
PC3	Patient-Centered Community Care
PEPReC	Partnered Evidence-Based Policy Resource Center
ProMES	Productivity Measurement and Enhancement System
QUERI	Quality Enhancement Research Initiative
RCT	Randomized controlled trial
SES	Socioeconomic status
SHEP	Survey of Healthcare Experiences of Patients
SGOT	Serum glutamic-oxaloacetic transaminase
SUD	Substance Use Disorder
TB	Tuberculosis
TEP	Technical Expert Panel
VA	Veterans Administration
VHA	Veterans Health Administration
VISN	Veterans Integrated Service Networks

EVIDENCE REPORT

INTRODUCTION

In pay-for-performance (P4P) programs, a portion of payments to providers, administrators, or health systems is linked to achievement of specific access to care, process of care, or patient outcome benchmarks. This strategy has become widespread in the Veterans Health Administration (VHA) since it was codified by law over a decade ago.[1] Centers for Medicare and Medicaid Services (CMS)' Merit-Based Incentive Payment System (MIPS) under the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) similarly heralds the establishment of P4P as a foundational strategy for health reform in the community.[2]

Even though P4P makes intuitive sense as a strategy to increase health care value, the empiric data are far from clear. In 2015, we completed a systematic review and key informant (KI) interviews broadly examining the benefits and harms associated with P4P in healthcare. We found that, while P4P programs may be associated with improved processes of care in ambulatory settings over the short-term, there was no consistent evidence of an effect on health outcomes. Through both the literature and KI interviews, we identified a number of potential unintended consequences associated with P4P, such as attention shift, gaming, and exacerbation of health care disparities; however, there were very few empiric data to firmly establish if and how commonly these occur. Information from both key informant interviews and a handful of studies suggested that P4P's balance of benefits and harms likely depends heavily on the nuances of program implementation.[3-5]

In 2016, there were 25.5 million Veteran appointments in the community.[6] In coming years, more and more Veterans are expected to receive care paid for by the VHA but provided in the community. The Commission on Care recommended that these payments be based on contemporary P4P strategies that incentivize quality and appropriate utilization.[7] In other words, the challenge is to integrate payment and care from the nation's largest health care system to a broad and diverse patchwork of community providers and health systems in a transparent and clinically meaningful way for a unique population of patients.

The purpose of the current project is to summarize the literature examining the benefits and harms of P4P specifically for Veterans, and to gain insight from experts in both P4P and Veteran care about the potential opportunities and pitfalls of P4P integration across health systems.

METHODS

TOPIC DEVELOPMENT

The key questions guiding this report were developed in collaboration with Clinton Greenstone MD, Deputy Clinical Director of Community Integration, VHA Office of Community Care, and with experts in the field:

Key Question 1: What are the effects of pay-for-performance programs on the quality of care and health of Veterans?

Key Question 2: In Veteran populations, what are the potential unintended consequences of pay-for-performance in healthcare?

Key Question 3: What metrics have been commonly incentivized in published literature examining P4P?

Key Question 4: In Veteran populations, what program features and implementation factors modify the effectiveness of pay-for-performance programs?

Key Question 5: What novel approaches and/or current or recently closed research projects funded by VA examine the effectiveness, implementation factors, or unintended consequences associated with pay-for-performance in Veteran populations?

SEARCH STRATEGY

We identified primary evidence examining P4P in Veteran populations and performance metrics reported in published literature from the studies identified in our prior review,[3] from an update search and a targeted search of known VA P4P and quality improvement researchers (PubMed, PsycINFO©, and CINAHL© [January 2014 to March 2017] see Appendix A for search strategy), and from a search of the VHA's website for unpublished studies. To identify current and recently closed studies funded by the VA Office of Research and Development (ORD), we searched the Health Services Research and Development (HSR&D) and Quality Enhancement Research Initiative (QUERI) websites. We used snowball sampling to identify additional studies and novel approaches currently being tested or implemented in the VHA, starting with members of our Technical Expert Panel (TEP; see Appendix B), VHA leadership, and Veterans Integrated Service Networks (VISN) Directors.

STUDY SELECTION

We included direct pay-for-performance programs targeting healthcare providers at the individual, group, managerial, or institutional level in VHA and Veterans Choice Program (VCP) settings (see study selection criteria in Appendices C and D). We excluded studies examining patient-targeted financial incentives, as well as payment models other than direct pay-for-performance (*eg*, managed care, capitation, bundled payments, and accountable care organizations). To assess the effectiveness of pay-for-performance on utilization, quality of care, and patient health outcomes (KQ1), we included all randomized and non-randomized controlled trials (RCTs and NRCTs), and observational studies that either a) had a comparison group, b) had 3 or more time points and reported a trend (*eg*, interrupted time series), or c) included 10,000 or more participants (*eg*, cross-sectional and uncontrolled before-after studies). All quantitative and qualitative study designs were included for key questions 2, 3, 4, and 5. To inform questions 2, 4, and 5, in addition to studies specifically examining performance pay, we also included relevant studies examining processes/variables that occur both upstream (*eg*, performance measures) and downstream (*eg*, audit and feedback) of P4P.

Two independent reviewers independently assessed studies for inclusion, and all discordant results were resolved through consensus or consultation with a third reviewer.

KEY INFORMANTS

We interviewed 17 individuals with extensive P4P research or administrative experience who were familiar with the VHA health system (see Appendix B for a list of key informants). The purpose of the interviews was to better understand the program features and implementation factors that might contribute to successful P4P programs in both VHA settings and those targeting Veteran care in the community.

We used conventional content analysis to guide protocol development and the analytic process.[8] We drafted a semi-structured interview that probed previously identified themes[3,4] which also allowed new themes and concepts to emerge. In addition, we asked specifically about anticipated challenges in constructing a P4P program guiding community care for Veterans (see Appendix E). All interviews were conducted by phone, were co-lead by 2 investigators, and averaged 60 minutes. Interviews were recorded and transcribed verbatim.

DATA ABSTRACTION

Data from each study was abstracted by one investigator and confirmed by a second. From each study, we abstracted data related to study design, sample size, observation period, program focus, the incentive (target, size, timing), comparison, implementation factors, unintended consequences, and findings. For key question 3, we abstracted incentivized metrics from identified studies and removed duplicates by program (*eg*, Quality and Outcomes Framework).

QUALITY ASSESSMENT

Two investigators independently assessed study quality using the Cochrane Risk-of-Bias tool[9] for RCTs and the Newcastle-Ottawa Scale[10,11] for observational studies (see Appendix F). We did not assess the quality of qualitative studies.

DATA SYNTHESIS AND ANALYSIS

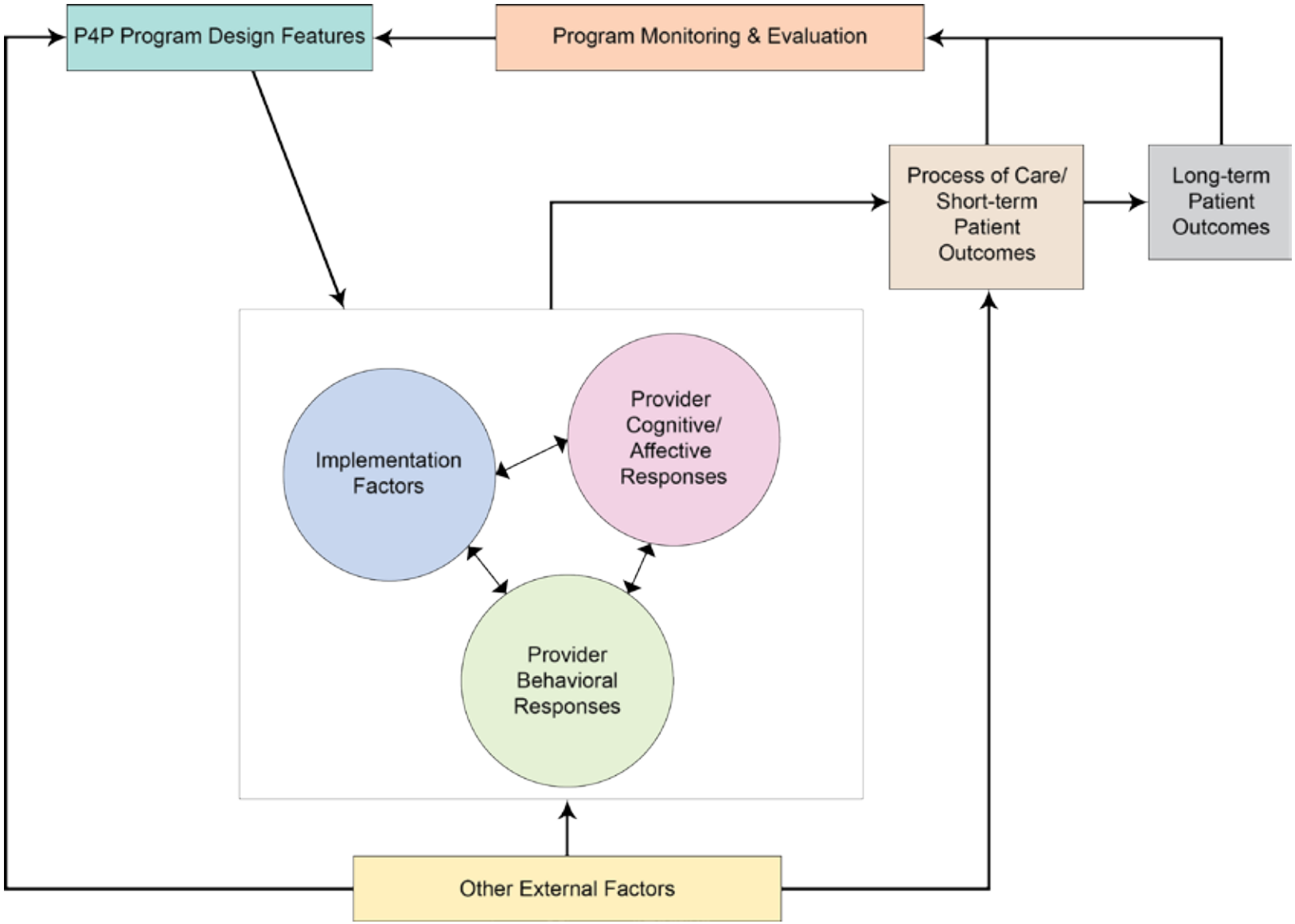
To analyze key informant data, 4 members of the team individually read each interview transcript and identified emergent themes and categories, which were then discussed with the full group. Following discussion, the group came to consensus regarding specific themes and categories to analyze further. A member of the research team reviewed all interview transcripts and tagged all quotes linked with identified themes and categories with a keyword, or "code." All quotes identified by the same code were then compared and contrasted, both across and within interviews, by 2 members of the research team. Through this process, key themes and findings were refined and elaborated.

We qualitatively synthesized the results of included studies and key informant interviews into an implementation framework developed in consultation with the Technical Expert Panel (TEP) for the 2015 ESP pay-for-performance review,[3] which was based largely on the Consolidated Framework for Implementation Research (CFIR).[12] The framework describes the relationship between the features of P4P programs, external factors, implementation factors, and provider cognitive/affective and behavioral responses – on processes of care and patient outcomes (see Figure 1). Table 2 describes each framework category. Due to heterogeneity among the studies, meta-analysis was not performed.

We separated incentivized metrics from published research into patient outcomes (*ie*, intermediate, health, and evaluations of care) and processes of care (including administrative metrics), and organized them broadly by clinical condition (*ie*, cardiovascular health, diabetes, neurology, mental health, pulmonary, women's health, and other). We further categorized processes of care metrics by type (*ie*, administrative, appointments, interventions/procedures, prescribing, screening, and vaccinations).

We briefly describe identified novel P4P programs, and provide a list of identified VA-funded ongoing and recently closed research examining P4P and related implementation factors.

Figure 1. Conceptual Framework



Note. Implementation Factors include implementation processes; outer setting; inner setting; and provider characteristics.

Table 1. Description of Implementation Framework Categories

Framework Category		Description
Program Design Features		Properties of the intervention itself such as the type of performance metric used or the size of the financial incentive
Implementation Factors	Implementation Processes	Actions taken to implement the P4P program such as planning, stakeholder engagement, academic detailing, audit and feedback, and whether the incentive was targeted at the team or individual level.
	Outer Setting	Refers to the broader health system context within which an intervention is implemented; the cultural and social norms at the state and federal level; and characteristics of the patient population.
	Inner Setting	Refers to characteristics of the institution or organization itself.
	Provider Characteristics	Refers to demographic characteristics (*eg,* age, gender, race/ethnicity), as well as other factors such as experience and specialization.
Provider Cognitive/Affective and Behavioral Responses		Refers to provider beliefs and attitudes. Includes cognitive response constructs such as biases, professionalism, heuristics, identification with one's organization. Also includes behavioral response constructs such as risk selection, gaming, systems improvement responses.
Process of Care and Short-term Patient Outcomes		Includes process of care outcomes such as performance of recommended screening or disease monitoring, as well as patient outcomes such as achieving target disease management goals (*eg,* blood pressure, cholesterol levels) and health outcomes.

RATING THE BODY OF EVIDENCE

We assessed the overall strength of evidence for the effectiveness of P4P and Veteran care (KQ1) using a method developed for the Agency for Healthcare Research and Quality's (AHRQ) Evidence-based Practice Centers (EPCs).[13] Ratings were based on the following criteria:

- High = Very confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has few or no deficiencies, the findings are stable, and another study would not change the conclusions.
- Moderate = Moderately confident that the estimate of effect lies close to the true effect for this outcome. The body of evidence has some deficiencies and the findings are likely to be stable, but some doubt remains.
- Low = Limited confidence that the estimate of effect lies close to the true effect for this outcome. The body of evidence has major or numerous deficiencies (or both). Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect.
- Insufficient = No evidence, unable to estimate an effect, or no confidence in the estimate of effect for this outcome. No evidence is available or the body of evidence has unacceptable deficiencies, precluding reaching a conclusion.

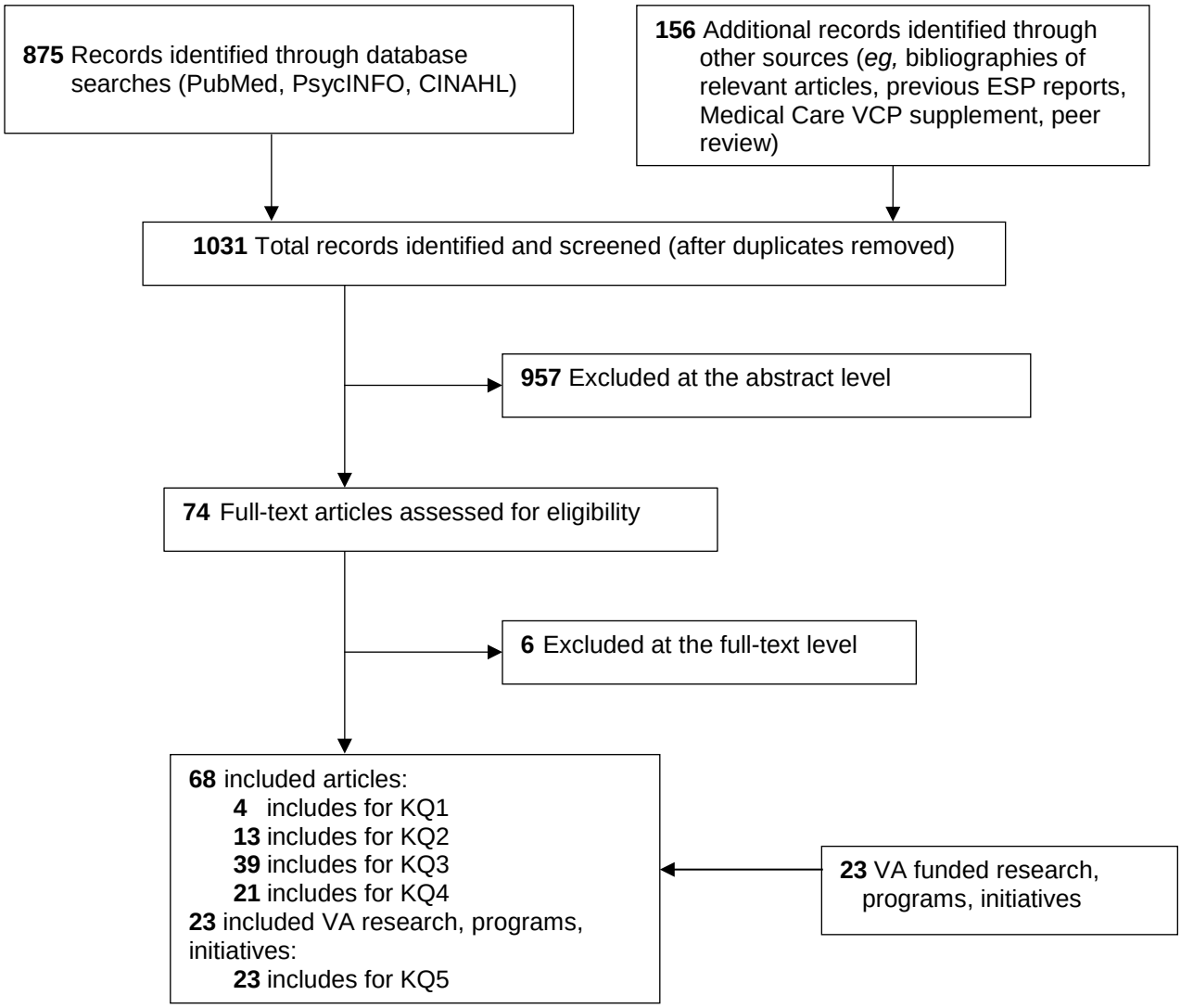
We did not assess the strength of evidence for other questions addressed in this report, given the heterogeneity inherent in research examining unintended consequences and implementation.

RESULTS

LITERATURE FLOW

We reviewed 1,031 titles and abstracts, and examined the full text of 74. We included 68 articles representing 62 studies; 7 articles informed more than one key question.[14-20] In addition, we identified 23 VA-funded research projects, programs, and initiatives working on topics related to P4P (see Figure 2).

Figure 2: Literature Flow Chart



Note. Seven articles provided data for more than one key question. Abbreviations: ESP = Evidence-based Synthesis Program, KQ = key question, VCP = Veterans Choice Program.

KEY QUESTION 1: What are the effects of pay-for-performance programs on the quality of care and health of Veterans?

Four articles[14-17] from 3 studies[14,15,17] provide data on the effects of P4P in VHA settings (see Table 2). Although the evidence shows promise, it is insufficient to draw firm conclusions about P4P's effectiveness for the improvement of the quality of care or health of Veterans in VHA settings. One RCT found that the combination of audit and feedback and physician-directed incentives resulted in a small, short-term positive effect on blood pressure control, but incentives directed at the practice, and at both the physician and practice, were not associated with improved outcomes.[17] 2 observational studies report evidence of positive effects on processes of care. However, it is possible that the findings of these studies may have been influenced by concomitant public reporting[14] and denominator management (see KQ2 for more detail).[15]

Table 2. The Effects of Pay-for-Performance in VHA Settings: Summary of Findings

Study · **Study Design** · **Sample Size** · **Observation Period/Follow-up**	**Program Focus** · **Incentive Target** · **Incentive Amount** · **Incentive Timing** · **Other Intervention Factors**	**Detailed Process of Care Findings**	**Detailed Patient Health Findings**	**Findings Summary**
Randomized Controlled Trial				
Petersen et al, 2013[17] · RCT · 83 physicians and 42 practice team members (*eg,* nurses, pharmacists) from 12 sites · 2007-2011	Primary care: BP · Physicians and/or Group · $9.10 per measure, $4270 average group total, $1648 average provider · 4-month performance period (up to 5 payments). Physicians with 2+ periods were included in the analysis. · Education, audit and feedback	Positive: Proportion of patients achieving BP control or receiving an appropriate provider response to uncontrolled blood pressure (group difference 8.36 [95% CI, 2.4 to 13]; P = .005) Null: Proportion of patients receiving guideline-recommended antihypertensive medication 4.72 [95% CI, -1.44 to 10.92]; P = .09).		A larger proportion of patients of providers receiving P4P achieved BP control or receive an appropriate response to high blood pressure. However, there was no difference in the use of guideline-recommended medications for high blood pressure.
Petersen et al, 2016[16] · RCT, Petersen et al, 2013[17] sub-study of AA/Black patients only · 67 physicians · 2007-2009		Positive: Proportion of patients achieving BP control or receiving an appropriate provider response to uncontrolled blood pressure (group difference 0.063 [95% CI, 0.008 to 0.117]; P = .03) Positive: Proportion of patients receiving an appropriate provider response to uncontrolled blood pressure (group difference 0.128 [95% CI, 0.003 to 0.253]; P = .05) Null: Proportion of patients receiving guideline-recommended antihypertensive medication (group difference 0.024 [95% CI, -0.028 to 0.075]; P = .37).	Null: Proportion of patients achieving BP control (group difference 0.021 [95% CI, -0.043 to 0.085]; P = 0.53).	P4P did not influence blood pressure control alone nor did the use of guideline-recommended medications for high blood pressure improve. However, patients of providers receiving P4P were more likely to achieve BP control or receive an appropriate response to high blood pressure.

Study · **Study Design** · **Sample Size** · **Observation Period/Follow-up**	**Program Focus** · **Incentive Target** · **Incentive Amount** · **Incentive Timing** · **Other Intervention Factors**	**Detailed Process of Care Findings**	**Detailed Patient Health Findings**	**Findings Summary**
Observational Studies				
Benzer et al, 2014[14] · Uncontrolled before/after · 128 VA medical centers · 2004-2010	Acute care: acute coronary syndrome, heart failure, and pneumonia · VISN and facility senior managers · Bonuses (amount NR) NR · Public Reporting	Positive: There were statistically significant improvements in 6 out of 7 measures related to acute coronary syndrome, heart failure and pneumonia (non-significant improvement for one heart failure measure).		Performance improved on 6 of 7 quality of care measures.
Harris et al, 2016[15] · Retrospective cohort · 133 VA Medical Centers · FY 2000 - 2009	Substance use · VISN Directors · Included as part of the performance contract - up to 10% of annual salary · Annual · Positive findings of denominator management	Positive: Percentage of patients entering a specialty SUD treatment program who were retained for at least 90 days increased from 23.1% just prior to when the Continuity of Care (CoC) measure was added to VISN Directors' performance contracts (FY 2002) to 48.3% (FY 2009), and increased at a rate of 0.7% per quarter (p<.001).		The percentage of patients entering a specialty SUD treatment program who were retained for at least 90 days increased significantly after the measure was added to VISN Directors' P4P contracts.

Abbreviations: AA = African American, BP = blood pressure, CoC = Coordination of Care, FY = fiscal year, NR = not reported, P4P = pay-for-performance, RCT = randomized controlled trial, SUD = substance use disorder, VA = Veterans Administration, VISN = Veterans Integrated Service Network

Detailed findings

To date, only one RCT has examined P4P in a VHA setting.[17] Petersen and colleagues compared audit and feedback plus incentives targeting 1) physicians, 2) practices, and 3) physicians and practices to audit and feedback alone (control) on guideline-recommended hypertension care. Across all subjects, physician-targeted, but not practice-level or combined incentives were significantly better than control in increasing the number of subjects either achieving blood pressure control or receiving an appropriate response to uncontrolled blood pressure. P4P, regardless of target, had no impact on the use of guideline-recommended medication.[17] Among African American/Black subjects only, P4P did not influence blood pressure control alone nor did the use of guideline-recommended medication for uncontrolled blood pressure improve. However, patients of providers receiving P4P (all incentivized groups combined) were more likely to achieve BP control or receive an appropriate response to uncontrolled blood pressure.[16]

Two studies examined the effect of P4P using VHA administrative data.[14,15] One study examined the effect of public reporting plus incentives targeting VISN- and facility-level senior managers on 7 inpatient guideline-concordant performance metrics for Veterans with acute coronary syndrome (ACS), heart failure, or pneumonia. Performance improved significantly for 6 of the 7 metrics, with improvements maintained through the two-to-four year incentivized period (varied by metric).[14] The second study, which also targeted VISN Directors, examined the effect of P4P on the continuity of care for Veterans with specialty substance use disorders (SUDs). Performance increased non-significantly from 20.8% to 23.1% during the 12-quarter pre-implementation period at a rate of 0.2% per quarter, then significantly a rate of 0.7% per quarter after P4P was implemented, reaching 48.3% twenty-eight quarters later. Of note, findings also indicate that after P4P was implemented, the proportion of Veterans with SUDs who qualified for the performance metric denominator decreased significantly (see KQ2 for more detail).[15]

KEY QUESTION 2: In Veteran populations, what are the potential unintended consequences of pay-for-performance in healthcare?

Thirteen articles from 11 studies (8 quantitative studies,[15-17,21-25] 2 qualitative articles from the same study,[26,27] and 3 others[19,20,28]) examined potential unintended consequences associated with pay-for-performance in VHA settings. In general, studies using administrative data and qualitative studies of VHA providers and leaders support the potential for overtreatment associated with performance metrics.[21,22,24,25] However, the sole RCT of P4P specifically found no association between P4P for hypertension and hypotension.[17] Furthermore, a qualitative sub-study of the same RCT found that despite no evidence of hypotension, a number of study participants reported concern for potential overtreatment.[28] Other studies found evidence of denominator management associated with a VISN Director-aimed incentive,[15] and no evidence of risk selection.[16,23] Qualitative studies explored provider perceptions of both negative and positive unintended consequences associated with performance metrics.[26,27] Figure 3 presents themes from KI interviews related to unintended consequences.

Detailed Findings

Overtreatment

Of the 7 articles providing data on overtreatment, 2 examined P4P specifically,[17,28] while the other 5 focused on the risk of overtreatment associated with performance metrics more generally (see Table 3).[21,22,24-26] A RCT that compared audit and feedback plus P4P to audit and feedback alone for guideline-concordant hypertension treatment found no relationship between P4P groups and control in the development of hypotension.[17] Furthermore, although there was no relationship between hypotension and P4P, a recently published qualitative sub-study of RCT participants found that participants were concerned that P4P would lead to overtreatment.[28]

Quantitative studies examining performance metrics more generally and using VHA administrative data[21,22,24,25] found that the upper age cutoff for colorectal cancer screening was associated with possible overuse in 40% of Veterans aged 70-75 with a Charlson comorbidity index ³ 4, that possible underuse in 16.5% of Veterans >75 with a Charlson comorbidity index of 0,[25] and that 23% of screening colonoscopies between 2011-2013 met the definition for probable or possible overuse.[24] In addition, among Veterans with diabetes, the facility rates for possible overtreatment associated with the clinical action metrics for hypertension and lipid management were 3-20%[22] and 8.5-18.4%[21] respectively, with higher rates of overtreatment significantly associated with higher rates of meeting the thresholds for both metrics ($p<.001$).[21,22]

In a qualitative study of unintended consequences associated with performance metrics, 32 of 59 VHA primary care providers and facility leaders interviewed expressed concern that performance metrics may be inappropriate for some patients, and 17 voiced specific concerns about the overuse of medication or treatment.[26]

Findings from Key Informant Interviews

Consistent with the literature, key informants voiced concern for potential overtreatment, particularly in facilities with metric-driven, rather than quality of care-driven cultures, and more commonly with metrics that vary, such as blood pressure. They also felt that one way to potentially mitigate overtreatment might be to include metrics focusing on prevention (*eg*, lifestyle counseling) and de-intensification.

Table 3. Unintended Consequences – Overtreatment: Summary of Findings

Study · **Study Design** · **Sample Size** · **Observation Period/Follow-up**	**Program Focus** · **Incentive Target** · **Incentive Amount** · **Incentive Timing**	**Comparison**	**Detailed Findings**	**Findings Summary**
Beard et al, 2013[21] · Retrospective Cohort · 964,818 · July 2010 to June 2011	Diabetes · NA · NA · NA	Examined rates of potential overtreatment in patients with diabetes associated with the clinical action measure for lipid management (proportion of patients with diabetes but without documented ischemic heart disease who were on high-dose statins).	13.7% of all diabetics (n = 131,722 were potentially overtreated. Excluding patients with cerebrovascular and peripheral vascular diseases reduced the rate of potential overtreatment to 11.5%. Facility rates varied from 8.5-18.4%. Facilities with higher rates of meeting the lipid threshold measure (LDL <100 mg/dL) had higher rates of potential overtreatment (p<0.001).	Facility potential overtreatment rates ranged from 8.5-18.4%. The predicted probability of potential overtreatment was higher in facilities with higher rates of meeting the lipid threshold measure.
Hysong et al, 2017[28] · Mixed-Methods Petersen et al, 2013[17] sub-study · 65 physicians from 12 sites · 2007-2011	Primary care: BP · Physicians and Group · $9.10 per measure, $4270 average group total, $1648 average provider · 4-month performance period (up to 5 payments)	Conducted 30-minute interviews with RCT participants to identify unintended consequences associated with financial incentives. Examined real/actual consequences vs voiced concerns.	Although the potential for overtreatment was voiced as a concern by participants, a chart audit found no evidence of overtreatment.	Although the potential for overtreatment was voiced as a concern by participants, a chart audit found no evidence of overtreatment.
Kerr et al, 2012[22] · Retrospective cohort · 977,282 · July 2009 to June 2010	Diabetes · NA · NA · NA	Examined rates of potential overtreatment in patients with diabetes associated with the clinical action measure for hypertension (proportion of patients with systolic blood pressure < 130 and diastolic blood pressure < 65 and receiving 3+ blood pressure medications or active intensification).	8% of all diabetics (n=80,903) were potentially overtreated. Facility rates varied from 3-20% (p<.001). Facilities with higher rates of meeting the hypertension threshold measure (<140/90 mm Hg) had higher rates of potential overtreatment (p<0.001).	Facility potential overtreatment rates ranged from 3-20%, with facilities in the highest quartile of meeting the hypertension threshold measure were 3.7 times more likely to be in the highest quartile of potential overtreatment.
Petersen et al, 2013[17] · RCT · 55 Physicians · 2007-2011	Primary care: BP · Physicians and Group · $9.10 per measure, $4270 average group total, $1648 average provider · 4-month performance period (up to 5 payments)	Compared P4P to control providers	There was no difference in rates of hypotension when comparing patients of P4P providers to controls (1.2% vs 1.4%; P=.18).	There was no evidence of overtreatment.

Study · **Study Design** · **Sample Size** · **Observation Period/Follow-up**	**Program Focus** · **Incentive Target** · **Incentive Amount** · **Incentive Timing**	**Comparison**	**Detailed Findings**	**Findings Summary**
Powell et al, 2012[26] · Qualitative · 59 VHA primary care providers and facility leaders at 4 VHAs · Feb-July 2009	Primary Care · NA · NA · NA	Conducted a semi-structured 60-90 min interview to identify unintended consequences associated with implementing a national performance measures into local primary care practices.	32 of 59 VHA primary care providers and facility leaders interviewed expressed concern that performance measures may be inappropriate for some patients, and 17 voiced specific concerns about the overuse of medications or treatments.	Participants identified a number of potential negative unintended consequences, including overtreatment.
Saini et al, 2014[25] · Retrospective cohort · 399,067 · 2008-2010	Colorectal Cancer · NA · NA · NA	Examined whether the upper age cutoff of the colorectal cancer screening measure was associated with overuse of screening among patients aged 70-75 in poor health and underuse in those over age 75 in good health.	Of Veterans 70-75 with a Charlson comorbidity index ≥ 4 (indicating poor health and shortened life expectancy), 40% underwent screening. Of Veterans >75 with a Charlson comorbidity of 0, 16.5% underwent screening. A 75-year-old Veteran who was unhealthy was more likely to undergo screening than a health 76-year-old (unadjusted RR = 1.64, 95% CI 91.36, 1.97].	Colorectal cancer screening was possibly overused in 40% of unhealthy Veterans 70-75, and possibly underused in 16.5% of Veterans >75.
Saini et al, 2016[24] · Cross-sectional · 88,754 · FY 2011 - FY 2013	Colorectal Cancer · NA · NA · NA	Examined screening colonoscopy overuse.	23% (n=20,530) met the definition for probable (17%) or possible (6%) overuse. Common reasons for overuse were that it was < 6 months after a negative FOBT (35%), < 9 years after a negative colonoscopy (31%), and age 40-49 (17%).	23% of screening colonoscopies between 2011-2011 met the definition for probable or possible overuse.

Abbreviations: BP = blood pressure, FOBT = fecal occult blood test, FY = fiscal year, LDL = low density lipoprotein, NA = not applicable, P4P = pay-for-performance, RCT = randomized controlled trial, VHA = Veterans Health Administration

Other Unintended Consequences

Three articles from 2 studies examined other potential unintended consequences specifically associated with P4P, and focused on denominator management,[15] risk selection,[16] and unintended consequences more generally (see Table 4).[28] Findings indicate that although a VISN Director-aimed incentive to increase the retention of Veterans enrolled in a specialty substance use disorder treatment program resulted in a significant increase after implementation, the percentage of patients meeting criteria for the metric denominator decreased significantly just both before and after P4P, with a more rapid decrease after P4P, and that facilities with higher pre-P4P denominators had steeper post-P4P denominator declines.[15] In contrast, a RCT[17] sub-study comparing audit and feedback plus P4P (aimed at providers, practices, or both) to audit and feedback alone for guideline-concordant care for African American/Black Veterans with hypertension found no evidence of risk selection.[16] A second (qualitative) sub-study from the same RCT[17] found that although participants voiced concerns for a wide range of negative unintended consequences, a chart audit revealed that with the exception of more time spent on data collection/clinical reminders, actual consequences were positive in nature.[28]

Related to performance metrics more generally, a retrospective cohort study examined comorbidity-related risk selection,[23] and 2 articles from a qualitative study focused on negative unintended consequences[26] and ancillary benefits of performance metrics for both providers and patients.[27] Findings suggested that contrary to the hypothesized disincentives, complex patients had higher odds of receiving appropriate follow-up, and there were no differences in patient-reported satisfaction by comorbidity.[23]

Qualitative studies found that participants felt performance metrics may lead to negative unintended consequences such as reduced focus on patient needs/concerns, unincentivized areas of care, and/or healthier patient populations (teaching to the test/attention shift),[19,20,26] and that they may negatively affect team dynamics, particularly if metrics are incentivized.[26] Participants also suggested that performance metrics may indirectly increase patient knowledge and motivation, may directly result in increased patient satisfaction and psychosocial benefits associated with increased patient-provider communication, and that providers may experience increased pride in individual and/or organizational performance and confidence that standards of care are aligned with evidence-based medicine.[27]

Findings from Key Informant Interviews

Consistent with findings in the literature, key informants were concerned about denominator management – and challenges related to the denominator in general, given the subjective nature and variability in some diagnoses and treatment recommendations. Other concerns included *risk selection* and *health disparities*, particularly for low-SES Veterans, the need mitigate the potential for *teaching to the test/attention shift*, by having a variety of actively monitored valid metrics covering different aspects of care, and *gaming*.

Gaming
The reason you have things like gaming the system isn't because people don't want to do the right thing, it's because they can't do the right thing. – Key Informant

They stressed the history of gaming, particularly within the context of P4P. In addition, they stressed the need for adequate resources with which to respond to and achieve incentivized metrics, and felt that any lack of adequate support is likely to increase the potential for gaming.

Finally, there was disagreement among KIs about the use of composite metrics to mitigate gaming. One KI felt that the lack of transparency makes composite metrics more difficult to game, while still targeting care processes. However, another KI felt that there isn't enough research to inform how to weight composite metrics, and that in addition, the goal of P4P should be to improve performance on specific metrics, rather than to improve a provider's average.

Table 4. Unintended Consequences – Other: Summary of Findings

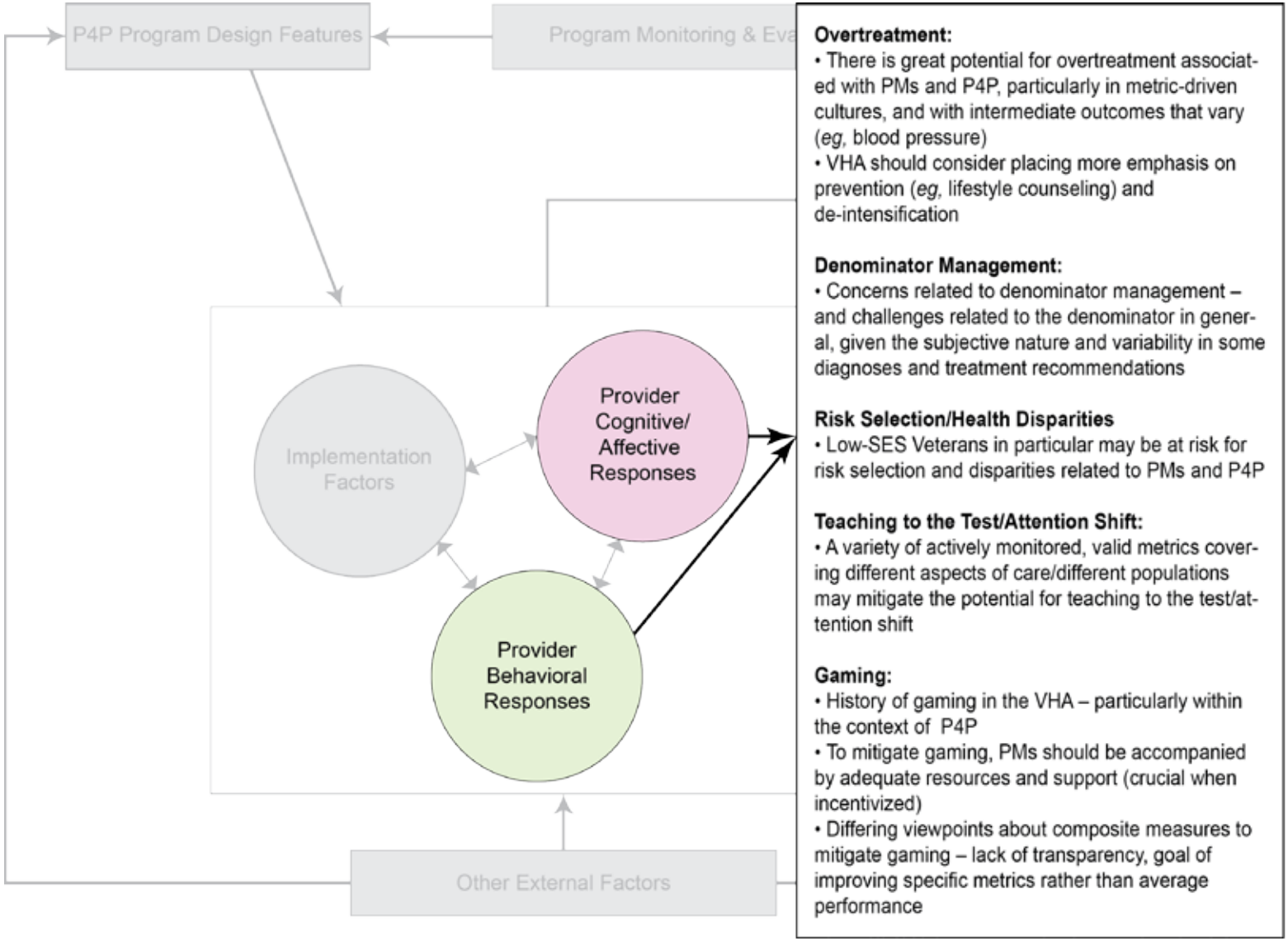
Type of Unintended Consequence	Study • Study Design • Sample Size • Observation Period/Follow-up	Program Focus • Incentive Target • Incentive Amount • Incentive Timing	Comparison	Detailed Findings	Findings Summary
Ancillary Benefits	Powell et al, 2014[27] • Qualitative • 59 VHA primary care providers and facility leaders at 4 VHAs • Feb-July 2009	Primary care • NA • NA • NA	Conducted a semi-structured 60-90 min interview to identify ancillary benefits associated with implementing a national performance measures into local primary care practices.	Performance measures may lead to educational interventions resulting in increased patient knowledge and motivation. Other benefits may include increased patient satisfaction and psychosocial benefits associated with increased patient-provider communication. Providers may experience increased pride in individual and/or organizational performance and confidence that standards of care are aligned with evidence-based medicine.	Participants identified numerous potential indirect and direct ancillary benefits for patients and providers associated with the implementation of performance measures.
Attention Shift	Damschroder et al, 2014[20] • Qualitative • 62 leaders, clinicians, and staff from 4 VHA facilities • 2012	Diabetes • NA • NA • NA	Described the perceived impact of implementation of diabetes performance measures on management actions and day-to-day clinical practice.	Participants describe a punitive mentality, with leaders focusing on metrics that aren't met, shifting attention from patient needs.	Participants describe a range of unintended consequences related to implementation, including attention shift.
	Kansagara et al, 2014[19] • Qualitative • 241 Clinical and Administrative Staff from 15 VHA Primary Care Clinics • December 2010 and February 2013	Primary Care • NA • NA • NA	Described primary care staff experiences with the use of performance metrics during the implementation of the VHA PACT model of care.	Primary care staff perceived that performance metrics are time consuming and represented an opportunity cost, with more attention paid to metrics that other areas of care, and to the patient populations performance metrics are targeting.	Primary care staff perceived responding to performance metrics as time-consuming and possibly distracting from unincentivized areas of care and/or populations.
Denominator Management	Harris et al, 2016[15] • Retrospective	Substance use • VISN Directors	Compared proportion of patients with SUD who	The percentage of patients who met the CoC measure denominator	The percentage of patients meeting criteria for the

Type of Unintended Consequence	Study • Study Design • Sample Size • Observation Period/Follow-up	Program Focus • Incentive Target • Incentive Amount • Incentive Timing	Comparison	Detailed Findings	Findings Summary
	cohort • 133 VA Medical Centers • FY 2000 - 2009	• Included as part of the performance contract - up to 10% of annual salary • Annual	qualified for the denominator of the CoC measure before and after it was added to VISN Directors' performance contract.	criteria decreased by 0.1% from 23.6% in FY 2000 to 22.1% in FY 2002 (p<.001), just before the measure was added to VISN Directors performance contracts. By the end of observation period (FY 2009), the percentage of patients meeting criteria for the denominator had fallen to 15.9% (p<.001), and the rate of decrease was more rapid after the measure was implemented (p=.02). In addition, facilities with higher pre-implementation prevalence had steeper declines after implementation (p<.001).	CoC measure denominator decreased significantly pre- and post-P4P with a more rapid decrease after P4P. Facilities with higher pre-P4P denominators had steeper post-P4P denominator declines.
Provider Behavior	Powell et al, 2012[26] • Qualitative • 59 VHA primary care providers and facility leaders at 4 VHAs • Feb-July 2009	Primary care • NA • NA • NA	Conducted a semi-structured 60-90 min interview to identify unintended consequences associated with implementing a national performance measures into local primary care practices.	Participants identified a number of potential negative unintended consequences, and felt that performance measures, particularly those linked to pay, may negatively affect team dynamics.	Participants identified a number of potential negative unintended consequences, including a negative impact on team dynamics.
Risk Selection	Petersen et al, 2016[16] • RCT Petersen et al, 2013[17] sub-study of AA/Black patients only • 67 physicians • 2007-2009	Primary care: BP • Physicians and/or Group • $9.10 per measure, • 4-month performance period (up to 5 payments)	Compared P4P to control providers	Null: Percentage of patients who switched providers (25.1% both groups; p = 0.98) Null: Percentage of patients no longer assigned to original provider (28.8% P4P vs 29.6 control; p = 0.61) Null: Number of patient visits (M = 9.1[SD = 8.1] P4P vs M = 9.0[SD = 8.4] control; p = 0.77)	There was no evidence of risk selection.
	Petersen et al, 2009[23] • Retrospective	Hypertension • NA	Compared appropriate follow-up received by hypertensive	Adjusted odds of receiving an appropriate follow-up were higher	Complex patients had higher odds of receiving

Type of Unintended Consequence	Study • Study Design • Sample Size • Observation Period/Follow-up	Program Focus • Incentive Target • Incentive Amount • Incentive Timing	Comparison	Detailed Findings	Findings Summary
	cohort • 141,609 • FY 2005	• NA • NA	patients and patient satisfaction among those with and without comorbid conditions.	for those with both concordant and discordant comorbid conditions, as compared to those with none (Adj OR = 1.64, 95% CI [1.57, 1.73] and Adj OR = 1.91, 95% CI [1.80, 2.02] respectively). There was no difference in the patients' responses of having received "very good" or "excellent" quality of care by comorbidity.	appropriate follow-up for diabetes. There were no differences in patient-reported satisfaction by comorbidity.
Multiple	Hysong et al, 2017[28] • Mixed-Methods Petersen et al, 2013[17] sub-study • 65 physicians from 12 sites • 2007-2011	Primary care: BP • Physicians and Group • $9.10 per measure, $4270 average group total, $1648 average provider • 4-month performance period (up to 5 payments)	Conducted 30-minute interviews with RCT participants to identify unintended consequences associated with financial incentives. Examined real/actual consequences vs voiced concerns.	Although participants voiced concerns for a wide range of negative unintended consequences including reduced morale, reduced clinical flexibility, pressure to treat incentivized conditions to the detriment of the patient, and attention shift, a chart audit revealed that with the exception of more time spent on data collection/clinical reminders, actual consequences were positive in nature (improved documentation, improved morale, learning and development, positive spillovers).	Many unintended consequences of financial incentives noted were either only concerns or attributable to ancillary quality-improvement initiatives. Actual unintended consequences included improved documentation of care without necessarily improving actual care, and positive unintended consequences.

Abbreviations: AA = African American, BP = blood pressure, CoC = Coordination of Care, FY = fiscal year, NA = not applicable, P4P = pay-for-performance, RCT = randomized controlled trial, SUD = substance use disorder, VA = Veterans Administration, VHA = Veterans Health Administration, VISN = Veterans Integrated Service Network

Figure 3. Key Informant Interviews: Themes – Unintended Consequences



Note. Abbreviations: P4P =pay-for-performance, PM = performance metric (PM), SES = socioeconomic status, VHA = Veterans Health Administration

KEY QUESTION 3: What metrics have been commonly incentivized in the published literature examining P4P programs?

Across 39 studies, we identified 82 process of care or administrative metrics, and 10 patient metrics (*ie,* intermediate and health outcomes, and evaluations of care) that have been reported as incentivized in published research examining P4P programs. Nineteen P4P programs in 8 countries are represented (see Table 5).

Table 5. P4P Programs Represented

Program	Process of Care	Patient Outcome
Australia's Practice Incentives Program[29]	X	
CMS Medicare Hospital Value-Based Purchasing[30]	X	X
English National Chlamydia Screening Programme [31]	X	
French National Health Insurance[32]	X	
Massachusetts General Physicians Organization [33]	X	
Medicare Advantage Prescription Drug Quality Compensation Program, TX[34]	X	
Netherlands GP P4P[35-37]	X	X
Ontario's Diabetes Management Incentive[38]	X	
Ontario's Family Health Network, Family Health Group, Comprehensive Care Model, and Family Health Organization[39]	X	
Partners HealthCare Inc., MA[40]	X	
Radiology P4P in a 751-bed, urban, tertiary care adult teaching hospital[41]	X	
Regional Health Care Evaluation Program, Lazio, Italy[42]	X	
Rochester Independent Practice Association (RIPA)[43]	X	
Taiwan's Diabetes Mellitus P4P[44-46]	X	
Taiwan's Obstetric P4P[47]		X
Taiwan's Pay-for-Performance on Tuberculosis [48]		X
UK's Premier Hospital Quality Incentive Demonstration[49]	X	
UK's Quality and Outcomes Framework[50-63]	X	X
Veterans Health Administration[14,17]	X	

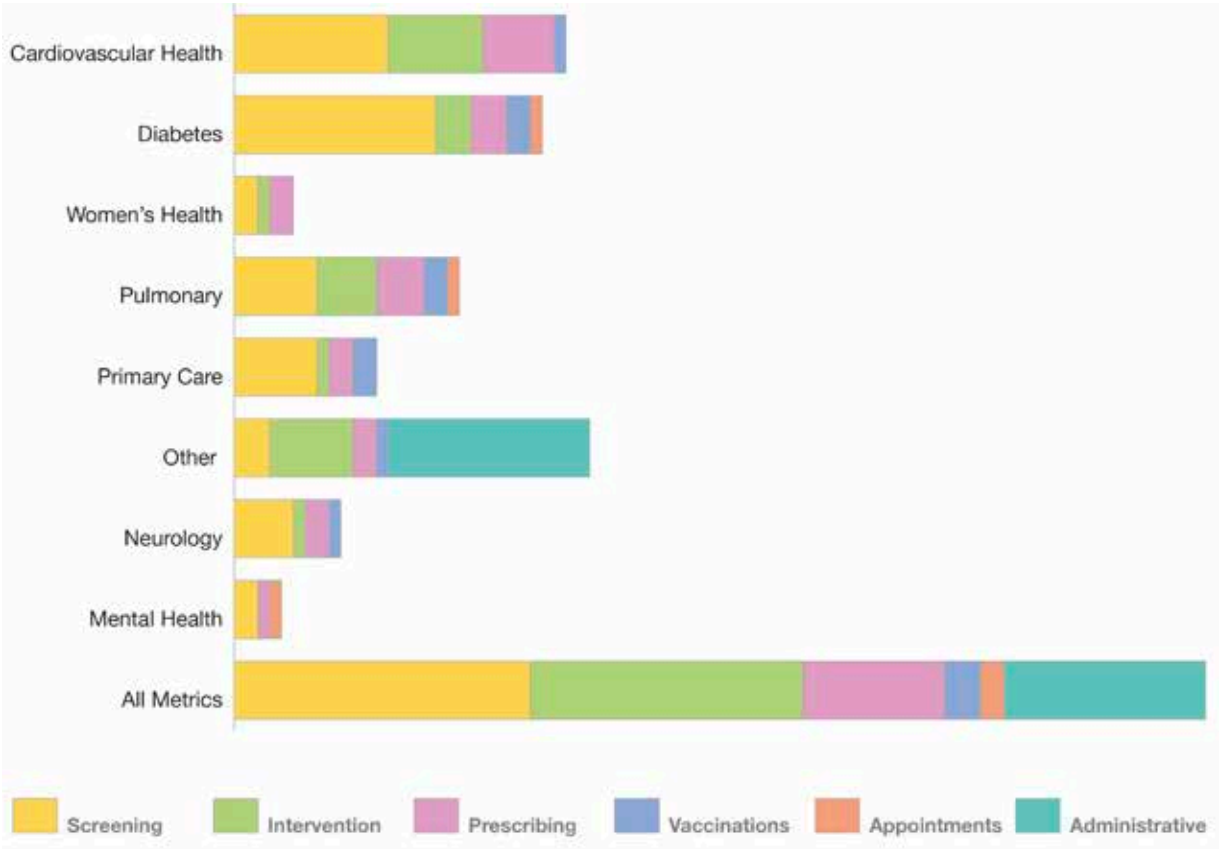
Processes of Care

Incentivized metrics most commonly targeted cardiovascular health and diabetes, followed by pulmonary conditions and primary care. Screenings were the most common type of incentivized metric, followed by interventions/procedures, and prescribing. Metrics classified as "other" were predominantly administrative in nature (*eg*, trainings and EHR use). Table 6 presents a list of identified metrics organized by type and clinical condition. Figure 4 illustrates the distribution of metric types overall, and by clinical condition.

Patient Outcomes

Very few patient outcome metrics were reported in published P4P research. The most common were intermediate measures such as blood pressure, cholesterol, and hemoglobin A1c targets, particularly in the United States and in the United Kingdom. 2 P4P programs in Taiwan were distinct from the others – one incentivizing stage-specific tuberculosis identification and cure rates,[48] and the other providing incentives for vaginal births (see Table 7).[47]

Figure 4. Incentivized Process of Care and Administrative Metrics Reported in Published Literature by Condition and Type



Note. "Other" includes incentives targeting all physicians, radiology, inpatient, surgery, hip fractures, and influenza.

Table 6. Process of Care and Administrative Metrics Incentivized in Published P4P Literature

	Cardio-vascular Health	Diabetes	Mental Health	Neurology	Primary Care	Pulmonary	Women's Health	Other
Screening								
Alcohol consumption recorded[64]	X	X		X		X		
BMI recorded[29,38,53,54,59,61,64]	X	X		X		X		
BP recorded[29,35,38,50,52,53,59,62,64]	X	X		X		X		X
Cervical cancer screening[35,39]				X				
Chlamydia screening[31]					X		X	
Colorectal cancer screening[39]				X				
Glucose recorded[35,44-46,54]	X	X						
HbA1C recorded[29,34,35,38,43-46,52,53,59,61]	X	X						
HDL recorded[29,44]		X						
LDL recorded[34,43,44]	X	X						
Lipids recorded[38]		X						
Lithium levels recorded[52,53]			X					
Microalbumin testing recorded[29,38,59,61]		X						
Neuropathy screening recorded[29,35,38,43,59,61]		X						
Oxygenation assessed[49]						X		
Peripheral pulses recorded[61]		X						
Retinal screening recorded[29,35,38,43,59,61]		X						
Serum creatinine recorded[34,35,38,44-46,53,61]	X	X	X					
Smoking status recorded[29,35,52,53,55,58,59,61,64]	X	X		X		X		
Spirometry/reversibility testing[37,53]						X		
Thyroid function test recorded[52]								X
Total cholesterol recorded[29,35,38,44,52-54,59,61,64]	X	X		X		X		
Triglycerides recorded[29,44]		X						
Urine albumin creatinine ratio recorded[50]								X

	Cardio-vascular Health	Diabetes	Mental Health	Neurology	Primary Care	Pulmonary	Women's Health	Other
Weight recorded[14,38]	X	X						
Interventions and Procedures								
ACEI or ARB for left ventricular systolic dysfunction[49]	X							
Aspirin administered on hospital arrival[49]	X					X		
Blood cultures performed in ED before initial antibiotics received[30,49]						X		
Cardiology involvement within 24 hours[14]	X							
Diagnostic catheterization[14]	X							
Discharge instructions provided[30,49]	X							
Fibrinolytic therapy within 30 min of hospital arrival[49]	X							X
Hand hygiene before/after patient contact[33]								
Initial program enrollment and check-up[44]		X						
Left ventricular systolic function assessed[49]	X							
Primary PCI within 90 minutes of hospital arrival[30]	X							
Prophylactic antibiotic within one hour prior to surgical incision[30]								X
Prophylactic antibiotics discontinued within 24 hours after surgery end time[30]								X
Proportion of surgeries for hip fracture performed within 48 hours of admission[42]								X
Self-management education[29]		X				X		
Smoking cessation advice[35,40,49,53,55,57,58,61]	X	X				X		X
Stage-based guideline-recommended breast cancer treatment[65]							X	
Surgery patients on a beta blocker prior to arrival that received a beta blocker during the perioperative period[30]								X
Surgery patients who received appropriate venous thromboembolism prophylaxis within 24 hours prior to surgery to 24 hours after surgery[30]								X
Surgery patients with recommended venous thromboembolism prophylaxis ordered[30]								X

	Cardio-vascular Health	Diabetes	Mental Health	Neurology	Primary Care	Pulmonary	Women's Health	Other
Troponin returned within 60 minutes of order[14]	X							
Timely antibiotic[14]								X
Prescribing								
Aspirin prescribed at discharge[49]	X							
Benzodiazepine prescription reduction[32]			X					
Beta blocker prescribed at discharge[49]	X							
Guideline concordant prescriptions[17,36,58]	X				X			
Initial antibiotic selection in immunocompetent patients[49]						X		
Medication review[29,49,53]		X		X		X		
Prescribed LARCs[51]							X	
Prescribed specific drugs/classes[14,36,37,50,52,60-62]	X	X			X			X
Prophylactic antibiotic selection[30]								X
Received appropriate initial antibiotic selection[49]						X		
Statins prescribed[35,37]	X	X						
Vaccinations								
Influenza[34,35,38,39,52,53,56,61]	X	X		X	X	X		X
Pneumococcal[14,38]		X						X
Toddler[39]					X			
Appointments								
Follow-up appointments[29,37]						X		
No annual review follow-up in 14 days (psychoses)[53]			X					
Administrative								
1+ prescription note in EHR 80%[33]								X
1+ problem note in EHR 80%[33]								X
Access to patient portal 10%[33]								X
CKD diagnosis recorded in EHR[50]								X
Communication training[33]								X

	Cardio-vascular Health	Diabetes	Mental Health	Neurology	Primary Care	Pulmonary	Women's Health	Other
EHR notes within _ days of outpatient visit[33]								X
EHR radiology order entry[33]								X
EHR training[33]								X
Electronic education for patients 50%[33]								X
Electronic prescriptions[33]								X
Joint Commission evaluation[33]								X
Joint Commission training[33]								X
Med allergies in EHR 80%[33]								X
Median signature time < 8 hours or 80% of reports signed within 16 hours[41]								X
Review Meaningful Use performance dashboard[33]								X
Smoking status in EHR 50%[33]								X
Vitals in EHR 50%[33]								X

Abbreviations: ACEI = angiotensin converting enzyme inhibitor, ARB = angiotensin II receptor blockers, CKD = Chronic Kidney Disease, ED = Emergency Department, EHR = Electronic Health Record, PCI = percutaneous coronary intervention

Table 7. Intermediate, Patient Experience, and Patient Health Outcome Metrics Incentivized in Published P4P Literature

	Cardio-vascular Health	Diabetes	Mental Health	Neurology	Primary Care	Pulmonary	Women's Health	Other
Intermediate Outcomes								
Blood Pressure Control[17,35,50,53,54,59,61-63]	X	X		X				X
Cholesterol Control[35,53,54,59,61,63]	X	X						
HbA1c Control[35,53,59,61,63]		X						
Glucose Control[54]		X						
Lithium in Therapeutic Range[53]			X					
Patient Experience[30,35]	X	X	X	X	X	X	X	X
Patient Outcomes								
Vaginal Delivery[47]							X	
No Exacerbation[35]						X		
TB Identified in months 1-3[48]						X		
TB Cured in months 4-12[48]						X		
Cardiac surgery patients with controlled 6am postoperative serum glucose[30]								X

Abbreviations: HbA1c = hemoglobin A1c, TB = tuberculosis

KEY QUESTION 4: In Veteran populations, what program features and implementation factors modify the effectiveness of pay-for-performance programs?

Program Features and Implementation Factors of Pay-for-Performance in VHA Settings

Sixteen articles from 13 studies (10 quantitative studies,[14,17,24,66-72] 2 sub-studies,[18,73] and 4 qualitative studies[19,20,74,75]) provide data examining program design features or implementation factors of pay-for-performance programs in VHA settings. In general, studies examining program design features found physician-targeted incentives to be more effective than those targeting groups/practices,[17] that the degree of agreement between EHR data and manual review varied by metric,[18,24] that the relationship between access metrics and patient satisfaction varied by access metric[71] and whether the patient was new or returning,[72] and that the difficulty of achieving multi-tasked metrics was not directly related to the number of tasks involved.[67] Studies examining implementation processes found no difference in the achievement of actively versus passively monitored metrics,[68] provide mixed evidence related to the impact of the removal of incentives on performance,[14,17] and suggest areas of improvement for implementing performance metrics at the local level.[19,20] One study examined provider affective/cognitive responses found that not only did P4P have no impact on goal commitment, but that physicians may perceive an external locus of control for hypertension care.[73] 2 studies examined audit and feedback, one finding that higher-performing facilities reported delivering more timely, individualized, and non-punitive feedback,[75] with the second reporting that the delivery of audit and feedback after Patient Aligned Care Team (PACT) implementation remained largely unchanged.[74] Themes from KI interviews are provided in Figure 5, and study details are provided in Tables 8 and 9.

Detailed Findings

Program Design Features

A RCT compared physician-targeted, practice-targeted, and a combination incentive to control, and found that patients of providers receiving physician-targeted, but not practice- or combination-targeted incentives were more likely to achieve blood pressure control or receive an appropriate response to uncontrolled blood pressure. The incentive target did not influence the receipt of guideline-recommended hypertensive medication.[17]

Studies examining features of performance metrics found that for Veterans with hypertension, the agreement between manual chart review and automatic processing electronic health record (AP-EHR) guideline-concordant metric data varied by metric, with lower percentages observed in the EHR,[18] and that as compared to manual chart review, using an electronic metric, colonoscopies could be correctly identified with high specificity, but with low sensitivity to overuse.[24] Studies also found that there was poor agreement between CMS' all cause hospital-wide 30-day readmission metric and condition specific metrics,[70] that Diagnostic Cost Groups (DCGs) were better predictors of mortality and long-term care hospitalization than Adjusted Clinical Groups (ACGs),[69] and that the number of tasks alone was not significantly related to the difficulty of completing complex clinical performance metrics; rather, that the difficulty of each step must be considered.[67] With regard to access and wait times, Veteran satisfaction varied widely by access measure (*eg*, telephone contact, face-to-face, primary versus specialty care),[71] and the relationship between patient satisfaction and specific wait time metrics (*eg*, first next

appointment, desired appointment date, *etc*) were different for new and returning patients.[72] Finally, when examining the relationship between mental health metrics and patient satisfaction, program reach and intensity, as well as treatment continuity, but not access to psychological services, were positively related to both access and encounter patient satisfaction.[66]

Findings from Key Informant Interviews

Larger Incentives
We're being encouraged to do a lot of things to meet a lot of different quality measures. Or to do the right thing or check off different boxes in the process of clinical care. Not all of those things are necessarily being incentivized. Unless you make that incentive more salient, it makes it hard to stand out among all of the other things that we're being encouraged to do. For example, for diabetic patients we're trying to get people below the performance measure for HbA1C to 9%, we're trying to get blood pressure controlled, we're trying to get the right people to be on statins, we're trying to get people to be taking aspirins. Depending on the medical center, there are different combinations of those things being incentivized, but we're just trying to do the right thing for the patient and making sure the patient gets the right care. – Key Informant

Key informants consistently felt that VHA incentives need to be larger to be meaningful to the provider. Clinician KIs understood the tradeoff between incentive size and frequency, and consistently felt that once-yearly incentives weren't frequent enough to influence behavior.

A number of key informants also thought that that the VHA should explore incentivizing teams, and other front-line staff.

Key informants also felt that performance metrics should be clinically meaningful. Although some KIs felt that a combination of process and outcomes metrics should be incentivized, others were less enthusiastic about incentivizing outcome metrics, because of the lack of provider control.

Consider Incentivizing Teams/Front-line Staff
Can you imagine giving a scheduling clerk the opportunity to earn an incentive? That might go a long way. – Key Informant

In addition, some KIs felt that the VHA should consider incentivizing administrative functions directly related to advancing quality. A number of KIs spoke about the challenges of incentivizing metrics related to access – as well as potential alternatives, with one KI suggesting that the VHA consider incentivizing efficiencies. Key informants stressed the importance of the feasibility of metric achievability, and felt that unrealistic metrics may result in adverse effects on staff.

Feasibility of attaining a measure is crucial
Case and point, if you say we want same-day access. Well that sounds great, but what if at your local climate your providers are at max 120% capacity and they are completely mismatched with supply and demand? And then you put it on the report card that you didn't achieve same-day access, which is completely unachievable. That's going to kill their morale. – Key Informant

A number of KIs also felt that more emphasis should be placed on de-intensification, and also mentioned potential barriers such as provider-patient buy-in, and the perception that it is more difficult to stop than start something.

Finally, KIs were interested in exploring alternative methods of identifying and validating performance metrics (*eg*, Productivity Measurement and Enhancement System [ProMES]), as well as the impact of the complexity of each step required to achieve a metric, and a number of them stressed the importance of validity – that the metrics accurately measure what they are intended to measure – regardless of whether or not they are incentivized.

Measures must be Valid
You're going to be in trouble if you're using metrics that are invalid or cease to be valid. You're going to be investing your management effort and your money in achieving things that you didn't set out to achieve. – Key Informant

Table 8. VHA Program Design Features – Summary of Findings

Type of Program Design Feature	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Incentive Target · Incentive Amount · Incentive Timing	Comparison	Detailed Findings	Findings Summary
Incentive: Target	Petersen et al, 2013[17] · RCT · 83 physicians and 42 practice team members (*eg,* nurses, pharmacists) from 12 sites · 2007-2011	Primary care: BP · Physicians and Group · $9.10 per measure, $4270 average group total, $1648 average provider · 4-month performance period (up to 5 payments)	Compared: 1) audit and feedback only, 2) physician incentives + audit and feedback, 3) group incentives + audit and feedback, 4) physician and group + audit and feedback.	A larger proportion of patients in the physician group achieved BP control or received an appropriate provider response to uncontrolled blood pressure (group difference 8.36 [95% CI, 2.4 to 13]; p=.005). There were no differences between control and practice incentives (group difference 3.24 [95% CI, -1.48 to 8.92]; p=.26) or physician + practice incentives (group difference 5.08 [95% CI, -0.04 to 10.56]; p=.09) There was no difference in the proportion of patients receiving guideline-recommended antihypertensive medication by incentive target.	Patients of providers receiving physician-targeted, but not practice-targeted or physician + practice-targeted incentives were more likely to achieve BP control or receive an appropriate response to high blood pressure. Incentive target did not influence the receipt of guideline-recommended hypertensive medication.
Metrics: Complexity	Hysong et al, 2016[67] · Cross-Sectional · 133 VA Medical Centers, 8 primary care providers · FY 2000-2008	Primary Care · NA · NA · NA	Examined the difficulty of achieving clinical performance metrics by examining the complexity of the component tasks using Functional Job Analysis.	Metrics varied in the component number of tasks. Number of tasks was not significantly related to measure difficulty. Measures of chronic care following AMI had significantly higher difficulty scores than diabetes and screening measures, but not immunizations (F[3, 186]=3.57, p=0.015). The number of steps and each step's difficulty must be considered when designing potential interventions for improving quality in a given metric or care area.	Both the number of steps and degree of difficulty of each step in a clinical performance metrics must be considered. Functional job analysis may useful in assessing the difficulty of clinical performance measures and variations among metrics and achievement.
Metrics: Data	Urech et al, 2015[18] · Cross-sectional,	· Physicians and Group	Examined the agreement between automatic	Data obtained via manual chart review resulted in higher percentages across	Agreement between manual chart review and

Type of Program Design Feature	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Incentive Target · Incentive Amount · Incentive Timing	Comparison	Detailed Findings	Findings Summary
Source	Petersen et al, 2013[17] sub-study · 52 physicians and 33 practice team members (*eg,* nurses, pharmacists) from 12 sites (2,840 patients) · April-July 2009	· $9.10 per measure, $4270 average group total, $1648 average provider · 4-month performance period (up to 5 payments)	processing electronic health record (AP-EHR) data to data collected via manual chart review, and the impact on performance pay.	all 3 outcomes (proportion of patients achieving BP control [overall agreement = 0.96, kappa = 0.87], proportion receiving an appropriate provider response to uncontrolled blood pressure [agreement = 0.64; kappa = 0.28], and proportion of patients receiving guideline-recommended antihypertensive medication [agreement = 0.85; kappa = 0.51]). AP-EHR data would result in mean payments being 10% less for guideline-recommended medication, and 8.3% less for BP control and appropriate provider response.	AP-EHR data varied by measure, with lower percentages observed in the EHR.
Metrics: Risk Adjustment	Petersen et al, 2005[69] · Cross-sectional · 3,069,168 · FY 2001	General · NA · NA · NA	Compared risk adjustment using DCGs and ACGs, both with and without age, as well as the combination of age + gender, using all cause mortality and long-term care hospitalization.	For both outcomes, DCG + age was the best predictor (c-statistic morality = 0.830, LTC hospitalization = 0.890), followed by DCG alone. ACG was the poorest predictor for mortality (c-statistic = 0.700), and age + gender alone was the poorest predictor of LTC hospitalization (c-statistic = 05.93).	DCGs were better predictors of both mortality and long-term care hospitalizations than ACGs.
Metrics: Type	Kansagara et al, 2014[19] · Qualitative · 241 Clinical and Administrative Staff from 15 VHA Primary Care Clinics · December 2010 and February 2013	Primary Care · NA · NA · NA	Described primary care staff experiences with the use of performance metrics during the implementation of the VHA PACT model of care.	Primary care staff perceived that performance metrics led to delivery changes that were not always aligned with PACT principles, such as efficient care delivery, and were not well-adapted to team-based care.	Primary care staff perceived that performance metrics were not consistently aligned with PACT principles of care.
Metrics: Validation	Frakt et al, 2017[66] · Cross-sectional · 6,990	Mental Health · NA · NA	Examined the extent to which facility-level administrative data-based performance metrics	Broad measures of program reach and intensity positively associated with both access and encounter patient satisfaction. No measure of access to	Program reach and intensity, and treatment continuity, but not access to psychological services

Type of Program Design Feature	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Incentive Target · Incentive Amount · Incentive Timing	Comparison	Detailed Findings	Findings Summary
	· FY 2013	· NA	(program reach, psychosocial service access, program intensity, and treatment continuity) predict individual-level access and encounter satisfaction.	psychosocial services and nearly all measures of continuity are positively associated with both kinds of satisfaction.	were positively related to both access and encounter patient satisfaction.
	Prentice et al, 2014[72] · Cross-sectional · 221,540 FY 2010	Access · Administrative data/SHEP	Examined the relationship between self-reported new and returning patient satisfaction (self-reported timely appointment, access to treatment and specialists, VA rating and satisfaction) and 1) first next available appointment, 2) retrospective create date, 3) prospective create date, 4) retrospective completed desired date, 5) prospective scheduled desired date	Appointment and prospective and retrospective create date were significantly related to all 5 measures of patient satisfaction. For returning patients, only the prospective scheduled desired date was consistently significantly associated with all 5 measures of satisfaction.	The relationship between access and patient satisfaction differs in new and returning patients, with first next appointment and appointment create dates better measures for new patients, and the prospective scheduled desired date a better metric for returning patients.
	Prentice et al, 2016[71] · Cross-sectional · 20,000 - 218,677 FY 2010, 2012	Access · Administrative data/SHEP	Examined the relationship between access metrics and 5 self-reported measures of patient satisfaction (timely appointment, access to treatment and specialists, VA rating and satisfaction) – 1) primary care third next appointment, 2) specialty care third next appointment, 3) speed of	Third next appointment metrics are less reliable indicators of patient satisfaction. Both speed of response and abandonment rate predict Veteran satisfaction, with more consistent results found for speed of response. Primary care telephone clinics were valid for general satisfaction (VA rating and satisfaction) but not access metrics.	Veteran satisfaction varies widely by type of access measure (telephone vs face to face, primary care, specialty care).

Type of Program Design Feature	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Incentive Target · Incentive Amount · Incentive Timing	Comparison	Detailed Findings	Findings Summary
			telephone response, 4) telephone abandonment rate, 5) primary care telephone clinics. Compared to previously established relationships between new patient create date and consult wait time metrics and patient satisfaction.		
	Rosen et al, 2016[70] · Retrospective cohort · 154 VA Medical Centers · FY 2007-2010	30-day readmission rates for patients with heart failure, AMI, and pneumonia · CMS Penalties · NR · NR	Compared CMS' hospital-wide all-cause readmission measure to condition-specific measures of readmission.	Agreement ranged from 73.1% (HF) to 79.6% (AMI). Weighted k values were low, ranging from 0.11 for comparisons between the HWR and HF measures, to 0.17 between the HWR and AMI measures. Among those hospitals classified as poor performers by the heart failure readmission measure, only 28.6% were similarly classified by the HWR measure. The HWR measure penalized only 60% of those hospitals that would have received penalties based on at least 1 of the condition-specific measures.	There was poor agreement between CMS' all-cause hospital wide readmission measure and the condition-specific readmission measures.
	Saini et al, 2016[24] · Cross-sectional · 2915 · FY 2011-2013	Colorectal Cancer · NA · NA · NA	Examined colonoscopy overuse, and compared manual review to an electronic measure.	The electronic measure was highly specific and correctly identified 97% of appropriate colonoscopies. However, the measure had low sensitivity, and correctly identified only 20% as overuse.	Colonoscopies can be correctly identified with high specificity, but with low sensitivity to overuse using electronic data.

Abbreviations: ACG = Adjusted Clinical Groups, AMI = acute myocardial infarction, AP- EHR = automated processing electronic health record, BP = blood pressure, CMS = Centers for Medicare and Medicaid Services, DCG = Diagnostic Cost Group, EHR = electronic health record, FY = fiscal year, HF = heart failure, HWR = hospital wide readmission, LTC = long term care, NA = not applicable, NR = not reported, PACT = Patient Aligned Care Team, VA = Veterans Administration, VHA = Veterans Health Administration

Implementation Factors

Studies focusing on implementation factors examined passive versus active performance metric monitoring[68] and incentive removal.[14,17] A retrospective cohort study examined the effect of changes in metric status from support indicators (passively monitored) to performance measures (actively monitored, and possibly incentivized) and vice-versa, and found that not only was there no significant relationship between metric performance and monitoring status, but that quality was either maintained or continued to improve.[68] Quality was similarly sustained for up to 3 years in an (uncontrolled) study that examined VISN and facility senior manager-incentivized acute care metrics before and after incentives were removed.[14] Conversely, Petersen and colleagues' P4P RCT found a significant reduction in performance on guideline-concordant hypertension metrics after a 12-month washout period.[17]

Petersen and colleagues' RCT also included audit and feedback via a website to both incentivized and non-incentivized participants, with participants receiving audit and feedback alone serving as controls. They found that incentivized participants were significantly more likely to access their feedback reports.[17]

Qualitative studies reported frustration with the implementation of metrics at the local level, feeling that implementation was guided by a "top down" approach,[19] that providers were being held accountable for gaps in performance that were beyond their control,[20] that metric implementation and clinical reminders weren't always aligned with the evidence and VHA priorities,[20] and that performance metrics were not always aligned with PACT principles of care.[19] 2 studies examined audit and feedback – one that found that as compared to facilities with poorer records of guideline adherence, high-performing facilities tended to deliver more timely, individualized, and non-punitive adherence feedback to providers.[75] The second examined audit and feedback after PACT implementation and found that clinical performance was still considered the responsibility of the physician, and that processes remained largely unchanged.[74] Related to incentives, physicians reported a lack of transparency in the calculation of their performance pay.[20]

Findings from Key Informant Interviews

P4P should be framed positively, and is just one aspect of quality improvement
(A negative frame) "You're not doing the right thing. I've got to dangle cash in front of you so that I get your attention and get you to start doing the right thing for patients." An alternative frame is, "You've got a lot of quality metrics to meet and this is really hard and I know you're doing the right thing, but there's a few of them that we're going to highlight. We really think that this is a tool that's going to help you do better with your patients more often and that's why we're doing this. In addition to this little reward that we're offering, we're also going to do X, Y, and Z to help you." – Key Informant

Improving the implementation of P4P and of performance metrics more generally in the VHA was a common theme among KIs. Related to implementation of performance metrics generally, KIs felt that some of the performance metrics are "opaque," and that the rollout of new metrics (as well as existing ones) has largely lacked adequate interpretation (including purpose and intent), support documentation, and technical assistance. In contrast, KIs believed that performance metric implementation should be considered within the context of existing performance metrics, should be risk-adjusted, and should be accompanied by the financial, human, and other resources necessary to achieve them – such as a menu of approaches to improve performance, technical assistance, feedback and implementation support, and opportunities to collaborate and share ideas with people from other facilities and VISN networks.

Implementation should be accompanied by support documentation
In other words, it's not, "Here's the measure. Here is how the numerator and the denominator are defined. Here's your goal." Maybe there is some description of it, but packaging it with resources for responding and how to respond to the measure. I know that possible and actionable responses will vary from site to site… On a national level, it's not going to be easy to generate that list, but it's something to consider including in a package along with the performance measure that helps decrease the cognitive load for front line clinicians to have an idea of where to start. – Key Informant

There was strong consensus across KIs that P4P be implemented within the context of a larger quality improvement program. KIs felt that other important components were frequent and constructive audit and feedback, and public reporting – and that public reporting of P4P metrics in particular provides additional incentives both at the provider (to improve performance) and the administrative levels (to select valid, attainable, and evidence-based metrics). KIs also stressed that performance pay be framed positively, and they discussed the importance of creating a culture of learning, performance improvement, and quality.

Transparent, bottom-up, goal-aligned, and achievable
[An important factor is] transparency about the process and the criteria that are chosen. I think that also there should be some abilities for the line staff to help choose, shape, and mold the criteria being used. I think that would be the best thing. Say, "Here's the a-la-cart menu. I'm going to propose we choose several of the following because we as a facility aren't doing as well as other places." Then we have people talk about that. You want the criteria to align with big VA goals that the Secretary and other leaders set. You have to align them with goals that you think are locally achievable based on the current climate at your facility. – Key Informant

Key informants also stressed that to succeed, providers must be given a "pathway to success," which would and should differ by facility. This pathway might require additional resources, such as structural or technical improvements, and it may determine who should be incentivized. KIs also suggested coordinating national-level priorities with context-driven local needs by implementing the type of metrics that would indicate whether the local infrastructure is able to meet national-level goals.

KIs felt that VHA performance metrics have often been implemented from the top down, and that going forward, facility leadership should consider a bottom-up approach. They felt that providers and other staff should be included in decisions about which metrics to incentivize, or that at the very least P4P should be implemented in a manner through which people perceive organizational and procedural justice – that the process and outcomes are "fair."

Related to the VHA's current P4P program, there was strong consensus that more transparency is needed. Regardless of VISN or division, KIs who were providers could not identify the performance metrics linked to their pay. Finally, a number of KIs raised issues related to the reliability of the incentive – with numerous KIs stating that the budgetary allocation for performance pay has been unpredictable, resulting in uncertainty about whether incentives would be paid even if metrics were achieved.

Providers don't know what their pay is linked to
For me, when I get money is random. Because of that I try to do a good job in general, but there's not necessarily a strong tie between that money and my performance. If those dots were more connected it would maybe make my behavior different. – Key Informant

P4P should be reliable
Another thing that happens with my incentive is that you're supposed to be eligible for some X% of your salary below some limit for incentives. My local VA has complete discretion over that depending on what their budget looks like, but when they pay me my incentive, they pay me some arbitrary amount of some total amount of what I could've gotten. It diffuses it even more. It makes it seem arbitrary, so why bother? – Key Informant

Provider Cognitive, Affective, and Behavioral Responses

Petersen and colleagues' RCT's audit and feedback process included an assessment of physician goal commitment to providing guideline-concordant hypertension care. They found no difference in goal commitment when comparing incentivized and non-incentivized physicians.[73] Furthermore, commonly reported barriers were patient non-adherence and inconsistent follow-up, suggesting that providers may perceived an external locus of control in managing hypertension.

Findings from Key Informant Interviews

A number of KIs expressed the belief that physicians are intrinsically motivated, and that for most, the driving factor in achieving evidence-based performance metrics that make clinical sense is intrinsic motivation. Some felt that P4P's extrinsic motivation to meet metrics that have less provider "buy-in" (*eg*, don't make clinical sense to providers) may be interpreted as reducing their autonomy, and as contrary to their values. One KI offered an idea for an alternative that could potentially be both extrinsically and intrinsically motivating:

Intrinsic and Extrinsic Motivation
If I could be offered an incentive for meeting quality metrics that would generate a donation to a Veteran service organization or to Veteran families in need. Because VA is a really important safety net provider in many communities, could there be a way that instead of that $1,000 going to me, could it be going towards Veterans in need? I think that potentially could be more attractive to providers. If you think you need to do something extrinsic, that could be a nice extrinsic motivation, which at the same time link to the autonomous motivations of providers who choose to work in the VHA. That would really be kind of a win-win that way. There are not many incentives that offer a carrot but also speak to someone's inner motivations. It just doesn't happen very often because it's usually one or the other. I think that if there could be creative incentives that could be tested in that way, I think that could be very exciting. – Key Informant

Table 9. VHA Implementation Factors and Provider Affective/Cognitive and Behavioral Response – Summary of Findings

Type of Implementation Factor or Provider Response	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Incentive Target · Incentive Amount · Incentive Timing	Comparison	Detailed Findings	Findings Summary
Implementation Factors					
Audit and Feedback	Damschroder et al, 2014[20] · Qualitative · 62 leaders, clinicians, and staff from 4 VHA facilities · 2012	Diabetes · NA · NA · NA	Described the perceived impact of implementation of diabetes performance measures on management actions and day-to-day clinical practice.	Providers were frustrated by feedback reports that were dissociated from a realistic capability to address performance gaps.	Participants describe a range of issues related to implementation, including frustration about responsibility for performance gaps that are outside of their control.
	Hysong et al, 2006[75] · Mixed-methods · 102 facility directors, managers, and outpatient personnel at 6 VAMCs · 2001	Ambulatory Care · NA · NA	Explored the differences between high- and low-performing facilities in the use of clinical audit data for feedback purposes.	High-performing facilities provided timely, individualized, non-punitive feedback to providers, with more variability and a higher reliance on standardized facility-level reports found in facilities that were low-performing. Actionable feedback emerged as the core category from the data, around which timeliness, individualization, non-punitiveness, and customizability can be hierarchically ordered.	Facilities with a successful record of guideline adherence tend to deliver more timely, individualized, and non-punitive feedback to providers about their adherence than facilities with a poor record of guideline adherence.
	Hysong et al, 2014[74] · Qualitative · 48 clinicians, department heads, and facility leadership from 16 VAMCs · NR	Primary Care (PACT) · NA · NA · NA	Described participants' experiences with audit and feedback in team-based primary care settings, with the goal of identifying differences in processes after PACT implementation.	Ownership of clinical performance still rested largely with the provider. A panel-management information tool emerged as the most prominent change to feedback dissemination. Existing tools were seen as most effective when monitored by the nurse members of the team. Facilities reported few, if any, appreciable changes to the assessment of clinical performance after PACT implementation.	Audit and feedback processes remained largely unchanged.

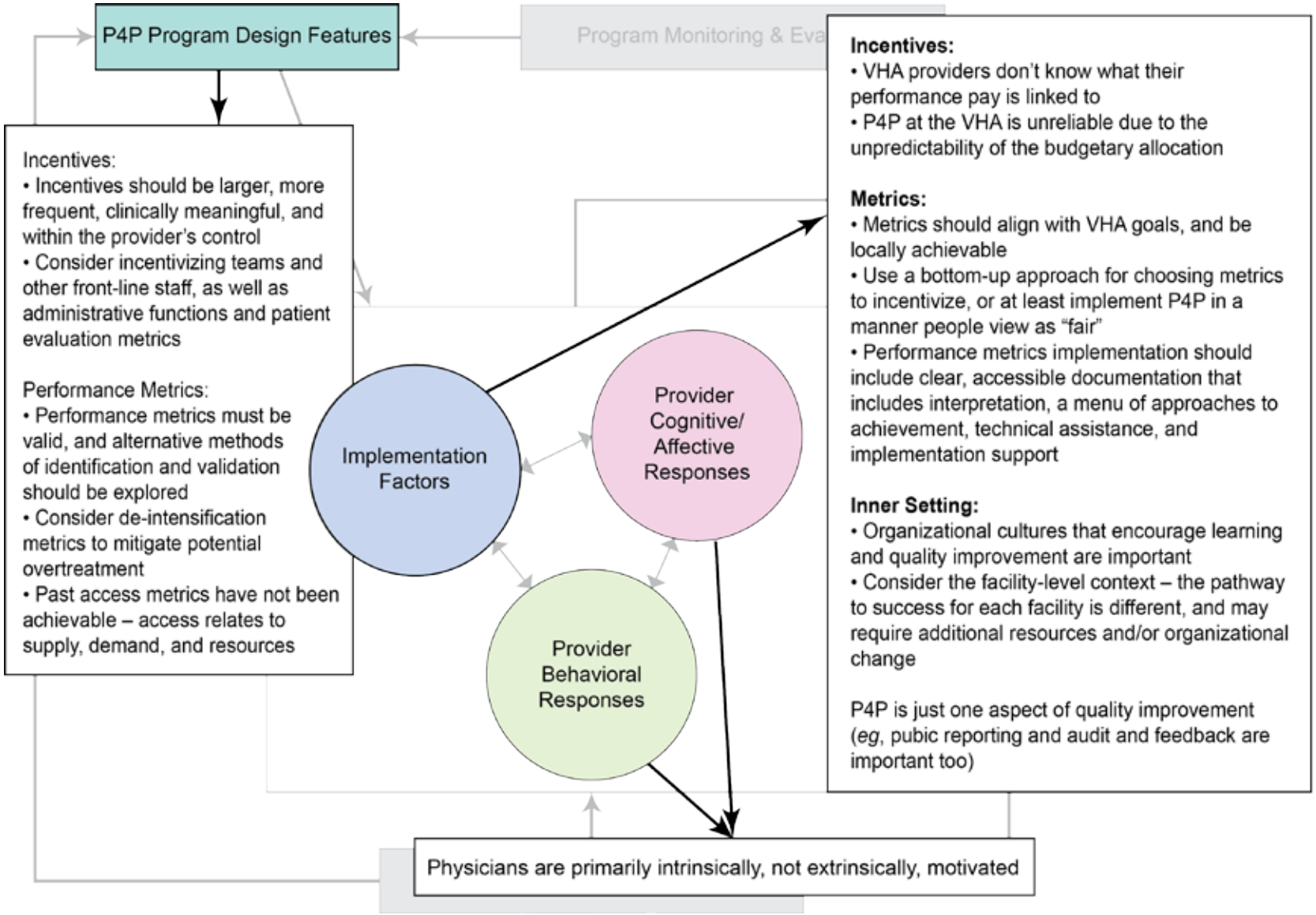
Type of Implementation Factor or Provider Response	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Incentive Target · Incentive Amount · Incentive Timing	Comparison	Detailed Findings	Findings Summary
	Petersen et al, 2013[17] · RCT · 83 physicians and 42 practice team members (*eg,* nurses, pharmacists) from 12 VAMCs · 2007-2011	Primary care: BP · Physicians and Group · $9.10 per measure, $4270 average group total, $1648 average provider · 4-month performance period (up to 5 payments)	Compared P4P to control providers	A significantly larger percentage of providers receiving P4P accessed their feedback reports via website, (67% P4P vs 25%; p=.001)	A significantly larger percentage of providers receiving P4P accessed their feedback reports
Incentive: Implementation	Damschroder et al, 2014[20] · Qualitative · 62 leaders, clinicians, and staff from 4 VHA facilities · 2012	Diabetes · NA · NA · NA	Described the perceived impact of implementation of diabetes performance metrics on management actions and day-to-day clinical practice.	Providers did not have a clear understanding of how their P4P was calculated.	Participants describe a range of issues related to implementation, including the lack of transparency about performance pay.
Incentive: Removal	Benzer et al, 2014[14] · Uncontrolled before/after · 128 VAMCs · 2004-2010	Acute care: acute coronary syndrome, heart failure, and pneumonia · VISN and facility senior managers · Bonuses (amount NR) · NR	Compared before/after incentive removal (up to 3 years)	Up to 3 years after removal of the incentive, latent growth models showed that performance was sustained for all 7 metrics (related to acute coronary syndrome, heart failure, and pneumonia), with no significant positive or negative slope (however weight monitoring showed a significant positive slope in the year following removal, then a significant negative slope the following year, and a non-significant slope in year 3 following removal).	Performance on quality of care metrics were maintained up to 3 years after incentive removal.
	Petersen et al, 2013[17] · RCT · 55 Physicians	Primary care: BP · Physicians and Group · $9.10 per measure,	Compared P4P to control providers	There was a significant reduction in performance.	There was a significant reduction in performance.

Type of Implementation Factor or Provider Response	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Incentive Target · Incentive Amount · Incentive Timing	Comparison	Detailed Findings	Findings Summary
	· 2007-2011	$4270 average group total, $1648 average provider · 4-month performance period (up to 5 payments)			
Metric: Implementation	Damschroder et al, 2014[20] · Qualitative · 62 leaders, clinicians, and staff from 4 VHA facilities · 2012	Diabetes · NA · NA · NA	Described the perceived impact of implementation of diabetes performance metrics on management actions and day-to-day clinical practice.	Local performance metrics/clinical reminder targets aren't always consistent with the evidence, the intent of the metric, and organizational priorities, and metric documentation and specifications aren't always followed.	Participants describe a range issues related to implementation, including metric implementation that is not consistent with evidence or organizational priorities.
Metric: Monitoring	Hysong et al, 2011[68] · Retrospective cohort · 133 VA Medical Centers · 2000-2008	Ambulatory care: screening, immunization, chronic care following AMI, diabetes, hypertension · VISN and facility senior managers (actively monitored measures only) · Bonuses (amount NR) · NR	Compared performance on metrics related to changes from passive monitoring (support indicators) to active monitoring (performance targets) or vice versa. Actively monitored measures may be incentivized.	Performance on all metrics improved or remained stable over time regardless of whether they changed from actively assessed to passively monitored or vice versa. After risk-adjusting for organizational characteristics, no organizational characteristics effects were found. 2/6 measures that changed from passive to active had significantly increased slopes after the change (HbA1c levels < 9, and colorectal cancer screening), indicating significant improvement in performance. 4/11 that changed from active to passive exhibited significant differences in slope; 2 exhibited positive slopes before the change, followed by negative slopes (lipid profile every 2 yrs; MDD screening), and 2 exhibited the opposite pattern (diabetic foot	Performance on all metrics improved or remained stable over time regardless of whether they changed from actively assessed to passively monitored or vice versa.

Type of Implementation Factor or Provider Response	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Incentive Target · Incentive Amount · Incentive Timing	Comparison	Detailed Findings	Findings Summary
				inspections; and pedal pulses). Remaining measures exhibited no significant changes, indicating sustained performance after changing from performance measure to support indicator.	
Metric: Transparency	Kansagara et al, 2014[19] · Qualitative · 241 Clinical and Administrative Staff from 15 VHA Primary Care Clinics · December 2010 and February 2013	Primary Care · NA · NA · NA	Described primary care staff experiences with the use of performance metrics during the implementation of the VHA PACT model of care.	Primary care staff perceived that performance metrics were imposed from the "top down," with little communication, transparency, or training.	Primary care staff perceived performance measures as being implemented from the "top down."
Provider Affective/Cognitive and Behavioral Response					
Goal commitment	Hysong et al, 2012[73] · Cross-sectional Petersen et al, 2013[17] sub-study · 83 physicians from 12 sites · 2007-2011	Primary care: BP · Physicians and Group · $9.10 per measure, $4270 average group total, $1648 average provider · 4-month performance period (up to 5 payments)	Compared provider goal commitment as part of the audit and feedback process by arm: 1) audit and feedback only, 2) physician incentives + audit and feedback, 3) group incentives + audit and feedback, 4) physician and group + audit and feedback.	There were no differences between goal commitment when comparing providers receiving financial incentives vs audit and feedback alone. In addition, patient non-adherence and inconsistent follow-up were cited as barriers to care.	No differences in goal commitment were found across all study arms or over time. Physicians may perceive an external locus of control for hypertension care.

Abbreviations: BP = blood pressure, MDD = Major Depressive Disorder, NA = not applicable, NR = not reported, PACT = Patient Aligned Care Team, P4P = pay-for-performance, RCT = randomized controlled trial, VAMC = Veterans Affairs Medical Centers, VHA = Veterans Health Administration, VISN = Veterans Integrated Service Network

Figure 5. Key Informant Interviews: Themes – Program Features and Implementation Factors in VHA Settings



Note. Implementation Factors include implementation processes; outer setting; inner setting; and provider characteristics. Abbreviations: P4P = pay-for-performance, VHA = Veterans Health Administration

Program Features and Implementation Factors of Pay-for-Performance in Community Settings

We identified 5 studies examining pay-for-performance or related implementation factors in Veteran populations in community settings.[76-80] In general, studies found that a number of survey instruments examining cross-system access and coordination exist,[78] and that Veterans, providers, and administrators expressed concern that the Veterans Choice Program (VCP) had resulted in fragmented care for patients, poor communication and coordination amongst providers, and that it placed an additional burden on VHA providers. Other concerns included barriers to sharing medical records,[77,79,80] and differences between providers interested in VCP participation and those who are not (see Table 10).[76]

In key informant interviews, a number of themes related to the use of P4P in the community as part of the Veterans Choice Program arose. Common themes related to identifying the overall goal of the VCP and the role of P4P, the types of metrics the VHA might consider incentivizing, tools to assist with care coordination, and suggestions for ways to ensure quality – such as how to choose providers/who to incentivize. In addition, KIs voiced concern for a number of potential challenges to implementing P4P in community settings.

Detailed Findings

Program Design Features

One study examined metrics relevant to community care. To address the question of patient experience with access and care coordination across multiple settings, researchers identified 31 existing survey instruments representing 6 access and 5 care coordination domains relevant to cross-system care. Domains frequently assessed included follow-up coordination, primary care access, cross-setting coordination, and continuity. They also interviewed a small sample of Veterans about their experience with community care. In addition to domains included in the existing instruments, Veterans also raised issues related to 1) acceptability of distance to care site given a patient's clinical situation, 2) burden on patients to access and coordinate care and billing, and 3) provider familiarity with Veteran culture and VHA processes.[78]

Differences between care in VHA and community settings

Internally, we have a fairly strong sense of values and effort, finding the things that are going to make the biggest difference and not going for things just to do them, but doing things because we know they are going to make a difference. Those are bigger factors than having a performance measure on, "what's your blood pressure?" Those are factors where VA care differs more from outside care…I think in general, if I were leading this, my first thought would be, "what is the clinical care that we want to change? What kind of care in the community would we want for it to be more like VA?" You start with that clinically, then you think, "what are the things we might do to influence that?" – Key Informant

Findings from Key Informant Interviews

A number of KIs stressed the importance of linking P4P in the community to the VHA's organizational and program goals – thus, the importance of considering what types of metrics to incentivize in the community based on the overarching goal of the VCP. Prioritizing increased access to healthcare for Veterans was clear to KIs. However, questions arose with regard to other important goals. Some suggested simply the receipt of quality care, while others felt strongly about coordination of care, cost effectiveness, and "conservative care" (*eg,* careful selection of surgical patients). Some KIs

wondered whether a goal of the VCP is or should be, to extend the type of care offered by the VHA to Veterans seeking care in the community, and if so, one KI suggested using the known differences between the VHA and community care as a guide to metric selection.

In terms of potential metrics to incentivize in the community, a number of KIs suggested that, particularly in the early stages of a program, P4P may be a strategy to address current challenges in the receipt and quality of patient records and documentation from community providers. Also, common among KIs was the notion that timely access, and coordination of care may be a good starting point. A few KIs also suggested that incentivizing guideline-based metrics at the (pooled) population level by comparing Veterans receiving care in VHA settings to those receiving care in the community may broadly address the question of the overall quality of care received by Veterans using the VCP, with the understanding that population-based incentives are unlikely to motivate behavior at the provider level.

Consider incentivizing documentation
One of the problems with Choice is that the records that we get back from the other health care systems aren't very detailed. They provide basic information about billing but not much about the clinical care that's been provided... One aspect of P4P might be in regards to getting good records…For example, for diabetes you'd want them to provide the tests being provided, the dates they were provided, and the data values for those lab tests. – Key Informant

Another question on the minds of KIs was how to choose community providers in a way that ensures quality care for Veterans. KIs stressed the importance of building relationships both at the national and local levels – that VHA providers should be talking to community providers. They also suggested using established networks, contracting only with board-certified physicians, and using providers' performance on already-established metrics, such as those reported publicly by the Centers for Medicare and Medicaid Services, (CMS) as a method of selection.

Implementation Factors

A cross-sectional study compared characteristics and attitudes of community-based primary care and mental health providers reporting interest or no interest in VCP participation during early implementation, and examined perceptions and experiences of the VCP among "early adopters". They found that providers who were Veterans were more likely to be interested in VCP, and that there were significant differences between interested and non-interested providers, including the willingness to accept Medicare rates and the willingness to submit copies of medical records.[76]

A qualitative study of pharmacists responsible for filling VCP prescriptions found that common challenges related to 1) non-formulary prescriptions, 2) opioids, and 3) HCV treatment – including lack of knowledge of VHA formularies, poor coordination and patient management, differences in standard and quality of care, and additional VHA burden.[77] Other qualitative studies examining experiences with community care found that patients, as well as VHA providers and administrators reported poor coordination and fragmented care, poor communication, and challenges related to the sharing of data/medical records.[79,80]

Findings from Key Informant Interviews

Key informants generated ideas for a number of quality improvement strategies to accompany P4P for community providers. They stressed the importance of transparency in P4P programs generally, and felt that the performance of community providers should be publicly reported. In addition, they voiced concern about coordination and quality of care, particularly given the lack of shared electronic health records, and suggested developing resources and tools to streamline the process for community providers. Suggestions included developing a system to allow community provider access to the same pop-up reminders available within the VHA, and providing access to VHA formularies – either through the development of a new mobile app, or the use of existing tools, such as Epocrates®.

Differences between care in VHA and community settings
Internally, we have a fairly strong sense of values and effort, finding the things that are going to make the biggest difference and not going for things just to do them, but doing things because we know they are going to make a difference. Those are bigger factors than having a performance measure on, "what's your blood pressure?" Those are factors where VA care differs more from outside care…I think in general, if I were leading this, my first thought would be, "what is the clinical care that we want to change? What kind of care in the community would we want for it to be more like VA?" You start with that clinically, then you think, "what are the things we might do to influence that?" – Key Informant

Limited Veterans per provider
As the VA becomes more like an insurance company, we need to start thinking like an insurance company. – Key Informant

A number of KIs discussed differences between Veterans and the general population – largely noting lower socioeconomic status (SES), mental health, substance use, and rural residence. KIs felt it important to account for SES when implementing P4P, and expressed concern for access and the availability of quality care for Veterans living in rural areas.

Limited Veterans per provider
Veterans are one percent of their patient population. Providers in the community are often working with 10 different insurers at once or more and the VA will literally probably be their smallest for a lot of them. For us to then say, "this is how you should practice differently" is a lot to ask under any circumstance, especially considering how poor the roll out has gone already. – Key Informant

In general, most KIs voiced concern for challenges the VHA might face in implementing P4P in community settings. Most commonly, KIs worried that Veterans accessing care through the VCP would be dispersed widely; thus, comprising a small percentage of any provider's patient population, and that the VCP would simply become one of – and for many providers the smallest of – insurers they are working with. This in turn may both reduce the likelihood of any potential impact of a VHA P4P program (unless the incentive is large enough to stand apart from the rest), particularly if incentivized metrics differ from other programs, and would likely result in providers with so few VCP patients that measured performance would vary widely, resulting in unreliable measures of quality. KIs reiterated the potential for incentives related to access or data, as well as population-based incentives, and suggested aligning incentivized metrics with larger P4P programs. Other KIs

Fragile relationship
You have to make sure that if you put these carrots out for Choice performance pay and the implementation of that program has a lot of wrinkles in it and the providers can never achieve to the point where they pay that money and they drop the VA, well than we have less options for community care and that is a detriment to our patients. That would be my biggest concern. – Key Informant

discussed the potential tradeoffs of utilizing narrow networks to increase the percentage of VCP patients per provider, and access to high-quality care, particularly for rural Veterans.

KIs were also concerned about the challenge presented by the already fragile relationship the VHA may have established with some community providers, citing slow payment and providers refusing Veteran patients as a result. They cautioned that the VHA should proceed with care, pay providers in a timely fashion, and reiterated the need for P4P to be achievable, fearing that an increased number of providers opting out may result in even poorer access for Veterans.

Challenges related to mental health
The culture in community mental health is "Big walls that are impermeable." They don't let data out. Its 2017 and there are people that are still handwriting their therapy notes. And of course, why wouldn't they be? It makes sense if you've worked in the field, but it would make sense to no other health care provider. There are very unique challenged on implementing CHOICE and being able to ensure that the health care provided is of high quality. – Key Informant

Mental health treatment was a concern for a number of KIs as well. They feared that sending Veterans to community providers for mental health services will reduce coordination and the quality of care received, particularly for those with combat-related PTSD, substance use disorders, and those who are experiencing homelessness. In addition, KIs were concerned that implementing and incentivizing metrics will present a barrier to entry because performance metrics are uncommon in community mental health settings, and in addition, obtaining treatment notes and other records from providers will be challenging.

Finally, KIs were concerned about the impact of VCP on current patients and VHA providers – that in time, resources may be diverted from Veterans receiving care in VHA settings, and that the VCP may influence the ability for VHA providers to maximize their own performance pay.

Community P4P may affect VHA P4P
In terms of VA providers and how Choice would influence their ability to achieve max performance pay. From a VA provider's perspective, you have to be somewhat mindful of the fact that with our current implementation of Choice, we have several times where there is failure to launch either because something gets dropped in the HealthNet referral process or we don't get the records. If more and more Choice is going to be used into the future and Choice is going to be used for things that end up being criteria for [VHA provider] P4P, then I think that facility leaders have to be mindful of putting people in situations where they can't succeed. – Key Informant

Provider Cognitive, Affective, and Behavioral Responses

KIs voiced concern for unintended consequences resulting from P4P in community settings – particularly overtreatment and overuse. They felt that not only might community providers be subject to the same dynamics that result in potential overtreatment in VHA settings, but that overtreatment may be even more common in the community, and that the lack of integration and coordination with VCP might place Veterans at increased risk.

Potential overtreatment
Even while we're building access for other patients, there are major overuse problems out in the private sector. I worry that we're opening the floodgates here a bit. I've seen that with a number of my patients where they've just gotten a number of things they don't need for a variety of reasons. The VA, because we are an integrated system, have been able to keep a pretty good explicit and implicit check. So, if I'm ordering a very (expensive) cat scan and I know someone can't get it for a month anyways, I may just not order it because it's just not worth waiting that long. These checks in the system for overuse in the VA, along with other hard stops that help prevent overuse, if we just send people out into the private sector I just worry that we're going to fuel that problem. – Key Informant

Table 10. Community Program Design Features and Implementation Factors – Summary of Findings

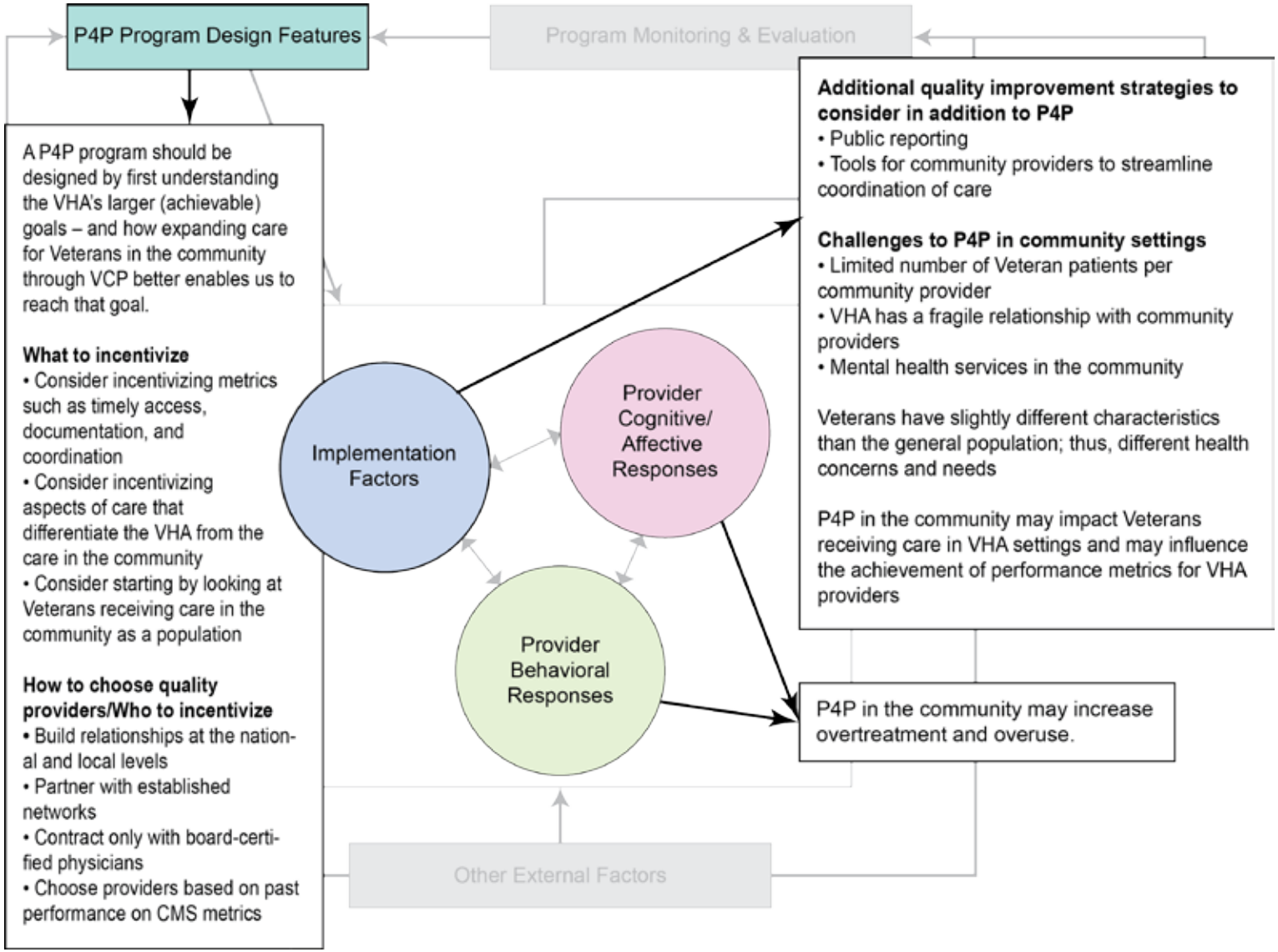
Type of Program Design Feature or Implementation Factor	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Description of Sample	Comparison	Detailed Findings	Findings Summary
Program Design Features					
Metrics: Development	Quinn et al, 2017[78] · Qualitative and literature search · 10 Veterans from a single VMAC who had been offered Community Care · NR	General · 8 of 10 Veterans accessed community care through VCP. Two opted for care in a VHA setting.	Map of existing patient experience survey instruments relevant to access and coordination of care across care settings, and a pilot evaluation of patient experiences conducted to help inform VHA development of a patient experience survey for Veterans eligible for the VCP.	31 existing survey instruments representing 6 access and 5 care coordination domains relevant to cross-system care were identified. Domains frequently assessed included follow-up coordination, primary care access, cross-setting coordination, and continuity. In addition, Veterans raised issues related to 1) acceptability of distance to care site given patient's clinical situation, 2) burden on patients to access and coordinate care and billing, 3) provider familiarity with Veteran culture and VHA processes.	Existing survey instruments assess many aspects of patient experiences with access and care coordination in cross-system care. However, additional factors such as patient burden and Veteran culture and VHA processes should be considered.
Implementation Factors					
Implementation Processes	Gellad et al, 2017[77] · Qualitative · 27 VHA Pharmacists · September 2015 to October 2015	Pharmacy · Pharmacists responsible for filling VCP prescriptions	Sought to understand the barriers and facilitators to safe and effective medication dispensing under VCP.	Three themes emerged: 1) non-formulary prescriptions from VCP providers places a burden on pharmacists to educate providers about the formulary, and may result in delayed care due to difficulty reaching VCP providers and lack of a streamlined process, 2) challenges related to opioid prescriptions (*eg,* communication/education with VCP providers, Veterans needing to hand-deliver prescriptions, differences in standards of care for opioid prescribing), 3) concerns for inconsistent HCV care, medication adherence, poor coordination and patient management.	Pharmacists filling VCP prescriptions described challenges related to 1) non-formulary prescriptions, 2) opioids, and 3) HCV treatment, including lack of knowledge of VHA formularies, poor coordination and patient management, differences in standard and quality of care, and additional burden.
Implementation Processes	Tsai et al, 2017[79] · Qualitative · 38 Veterans and	HCV · Patients and providers involved	Evaluated the implementation of the VCP and Choice First for	Themes included difficulties with the approval, coordination, and communication. Patients and Providers	Veterans and VHA providers report challenges with Choice

Type of Program Design Feature or Implementation Factor	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Description of Sample	Comparison	Detailed Findings	Findings Summary
	10 VHA Providers from 3 VAMCs · October 2015 to May 2016	in the VCP's First Choice Initiative	HCV treatment.	felt they didn't have a choices or control over care, with additional burden placed on the VHA provider. Patients and providers reported fragmented care, and providers reported barriers to sharing medical records. Other themes included provider reservations about community providers, and reservations about community providers having the expertise to meet the needs of unique Veteran populations.	First for HCV, including fragmented care, barriers to sharing medical records, increased burden on VHA providers, and concerns about the unique needs of Veterans being met in the community.
Implementation Processes	Zuchowski et al, 2017[80] · Qualitative · 23 VHA Women's Health providers, staff, and administrators at 15 VAMCs · NR	Women's Health Gynecologists, Women's Health Medical Directors, and other staff responsible for coordinating care with the community from 5 facilities with 0 FTEs, 5 with part-time FTE, and 5 with ≥1 full-time FTEs	Examined challenges to women's health VHA and community care coordination.	Key informants reported substantial challenges in coordinating women's health care between VHA and community settings. KIs reported that care was fragmented, that there is a lack of provider role clarity and tracking, poor VHA/community provider communication, and that challenges with health record exchange (*eg,* not timely, incomplete) results in a significant barrier to coordination. KIs provide suggestions such as the establishment of roles such as a "care tracker," a provider point of contact, a patient liaison, and a records administrator.	Women's health providers, administrators, and staff report significant barriers to coordinating VHA and community care, such as fragmented care, poor communication, lack of role clarity, and challenges with medical record exchanges.
Provider Characteristics	Finley et al, 2017[76] · Cross-sectional · 553 Community and 115 VCP Authorized Providers in TX and VT · Summer/Fall 2015	Mental Health: PTSD · Prescribing and psychotherapy providers. VHA authorized providers were enrolled in either VCP or PC3.	Compared characteristics and attitudes of community-based primary care and mental health providers reporting interest or no interest in VCP participation during early implementation; and to examine perceptions and	Interest in VCP participation was associated with factors including being a Veteran and receiving VA reimbursement. Community and VCP authorized samples reported not knowing enough about the VCP, being unwilling to deal with claims/billing, and feeling that the reimbursement rate was inadequate. Of note, there was a significant difference between those interested and not	Provider interest in the VCP was associated with factors such as being a Veteran. Community and providers cite reasons for unwillingness to participate in VCP. Significant differences between interested and

Type of Program Design Feature or Implementation Factor	Study · Study Design · Sample Size · Observation Period/Follow-up	Program Focus · Description of Sample	Comparison	Detailed Findings	Findings Summary
			experiences of the VCP among "early adopters."	interested in VCP participation in willingness to both accept Medicare rates and submit copies of medical records for the VHA record -- and even among those interested in VCP less than 65% were willing to submit records. Among the 21 providers who were attempting to become a VCP provider, mean satisfaction with the process was 4.85/10. Mean satisfaction among the 12 current VCP providers was 6/10.	non-interested providers include willingness to accept Medicare rates and willingness to submit copies of medical records.

Abbreviations: HCV = Hepatitis C Virus, NR = Not Reported, PC3 = Patient-Centered Community Care, SHEP = Survey of Healthcare Experiences of Patients, TX = Texas, VCP = Veterans Choice Program, VHA = Veterans Health Administration, VT = Vermont

Figure 6. Key Informant Interviews: Themes – Program Design Features and Implementation Factors in Community Settings



Note. Implementation Factors include implementation processes; outer setting; inner setting; and provider characteristics. Abbreviations: CMS = Centers for Medicare and Medicaid Services; P4p = pay-for-performance; VCP = Veterans Choice Program; VHA = Veterans Health Administration.

KEY QUESTION 5: What novel approaches and/or current or recently closed research projects funded by VA examine the effectiveness, implementation factors, or unintended consequences associated with pay-for-performance in Veteran populations?

We identified no novel approaches to P4P being tested in clinical VHA settings. However, we identified 23 current and recently closed (2016 – present) projects, initiatives, and programs funded by VA (see Table 11). To our knowledge, only the Partnered Evidence-Based Policy Resource Center (PEPReC) is currently engaged in work directly related to P4P. Along with the Office of Community Care, they are developing performance standards for P4P in community-based treatment for opioid use disorder, and in addition, are designing a randomized evaluation of the resulting pilot program. All identified Quality Enhancement Research Initiative (QUERI) activities and one additional project (SDR 17-157, Wagner, T) relate to community care.

Table 11. Ongoing and Recently Closed VA-funded Studies and Programs

Project Number	Title	Principal Investigator	Location	Funding End
HSR&D and QUERI	Partnered Evidence-Based Policy Resource Center (PEPReC)	Frakt, A	Boston, MA	NR
QUERI	Factors Affecting Choice Act Implementation and Quality for Veterans with PTSD	Finley, E	San Antonio, TX	NR
QUERI	Triple Aim QUERI Program	Ho, Michael	Denver, CO	NR
QUERI	A Mixed-Method, Multi-Site Evaluation of the Implementation of the Veterans Choice Act	Ho, Michael	Denver, CO Cleveland, OH Seattle, WA	NR
QUERI	Ensuring Quality and Care Coordination in the Era of Veterans Choice	Kerr, E	Ann Arbor, MI	NR
QUERI	Examining VACAA Implementation and Care Coordination for Women Veterans	Mattocks, K	Northampton, MA	NR
QUERI	Partnered Evaluation of the Clinic Management Training Program/Center for Access Policy Evaluation and Research (CAPER)	Prentice, J	Boston, MA	NR
QUERI	The Care Coordination QUERI: Improving Patient-Centered Care Coordination for High Risk Veterans in PACT	Rubenstein, L	Los Angeles, CA	NR
QUERI	Differences in Satisfaction with Choice: Laying the Foundation for the Evaluation of the Choice Act	Zickmund, S	Pittsburgh, PA	NR
CRE 12-023	SUD Treatment Staffing and Handbook Implementation: Impact on Patient Outcomes	Frakt, A	Boston, MA	5/31/2016
PPO 14-372	Measuring Quality of Palliative Care for Patients with End Stage Liver Disease	Walling, A	West Los Angeles, CA	9/30/2016

Project Number	Title	Principal Investigator	Location	Funding End
IIR 15-438	Improving the Measurement of VA Facility Performance to Foster a Learning Healthcare System	Petersen, L	Houston, TX	8/31/2017
IIR 12-383	Linking Clinician Interaction and Coordination to Clinical Performance in VA PACT	Hysong, S	Houston, TX	9/30/2017
CRE 12-035	Identifying and Delivering Point-of-care Information to Improve Care Coordination	Petersen, L	Houston, TX	9/30/2017
PPO 16-100	Can Electronic Data and Natural Language Processing Accurately Reproduce a Surgical Quality Measure?	Richman, J	Birmingham, AL	1/31/2018
SDR 17-157	Planning for a New Era in Veterans Health Care: Community Care, Information Exchange and Multi-system Use	Wagner , T	Palo Alto, CA	9/30/2018
IIR 14-082	Glycemic Control: Overtreatment, Hypoglycemia, Mortality and De-Intensification	Tseng, C-L	East Orange, NJ	2/29/2019
IIR 15-131	Identifying, Measuring, and Facilitating Opportunities for De-intensification of Medical Services	Kerr, E	Ann Arbor, MI	10/31/2019
IIR 15-436	Using Data Integration and Predictive Analytics to Improve Diagnosis-Based Performance Measures	Hoggatt, K	Sepulveda, CA	12/31/2019
IIR 15-292	Risk-Adjusting Hospital Outcomes for Veteran's Socioeconomic Status	Trivedi, A	Providence, RI	1/31/2020
IIR 14-345	Incorporating Treatment Outcomes into Quality Measurement of Depression Care	Pfeiffer, P	Ann Arbor, MI	3/30/2020
IIR 15-432	Developing Benefit-Based Performance Measurement for VHA	Sussman, J	Ann Arbor, MI	6/30/2020
IIR 15-438	Improving the Measurement of VA Facility Performance to Foster a Learning Healthcare System	Petersen, L	Houston, TX	8/31/2020

Abbreviations: PACT = Patient Aligned Care Team, SUD = Substance Use Disorders, VA = Veterans Administration, VHA = Veterans Health Administration

In the course of the project, we also learned of 2 relevant unpublished studies, examining 1) the impact of the difficulty of coordination, documentation, and accomplishment of performance metrics on clinical outcomes (Sylvia Hysong, PhD, phone call, May 19, 2017) and 2) whether the mental models individuals and organizations hold regarding audit and feedback serve to moderate its effectiveness (Sylvia Hysong, PhD, email, September 9, 2017), as well as a proposal for research to assess the feasibility, accuracy, and interpretability of template matching

versus conventional regression approaches for comparing quality of care across the VHA's diverse settings (Hallie Prescott, email communication, August 2, 2017).

SUMMARY AND DISCUSSION

We examined 68 articles and conducted interviews with 17 key informants to help inform the implementation of pay-for-performance programs for Veterans in the VHA and in community settings. While we found insufficient evidence to determine whether and how much P4P affects Veteran outcomes, we did find information in the literature and through KI interviews that may help guide the implementation of P4P and maximize potential benefits while minimizing negative unintended consequences

P4P programs in published literature have focused primarily on incentives for process of care/quality metrics. Other than intermediate targets such as blood pressure and cholesterol for cardiovascular health and diabetes, we identified very few incentivized patient outcomes.

PAY-FOR-PERFORMANCE IN VHA SETTINGS

Several themes related to general issues with P4P in the VHA emerged from key informant interviews that are consistent with the findings from published literature (see Table 12):

- *Regardless of whether performance metrics are incentivized, they should be valid, achievable, and within a provider's control.* Consistent with previous findings the importance of evidence-based metrics,[3,4] key informants felt that performance metrics should be valid and well-designed and cited a need for further research evaluating alternate validation methods.
- *Potential overtreatment and overuse may be an unintended consequence of performance metrics, and de-intensification metrics should be considered.* Findings from a handful of included studies[21,22,24-26] coupled with concerns voiced by key informants, suggest that potential overtreatment and overuse may be an unintended consequence of performance metrics, regardless of whether they are incentivized – and particularly in VHA facilities that are metric-oriented.
- *Consider re-evaluation of the size (monetary), frequency, and target (provider versus team) of performance pay in the VHA.* Consistent with research examining P4P in the VHA,[20] provider KIs consistently stated that they did not know which metrics were incentivized and did not feel that the current structure influences their behavior. Additionally, despite a study that found otherwise,[17] a number of KIs proposed that the VHA consider implementing team-based incentives and incentives for other front-line staff.
- *Use a transparent, bottom-up approach for selecting and implementing metrics, and secure provider and staff buy-in.* Despite previous research stressing the importance of bottom-up, realistic metrics,[3,4] VA staff describe performance metrics as poorly implemented in the VHA,[20] and voice frustration with the current top-down methods.[19] There was strong consensus among KIs that incentivized metrics be achievable and accompanied by the local resources necessary for achievement, that decisions regarding

what to incentivize are perceived as equitable, and that incentive payments are predictable and reliable.

- *Foster overall and local-level cultures that encourage learning and value quality improvement.* Included studies found that metric-driven cultures were more prone to potential overtreatment.[21,22] Similarly, one KI said, "We think of the VA as a learning organization, so a pathway is learning to do better and learning to do things that fit with the goals of the VA as an institution. You have to figure out a way of developing the learning organization in all of our facilities. If we improve that they can be successful."

- *Gaming will likely be mitigated by providing the resources support necessary for achievement.* According to one KI, "Inadequate resources, unrealistic expectations, and the opportunity to cheat all are factors in gaming," and another, "The reason you have things like gaming the system isn't because people don't want to do the right thing, it's because they can't do the right thing." Overtreatment may also be mitigated by incentivizing appropriate care, rather than treatment or targets, as demonstrated by the single included RCT.[17]

Table 12. Evidence and Policy Implications – P4P in VHA Settings

Implementation Framework Category	Study Evidence	Themes from KI Interviews	Policy Implications
Program Design Features	Ten studies examined program design features.[17-19,24,66,67,69-72] In general, studies found: · Physician-targeted incentives to be more effective than those targeting practices · The degree of agreement between EHR data and manual review varies by metric · The difficulty of achieving multi-tasked metrics is not directly related to the number of tasks involved · The relationship between access and patient satisfaction varies by measure and in new vs returning patients	Incentives: · Incentives should be larger, more frequent, clinically meaningful, and within the provider's control · Consider incentivizing teams and other front-line staff, as well as administrative functions and patient evaluation metrics Performance Metrics: · Performance metrics must be valid, and alternative methods of identification and validation should be explored · Consider de-intensification metrics to mitigate potential overtreatment · Past access metrics have not been achievable – access relates to supply, demand, and resources	· Regardless of whether performance metrics are incentivized, they should be valid, achievable, and within a provider's control · Consider re-evaluation of the size (monetary), frequency, and target (provider vs team) of performance pay in the VHA. · Potential overtreatment and overuse may be an unintended consequence of performance metrics, and de-intensification metrics should be considered.
Implementation Factors	The 8 studies[14,17,19,20,68,73-75] examining implementation factors found: · There was no difference in the achievement of actively	P4P is just one aspect of quality improvement (*eg*, public reporting and audit and feedback are important too)	· Use a transparent, bottom-up approach for selecting and implementing metrics, and secure provider and

Implementation Framework Category	Study Evidence	Themes from KI Interviews	Policy Implications
	vs passively monitored metrics · The evidence related to the impact of the removal of incentives on performance is mixed · Providers express frustration for current top-down implementation strategies, and suggest areas of improvement for implementing performance metrics at the local level · Facilities with high adherence to clinical guidelines were more likely to deliver more timely, individualized, and non-punitive feedback · Audit and feedback processes remained largely unchanged after PACT implementation	Incentives: · VHA providers don't know what their performance pay is linked to · P4P at the VHA is unreliable due to the unpredictability of the budgetary allocation Metrics: · Metrics should align with VHA goals, and be locally achievable · Use a bottom-up approach for choosing metrics to incentivize, or at least implement P4P in a manner people view as "fair" · Performance metrics implementation should include clear accessible documentation that includes interpretation, a menu of approaches to achievement, technical assistance, and implementation support Inner Setting: · Organizational cultures that encourage learning and quality improvement are important · Consider the facility-level context – the pathway to success for each facility is different, and may require additional resources and/or organizational change	staff buy-in. · Foster overall and local-level cultures that encourage learning and value quality improvement
Provider Cognition, Affect, and Behavior	One study examined provider affective/cognitive responses, and found that P4P had no impact on goal commitment.[73] Twelve articles[15-17,19-27] examined unintended consequences and found: · Administrative data support the potential for over-treatment associated with performance metrics. However, a RCT found no association between P4P	Physicians are primarily intrinsically, not extrinsically motivated Overtreatment: · There is great potential for overtreatment associated with PMs and P4P, particularly in metric-driven cultures, and with intermediate outcomes that vary (*eg,* blood pressure) · VHA should consider placing more emphasis on prevention (*eg,* lifestyle	· Potential overtreatment and overuse may be an unintended consequence of performance metrics, and de-intensification metrics should be considered. · Gaming will likely be mitigated by providing the resources support necessary for achievement

Implementation Framework Category	Study Evidence	Themes from KI Interviews	Policy Implications
	for hypertension and hypotension · There was evidence of denominator management associated with a VISN Director-aimed incentive · There was no evidence of risk selection · Providers perceived both negative and positive unintended consequences associated with performance metrics	counseling) and de-intensification Denominator Management: · Concerns related to denominator management – and challenges related to the denominator in general, given the subjective nature and variability in some diagnoses and treatment recommendations Risk Selection/Health Disparities · Low-SES Veterans in particularly, may be at risk for risk selection and disparities related to PMs and P4P Teaching to the Test/Attention Shift: · A variety of actively monitored, valid metrics covering different aspects of care/different populations may mitigate the potential for teaching to the test/attention shift Gaming: · History of gaming in the VHA – particularly within the context of P4P · To mitigate gaming, PMs should be accompanied by adequate resources and support (crucial when incentivized) · Differing viewpoints about composite measures to mitigate gaming – lack of transparency, goal of improving specific metrics rather than average performance	

Abbreviations: EHR = electronic health record, PACT = Patient Aligned Care Team, P4P = pay-for-performance, PM = performance metric, RCT = randomized controlled trial, VHA = Veterans Health Administration, VISN = Veterans Integrated Service Network

PAY-FOR-PERFORMANCE IN COMMUNITY SETTINGS

A number of themes related to the design and implementation of P4P in community settings also emerged (see Table 13).

- *Initially target areas in need of improvement such as documentation and coordination (eg, – receipt of records from community providers).* Overwhelmingly, KIs felt that given known challenges related to receipt of documentation,[79,80] data, and care coordination may be appropriate initial areas for P4P to target.

- *Develop relationships with providers and health systems with records of strong performance on commonly used, well-validated, and well-established metrics.* KIs stressed the importance of establishing relationships with local providers, and suggested a number of ways for selecting providers with track records of providing quality care.

- *The likely small number of Veteran patients per community provider may pose a challenge, both in terms of accurately assessing quality and the potential for an incentive to influence behavior. Consider beginning with alternate approaches, such population-based incentives.* Several KIs expressed concern aboutVA's ability to influence provider behavior using P4P or to accurately estimate quality at the provider level, given that Veterans will compose a small percentage of any one provider's patient population.

- *Use strategies such as public reporting to complement P4P.* Consistent with the findings from previous research,[3] KIs stressed that P4P is just one part of a quality improvement strategy.

- *Developing tools and resources to streamline the data-sharing and coordination necessary to inform a cross-system P4P program.* Similar to findings from included studies,[77,79] KIs noted on-going challenges in coordinating care with community providers, and suggested the development of tools to facilitate coordination.

- *Consider how funding expanded care in the community might affect funding for Veterans receiving care in VHA settings.* KIs voiced both concern for and uncertainty about how the Veterans Choice Program (VCP) may affect Veterans who continue to receive care in VHA.

- *Consider how performance by community providers might impact measured performance for VHA providers.* Several KIs noted that there may be Veterans who receive care both in the community and in VHA settings, and voiced concern for the potential impact on the achievement of VHA performance metrics, and possibly on the performance pay of VHA providers.

- *Be vigilant for overtreatment and for differences in standards of care (eg, opioid prescriptions).* KIs noted that one fundamental difference between care received in the VHA as compared to the community is that the VHA tends to be more conservative, and despite evidence of potential overtreatment in VHA settings,[21,22] overtreatment is more common in community settings. In addition, consistent with included research,[77] KIs

voiced concern that community providers may be more prone to prescribing opioids than providers in the VHA.

Table 13. Evidence and Policy Implications – P4P in Community Settings

Implementation Framework Category	Study Evidence	Themes from KI Interviews	Policy Implications
Program Design Features	One study[78] examined program design features relevant to P4P in community settings, and found: · A number of survey instruments examining cross-system access and coordination exist	A P4P program should be designed by first understanding the VHA's larger (achievable) goals – and how expanding care for Veterans in the community through VCP better enables us to reach that goal. What to incentivize · Consider incentivizing metrics such as timely access, documentation, and coordination · Consider incentivizing aspects of care that differentiate the VHA from the care in the community · Consider starting by looking at Veterans receiving care in the community as a population How to choose quality providers/Who to incentivize · Build relationship at the national and local levels · Partner with established networks · Contract only with board certified physicians · Choose providers based on past performance on CMS metrics	· Initially target areas in need improvement such as documentation and coordination (*eg,* receipt of records from community providers) · Develop relationships with providers and health systems with records of strong performance on commonly used, well-validated, and well-established metrics
Implementation Factors	The 4 studies[76,77,79,80] examining implementation factors found: · Veterans, providers, and administrators reported VCP-related challenges such as fragmented care, poor communication and coordination, additional burden on VHA providers, and barriers to sharing medical records.	Additional quality improvement strategies to consider in addition to P4P · Public reporting · Tools for community providers to streamline coordination of care Challenges to P4P in community settings · Limited number of Veteran patients per community provider · VHA has a fragile relationship with community providers · Mental health services in the	· The likely small number of Veteran patients per community provider may pose a challenge, both in terms of accurately assessing quality and the potential for an incentive to influence behavior. · Use tools such as public reporting to complement P4P · Developing tools and

Implementation Framework Category	Study Evidence	Themes from KI Interviews	Policy Implications
	· There are differences between providers interested in VCP participation and those who are not, such as Veteran status and willingness to provide patient medical records.	community Veterans have slightly different characteristics than the general population; thus, different health concerns and needs P4P in the community may impact Veterans receiving care in VHA settings and may influence providers and may influence the achievement of performance metrics for VHA providers	resources to streamline the data-sharing and coordination necessary to inform a cross-system P4P program · Consider how funding expanded care in the community might affect funding for Veterans receiving care in VHA settings · Consider how performance by community providers might impact measured performance for VHA providers
Provider Cognition, Affect, and Behavior		P4P in the community may increase overtreatment and overuse	· Be vigilant for overtreatment and for differences in standards of care (*eg,* opioid prescriptions)

Abbreviations: P4P = pay-for-performance, VCP = Veterans Choice Program, VHA = Veterans Health Administration

LIMITATIONS

Our report has a number of limitations. Given the limited research directly examining P4P in VHA settings and the heterogeneity in the way that P4P is implemented in the VHA, we were unable to form conclusions about the effectiveness of VHA performance pay on quality or health outcomes, and instead focus primarily on describing factors related to program design features, implementation factors, and unintended consequences. Similarly, research examining VCP is just starting to emerge, and as a result, our findings related to P4P in community settings are influenced heavily by findings from our key informant interviews. The breadth of topics and outcomes related to program design features, implementation factors, and unintended consequences made it difficult to restrict our criteria by study design. Thus, we included studies that utilized less-rigorous methodology, some of which had small samples. We interviewed 17 key informants to gain insight into factors important to the design and implementation of P4P in VHA and community settings. Although we aimed for a broad range of stakeholders, we recognize that a larger sample or different mix of KIs may have resulted in a different subset of themes identified. We also recognize that our aim to identify performance metrics that had been used in published research captures only a fraction of performance metrics used in P4P programs worldwide.

RESEARCH GAPS/FUTURE RESEARCH

Despite the fact that performance pay has been a part of VHA providers' contracts for more than a decade, very little research has evaluated its effectiveness, and to our knowledge, no research has explored alternatives. The nature of the VHA as an integrated yet closed system provides opportunities to compare program design features and implementation factors, given the control a closed system provides.

Although Veterans seeking care in the community is not a new phenomenon, continued funding for the Veterans Choice Program necessitates the need for more comprehensive evaluation of the quality of care received by Veterans participating in VCP. Current research projects, programs, and initiatives funded largely by QUERI are evaluating metrics, quality, and P4P programs directly within the context of community care, but more research is needed to get a better sense of how expanded care in the community might impact Veterans receiving care in VHA settings – in particular vulnerable populations such as Veterans of color, low-income Veterans, and Veterans living in rural areas, for whom even community providers may be limited.

CONCLUSIONS

The effectiveness of pay-for-performance in VHA settings has been largely understudied, but we highlight a number of key lessons learned from the implementation of programs that may help guide future P4P program improvements in VHA. In P4P programs targeting Veteran health in community settings, care should be taken to establish relationships with providers with track records of quality; consideration should be given to the impact of the small number of Veterans that will receive care in a given community setting per community provider; efforts should be made to develop resources and tools to better enable coordination of care, data-sharing, and record transfer; and special attention should be paid to mitigate the potential for overtreatment and ensure quality care for all Veterans.

REFERENCES

1. Senate: Veterans' Affairs Committee. *S. 2484 (108th Congress): Department of Veterans Affairs Health Care Personnel Enhancement Act of 2004.* December 3 2004.
2. Clough JD, McClellan M. Implementing MACRA: Implications for Physicians and for Physician Leadership. *JAMA.* 2016;315(22):2397-2398.
3. Kondo K, Damberg C, Mendelson A, et al. *Understanding the Intervention and Implementation Factors Associated with Benefits and Harms of Pay for Performance Programs in Healthcare.* Washington, DC: Department of Veterans Affairs;2015.
4. Kondo KK, Damberg CL, Mendelson A, et al. Implementation processes and pay for performance in healthcare: A systematic review. *J Gen Intern Med.* 2016;31(Suppl 1):61-69.
5. Mendelson A, Kondo K, Damberg C, et al. The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care: A Systematic Review. *Ann Intern Med.* 2017;166(5):341-353.
6. Department of Veterans Affairs. VHA Office of Community Care. Available at https://www.va.gov/purchasedcare/. Accessed March 24, 2017.
7. Commission on Care. *Final Report of the Commission on Care.* 2016.
8. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res.* 2005;15(9):1277-1288.
9. Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0.* 2011; http://handbook.cochrane.org/. Accessed March 24, 2017.
10. Wells GA, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses. Available at http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed March 24, 2017.
11. Herzog R, Alvarez-Pasquin MJ, Diaz C, Del Barrio JL, Estrada JM, Gil A. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. *BMC Public Health.* 2013;13(1):17.
12. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci.* 2009;4:50.
13. Berkman N, Lohr K, Ansari M, et al. *Grading the Strength of a Body of Evidence When Assessing Health Care Interventions for the Effective Health Care Program of the Agency for Healthcare Research and Quality: An Update*. 2013; http://www.effectivehealthcare.ahrq.gov/ehc/products/457/1752/methods-guidance-grading-evidence-131118.pdf. Accessed December 28, 2016.
14. Benzer JK, Young GJ, Burgess JF, et al. Sustainability of quality improvement following removal of pay-for-performance incentives. *J Gen Intern Med.* 2014;29(1):127-132.

15. Harris AH, Chen C, Rubinsky AD, Hoggatt KJ, Neuman M, Vanneman ME. Are Improvements in Measured Performance Driven by Better Treatment or "Denominator Management"? *J Gen Intern Med.* 2016;31 Suppl 1:21-27.

16. Petersen LA, Ramos KS, Pietz K, Woodard LD. Impact of a Pay-for-Performance Program on Care for Black Patients with Hypertension: Important Answers in the Era of the Affordable Care Act. *Health Serv Res.* 2016.

17. Petersen LA, Simpson K, Pietz K, et al. Effects of individual physician-level and practice-level financial incentives on hypertension care: a randomized trial. *JAMA.* 2013;310(10):1042-1050.

18. Urech TH, Woodard LD, Virani SS, Dudley RA, Lutschg MZ, Petersen LA. Calculations of Financial Incentives for Providers in a Pay-for-Performance Program: Manual Review Versus Data From Structured Fields in Electronic Health Records. *Med Care.* 2015;53(10):901-907.

19. Kansagara D, Tuepker A, Joos S, Nicolaidis C, Skaperdas E, Hickam D. Getting performance metrics right: a qualitative study of staff experiences implementing and measuring practice transformation. *J Gen Intern Med.* 2014;29 Suppl 2:S607-613.

20. Damschroder LJ, Robinson CH, Francis J, et al. Effects of performance measure implementation on clinical manager and provider motivation. *J Gen Intern Med.* 2014;29 Suppl 4:877-884.

21. Beard AJ, Hofer TP, Downs JR, et al. Assessing appropriateness of lipid management among patients with diabetes mellitus: moving from target to treatment. *Circ Cardiovasc Qual Outcomes.* 2013;6(1):66-74.

22. Kerr EA, Lucatorto MA, Holleman R, Hogan MM, Klamerus ML, Hofer TP. Monitoring performance for blood pressure management among patients with diabetes mellitus: too much of a good thing? *Arch Intern Med.* 2012;172(12):938-945.

23. Petersen LA, Woodard LD, Henderson LM, Urech TH, Pietz K. Will hypertension performance measures used for pay-for-performance programs penalize those who care for medically complex patients? *Circulation.* 2009;119(23):2978-2985.

24. Saini SD, Powell AA, Dominitz JA, et al. Developing and Testing an Electronic Measure of Screening Colonoscopy Overuse in a Large Integrated Healthcare System. *J Gen Intern Med.* 2016;31 Suppl 1:53-60.

25. Saini SD, Vijan S, Schoenfeld P, Powell AA, Moser S, Kerr EA. Role of quality measurement in inappropriate use of screening for colorectal cancer: retrospective cohort study. *BMJ.* 2014;348:g1247.

26. Powell AA, White KM, Partin MR, et al. Unintended consequences of implementing a national performance measurement system into local practice. *J Gen Intern Med.* 2012;27(4):405-412.

27. Powell AA, White KM, Partin MR, et al. More than a score: a qualitative study of ancillary benefits of performance measurement. *BMJ Qual Saf.* 2014;23(8):651-658.

28. Hysong SJ, SoRelle R, Broussard Smitham K, Petersen LA. Reports of unintended consequences of financial incentives to improve management of hypertension. *PLoS One.* 2017;12(9):e0184856.

29. Greene J. An examination of pay-for-performance in general practice in Australia. *Health Serv Res.* 2013;48(4):1415-1432.

30. Ryan AM, Burgess JF, Jr., Pesko MF, Borden WB, Dimick JB. The early effects of Medicare's mandatory hospital pay-for-performance program. *Health Serv Res.* 2015;50(1):81-97.

31. Kalwij S, French S, Mugezi R, Baraitser P. Using educational outreach and a financial incentive to increase general practices' contribution to chlamydia screening in South-East London 2003-2011. *BMC Public Health.* 2012;12:802.

32. Rat C, Penhouet G, Gaultier A, et al. Did the new French pay-for-performance system modify benzodiazepine prescribing practices? *BMC Health Serv Res.* 2014;14:301.

33. Torchiana DF, Colton DG, Rao SK, Lenz SK, Meyer GS, Ferris TG. Massachusetts General Physicians Organization's quality incentive program produces encouraging results. *Health Aff (Millwood).* 2013;32(10):1748-1756.

34. Esse T, Serna O, Chitnis A, Johnson M, Fernandez N. Quality compensation programs: are they worth all the hype? A comparison of outcomes within a Medicare advantage heart failure population. *J Manag Care Pharm.* 2013;19(4):317-324.

35. Kirschner K, Braspenning J, Akkermans RP, Jacobs JE, Grol R. Assessment of a pay-for-performance program in primary care designed by target users. *Fam Pract.* 2013;30(2):161-171.

36. Martens JD, Werkhoven MJ, Severens JL, Winkens RA. Effects of a behaviour independent financial incentive on prescribing behaviour of general practitioners. *J Eval Clin Pract.* 2007;13(3):369-373.

37. Pechlivanoglou P, Wieringa JE, de Jager T, Postma MJ. The effect of financial and educational incentives on rational prescribing. A state-space approach. *Health Econ.* 2015;24(4):439-453.

38. Kantarevic J, Kralj B. Link between pay for performance incentives and physician payment mechanisms: evidence from the diabetes management incentive in Ontario. *Health Econ.* 2013;22(12):1417-1439.

39. Li J, Hurley J, DeCicca P, Buckley G. Physician response to pay-for-performance: evidence from a natural experiment. *Health Econ.* 2014;23(8):962-978.

40. Kruse GR, Chang Y, Kelley JH, Linder JA, Einbinder JS, Rigotti NA. Healthcare system effects of pay-for-performance for smoking status documentation. *Am J Manag Care.* 2013;19(7):554-561.

41. Andriole KP, Prevedello LM, Dufault A, et al. Augmenting the impact of technology adoption with financial incentive to improve radiology report signature times. *J Am Coll Radiol.* 2010;7(3):198-204.

42. Colais P, Pinnarelli L, Fusco D, Davoli M, Braga M, Perucci CA. The impact of a pay-for-performance system on timing to hip fracture surgery: experience from the Lazio Region (Italy). *BMC Health Serv Res.* 2013;13:393.

43. Young GJ, Beckman HB, Baker E. Financial incentives, professional values and performance: A study of pay-for-performance in a professional organization. *Journal of Organizational Behavior.* 2012;33(7):964–983.

44. Chang RE, Lin SP, Aron DC. A pay-for-performance program in Taiwan improved care for some diabetes patients, but doctors may have excluded sicker ones. *Health Aff (Millwood).* 2012;31(1):93-102.

45. Lai CL, Hou YH. The association of clinical guideline adherence and pay-for-performance among patients with diabetes. *J Chin Med Assoc.* 2013;76(2):102-107.

46. Lee TT, Cheng SH, Chen CC, Lai MS. A pay-for-performance program for diabetes care in Taiwan: a preliminary assessment. *Am J Manag Care.* 2010;16(1):65-69.

47. Chen CS, Liu TC, Chen B, Lin CL. The failure of financial incentive? The seemingly inexorable rise of cesarean section. *Soc Sci Med.* 2014;101:47-51.

48. Li YH, Tsai WC, Khan M, et al. The effects of pay-for-performance on tuberculosis treatment in Taiwan. *Health Policy Plan.* 2010;25(4):334-341.

49. Kristensen SR, Meacock R, Turner AJ, et al. Long-term effect of hospital pay for performance on mortality in England. *N Engl J Med.* 2014;371(6):540-548.

50. Karunaratne K, Stevens P, Irving J, et al. The impact of pay for performance on the control of blood pressure in people with chronic kidney disease stage 3-5. *Nephrol Dial Transplant.* 2013;28(8):2107-2116.

51. Arrowsmith ME, Majeed A, Lee JT, Saxena S. Impact of pay for performance on prescribing of long-acting reversible contraception in primary care: an interrupted time series study. *PLoS One.* 2014;9(4):e92205.

52. Doran T, Kontopantelis E, Valderas JM, et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. *BMJ.* 2011;342:d3590.

53. Kontopantelis E, Springate D, Reeves D, Ashcroft DM, Valderas JM, Doran T. Withdrawing performance indicators: retrospective analysis of general practice performance under UK Quality and Outcomes Framework. *BMJ.* 2014;348:g330.

54. Millett C, Bottle A, Ng A, et al. Pay for perfomance and the quality of diabetes management in individuals with and without co-morbid medical conditions. *J R Soc Med.* 2009;102(9):369-377.

55. Millett C, Gray J, Saxena S, Netuveli G, Majeed A. Impact of a pay-for-performance incentive on support for smoking cessation and on smoking prevalence among people with diabetes. *CMAJ.* 2007;176(12):1705-1710.

56. Norbury M, Fawkes N, Guthrie B. Impact of the GP contract on inequalities associated with influenza immunisation: a retrospective population-database analysis. *Br J Gen Pract.* 2011;61(588):e379-385.

57. Szatkowski L, McNeill A, Lewis S, Coleman T. A comparison of patient recall of smoking cessation advice with advice recorded in electronic medical records. *BMC Public Health.* 2011;11:291.

58. Taggar JS, Coleman T, Lewis S, Szatkowski L. The impact of the Quality and Outcomes Framework (QOF) on the recording of smoking targets in primary care medical records: cross-sectional analyses from The Health Improvement Network (THIN) database. *BMC Public Health.* 2012;12:329.

59. Tahrani AA, McCarthy M, Godson J, et al. Diabetes care and the new GMS contract: the evidence for a whole county. *Br J Gen Pract.* 2007;57(539):483-485.

60. MacBride-Stewart SP, Elton R, Walley T. Do quality incentives change prescribing patterns in primary care? An observational study in Scotland. *Fam Pract.* 2008;25(1):27-32.

61. Kontopantelis E, Reeves D, Valderas JM, Campbell S, Doran T. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. *BMJ Qual Saf.* 2013;22(1):53-64.

62. Simpson CR, Hannaford PC, Ritchie LD, Sheikh A, Williams D. Impact of the pay-for-performance contract and the management of hypertension in Scottish primary care: a 6-year population-based repeated cross-sectional study. *Br J Gen Pract.* 2011;61(588):e443-451.

63. Vaghela P, Ashworth M, Schofield P, Gulliford MC. Population intermediate outcomes of diabetes under pay-for-performance incentives in England from 2004 to 2008. *Diabetes Care.* 2009;32(3):427-429.

64. Sutton M, Elder R, Guthrie B, Watt G. Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health Econ.* 2010;19(1):1-13.

65. Kuo RN, Chung KP, Lai MS. Effect of the Pay-for-Performance Program for Breast Cancer Care in Taiwan. *J Oncol Pract.* 2011;7(3 Suppl):e8s-e15s.

66. Frakt AB, Trafton J, Pizer SD. The association of mental health program characteristics and patient satisfaction. *Am J Manag Care.* 2017;23(5):e129-e137.

67. Hysong SJ, Amspoker AB, Petersen LA. A Novel Method for Assessing Task Complexity in Outpatient Clinical-Performance Measures. *J Gen Intern Med.* 2016;31 Suppl 1:28-35.

68. Hysong SJ, Khan MM, Petersen LA. Passive monitoring versus active assessment of clinical performance: impact on measured quality of care. *Med Care.* 2011;49(10):883-890.

69. Petersen LA, Pietz K, Woodard LD, Byrne M. Comparison of the predictive validity of diagnosis-based risk adjusters for clinical outcomes. *Med Care.* 2005;43(1):61-67.

70. Rosen AK, Chen Q, Shwartz M, et al. Does Use of a Hospital-wide Readmission Measure Versus Condition-specific Readmission Measures Make a Difference for Hospital Profiling and Payment Penalties? *Med Care.* 2016;54(2):155-161.

71. Prentice J, Legler A, Li D, Pizer SD. Optimizing Access Metrics in the VA. https://wwwhsrdresearchvagov/for_researchers/cyber_seminars/archives/1166-notespdf.

72. Prentice JC, Davies ML, Pizer SD. Which outpatient wait-time measures are related to patient satisfaction? *Am J Med Qual.* 2014;29(3):227-235.

73. Hysong SJ, Simpson K, Pietz K, SoRelle R, Broussard Smitham K, Petersen LA. Financial incentives and physician commitment to guideline-recommended hypertension management. *Am J Manag Care.* 2012;18(10):e378-391.

74. Hysong SJ, Knox MK, Haidet P. Examining clinical performance feedback in Patient-Aligned Care Teams. *J Gen Intern Med.* 2014;29 Suppl 2:S667-674.

75. Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci.* 2006;1:9.

76. Finley EP, Noel PH, Mader M, et al. Community Clinicians and the Veterans Choice Program for PTSD Care: Understanding Provider Interest During Early Implementation. *Med Care.* 2017;55 Suppl 7 Suppl 1:S61-s70.

77. Gellad WF, Cunningham FE, Good CB, et al. Pharmacy Use in the First Year of the Veterans Choice Program: A Mixed-methods Evaluation. *Med Care.* 2017;55 Suppl 7 Suppl 1:S26-s32.

78. Quinn M, Robinson C, Forman J, Krein SL, Rosland AM. Survey Instruments to Assess Patient Experiences With Access and Coordination Across Health Care Settings: Available and Needed Measures. *Med Care.* 2017;55 Suppl 7 Suppl 1:S84-s91.

79. Tsai J, Yakovchenko V, Jones N, et al. "Where's My Choice?" An Examination of Veteran and Provider Experiences With Hepatitis C Treatment Through the Veteran Affairs Choice Program. *Med Care.* 2017;55 Suppl 7 Suppl 1:S13-s19.

80. Zuchowski JL, Chrystal JG, Hamilton AB, et al. Coordinating Care Across Health Care Systems for Veterans With Gynecologic Malignancies: A Qualitative Analysis. *Med Care.* 2017;55 Suppl 7 Suppl 1:S53-s60.

APPENDIX A. SEARCH STRATEGIES

Main Search: March 14, 2017

(((((((((("Minority Health"[Mesh]) OR ("Healthcare Disparities"[Mesh] OR "Health Status Disparities"[Mesh])) OR "Racism"[Mesh]) OR "Sexism"[Mesh]) OR "Homophobia"[Mesh]) OR "Observer Variation"[Mesh])) OR ((((((unintended[Title/Abstract] OR unanticipated[Title/Abstract] OR unforeseen[Title/Abstract] OR Unexpected[Title/Abstract] OR unplanned[Title/Abstract] OR undesired[Title/Abstract] OR unconscious[Title/Abstract])) OR exception reporting[Title/Abstract]) OR preferential reporting[Title/Abstract]) OR bias[Title/Abstract]) OR (gaming[Title/Abstract] OR spillover[Title/Abstract] OR "risk selection"[Title/Abstract] OR unincentivised[Title/Abstract] OR nonincentivised[Title/Abstract])))) AND ((("Physician Incentive Plans"[Mesh]) OR "Reimbursement, Incentive"[Mesh]) OR "economics"[Subheading]) AND VA filter: (("Veterans Health"[Mesh])) OR (((VA OR Veteran OR VAMC OR Veterans)) OR ("Veterans"[Mesh] OR "United States Department of Veterans Affairs"[Mesh] OR "Hospitals, Veterans"[Mesh]))	
Limited to Publication Date after 1/1/2014	
PubMed	N = 107
CINAHL (after deduplication with PubMed search)	N = 34
PsycINFO (after deduplication with PubMed and CINAHL)	N = 116

Author Search: March 28, 2017

(((((Petersen L[Author]) OR Kerr E[Author]) OR Hofer T[Author]) OR Benzer J[Author]) OR Werner R[Author]) OR Volpp K[Author]	
PubMed	N = 618

APPENDIX B. TECHNICAL EXPERT PANEL AND KEY INFORMANTS

TECHNICAL EXPERT PANEL MEMBER AND KEY INFORMANT

Justin Benzer, PhD
Research Health Scientist; Department of Veterans Affairs
Associate Professor; University of Texas at Austin Dell Medical School
Research Associate Professor; Public Health, Texas A&M University

Laura A. Petersen, MD, MPH, FACP
Associate Chief of Staff for Research; Michael E. DeBakey VA Medical Center
Director; Center for Innovations In Quality, Effectiveness, and Safety (IQuESt)
Professor of Medicine, Chief of Section of Health Services Research; Baylor College of Medicine

Laura Damschroder, MS, MPH
Investigator; Center for Clinical Management Research, VA Ann Arbor Healthcare System

Rachel Werner, MD, PhD
Core Investigator; VA Center for Health Equity Research and Promotion (CHERP)
Professor of Medicine; Division of General Internal Medicine, University of Pennsylvania Perelman School of Medicine
Director of Health Policy and Outcomes Research, Medicine; University of Pennsylvania Perelman School of Medicine
Senior Fellow; Leonard Davis Institute of Health Economics, University of Pennsylvania
Attending Physician; Philadelphia VA Medical Center

KEY INFORMANTS

Alvaro Sanchez, MD
Chief Medical Officer; VA Midwest Healthcare Network (VISN 23)

Gary Young, JD, PhD
Director; Northeastern University Center for Health Policy and Healthcare Research
Professor of Strategic Management and Healthcare Systems; Northeastern University

Hallie Prescott, MD, MSc
Assistant Professor; Department of Internal Medicine, University of Michigan Medical School

Ilse Wiechers, MD, MPP, MHS
Assistant Professor of Psychiatry; Yale University School of Medicine
Associate Director of the Northeast Program Evaluation Center; VA Office of Mental Health Operations (OMHO)
National Program Director; Psychotropic Drug Safety Initiative

Jeff Kullgren, MD, MS, MPH
Research Scientist; VA Center for Clinical Management Research, VA Ann Arbor Healthcare System
Assistant Professor; Department of Internal Medicine, University of Michigan Medical School and Institute for Healthcare Policy and Innovation

Jeremy Sussman, MD, MS
Research Scientist; Center for Clinical Management Research, VA Ann Arbor Health System
Assistant Professor; Department of Internal Medicine, University of Michigan Medical School

James F. Burgess, Jr., PhD
Senior Investigator; HSR&D's Center for Healthcare Organization and Implementation Research (CHOIR)
Director of the Health Economics Program; Boston University

Joan E. McInerney, MD, MBA, MA, FACEP
Network Director; VISN 2

Praveen Mehta, MD
Chief Medical Officer; VISN 12

Scott Pawlikowski, MD
Primary Care Lead Physician; VA Great Lakes Healthcare System, VISN 12

Steven D. Pizer, PhD
Director of Health Care Financing and Economics; VA Boston Healthcare System
Associate Professor; Department of Pharmacy Practice and Department of Economics, Northeastern University

Sylvia J. Hysong, PhD
Investigator and Ph.D. Fellowship Program Director; Health Policy, Quality & Informatics Program, Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey VA Medical Center
Associate Professor; Department of Medicine, Section of Health Services Research, Baylor College of Medicine
Adjunct Assistant Professor; University of Houston

Tim Hofer, MD, MSc
Professor; Division of General Medicine, University of Michigan
Associate Director for Analytic and Information Resources; VA Center for Practice Management and Outcomes Research

APPENDIX C. STUDY SELECTION CRITERIA – PICOTS

Key Questions	KQ1. What are the effects of pay-for-performance programs on the quality of care in veteran populations?	KQ2. In Veteran populations, what are the potential unintended consequences of pay-for-performance in healthcare?	KQ3. What metrics have been commonly incentivized in the published literature examining P4P?	KQ4. In Veteran populations, what program features and implementation factors modify the effectiveness of pay-for-performance programs?	KQ5. What novel approaches and/or current or recently closed research projects funded by the VA examine the effectiveness, implementation factors, or unintended consequences associated with pay-for-performance in Veteran populations?
Population	Healthcare providers at the individual, managerial (*eg,* VISN directors), group, and institutional levels. Veteran populations whose providers/health care systems are being incentivized based on performance metrics.				
Interventions	Pay-for-performance programs targeting providers, managers, health care systems				
Comparators	Other financial incentive models; other payment models (*eg,* fee for service, bundled payments)				
Outcomes	Utilization of healthcare services Quality of healthcare services Intermediate outcomes/process of care metrics Patient evaluations of care Patient health outcomes				
Timing	Long- or short-term				
Study design	RCTs, NRCTs. Eligible observational studies: have a comparator, ³ 3 time points and report a trend (*eg,* ITS), or n³ 10,000 cross-sectional or uncontrolled before–after study	All quantitative and qualitative study designs.			
Setting	VHA, CBOCs, or community sites serving Veteran populations		VHA or healthcare systems similar to the VHA or the US	VHA, CBOCs, or community sites serving Veteran populations	

APPENDIX D. STUDY SELECTION CRITERIA – INCLUSION/EXCLUSION CRITERIA

1. Population: Is the study population Veterans?
   Yes.....Proceed to #2
   No ..... Code **X1 for KQ1**. STOP

2. Financial Incentives: Does the article report outcomes and report outcomes separately for groups with and without financial incentives at the provider, managerial, group, or institutional level.
   Yes ..... Proceed to #3
   No .....Code **X2 for KQ1.** Add code **B** if retaining for background/discussion. STOP

3. KQ2: Does the article discuss or report implementation factors that modify the effectiveness of financial incentives?
   Yes .....Code **I for KQ2**. Proceed to #4
   No .....Code **X for KQ2.** Proceed to #4

4. KQ3: Does the article report unintended consequences/health disparities related to financial incentives?
   Yes .....Code **I for KQ3**. Proceed to #5
   No .....Code **X for KQ3.** Proceed to #5

5. Outcomes Does the article report utilization, quality of care (*eg,* intermediate, patient evaluations of care), or patient health outcomes (modeling studies are not included)?
   Yes ..... Proceed to #6
   No .....Code **X3 for KQ1.** Add code **B** if retaining for background/discussion. STOP

6. Publication Type: Does the article present original study data, a systematic review, or meta-analysis? Narrative or non-systematic reviews, letters, editors, and commentaries are excluded.
   Yes ..... Proceed to #7
   No .....Code **X4 for KQ1.** Add code **B** if retaining for background/discussion. STOP

7. Systematic Review: Is the article a systematic review or meta-analysis of primary studies?
   Yes .....Code **SR for KQ1**. STOP
   No .....Proceed to # 8

8. Case Studies/Case Series: Does the article present a case study, case series, or case report?
   Yes .....Code **X5 for KQ1**. STOP
   No .....Proceed to # 9

9. Comparator/Study design: Is the article a primary study that compares a financial incentive to another financial incentive model or no financial incentive/usual care, or does it report 3 or more time points (and trend data), or have more than 10,000 participants
   Yes ..... STOP
   No .....Code **X7 for KQ1**. Add code **B** if retaining for background/discussion. STOP.

APPENDIX E. KEY INFORMANT INTERVIEW GUIDE

Portland Evidence-based Synthesis Program

Pay-for-Performance and Veteran Care: Effects, Implementation, and Unintended Consequences

Introduction

- § ESP
- § Project
- § Project Description

1. Given your experience, what factors do you think are most important for the VA to consider in implementing pay for performance programs?

   b. Can you give an example of a situation where [the factor/s mentioned above] made a difference in the success/failure of implementing a P4P program within VA?

   c. Do you think factors influencing implementation success of P4P in the CHOICE Program would be similar to, or different, from those you mentioned above? Why?

Next I want to talk more specifically about aspects of P4P programs that you believe are important.

Measures

2. Are there types of measures that you believe the VA should prioritize?
   a. Can you give us an example of such a measure?

3. Are there types of measures you have found problematic/think might be problematic within a VA context?
   a. Can you provide an example?

4. How might measures differ when used in the context of P4P versus the CHOICE Program?

Incentives

5. Can you tell us about different types of incentives that you think are important (*eg*, rewards vs penalties, type/nature, frequency/duration, certainty)?

6. Do you think incentives used in the context of the CHOICE Program should differ from those used within VA? Why or why not?

Implementation factors

7. What other implementation factors should the VA consider as they set up partnerships with the community?

Probe: Implementation processes:

- measure monitoring/evaluation,
- incentive removal,
- stakeholder engagement

8. How should the VA engage stakeholders as they start setting up partnerships in the community?

Probe:

- Inner setting (Institutional)
- Outer setting (economic, political, social contexts)
- Providers
- Cognitive/affective responses
  - beliefs, attitudes
  - cognitive response constructs such as biases, professionalism, heuristics, identification with one's organization
  - behavioral response constructs such as risk selection, gaming, systems improvement responses

Unintended Consequences

9. As the VA moves forward with implementing P4P in the community, how can they minimize negative unintended consequences?

- Examples:
  - Risk selection
  - Deterioration of un-incentivized care
  - Impairment of intrinsic motivation/professionalism
  - Gaming
  - Teaching to the test and/or overtreatment

10. Theoretically, community care can be very costly – and with a fixed budget, do you think it could potentially affect health disparities in Veteran populations? If yes, how so? Are these 2 populations likely to be different? And is there anyone working on this?
    - Examples:
      - Low income
      - Racial/ethnic minorities
      - Rural/distance from VHA
      - Homeless
      - Mental health
      - Disabilities
      - Women
      - LGBT

11. Do you know of anyone in the VA who is piloting novel performance metrics, novel approaches to P4P, etc. If so, who? (And can you tell us about these metrics or approaches?)

12. Is there anyone else you think it would be important for us to talk to?

13. Attached to the original email was a list of studies we identified for inclusion. Are there any others you suggest?

14. Is there anything else that you think it is important that we know?

APPENDIX F. RISK OF BIAS ASSESSMENT

Risk of Bias of Randomized Controlled Trials

Trial	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting
Petersen et al, 2013,[17] 2016[16]	Low	Low	Low	Low	Unclear	Unclear

Risk of Bias of Cohort Studies

Study	Representative-ness of the exposed cohort	Selection of the non-exposed cohort	Ascertain-ment of exposure	Description of concurrent QI Initiatives	Demonstration that the outcome of interest was not present at the start of the study	Comparability of cohorts on the basis of the design or analysis	Assess-ment of outcome	Follow-up long enough for outcomes to occur	Adequacy of follow up cohorts
Beard et al, 2013[21]	Yes	NA	Yes	No	Yes	Yes	Yes	Yes	Yes
Benzer et al, 2014[14]	Yes	NA	Yes	Yes	NA	NA	Yes	Yes	Yes
Harris et al, 2015[15]	Yes	NA	Yes	Yes	Yes	NA	Yes	Yes	NA
Hysong et al, 2011[68]	Yes	NA	Yes	No	NA	NA	Yes	Yes	Yes
Kerr et al, 2012[22]	Yes	NA	Yes	No	Yes	Yes	Yes	Yes	Yes
Petersen et al, 2009[23]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	NA
Prentice et al, 2014[72]	Yes	Yes	Yes	No	Yes	Yes	No	Yes	NA
Rosen et al, 2016[70]	Yes	NA	Yes	Yes	Yes	Yes	Yes	Yes	NA
Saini et al, 2014[25]	Yes	NA	Yes	No	Yes	Yes	Yes	Yes	Yes

Risk of Bias of Cross-sectional Studies

Study	Representative-ness of the sample	Sample size	Non-respondents	Ascertain-ment of exposure/risk factor	The subjects in different outcome groups are comparable, based on the study design or analysis. Confounding factors are controlled.	Assessment of the outcome	Statistical test
Finley et al, 2017[76]	Yes	No	No	Yes	Yes	Yes	Yes
Frakt et al, 2017[66]	Yes	Yes	No	Yes	Yes	Yes	Yes

Hysong et al, 2012[73]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Hysong et al, 2016[67]	Yes	Yes	NA	Yes	NA	Yes	Yes
Petersen et al, 2005[69]	Yes	Yes	NA	Yes	Yes	Yes	Yes
Prentice et al, 2016[71]	Yes	Yes	NA	Yes	Yes	Yes	Yes
Saini et al, 2016[24]	Yes	Yes	NA	Yes	Yes	Yes	Yes
Urech et al, 2015[18]	Yes	Yes	NA	Yes	Yes	Yes	Yes

APPENDIX G. PEER REVIEW DISPOSITION TABLE

Reviewer Number	Comment	Response
Are the objectives, scope, and methods for this review clearly described?		
1, 2, 3, 7	Yes	Thank you.
Are there any published or unpublished studies that we may have overlooked?		
1	No	Noted
2	Yes - You cite some presentations that led at least partly to publications: Frakt, Austin B., Jodie Trafton, and Steven D. Pizer. "The association of mental health program characteristics and patient satisfaction." The American journal of managed care 23.5 (2017): e129. Prentice, Julia C., Michael L. Davies, and Steven D. Pizer. "Which outpatient wait-time measures are related to patient satisfaction?." American Journal of Medical Quality 29.3 (2014): 227-235.	Thank you. We have replaced the presentation Frakt et al, 2016 with the suggested Frakt et al, 2017 publication, and we have added Prentice et al, 2014.
3	No	Noted
7	Yes - Hysong SJ, SoRelle R, Broussard Smitham KK, and Petersen LA (in press). Reports of Unintended Consequences of Financial Incentives To Improve Management of Hypertension. PLoS One. Hysong SJ, Knox MK, Haidet P. Examining clinical performance feedback in patient-aligned care teams. Journal of general internal medicine. 2014 Jul 1;29(2):667-74. Hysong SJ, Broussard Smitham KK, SoRelle R, Knox MK, Amspoker AB, Hughes A, Haidet P. Mental Models of Audit and Feedback in Primary Care Settings. In preparation, (target journal, Implementation Science)	Thank you. The in press article was published prior to the finalization of our report; thus, it was formally added to our synthesis. We also added Hysong et al, 2014. As the manuscript examining mental models remains unpublished, we added a brief description of the purpose to our section that describes unpublished work.
Is there any indication of bias in our synthesis of the evidence?		
1, 2, 3, 7	No	Noted
Additional suggestions or comments can be provided below. If applicable, please indicate the page and line numbers from the draft report.		
1	This is a clearly written sequel to 2 prior reviews of P4P conducted by this ESP Center. It frames the question in terms of how the findings might apply to the Veterans Choice Program. While it hardly presents earth-shattering conclusions, it represents the literature fairly and provides a timely refresh of published studies that one hopes will inform policymaking, if not immediately, hopefully sometime in the future. The questions that are unanswered are those that remain unstudied - how do health outcomes change? What is the impact on physician morale esp burnout? beyond anecdote and qualitative assessments, how prevalent and important are the unintended consequences? Hopefully some of these questions will be addressed in future HSR&D funded research.	Noted, thank you.

Reviewer Number	Comment	Response
2	The content of the report is well organized and presented, but the text needs thorough proofreading and editing for grammar and clarity with issues on almost every page. These typically do not detract seriously from the content, but combine to undermine the credibility of the report.	Noted. We have thoroughly proofread and edited the report.
2	Regarding substance, there is some tension between the idea of targeting incentives to providers instead of to systems or clinics and using population-based metrics. We want to target providers to improve the chance they will respond to the incentive, but we want population-based metrics to improve precision and encourage a population-based approach to care. It seems to me that this tension should at least be acknowledged.	Thank you. We have added a statement clarifying that population-based methods are unlikely to motivate behavior at the provider level.
2	The boxed quote on page 56 is potentially very confusing. Perhaps the key informant intended to say "I'm ordering a very expensive cat scan," but s/he said "valuable," which suggests undertreatment instead of overtreatment.	Thank you for pointing that out. We have deleted “valuable” and replaced it with “(expensive).”
3	This is an excellent document that I think will be quite helpful to VA as they try to apply their strengths to community care, where may of those strengths will be less usable. While structured in 5 key questions, there were really 2 projects here. In the first, they essentially described the research on P4P in VHA historically. In the second, they interviewed VA researchers who could help inform how VHA should approach this going forward.	Noted, thank you.
3	Both projects were well-performed and their results were clearly presented. For a long document and a slightly diffuse project, I thought the document itself was readable and the findings were clear. While there are a few research decisions that I might have made differently, it was clear what decisions they had made and why.	Noted, thank you.
3	For the question about how P4P has worked in VHA in the past, the answer was that the research isn’t very clear, they’ve used a lot of different measures, and there have been some unintended consequences, particularly overuse. The research has shown particular reliability questions with a clear rise in metric gaming and “denominator management.”	Noted.
3	For the question of how P4P can work in VCA and in the future, I thought the interviewees overall gave a fairly large number of potentially useful ideas. The most important of these is to essentially remember that P4P choices are intended as tools to help implement whatever is important to the organization. The major initial concerns about VCA are probably more about access, coordination with VA providers, and making sure community providers access the services VA truly excels at (like PTSD and rehab care), than more traditional measures like blood pressure control. Relatedly, P4P should always be seen as one tool in an integrated implementation system that would also include decision support, education, and audit and feedback, among others.	Noted.
3	My biggest problems with the document are not really addressable at this stage. My major problem was that the review work really focused on P4P in VA. However, if the question is how to use P4P to improve VCA, this is not ideal. I’d be more curious how individual insurers with non-dominant market shares have attempted to use P4P in the wild west of the community than how the centralized, mission-driven VA providers have used it. The difficulties of P4P in VCA will resemble those of private	Thank you. Given time limitations (this project had a six-month timeline) and the system-level and cultural differences between the VHA and the private sector, we focused on the VHA as a system due. We agree that

Reviewer Number	Comment	Response
	insurers more than that of internal VA historical issues.	future research should examine the issue of insurers with small market shares.
7	p. 45, lines 52-53. The goal commitment study referenced did find no difference in goal commitment between incentivized and non-incentive physicians. More importantly, though, they found goal commitment to be modest at best, suggesting there were stronger situational factors that were impacting both physician types.	Thank you. We have edited the statement to read, “One study examined a provider affective/cognitive response, and found that not only did P4P have no impact on goal commitment, but that physicians may perceive an external locus of control for hypertension care.” To the detailed findings in Table 9, we have added “In addition, patient non-adherence and inconsistent follow-up were cited as barriers to care.”
7	Figure 3. The Hysong et al. (in press) study referenced in the previous question provides new findings that many of the commonly discussed unintended consequences reflect concerns, rather than actual instances of unintended consequences occurring.	Thank you. Figure 3 highlights key informant themes, so we did not add findings from the recently published study to the figure. However, we have added the study to the report and discuss its findings.
7	P. 21, Lines 29-31. Another way to potentially mitigate overtreatment is to incentivize appropriate care rather than treatment. In other words, follow a patient over time and see whether the series of decisions made for that patient led to patient improvement (including choosing not to treat). It would be more labor intensive but would not only mitigate over treatment, it would likely have more face validity with the clinicians.	Thank you. We have added the following statement to the summary and discussion section on overtreatment: However, it is possible that overtreatment may be mitigated by incentivizing appropriate care, rather than treatment or targets, as demonstrated by the single included RCT.
7	Table 2 (p. 18) line 43 -- Peterson should be spelled "Petersen"	Thank you. Corrected.
7	P. 15 -- Rating the body of evidence; did the reviewers only include high evidence strength studies? If not, what was the criteria for including the others? Too late now, but perhaps a better approach would be to code specific characteristics of the evidence in order to better assess the strength of the evidence. This method is generally considered stronger than blanket ratings of strength.	Thank you. In addition to rating the strength of evidence for the question of effectiveness (KQ1), we did quality assess all studies using the Cochrane Risk of Bias tool for RCTs and the Newcastle Ottawa Scale for observational studies. Detailed ratings can be found in Appendix F. Given that much of the report was descriptive, other than in our methods section, we did not refer to the ratings in the body of the report.