

# Pay-for-Performance and Veteran Care in the VHA and the Community: a Systematic Review

Karli K. Kondo, PhD<sup>1,2</sup>, Jessica Wyse, PhD<sup>1</sup>, Aaron Mendelson, BA<sup>2</sup>, Gabriella Beard<sup>1</sup>, Michele Freeman, MPH<sup>1</sup>, Allison Low, BA<sup>1</sup>, and Devan Kansagara, MD, MCR<sup>1,2</sup>

<sup>1</sup>Portland VA Health Care System, Evidence-based Synthesis Program, Portland, OR, USA; <sup>2</sup>Oregon Health and Science University, Portland, OR, USA.

**BACKGROUND:** Although pay-for-performance (P4P) strategies have been used by the Veterans Health Administration (VHA) for over a decade, the long-term benefits of P4P are unclear. The use of P4P is further complicated by the increased use of non-VHA healthcare providers as part of the Veterans Choice Program. We conducted a systematic review and key informant interviews to better understand the effectiveness and potential unintended consequences of P4P, as well as the implementation factors and design features important in both VHA and non-VHA/community settings.

**METHODS:** We searched PubMed, PsycINFO, and CINAHL through March 2017 and reviewed reference lists. We included trials and observational studies of P4P targeting Veteran health. Two investigators abstracted data and assessed study quality. We interviewed VHA stakeholders to gain further insight.

**RESULTS:** The literature search yielded 1031 titles and abstracts, of which 30 studies met pre-specified inclusion criteria. Twenty-five examined P4P in VHA settings and 5 in community settings. There was no strong evidence supporting the effectiveness of P4P in VHA settings. Interviews with 17 key informants were consistent with studies that identified the potential for overtreatment associated with performance metrics in the VHA. Key informants' views on P4P in community settings included the need to develop relationships with providers and health systems with records of strong performance, to improve coordination by targeting documentation and data sharing processes, and to troubleshoot the limited impact of P4P among practices where Veterans make up a small fraction of the patient population.

**DISCUSSION:** The evidence to support the effectiveness of P4P on Veteran health is limited. Key informants recognize the potential for unintended consequences, such as overtreatment in VHA settings, and suggest that implementation of P4P in the community focus on relationship building and target areas such as documentation and coordination of care.

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11606-018-4444-4>) contains supplementary material, which is available to authorized users.

---

Received January 17, 2018

Revised March 9, 2018

Accepted April 10, 2018

Published online April 26, 2018

**KEY WORDS:** Veterans; pay for performance; financial incentives; implementation; performance metrics; systematic review.

J Gen Intern Med 33(7):1155–66

DOI: 10.1007/s11606-018-4444-4

© Society of General Internal Medicine (This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply) 2018

---

## INTRODUCTION

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

Although P4P aims to increase health care value, the empiric data are far from clear. A recent systematic review 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 patient health. The review also found that P4P was associated with potential unintended consequences, and that ultimately, P4P's balance of benefits and harms depends heavily on the nuances of program implementation.<sup>3–5</sup>

In 2016, there were 25.5 million Veteran appointments with non-VHA providers in the community,<sup>6</sup> and this number is expected to rise with the recent extension of the Veterans Choice Program (VCP). The VA Commission on Care recommended that payments to community providers be based on P4P incentives on quality and appropriate utilization.<sup>7</sup> Yet, how 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, without encouraging unintended consequences, remains largely unknown.

This report, which was part of a larger report commissioned by the VHA, presents the results of a systematic review and key informant interviews on (1) the effects of P4P programs

on the quality of care and health of Veterans, (2) potential unintended consequences of P4P targeting Veteran health, and (3) program design features and implementation factors that might modify the effectiveness of P4P targeting Veterans, in both VHA and community settings.

## METHODS

### Data Sources and Strategy

We searched PubMed, PsycINFO®, and CINAHL® (January 2014 to March 2017) for studies examining P4P in Veteran populations (search strategy in online Appendix 1), updating our previous P4P review<sup>3, 5</sup>; from a targeted search of known VA P4P and quality improvement researchers; and from a search of the VHA's website for unpublished studies.

### Study Selection

We included studies examining direct P4P programs targeting healthcare providers in VHA and VCP settings (study selection criteria in online Appendices 2 and 3). We excluded studies examining other payment models and patient-targeted incentives. To assess the effectiveness of P4P, we included trials and observational studies that either (a) had a comparison group, (b) had three or more time points and reported a trend (e.g., interrupted time series), or (c) included 10,000+ participants. All study designs were included for questions related to unintended consequences and community care. We included studies examining processes occurring both upstream (e.g., performance measures) and downstream (e.g., audit and feedback) of P4P. Two independent reviewers assessed studies, and all discordant results were resolved through consensus or consultation with a third reviewer.

### Data Extraction and Quality Assessment

Data from each study were abstracted by one investigator and confirmed by a second. We abstracted information on study design, sample size, observation period, program focus, incentive (target, size, timing), comparator, implementation factors, unintended consequences, and findings. Two investigators independently assessed study quality using the Cochrane Risk-of-Bias tool<sup>8</sup> for RCTs and the Newcastle-Ottawa Scale<sup>9, 10</sup> for observational studies (see online Appendix 4). We did not assess the quality of qualitative studies.

### Key Informants

We engaged VHA stakeholders and technical experts experienced with P4P as key informants to better understand the program features and implementation factors that might contribute to successful P4P programs in VHA and community settings. We identified key informants through snowball sampling. We used conventional content analysis<sup>11</sup> and developed a semi-structured interview that probed previously identified themes<sup>3, 4</sup>

and explored emerging themes (see online Appendix 5). Two investigators co-led telephone interviews (June–August 2017). Interviews were approximately 60 min and were recorded and transcribed verbatim. Four investigators reviewed each transcript and identified emergent themes and categories. Key themes were determined by consensus, and two investigators compared all related quotes across and within interviews.

### Data Synthesis and Analysis

We qualitatively synthesized the results of included studies and interviews according to an implementation framework developed for our previous P4P review.<sup>3</sup> The framework describes the relationship between P4P program features, external and implementation factors, and provider cognitive/affective and behavioral responses on processes of care and patient outcomes (see Fig. 1). Table 1 describes each of these categories. Due to heterogeneity among the studies, we did not perform meta-analysis.

We assessed the overall strength of evidence for the effectiveness of P4P on Veteran care using a method developed by the Agency for Healthcare Research and Quality.<sup>12</sup>

## RESULTS

We reviewed 1031 titles and abstracts and selected 74 articles for full-text review. Thirty met inclusion criteria and provided evidence addressing the key questions (Fig. 2). We invited 29 individuals for key informant interviews. Seventeen participated (see online Appendix 6). Tables 2 and 3 provide quotes from the interviews.

### P4P in VHA Settings

**Effectiveness of P4P.** Four articles<sup>13, 23, 29, 30</sup> from three studies<sup>13, 23, 29</sup> provide data on the effectiveness of P4P in VHA settings (see online Appendix 7 for detail). Overall, the evidence is insufficient to determine whether P4P results in durable improvements in the quality of care or health of Veterans. The sole 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.<sup>13</sup> Two observational studies reported evidence of positive effects of P4P on processes of care. However, it is possible that these findings may have been influenced by concomitant public reporting<sup>23</sup> and denominator management (i.e., a decrease in the number of patients eligible for a performance measure that may be positive, resulting in improvements in identification; or negative, resulting from gaming).<sup>29</sup>

**Unintended Consequences.** Eleven studies published in 13 articles<sup>13–15, 24, 29–36, 42</sup> examined potential unintended consequences in VHA settings (see online Appendices 8 and 9). In general, qualitative studies and those using administrative data identify the potential for overtreatment associated with performance measures.<sup>14, 31, 32, 34</sup> However, one RCT found that P4P for hypertension did not increase the risk of

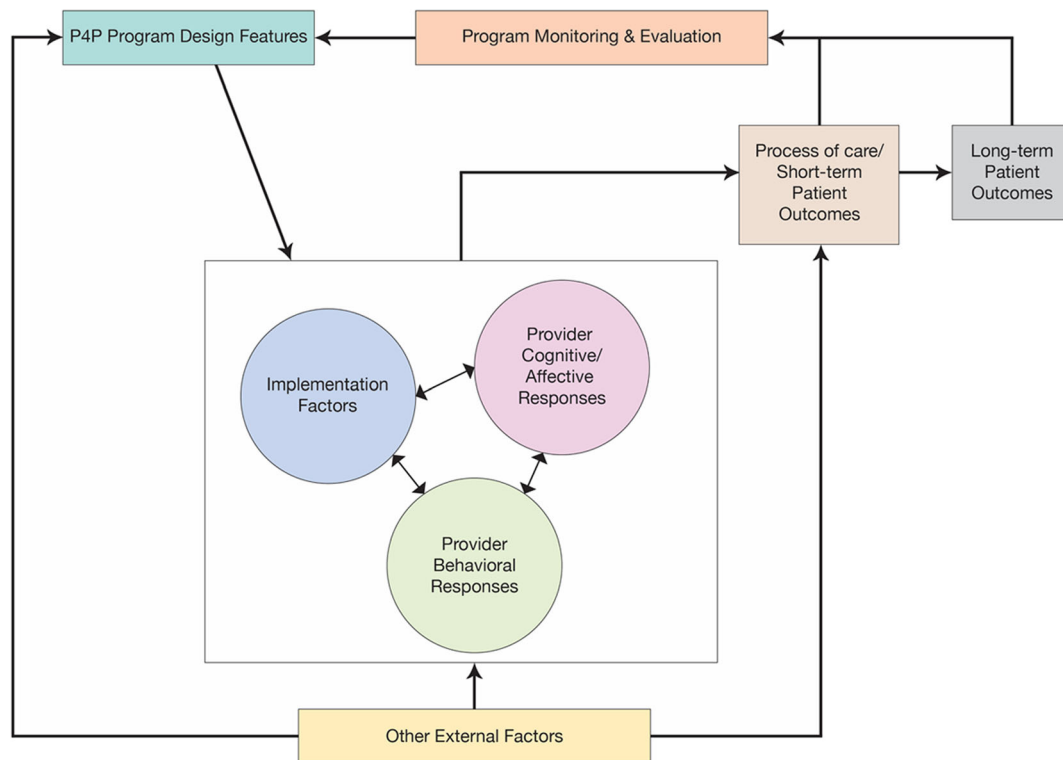


Fig. 1 Conceptual framework. P4P pay-for-performance.

hypotension despite findings from a sub-study that suggested subjects were concerned about the risk of overtreatment.<sup>13, 42</sup> Other studies examining unintended consequences reported findings congruent with denominator management,<sup>29</sup> but no evidence of risk selection.<sup>30, 33</sup> Qualitative studies found that participants felt performance measures 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),<sup>15, 24, 35</sup> and that they may negatively affect team dynamics, particularly if metrics are incentivized.<sup>35</sup>

**Findings from Key Informant Interviews (See Figure 3 for Themes, and Table 2 for Quotes).** Consistent with the literature, key informants voiced concern for potential overtreatment, particularly in facilities with metric-driven cultures, and more commonly with metrics that vary (e.g., blood pressure). Other concerns included *denominator management*, *gaming*, *risk selection* and *health disparities*—particularly for low SES Veterans, and the need to mitigate against *teaching to the test/attention shift* by having a variety of actively monitored *valid* metrics covering different aspects of care.

**Implementation of P4P in VHA Settings.** Thirteen studies reported in 16 articles<sup>13–28</sup> provide data examining program design features or implementation factors and/or provider cognitive or affective responses related to pay-for-performance programs in VHA settings (see online Appendices 10 and 11

for detail). In general, studies found physician-targeted incentives to be more effective than those targeting groups or practices<sup>13</sup>; the agreement between EHR data and manual review varied by metric<sup>14, 22</sup>; the relationship between access and patient satisfaction varied by the access metric used<sup>20</sup> as well as whether the patient was new or returning<sup>21</sup>; and the difficulty of achieving multi-tasked metrics was not directly related to the number of tasks involved.<sup>17</sup> Studies also found no difference in the achievement of actively vs. passively monitored metrics,<sup>25</sup> and were mixed on the impact of the removal of incentives on performance.<sup>13, 23</sup> Areas of improvement for implementing performance measures at the local level were suggested.<sup>15, 24</sup> One study examined provider affective/cognitive responses and found that P4P had no impact on goal commitment.<sup>26</sup>

**Findings from KI Interviews (See Figure 4 for Themes, and Table 2 for Quotes). Program Design Features.** Key informants consistently stressed the need for larger and more frequent incentives attached to clinically meaningful metrics that are within provider control. Other key themes included the potential benefit of incentivizing teams or front-line staff, placing greater emphasis on patient evaluation metrics, establishing the validity of performance measures, the feasibility of achieving performance measures at the local level, and the importance of de-intensification metrics.

**Implementation Factors.** Common among key informants was a belief that the implementation of P4P and performance

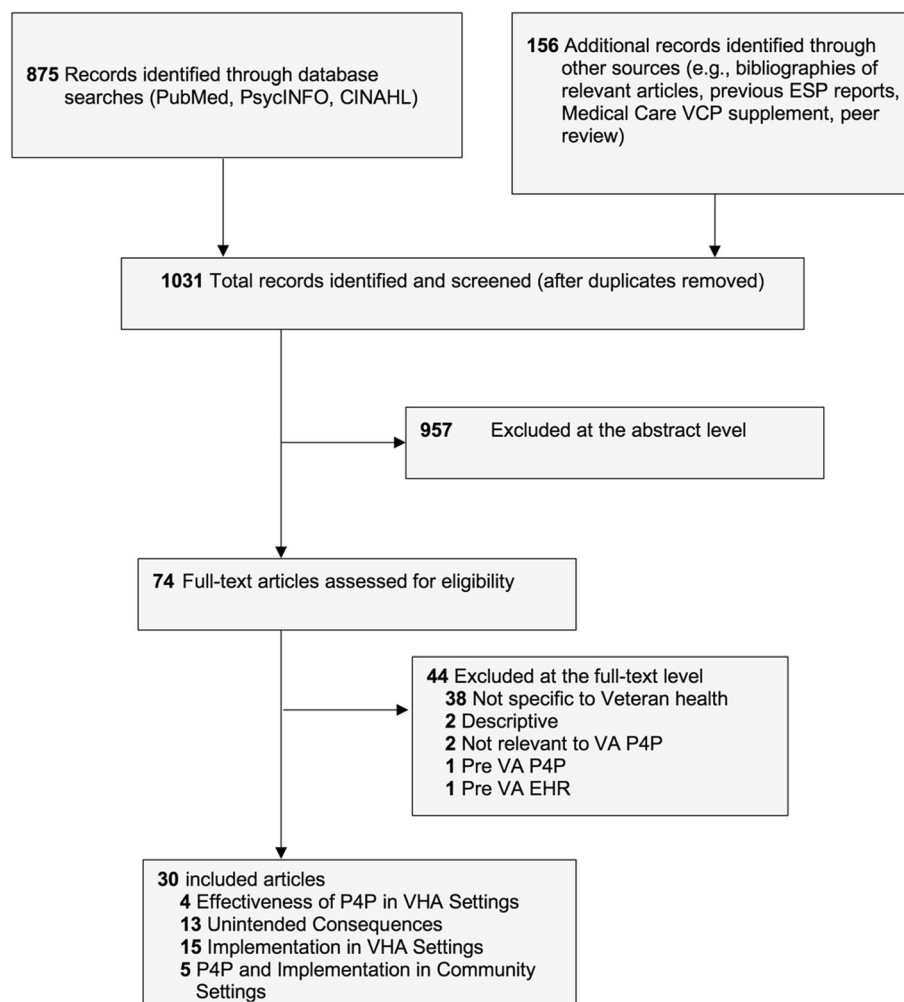
**Table 1 Description of Implementation Framework Categories**

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	<p>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.</p> <p>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.</p> <p>Inner setting: Refers to characteristics of the institution or organization itself.</p> <p>Provider characteristics: Refers to demographic characteristics (e.g., age, gender, race/ethnicity), as well as other factors such as experience and specialization.</p>
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 (e.g., – blood pressure, cholesterol levels) and health outcomes.

measures in the VHA needs improvement. Key informants felt that VHA physicians are not able to identify their P4P-linked metrics; that the implementation of metrics has historically lacked interpretation, documentation, and support; and that implementation should include the resources necessary to ensure success and take into consideration facility-level contextual factors.

### *Provider Cognitive, Affective, and Behavioral Responses.*

Key informants expressed the belief that the intrinsic motivation of physicians is the driving factor in achieving evidence-based performance metrics that make clinical sense.



**Fig. 2 Literature flowchart. EHR electronic health record, ESP evidence-based synthesis program, P4P pay-for-performance, VA Veterans Administration, VCP Veterans Choice Program, VHA Veterans Health Administration.**

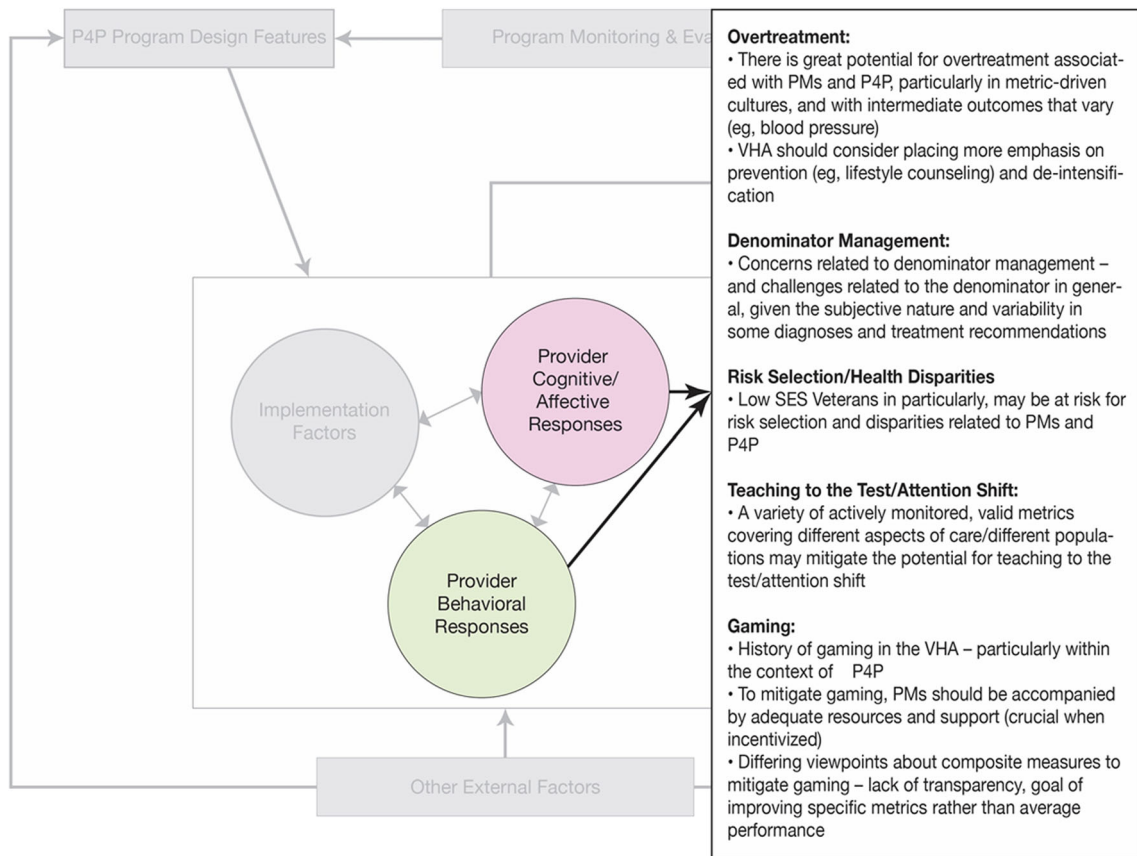


Fig. 3 Key informant interviews: themes—unintended consequences.

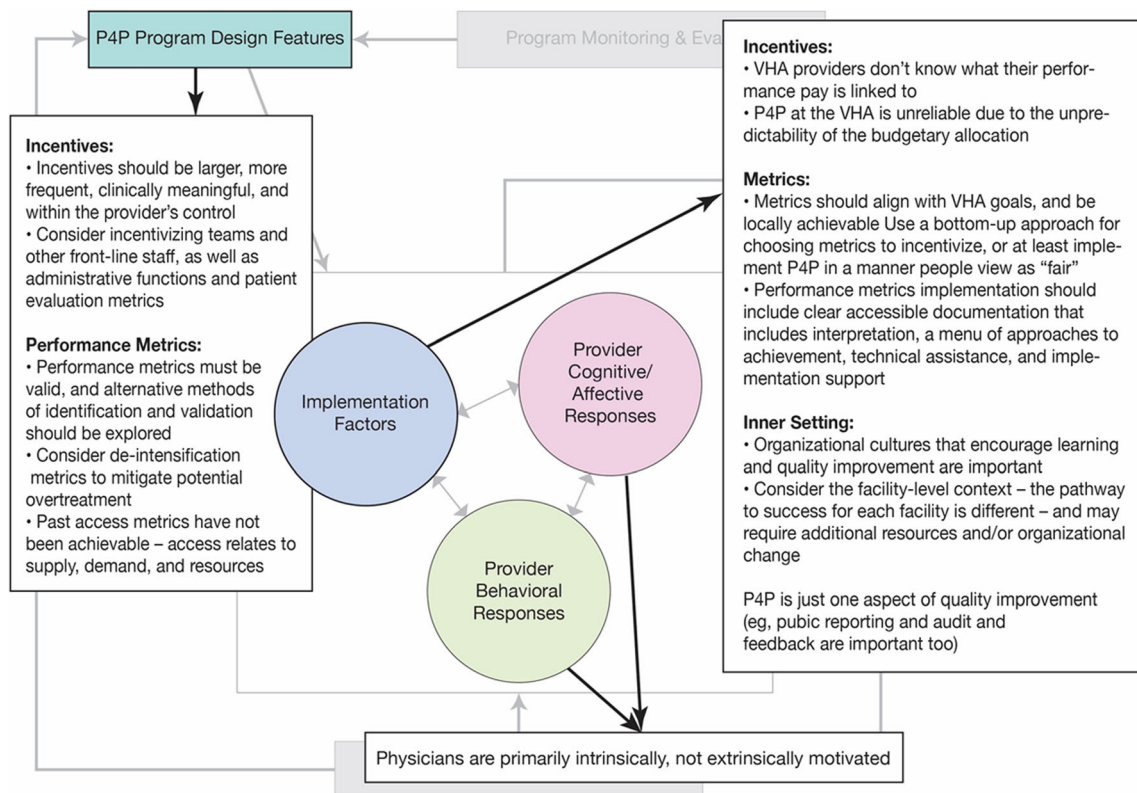


Fig. 4 Key informant interviews: themes—program design features and implementation factors in VHA settings.

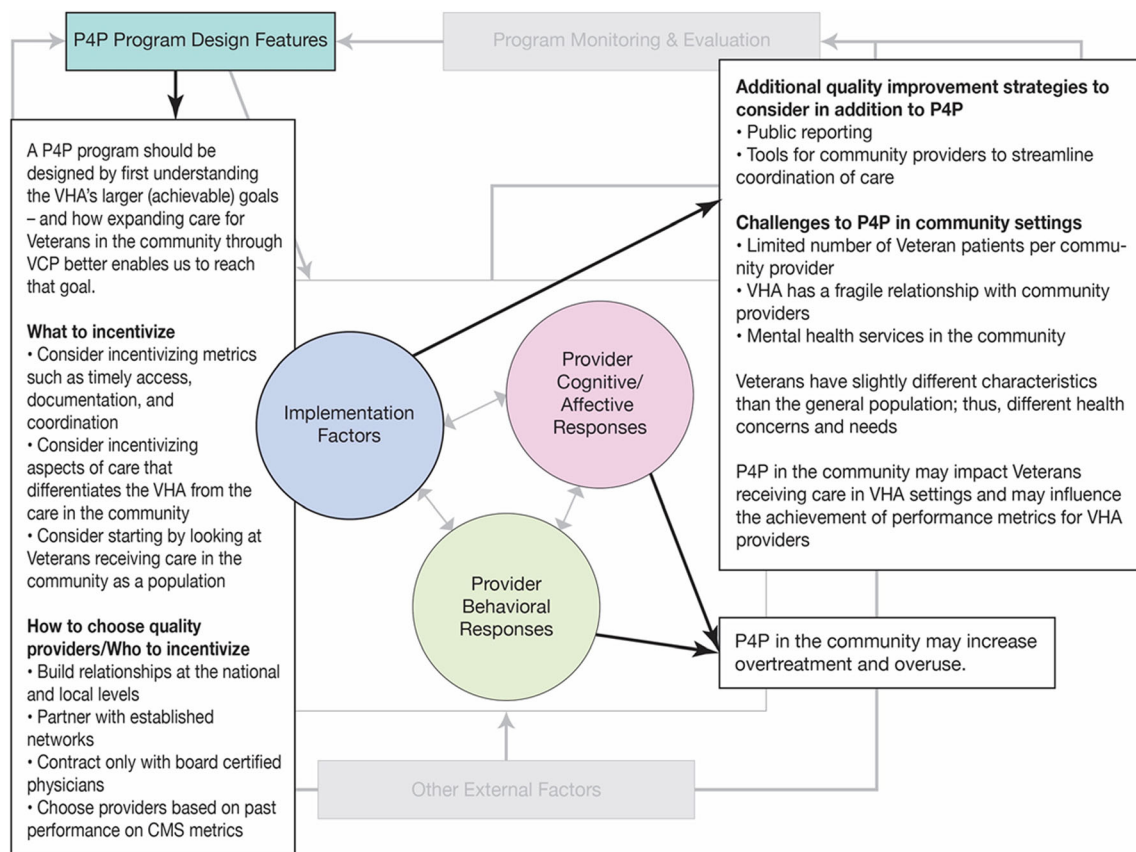


Fig. 5 Key informant interviews: themes—program design features and implementation factors in non-VHA/community settings.

## P4P at the Intersection of VHA and Community Care

**Implementation of Pay-for-Performance in Community Settings. Findings from the Literature.** We identified five studies examining P4P or related implementation factors in Veteran populations in community settings.<sup>37–41</sup> One study identified published survey instruments examining cross-system access and coordination.<sup>37</sup> Across studies, findings suggest that Veterans, providers, and VHA administrators are concerned that VCP already has and will continue to result in fragmented care, poor communication and coordination among providers, and places an additional burden on and VHA providers and on Veterans.<sup>37, 40, 41</sup> Other concerns include barriers to sharing medical records,<sup>39–41</sup> and differences between providers who are interested in VCP participation and those who are not (see online Appendix 12 for detail).<sup>38</sup>

**Findings from Key Informant Interviews (See Figure 5 for Themes, and Table 3 for Quotes). Program Design Features.** Key informants stressed the importance of considering the overarching goal of the VCP in decisions about the metrics to incentivize. Although key informants recognized the need for increased access to healthcare for Veterans, they also suggested goals including the receipt of quality care, coordination of care, cost effectiveness, and “conservative

care” (e.g., restrictive selection of surgical patients). Some key informants suggested that known differences between VHA and community care be used to guide metric selection.

Several key informants suggested that incentives might help to address known challenges related to the receipt of documentation and the overall quality of records received from community providers—particularly early in the program. Key informants also suggested the possibility of pooled population guideline-based metrics to compare the outcomes of Veterans receiving care in VHA to VCP, acknowledging that population-based incentives are unlikely to motivate provider behavior.

Key informants stressed the importance of building relationships between VHA and community providers at both national and local levels, and raised the question of how to select high-quality providers. Suggestions included contracting with established networks and/or only with board certified physicians; as well as using providers’ performance on established metrics (e.g., Centers for Medicare and Medicaid Services) for selection.

**Implementation Factors.** Key informants suggested a number of quality improvement strategies to accompany P4P in the community and stressed the importance of transparency and public reporting. To improve coordination of care, they suggested implementing systems that would provide community providers with the pop-up reminders available in

**Table 2 Evidence and Policy Implications—P4P in VHA Settings**

Study evidence	Quotes from KI interviews	Themes	Policy implications
<p><b>Program design features</b>                      Ten studies examined program design features.<sup>13–22</sup> In general, studies found:</p> <ul style="list-style-type: none"> <li>• Physician-targeted incentives to be more effective than those targeting practices</li> <li>• The degree of agreement between EHR data and manual review varies by metric</li> <li>• The difficulty of achieving multi-tasked metrics is not directly related to the number of tasks involved</li> <li>• The relationship between access and patient satisfaction varies by measure and in new vs. returning patients</li> </ul>	<p><i>We're being encouraged to do a lot of things to meet a lot of different quality measures...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.</i></p> <p><i>I want to increase access. I want to make sure that veterans don't have to wait a day to be seen. I want same-day access for everything. Then I would need to have more appointments but I don't have more doctors. What I would like to incentivize then is how to become a more efficient group.</i></p> <p><i>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.</i></p> <p><i>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.</i></p>	<p>Incentives:</p> <ul style="list-style-type: none"> <li>• Incentives should be larger, more frequent, clinically meaningful, and within the provider's control</li> <li>• Consider incentivizing teams and other front-line staff, as well as administrative functions and patient evaluation metrics</li> </ul> <p>Performance Metrics:</p> <ul style="list-style-type: none"> <li>• Performance metrics must be valid, and alternative methods of identification and validation should be explored</li> <li>• Consider de-intensification metrics to mitigate potential overtreatment</li> <li>• Past access metrics have not been achievable—access relates to supply, demand, and resources</li> </ul>	<ul style="list-style-type: none"> <li>• Regardless of whether performance metrics are incentivized, they should be valid, achievable, and within a provider's control</li> <li>• Consider re-evaluation of the size (monetary), frequency, and target (provider vs team) of performance pay in the VHA.</li> <li>• Potential overtreatment and overuse may be an unintended consequence of performance metrics, and de-intensification metrics should be considered.</li> </ul>
<p><b>Implementation factors</b>                      The eight studies<sup>13, 15, 23–28</sup> examining implementation factors found:</p> <ul style="list-style-type: none"> <li>• There was no difference in the achievement of actively vs. passively monitored metrics</li> <li>• The evidence related to the impact of the removal of incentives on performance is mixed evidence related to the impact of the removal of incentives on performance</li> <li>• Providers express frustration for current top-down implementation strategies, and suggest areas of improvement for implementing performance metrics at the local level</li> <li>• Facilities with high adherence to clinical guidelines were more likely to deliver more timely, individualized and non-punitive feedback</li> <li>• Audit and feedback processes remained largely unchanged after PACT implementation</li> </ul>	<p><i>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.</i></p> <p><i>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?</i></p> <p><i>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.</i></p> <p><i>[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.</i></p>	<p>P4P is just one aspect of quality improvement (e.g., public reporting and audit and feedback are important too)</p> <p>Incentives:</p> <ul style="list-style-type: none"> <li>• VHA providers do not know what their performance pay is linked to</li> <li>• P4P at the VHA is unreliable due to the unpredictability of the budgetary allocation</li> </ul> <p>Metrics:</p> <ul style="list-style-type: none"> <li>• 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"</li> <li>• Performance metrics implementation should include clear accessible documentation that includes interpretation, a menu of approaches to achievement, technical assistance, and implementation support</li> </ul> <p>Inner setting:</p> <ul style="list-style-type: none"> <li>• Organizational cultures that encourage learning and quality improvement are important</li> <li>• Consider the facility-level context—the pathway to success for each facility is different—and may require additional resources and/or organizational change</li> </ul>	<ul style="list-style-type: none"> <li>• Use a transparent, bottom-up approach for selecting and implementing metrics, and secure provider and staff buy-in.</li> <li>• Foster overall and local-level cultures that encourage learning and value quality improvement</li> </ul>

(continued on next page)

Table 2. (continued)

Study evidence	Quotes from KI interviews	Themes	Policy implications
<p><b>Provider cognition, affect, and behavior</b></p> <p>One study examined provider affective/cognitive responses, and found that P4P had no impact on goal commitment.<sup>26</sup> Twelve articles<sup>13–15, 24, 29–36</sup> examined unintended consequences and found:</p> <ul style="list-style-type: none"> <li>• Administrative data support the potential for over-treatment associated with performance metrics.</li> <li>• However, an RCT found no association between P4P for hypertension and hypotension</li> <li>• There was evidence of denominator management associated with a VISN Director-aimed incentive</li> <li>• There was no evidence of risk selection</li> <li>• Providers perceived both negative and positive unintended consequences associated with performance metrics</li> </ul>	<p><i>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 \$1000 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.</i></p> <p><i>I think that de-intensification and intensification are mirror images of each other. There's simple and then there are more complex ones on both sides. You start with the simpler one and you may end with the simpler one, depending on how the performance measures are used. I think the goal is to identify those things that can be operationalized in some way. I think people's natural assumption is that it's much harder to stop things than it is to start things, which I think isn't true. It has become acclimatized that there is more risk in not treating people than there is in treating people, which is debatable I think. That is one of the barriers you have to overcome.</i></p> <p><i>There are two kinds of unintended consequences that I would worry about with SES. One is that physicians may not want to take patients that are lower SES because they may be harder to achieve high quality health care delivery. So, you have to worry about them getting rid of low SES patients. The other is that if you don't risk adjust, facilities or physicians that take low SES patients tend to look worse on average and so then when you start tying financial incentives to it, they start getting a smaller piece of the financial pie.</i></p> <p><i>I think sometimes P4P works better at the level if you can actually get higher-level managers that have control of resources to deploy the resources to improve areas that are problematic. They can also backfire spectacularly as we've found in the Arizona waitlist. When there aren't enough resources and you put strong incentives on the measure, you're going to get gaming of the system.</i></p> <p><i>Inadequate resources, unrealistic expectations, and the opportunity to cheat all are factors in 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.</i></p>	<p>Physicians are primarily intrinsically, not extrinsically motivated</p> <p>Overtreatment:</p> <ul style="list-style-type: none"> <li>• There is great potential for overtreatment associated with PMs and P4P, particularly in metric-driven cultures, and with intermediate outcomes that vary (e.g., blood pressure)</li> <li>• VHA should consider placing more emphasis on prevention (e.g., lifestyle counseling) and de-intensification</li> </ul> <p>Denominator Management:</p> <ul style="list-style-type: none"> <li>• 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</li> </ul> <p>Risk Selection/Health Disparities</p> <ul style="list-style-type: none"> <li>• Low SES Veterans in particular, may be at risk for risk selection and disparities related to PMs and P4P</li> </ul> <p>Teaching to the Test/Attention Shift:</p> <ul style="list-style-type: none"> <li>• 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</li> </ul> <p>Gaming:</p> <ul style="list-style-type: none"> <li>• History of gaming in the VHA—particularly within the context of P4P</li> <li>• To mitigate gaming, PMs should be accompanied by adequate resources and support (crucial when incentivized)</li> <li>• Differing viewpoints about composite measures to mitigate gaming—lack of transparency, goal of improving specific metrics rather than average performance</li> </ul>	<ul style="list-style-type: none"> <li>• Potential overtreatment and overuse may be an unintended consequence of performance metrics, and de-intensification metrics should be considered.</li> <li>• Gaming will likely be mitigated by providing the resources support necessary for achievement</li> </ul>

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

the VHA and VHA formulary lists by adapting existing tools (e.g., Epocrates®) or creating new ones.

Key informants discussed differences between Veterans and the general population, largely noting lower socioeconomic status (SES) among Veterans, as well as greater mental health needs, higher rates of substance use, and a large rural population. Key informants felt it important to account for SES when implementing P4P and expressed concern for the limited availability of quality care for Veterans living in rural areas.

Key informants identified potential challenges the VHA might face in implementing P4P in community settings. Most commonly, key informants worried that because Veterans accessing care through VCP would be dispersed widely (comprising a small percentage of a provider's patient population), community providers would view VCP as just one of many insurers—and for many providers, the smallest. This may inhibit the potential impact of P4P in community care, particularly if incentivized metrics do not align with those of other



**Table 3 Evidence and Policy Implications—P4P in Non-VHA/Community Settings**

Study evidence	Quotes from KI interviews	Themes	Policy implications
<p><b>Program design features</b>                      One study<sup>37</sup> examined program design features relevant to P4P in community settings, and found:                      • A number of survey instruments examining cross-system access and coordination exist</p>	<p><i>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?”</i>  <i>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.</i></p>	<p>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 differentiates 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</p>	<ul style="list-style-type: none"> <li>• Initially target areas in need improvement such as documentation and coordination (e.g., receipt of records from community providers)</li> <li>• Develop relationships with providers and health systems with records of strong performance on commonly used, well-validated, and well-established metrics</li> </ul>
<p><b>Implementation factors</b>                      The four studies<sup>38-41</sup> 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,                      • 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.</p>	<p><i>As the VA becomes more like an insurance company, we need to start thinking like an insurance company.</i>  <i>Veterans are one percent of their patient population. Providers in the community are often working with ten 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.</i>  <i>You have to make sure that if you put these carrots out for Choice performance pay and the 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.</i>  <i>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.</i>  <i>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.</i></p>	<p>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 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</p>	<ul style="list-style-type: none"> <li>• 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.</li> <li>• Use tools such as public reporting to complement P4P</li> <li>• Developing tools and resources to streamline the data-sharing and coordination necessary to inform a cross-system P4P program</li> <li>• Consider how funding expanded care in the community might affect funding for Veterans receiving care in VHA settings</li> <li>• Consider how performance by community providers might impact measured performance for VHA providers</li> </ul>

(continued on next page)

Table 3. (continued)

Study evidence	Quotes from KI interviews	Themes	Policy implications
<b>Provider cognition, affect, and behavior</b>	<p><i>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.</i></p>	P4P in the community may increase overtreatment and overuse	<ul style="list-style-type: none"> <li>• Be vigilant for overtreatment and for differences in standards of care (e.g., opioid prescriptions)</li> </ul>

insurers. Furthermore, if providers have only a handful of VCP patients, their measured performance may vary widely and result in unreliable measures of quality. Key informants 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 key informants discussed the potential tradeoffs of using narrow networks to increase the percentage of VCP patients per provider and access to high quality care, particularly for rural Veterans.

There was concern among key informants that the VHA may have already developed a fragile relationship with community providers due to slow payment, with providers refusing to accept Veteran patients. They advised that the VHA pay providers in a timely fashion and reiterated that P4P metrics must be achievable, or risk additional providers opting out and resulting in even poorer access for Veterans otherwise.

Concerns related to mental health treatment were raised frequently. Key informants cautioned that sending Veterans to community mental health providers will likely reduce the quality of care and coordination Veterans receive, especially for those with combat related PTSD, substance use disorders, and those experiencing homelessness. Key informants were also concerned that implementing P4P metrics would present a barrier to entry for providers, as the use of performance metrics is uncommon in community mental health. In addition, they felt strongly that providers would resist sharing treatment notes and other records.

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

#### **Provider Cognitive, Affective, and Behavioral Responses.**

Key informants voiced concern for unintended consequences resulting from P4P in community settings, particularly

overtreatment and overuse. They felt that overtreatment may be more common in the community than in VHA settings, and that the lack of integration and coordination with VCP might place Veterans at increased risk.

## DISCUSSION

We examined 30 articles and conducted interviews with 17 key informants to help inform the implementation of pay-for-performance programs for Veterans in VHA and community settings. Although we found insufficient evidence to determine the degree to which P4P affects Veteran outcomes, we identified information in the literature and through key informant interviews that may help guide the implementation of P4P and maximize potential benefits while minimizing negative unintended consequences.

Several themes emerged from the interviews related to general issues with P4P in VHA that are consistent with the findings from published literature (see Table 2).<sup>3, 4</sup> First, key informants felt that performance measures should be valid and well-designed and cited a need for further research evaluating alternate validation methods. Second, findings from a handful of included studies<sup>14, 31, 32, 34, 35</sup> combined 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. Third, consistent with qualitative findings,<sup>24</sup> provider key informants consistently stated that they did not know which metrics were incentivized and did not feel that the current P4P structure influences their behavior. Fourth, despite previous research stressing the importance of bottom-up, realistic metrics,<sup>3, 4</sup> qualitative findings illustrate VHA staff are frustrated with current implementation practices.<sup>15, 24</sup> There was strong consensus among key informants that incentivized metrics need to be achievable, that local resources are necessary for

achievement, that incentivization decisions are perceived as equitable, and that incentive payments are predictable and reliable. Fifth, included studies found that metric-driven cultures were more prone to potential overtreatment,<sup>31, 32</sup> and that overtreatment may be mitigated by incentivizing appropriate care rather than treatment or targets.<sup>13</sup>

Several themes related to P4P in community settings also emerged (see Table 3). First, key informants expressed that, given known challenges related to receipt of documentation,<sup>40, 41</sup> data and care coordination may be an initial area for P4P to target. Secondly, they stressed the importance of establishing relationships with local providers and suggested ways to select providers with demonstrated records of quality care. Third, there was concern about the VA's ability to influence provider behavior using P4P and to accurately estimate quality at the provider level, given that Veterans may comprise a small percentage of an individual provider's patient population. Fourth, consistent with the findings from previous research,<sup>3</sup> key informants stressed that P4P is only one part of a quality improvement strategy. Fifth, along with findings from included studies,<sup>39, 40</sup> key informants cited ongoing challenges in coordinating care with community providers, and suggested the development of tools to facilitate coordination. Sixth, there was concern for and uncertainty about how VCP may affect Veterans who continue to receive care in VHA. Seventh, key informants 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 VHA provider metrics and performance pay. Finally, key informants stressed that a fundamental difference between VHA and community care is that the VHA tends to be more conservative. They felt that despite evidence of potential overtreatment in VHA settings,<sup>31, 32</sup> overtreatment is even more common in community settings and community providers may be more prone to prescribing opioids than VHA providers.<sup>39</sup>

Our approach to the topic of P4P and Veteran health has several strengths. To our knowledge, this is the first paper to examine P4P specific to Veteran care. The VHA is a large integrated system that differs significantly from others in the United States, and the recent expansion of community care adds additional complexity. We recognized early that much of the information we sought related to the implementation of P4P would not be found in published research—particularly related to the intersection of VHA and community care. We interviewed VHA stakeholders with P4P expertise as researchers, clinicians, and administrators to provide informed insight into the implementation factors and program design features important to P4P success in the community.

Our review is limited by the paucity of research directly assessing the effectiveness of P4P in VHA settings, and the heterogeneity in the way that P4P is implemented in VHA settings. We therefore focused primarily on examining program design features, implementation factors, and unintended consequences. As research examining VCP is just beginning

to emerge, our findings regarding P4P in community settings are influenced heavily by our key informant interviews. The breadth of topics and outcomes made it difficult to apply strict study design criteria. Thus, we included studies with less-rigorous methodology, some of which had small samples. We conducted 17 interviews 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 key informants could yield a different subset of themes.

Although performance pay has been a part of the VHA for more than a decade, little research has evaluated its effectiveness, and no research has explored alternatives. The nature of the VHA as an integrated yet closed system provides a unique opportunity for research comparing P4P program design and implementation.

Although Veterans seeking care in the community is not a new phenomenon, continued funding for VCP necessitates the need for more comprehensive evaluation. Current research, programs, and initiatives funded largely by QUERI are evaluating metrics, quality, and P4P programs directly within the context of community care. More research is needed to identify how expanded care in the community may 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.

## CONCLUSION

While the effectiveness of P4P in VHA settings is understudied, we highlight 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 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.

---

### Acknowledgements:

We would like to acknowledge Rose Relevo for conducting literature searches and also the contributions of our stakeholders and Technical Expert Panel.

**Prior Presentation:** The contents of this manuscript have not been presented at any conference.

**Corresponding Author:** Karli K. Kondo, PhD; Portland VA Health Care System Evidence-based Synthesis Program, Portland, OR, USA (e-mail: Karli.Kondo@va.gov).

**Funding** This project was funded by the US Department of Veterans Affairs, Veterans Health Administration (VHA) ESP Project #05-225.

### Compliance with Ethical Standards:

**Conflict of Interest:** The authors declare that they do not have a conflict of interest.

## REFERENCES

- Senate: Veterans' Affairs Committee. S. 2484 (108th Congress): Department of Veterans Affairs Health Care Personnel Enhancement Act of 2004. December 3 2004.
- Clough JD, McClellan M. Implementing MACRA: Implications for Physicians and for Physician Leadership. *JAMA* 2016;315(22):2397-2398.
- 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.
- 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.
- 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.
- Department of Veterans Affairs. VHA Office of Community Care. Available at <https://www.va.gov/purchasedcare/>. Accessed 24 March 2017.
- Commission on Care. Final Report of the Commission on Care. 2016.
- Higgins J, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0. 2011; <http://handbook.cochrane.org/>. Accessed 24 March 2017.
- 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](http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp). Accessed 24 March 2017.
- 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.
- Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005;15(9):1277-1288.
- 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 28 Dec 2016.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Prentice J, Legler A, Li D, Pizer SD. Optimizing Access Metrics in the VA. [https://www.hsrdresearchvagov/for\\_researchers/cyber\\_seminars/archives/1166-notes.pdf](https://www.hsrdresearchvagov/for_researchers/cyber_seminars/archives/1166-notes.pdf).
- 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.
- 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.
- 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.
- Damschroder LJ, Robinson CH, Francis J, et al. Effects of performance measure implementation on clinical management and provider motivation. *J Gen Intern Med* 2014;29 Suppl 4:877-884.
- 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.
- 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.
- 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.
- Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci* 2006;1:9.
- 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.
- 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.
- 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.
- Kerr EA, Lucatoro MA, Holleman R, Hogan MM, Klammer 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.