

Evidence-based Synthesis Program (ESP)

Understanding the Intervention and Implementation Factors Associated with Benefits and Harms of Pay for Performance in Healthcare

A Systematic Review

Portland VA Medical Center

November 16, 2015

Evidence-based Synthesis Program (ESP)

Acknowledgements

Report Authors

- Karli Kondo, PhD, MA
- Cheryl Damberg, PhD, MPH
- Aaron Mendelson, BA
- Makalapua Motu'apuaka, BS
- Michele Freeman, MPH
- Maya O'Neil, PhD
- Rose Relevo, MILS, MS
- Devan Kansagara, MD, MCR

Presenters & Discussants

- Karli Kondo, PhD, MA
- Aaron Mendelson, BA
- Devan Kansagara, MD, MCR
- Cheryl Damberg, PhD, MPH
- David Atkins, MD, MPH
- Joe Francis, MD

Report Nominators & Reviewers

- David Atkins, MD, MPH
- Joe Francis, MD
- Carolyn Clancy, MD
- Edward Chow, MD
- Laura Damschroder, MS, MPH
- Laura Dimmler, MPA, PhD
- John McConnell PhD
- Richard Stenson, MHA, MBA
- Kevin Volpp, MD, PhD

Evidence-based Synthesis Program (ESP)

Disclosure

This report is based on research conducted by the Evidence-based Synthesis Program (ESP) Center located at the Portland VA Medical Center funded by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Health Services Research and Development. The findings and conclusions in this document are those of the author(s) who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the Department of Veterans Affairs or the United States government. Therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs. No investigators have any affiliations or financial involvement (e.g., 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.

Evidence-based Synthesis Program (ESP)

VA Evidence-based Synthesis (ESP) Program Overview

- Sponsored by VA Quality Enhancement Research Initiative (QUERI) Program.
- Established to provide timely and accurate syntheses/reviews of healthcare topics identified by VA clinicians, managers and policy-makers, as they work to improve the health and healthcare of Veterans.
- Builds on staff and expertise already in place at the Evidence-based Practice Centers (EPC) designated by AHRQ. Four of these EPCs are also ESP Centers:
 - Durham VA Medical Center; VA Greater Los Angeles Health Care System; Portland VA Medical Center; and Minneapolis VA Medical Center.

Evidence-based Synthesis Program (ESP)

- Provides evidence syntheses on important clinical practice topics relevant to Veterans, and these reports help:
 - develop clinical policies informed by evidence,
 - the implementation of effective services to improve patient outcomes and to support VA clinical practice guidelines and performance measures, and
 - guide the direction for future research to address gaps in clinical knowledge.
- Broad topic nomination process – e.g. VACO, VISNs, field – facilitated by ESP Coordinating Center (Portland) through online process:

<http://www.hsrd.research.va.gov/publications/esp/TopicNomination.cfm>

Evidence-based Synthesis Program (ESP)

- Steering Committee representing research and operations (PCS, OQP, ONS, and VISN) provides oversight and guides program direction.
- Technical Advisory Panel (TAP)
 - Recruited for each topic to provide content expertise.
 - Guides topic development; refines the key questions.
 - Reviews data/draft report.
- External Peer Reviewers & Policy Partners
 - Reviews and comments on draft report
- Final reports posted on VA HSR&D website and disseminated widely through the VA.

<http://www.hsrd.research.va.gov/publications/esp/reports.cfm>

Evidence-based Synthesis Program (ESP)

Current Report

Understanding the Intervention and Implementation Factors Associated with Benefits and Harms of Pay for Performance in Healthcare

(May 2015)

Full-length report available on the ESP website (intranet only, release date November 20, 2015):

<http://vawww.hsrd.research.va.gov/publications/esp/financialincentives.cfm>

Evidence-based Synthesis Program (ESP)

Overview of Today's Presentation

- Background
- Scope of the review
- Results (94 studies)
 - Summary of results by key question
- Future research
- Implications for VA
- Discussion/Q&A

Evidence-based Synthesis Program (ESP)

Background

- Over the last decade, pay for performance (P4P) programs have been implemented in a variety of settings, including the Veterans Health Administration (VHA) as a means to improve efficiency and quality.
- A number of recent reviews have recently summarized the literature, and have generally found insufficient evidence to broadly characterize the balance of benefits and harms.
- P4P programs are complex interventions, and the effects may depend in part on the setting in which they are implemented, methods of implementation, patient populations, and program characteristics.

Evidence-based Synthesis Program (ESP)

Scope of the Review

- Key Question 1: What are the effects of pay for performance programs on patient outcomes and processes of care?
- Key Question 2: What implementation factors modify the effectiveness of pay for performance?
- Key Question 3: What are the positive and negative unintended consequences, including any effect on health disparities, associated with pay for performance?

Evidence-based Synthesis Program (ESP)

Scope of the Review

- **Criteria:**

- **Population**: Healthcare providers at the individual, managerial (eg, VISN directors), group, and institutional levels. General patient populations that are part of existing performance measures.
- **Intervention**: Financial incentives/pay-for performance programs
- **Comparator**: Other financial incentive models; non-financial incentives; usual care
- **Outcome**: Patient outcome measures (e.g., utilization such as ER visits, intermediate physiological markers such as blood pressure, HbA1c, and cholesterol), process of care measures (e.g., screening)
- **Timing**: Any
- **Setting**: VHA or other large managed care institutions, other healthcare systems in the US, and healthcare systems in countries with health systems similar to the VHA.

Evidence-based Synthesis Program (ESP)

Methods

- Search strategies based on recent RAND report. Conducted a search to update their report and avoid overlap.
- Searched MEDLINE (PubMed), Cochrane Database of Systematic Reviews (Ovid), CINAHL (EBSCO), and PsycINFO (Ovid) through April 2014
- Additional articles and reviews considered for inclusion were obtained from reference lists and reviewer suggestions
- Grey literature search of additional databases and websites (e.g., AHRQ, NICE, Kaiser Permanente's Center for Health Research)
- Targeted PubMed and Google searches for specific P4P programs (e.g., Quality and Outcomes Framework [QOF])

Evidence-based Synthesis Program (ESP)

Methods

- Used a “best evidence” approach.
 - Included direct P4P programs targeting providers.
 - Studies with $N \geq 10,000$ with a comparison group or at least 3 time points analyzing trends for KQ1
- Summarized RAND’s findings in collaboration with primary author
- Key informant interviews
 - Experienced P4P researchers
 - 60 minute Semi-structured phone interview
 - Questions related to implementation, unintended consequences, future research needs

Evidence-based Synthesis Program (ESP)

Overall Results

- Search yielded 1,363 citations; 509 were selected for full-text review
- 93 included studies plus one recommended by a peer reviewer
 - Key Question 1 (47)
 - Key Question 2 (41)
 - Key Question 3 (42)
- 14 Key informant interviews

Evidence-based Synthesis Program (ESP)

KQ 1 Results:

By Type of Outcome

- *Process of Care*

- 42 Studies

- Ambulatory (36 studies)

- Modest improvements associated with UK's Quality and Outcomes Framework (QOF), with the largest improvements in years 1 and 2, followed by a plateau or slowing of improvement rates (17 studies)
- Mixed findings in US and other countries, with some studies reporting modest short term improvements (e.g., Taiwan's Diabetes Mellitus P4P), and others, particularly longer-term studies reporting a slowing of improvement or little to no association

Evidence-based Synthesis Program (ESP)

KQ 1 Results:

- ***Process of Care***
 - 42 Studies
 - Hospital (6 studies)
 - No significant improvement associated with the Premier Hospital Quality Incentive Demonstration (HQID) and the Hospital Value-Based Purchasing (HVBP)
 - VHA study targeting acute coronary syndrome, heart failure, and pneumonia process measures reported significant improvement for 6 of the 7 measures examined
 - Internationally, studies report generally positive effects, with a slowing of improvements or a plateau over time

Evidence-based Synthesis Program (ESP)

KQ 1 Results:

By Type of Outcome

- ***Patient Outcomes***

- 23 Studies

- Ambulatory (19 studies)

- No clear evidence that the QOF increases clinical target achievement, with achievement for some (e.g, HbA1c) lower than the pre-QOF trend. In areas that improved, the greatest improvement was in the first year, with a plateau or slowing of improvement over time (11 studies)
- Little to no evidence in the US and Taiwan, with Taiwan's DM-P4P associated with no significant short term effect, but marginally fewer diabetes-related complications and hospitalizations in the long-term, and studies in the US reporting fewer ED visits but marginally higher acute and ambulatory care-sensitive hospital admissions

Evidence-based Synthesis Program (ESP)

KQ 1 Results:

By Type of Outcome

- ***Patient Outcomes***

- 23 Studies

- Hospital (4 studies)

- Studies in Taiwan report higher 5-year breast cancer survival and lower recurrence rates, as well as higher tuberculosis cure rates
- No improvement in patient experience associated with the HVBP
- UK's HQID associated with short term improvement, with no difference from controls in the long term

Evidence-based Synthesis Program (ESP)

Key Question 2: Implementation Framework Categories

Table 1. Description of Implementation Framework Categories

Framework Category		Description
Program design features		Properties of the intervention itself such as the type of quality measure 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 (e.g., 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 (e.g., – blood pressure, cholesterol levels) and health outcomes.

Evidence-based Synthesis Program (ESP)

KQ 2 Results:

By Implementation Framework Category

- ***Program Design Features***

- 13 studies

- Benefits associated with clinical quality, patient experience measures, and measures that aligned with institutional goals, but not productivity and efficiency measures
- Latent variable composites are more reliable than raw sum scores
- Financial salience of incentive amount important, but no clear amount/percentage predicted participation in P4P or program success

Evidence-based Synthesis Program (ESP)

KQ2 Results

- ***Program Design Features***

- Key Informants

- Broad but manageable number of combination of processes of care and patient outcomes important
- Measures should be clinically significant, realistically attainable, reflect institutional priorities, evidence-based, clear, simple
- Incentives should be large enough to motivate, but not so large as to encourage gaming
- Penalties may be more successful than incentives
- Team based measures may increase buy-in of non-clinical staff
- Timing should be frequent, but balanced with payment size

Evidence-based Synthesis Program (ESP)

KQ 2 Results

- ***Implementation Processes***
 - 8 studies
 - Quality continued to increase after increases in QOF maximum thresholds, particularly for lower performing providers (3 studies)
 - Quality was maintained after incentive was removed from measures (3 studies [2 VA])
 - Limited improvement found in the VA for measures changed from passive (no incentive) to active (incentivized) (1 study)

Evidence-based Synthesis Program (ESP)

KQ 2 Results

- ***Implementation Processes***
 - Key Informants
 - Evaluate measures regularly/yearly, possibly increasing thresholds or removing incentives once achievement is high
 - Processes should be transparent, provide resources to link to clinical quality and guidance on how to achieve success
 - Stakeholder engagement, “bottom up approach,” regular performance feedback

Evidence-based Synthesis Program (ESP)

KQ 2 Results

- ***Outer Setting***
 - 6 studies
 - No clear evidence related to region, population density, patient population
 - Key Informants
 - Consider the needs of patient population in design
 - Large, multi-site programs should allow for flexibility to meet local patient population needs

Evidence-based Synthesis Program (ESP)

KQ 2 Results

- ***Inner Setting***
 - 18 studies
 - Larger QOF practices performed better, but varied by condition, condition, and indicator
 - Inconsistent results in the US related to practice size, setting, patient volume
 - In the US, culture change interventions, clinical support tools, and possibly quality improvement visits and trainings associated with higher quality/greater improvement
 - Key Informants
 - P4P is just one part of a quality improvement program
 - Other important factors include a strong infrastructure/infrastructure support, organizational culture, alignment/allocation of resources to P4P, and public reporting

Evidence-based Synthesis Program (ESP)

KQ 2 Results

- ***Provider Characteristics***
 - 5 studies
 - No strong evidence that provider characteristics relate to P4P program performance

Evidence-based Synthesis Program (ESP)

KQ 3 Results

- ***Health Disparities***

- 42 Studies

- No strong consistent evidence of differential effects on patient subgroups (race/ethnicity, SES, other demographic characteristics)
- Groups with lower baseline levels of care tend to increase/improve over the short term

- Key Informants

- First 2 years of QOF showed a reduction in disparities, but this was due to lower baselines. Once high deprivation practices were high performing, the costs associated with eliminating the remaining gaps were higher
- Relationship of P4P to health disparities not well studied in the US. Formal evaluation, along with consistent measures, and demographic/cultural variable collection needed

Evidence-based Synthesis Program (ESP)

KQ 3 Results

- ***Gaming***
 - 3 Studies
 - No clear evidence related to gaming
 - Key Informants
 - Gaming is likely to occur, and programs should be designed with this in mind
 - Stakeholder involvement, input, and buy-in, and precise, simple, evidence-based, realistic measures may reduce the likelihood of gaming

Evidence-based Synthesis Program (ESP)

KQ 3 Results

- ***Risk Selection***
 - 8 Studies
 - 6 studies examined exception reporting in the QOF and found:
 - Higher exclusion rates for non-white, low income patients, and those with comorbid conditions
 - Positive relationship between rates of exception reporting and total QOF score, and higher levels of quality in non-excluded vs. all patients
 - In Taiwan, non-enrolled patients were older, had higher diabetes risk scores, and more comorbid conditions
 - Key Informants
 - Exception reporting is likely not being abused in the UK
 - Concern that high-risk patients can be easily identified using algorithms
 - Incentive payments should be risk-adjusted

Evidence-based Synthesis Program (ESP)

KQ 3 Results

- ***Spillover Effects***

- 11 Studies

- Some evidence of “attention shift,” or “teaching to the test,” with achievement rates/quality lower for non-incentivized measures (3 studies)
- Positive spillover effects found in the QOF, and in UK/CMS hospital programs (7 studies)
- Spillover mechanisms are unclear, may be due to provider behavior change, public reporting, or other quality improvement factors (e.g., electronic medical records)

- Key Informants

- Lack of significant differences in P4P vs. comparators may be due to positive spillovers
- Positive spillover in VA P4P resulted in improvements in nursing staff professionalism

Evidence-based Synthesis Program (ESP)

Discussion

- Despite numerous examples of P4P programs heterogeneity precludes us from drawing strong conclusions that can be broadly applied
- Findings from both the literature and KI interviews support the use of evidence-based measures that are congruent with providers expectations for clinical quality, and there was a strong agreement among KIs that provider buy-in is crucial
- Incentive structure should carefully consider several factors, including incentive size, frequency, and target
- Programs should have the capacity to change over time in response to ongoing measurement of data and provider input
- P4P programs should target areas of poor performance and consider de-emphasizing areas that have achieved high performance

Evidence-based Synthesis Program (ESP)

Limitations

- Heterogeneity
- Largely poor quality studies
- Large number of ambulatory studies, with very few included studies of hospital-based programs

Evidence-based Synthesis Program (ESP)

Future Research

- **Studies examining P4P have been largely observational and primarily retrospective, or lack good matched comparison groups**
 - More trials examining P4P
- **Very few studies good quality examining implementation factors**
 - More good quality studies examining implementation factors such as public reporting, the number and focus of measures, incentive size, structure, and target
- **Limited research examining subpopulations, particularly in the US**
 - More trials and high quality observational studies

Evidence-based Synthesis Program (ESP)

**David Atkins, MD, MPH, Director,
Health Services Research & Development,
Department of Veterans Affairs**



**Joe Francis, MD, MPH
Director, Clinical Analytics and Reporting
Office of Analytics and Business Intelligence
Veterans Health Administration**



Evidence-based Synthesis Program (ESP)

Questions?

**If you have further questions,
feel free to contact:**

Karli Kondo, PhD
Karli.Kondo@VA.gov
kondo@ohsu.edu

The full report and cyberseminar presentation is available on the ESP website:

<http://www.hsrd.research.va.gov/publications/esp/>