QUERI

Evidence Brief: Use of Performance Measures as Criteria for Selecting Community Cardiac and Orthopedic Surgical Providers for the Veterans Choice Program

June 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)
Coordinating Center
Portland VA Medical Center
Portland, OR
Mark Helfand, MD, MPH, MS, Director

Investigators:
Kim Peterson, MS
Johanna Anderson, MPH
Donald Bourne, MPH
Erin Boundy, MS
Mark Helfand, MD, MS, MPH

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 four 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: Peterson K, Anderson J, Bourne D, Boundy E, Helfand M. Evidence Brief: Use of Performance Measures as Criteria for Selecting Community Cardiac and Orthopedic Surgical Providers for the Veterans Choice Program. VA ESP Project #09-199; 2017.

This report is based on research conducted by the Evidence-based Synthesis Program (ESP) Coordinating Center located at the **Portland VA Health Care System, Portland, OR**, 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

FIGURES AND TABLES

EXECUTIVE SUMMARY

Background

The ESP Coordinating Center (ESP CC) is responding to a request from the Office of Community Care for an evidence brief on the use of performance measures for selective contracting with community cardiac and orthopedic surgical providers. Findings from this evidence brief will be used to inform the development of value-based community care purchasing pilots.

Methods

To identify studies, we searched MEDLINE®, CINAHL, CCRT, CDSR, NHS Economic Evaluation, and other sources through April 2017. We used prespecified criteria for study selection, data abstraction, and rating internal validity and strength of the evidence.

After a reported access crisis in the VHA involving long wait times, the 2014 Veterans Access, Choice, and Accountability Act (VACCA) was passed. It established the Veterans Choice Program or "Choice" to expand Veterans' access to community providers when VHA medical facilities have long wait times or when geographic accessibility is excessively burdensome. Choice provider network inadequacy is one challenge that stemmed from implementing a complex program under an aggressive timeline. Accordingly, the VHA is planning to replace and expand existing community care networks.

Quality, efficiency, and costs of surgeries still vary greatly among community providers. The Institute of Medicine (IOM) defines quality of care as "the degree to which health services for individuals increase the likelihood of desired health outcomes and are consistent with current professional knowledge." Performance measurement is the regular collection of data on health care processes, experiences, and/or patient outcomes for use in characterizing variation in quality and efficiency across providers and facilities. The Office of Community Care is considering adopting performance measures to identify eligible community orthopedic and cardiac surgery providers that meet certain minimum requirements.

A significant barrier to the Office of Community Care in selecting performance measures for determing community provider eligibility is that there are a large number of measures meant as indirect indicators of health outcomes (*eg*, readmissions, process measures, *etc*) but uncertainty about their actual association with health outcomes (*eg*, mortality, quality of life, or function). The purpose of our review was to determine whether such performance measures are associated with health outcomes and compare their measurement burden and unintended consequences.

Summary of Findings

All-cause 30-day readmission is a moderate-strength indicator of 30-day mortality for coronary artery bypass grafting (CABG). Thirty-day readmission is a weaker indicator of early mortality for hip replacement. The correlations with early mortality are modest (*eg*, *R* range, 0.32 to 0.38; OR 1.14), but consistent. Thirty-day readmission would likely be feasible to collect as all community Medicare providers are already mandated to collect this information, but several potential limitations must be considered, such as lack of consensus on case-mix adjustment methods, whether to include socioeconomic status (SES), and difficulties capturing readmissions to non-index hospitals. Although the use of 30-day mortality as a measure of quality has recently been questioned, a recent VHA study reinforced its usefulness as a surrogate for long-term outcomes and disputed its susceptibility to gaming.

Adherence to standardized CABG wait time protocols, a set of cardiac surgery process measures, and a specific cardiac surgery antibiotic prophylaxis guideline-based protocol, but not other

studied individual process measures, has decreased the likelihood of mortality and complications in cardiac surgery but not orthopedic surgery. We did not find any studies of performance measures collected in the inpatient setting that directly assessed the quality of postoperative bundles of care.

Implications for Policy and Implementation

Possible minimum requirements for Choice providers include acceptable performance on national rankings, compatible operational infrastructure, and ability to comply with an agreed-upon wait time threshold. In addition to these minimum requirements, the Office of Community Care could consider the added value of the performance measures that this review has identified as being indicators of desirable health outcomes. These include the single measures of 30-day mortality, a direct health outcome measure, and/or 30-day readmission, the indirect measure with the strongest association with mortality – both of which are commonly measured by surgery programs. Another option is to use a composite performance measure that includes mortality, readmission, and other process measures – such as the Society of Thoracic Surgeons' (STS) composite CABG measure. Also, we recommend use of public reporting program participation and measures of efficiency as additional considerations for Choice community provider minimum standards.

However, an unintended effect of stricter performance measure-based criteria for selective contracting with Choice providers may be an undersupply of providers, which could diminish Choice's effect on reducing Veterans' wait times.

Executive Summary Table 1. Summary of Findings

Population: Evidence base	Summary of findings on validity as indicators of quality
Early all-cause mortality	
CABG: 1 fair-quality retrospective cohort study[1]	No correlation between all-cause risk-adjusted in-hospital mortality with *preventable* mortality (*r* = -0.42, P = 0.26) « «
Cardiac and orthopedic surgery: 1 good-quality retrospective cohort study[2]	30-day mortality is a valid surrogate of 365-day survival in the VHA; 365-day survival in highest 30-day mortality risk decile vs others for cardiac (81% vs 91-98%) and orthopedic (77% vs 96-99%)‡ « «
Readmission	
CABG: 6 fair- to good-quality retrospective cohort studies[3-8]	Consistent significant association with 30-day mortality across 3 of most rigorous studies (*r* range, 0.32 to 0.38; OR 1.14†), but not with process measures.[4] « « «
Hip replacement: 1 fair-quality retrospective cohort[7]	Higher risk of readmission in highest vs lowest mortality quartile (11.7% vs 10.2%-10.9%; P < 0.001), but no association with process measures « «
Any orthopedic surgery: 1 good-quality retrospective cohort[9]	Significant association with 2 aggregate process compliance measures (*r* range, -0.05 to -0.06) « «
Surgical standard adherence	
CABG: 5 fair- to good-quality retrospective studies.[10-14]	Low 'total care quality' scores*, but not single process measures associated with increased risk of 30-day mortality and complications (OR range, 1.51 to 1.91)[12] « « Antibiotic protocol violation associated with increased mortality and complications (OR range, 7.03 to 10.16)[14] « «

Hip or knee replacement: 2 fair-quality retrospective cohort study[15,16]	Nonsignificant trend toward higher rate of 30-day mortality for hospitals at bottom 50% of performance on a composite of process measures (r = 0.116, P = 0.088) « « No association of adherence to multiple individual SCIP measures and post-operative infection rates or hospital-level SSI rates, with the exception of *increased* infection rates with higher adherence to VTE prophylaxis (OR range, 1.50 to 1.91) « «
Wait times	
CABG: 1 fair-quality prospective[17] and 1 fair-quality retrospective[18] cohort study.	Adherence to a short delay wait time protocol significantly reduced risk of in-hospital mortality (OR 0.32, 95% CI 0.20 to 0.51) « « Use of a wait-time triage system to assign standard wait times of 0 days for emergent cases, 7 days for in-hospital urgent, 21 days for out-of-hospital semi-urgent A, and 56 days for semi-urgent B resulted in equitable composite mortality and morbidity score across waiting groups. Emergent: (OR 2.5, 95% CI 0.95 to 6.5) In-hospital urgent: (OR 0.9, 95% CI 0.4 to 1.9) Out-of-hospital semi-urgent A: (OR 0.7, 95% 0.3 to 1.6) « «

* The "total care quality" score ranges from 0 to 5-6, where 0 means no process measures were met and 5-6 is the total number of process measures used
†ESP calculated OR using online calculator available at https://www.medcalc.org/calc/odds_ratio.php
‡ESP calculated percentages using an online tool available at http://arohatgi.info/WebPlotDigitizer/app/
Abbreviations: CABG=coronary artery bypass grafting; SCIP=Surigical Care Improvement Project; VTE=Venous Thromboembolism
Strength of evidence: «=Insufficient, « «=low, « « «=moderate, « « « «=high

EVIDENCE BRIEF

INTRODUCTION

PURPOSE

The ESP Coordinating Center (ESP CC) is responding to a request from the Office of Community Care for an evidence brief on the use of performance measures for selecting community cardiac and orthopedic surgical providers for the Veterans Choice Program. Findings from this evidence brief will be used to inform development of value-based community care purchasing pilots.

BACKGROUND

In 2014, after reports of long waiting times for VA clinical services,[19] Congress passed the Veterans Access, Choice, and Accountability Act (VACCA), which expanded the criteria through which Veterans can access civilian providers.[20] VACCA mandated the VHA to implement a new national program within 90 days that allowed Veterans to pursue care from a third-party administrator's (TPA) network provider in their community when VHA medical facilities have long wait times or when geographic accessibility is excessively burdensome (Veterans Choice Program or "Choice"). Important aspects of VACCA are that community providers must (1) "maintain the same or similar credentials and licenses as VA providers", and (2) enter agreements with VA to furnish care rather than an open market or 'voucher-based' system.

Implementation of a new national program is a complex process, and Choice's aggressive implementation timeline created challenges. Community provider network inadequacy was among a number of specific challenges identified by the VA Office of Inspector General review of the first 11 months of Choice implementation (November 1, 2014 through September 30, 2015).[21] The VHA is planning to replace and expand existing community care networks to improve the quality of care for some services, such as general and vascular surgeries[22] and Percutaneous Coronary Intervention (PCI).[23] Possible minimum requirements for Choice providers might include acceptable performance on national rankings, compatible operational infrastructure (*eg*, billing systems that can process bundled payments), and ability to comply with an agreed-upon wait time threshold. In addition to such requirements, the Office of Community Care is considering using performance measures to select and monitor community providers.[24] Initial pilots are targeting orthopedic and cardiac surgery because they are frequently performed and relatively standardized.

Health Care Quality and Purpose of Performance Measurement

A challenge in performance measurement is the variation in how quality is defined.[25] The Institute of Medicine (IOM) defines quality of care as "the degree to which health services for individuals increase the likelihood of desired health outcomes and are consistent with current professional knowledge."[26] Health care quality is a complex concept that encompasses clinical processes and health outcomes[27] which are susceptible to influence by provider, system, social, socioeconomic, and patient risk factors and adherence to care.[28] Performance measurement, a critical tool for quality improvement, is "the regular collection of data to assess whether the

correct processes are being performed and desired results are being achieved."[29] As part of a quality improvement system, performance measures are used to characterize variation in quality across providers and facilities, identify substandard care, and monitor improvements in the context of quality improvement efforts.[15] To better align quality with patients' values, arguments have been made to focus performance assessment on survival beyond 30 days[30] and/or the subset of preventable deaths that could have been avoided if optimal care had been delivered.[1,31]

Proliferation of Performance Measures and Variation in Their Use

There is no gold standard for performance measurement,[32] and development of performance measures has proliferated over the past 2 decades. Different organizations use different sets of measures (supplemental materials Appendix A). National Quality Forum (NQF), Surgical Care Improvement Project (SCIP), American College of Surgeons' National Surgical Quality Improvement Program (ACS NSQIP), and the VA Surgical Quality Improvement Program (VASQIP) have developed and maintain various measure sets. The number of measures ranges from 15 in SCIP[33] to 135 in NSQIP.[34] Most measures sets are developed by expert consensus, although the VASQIP was developed by the National VA Surgical Risk Study (NVASRS) from 1991-1993.[35] Performance measure sets most often include both process and outcome measures. Common individual process measures include receipt and timing of prophylactic antibiotic, and common outcome measures include 30-day mortality and 30-day readmissions. Composite performance measures have also been created to broadly encapsulate the overall quality of care by combining information from multiple individual outcome and process performance measures into a single, comprehensive, multidimensional measure.[36]

To evaluate the quality of surgery services within VA, the National Surgery Office evaluates all-cause 30-day unadjusted mortality and risk-adjusted observed-to-expected (O/E) ratios for 30-day mortality and morbidity for VASQIP-assessed procedures. Expected figures are calculated using VASQIP models that consider the patient populations' associated risk factors. Surgical care delivery is evaluated by case review and site visit for any program whose outcomes deviate with statistical significance from national program O/E ratios.[2] Other organizations base their evaluation of surgical quality on other measures. For example, the Society of Thoracic Surgeons (STS)[37-39] evaluates quality of cardiac surgery using risk-adjusted mortality, risk-adjusted morbidity, optimal surgical techniques, and NQF recommended medications.[40] Consumer Reports publishes rankings based on these STS measures. Hospital Compare, from the Centers for Medicare & Medicaid (CMS), reports results from the American College of Surgeons' National Surgery Quality Improvement Program (ACS-NSQIP), which, like VASQIP, relies on review of medical records rather than on insurance claims.[41] US News & World Report rankings, which are developed by RTI International, uses the 3M Health Information Systems Medicare Severity Grouper, reputational surveys, and hospital indicators such as staffing ratios and surgical volume to rate surgery centers.[42]

Use of Performance Measures for Competency Assessment

Performance measures were originally designed for various internal uses, such as to monitor and improve quality and efficiency. In this context, it is sufficient to subjectively view performance measures as relevant indicators of health care quality. Performance measures have been adapted for use in external competency assessment and to reduce payment and contracting opportunities for poor performers.[43,44] For example, CMS's Hospital Readmissions Reduction Program (HRRP) uses Excess Readmission Ratios (ERRs) to determine hospitals' reimbursement levels

for Coronary Artery Bypass Graft (CABG) and elective primary Total Hip and/or Total Knee Arthroplasty (THA/TKA).[45] A 2014 RAND Corporation research report found that value-based purchasing programs (*eg*, pay-for-performance, accountable care organizations, bundled payments) typically determine reimbursement based on measures of clinical process and intermediate outcomes, patient safety measures, utilization, patient experience, and, to a more limited degree, outcomes and structural elements.[46] Because the use of performance measures in these ways has potential financial consequences for providers, it is important that the measures that are used can be applied consistently, reliably, and equitably; are feasible to collect in community practice settings; and are valid quality indicators – that is, have been shown to be predictive of or associated with health outcomes.

How to Choose Performance Measures for Selective Contracting

With the proliferation of performance measures, it is important to have a framework for deciding (1) which measures to use to select and monitor community providers, and (2) how to use them. One clear principle is to use a small set of measures that focuses on the highest-priority clinical goals.[47] After a period of time in which the number of new performance measures adopted by the VA and non-VA organizations grew, we have begun to reduce the number and focus on the most important measures.[48] Secondly, use valid measures – measures that directly measure or have been proven to be strong indicators of health outcomes. Applying these principles may reduce feelings of measurement overload and measurement irrelevance. A third principle is to apply measures judiciously, rather than formulaically. This means adjusting for factors outside the hospital's control, such as social and economic conditions.[43,49,50] More detailed criteria for equitable performance measurement have been proposed. Equitable performance measures should: (1) have established standards for satisfactory performance, (2) be collected in a standardized and reliable way, (3) be applicable to a group of patients of sufficient size to provide reliable estimates for individual physicians, (4) be adjusted for confounding patient factors, (5) be attributable to individual physicians, (6) be feasible to collect, (7) be representative of the activities of the specialty, and (8) have minimal unintended consequences.[15,25,44,51]

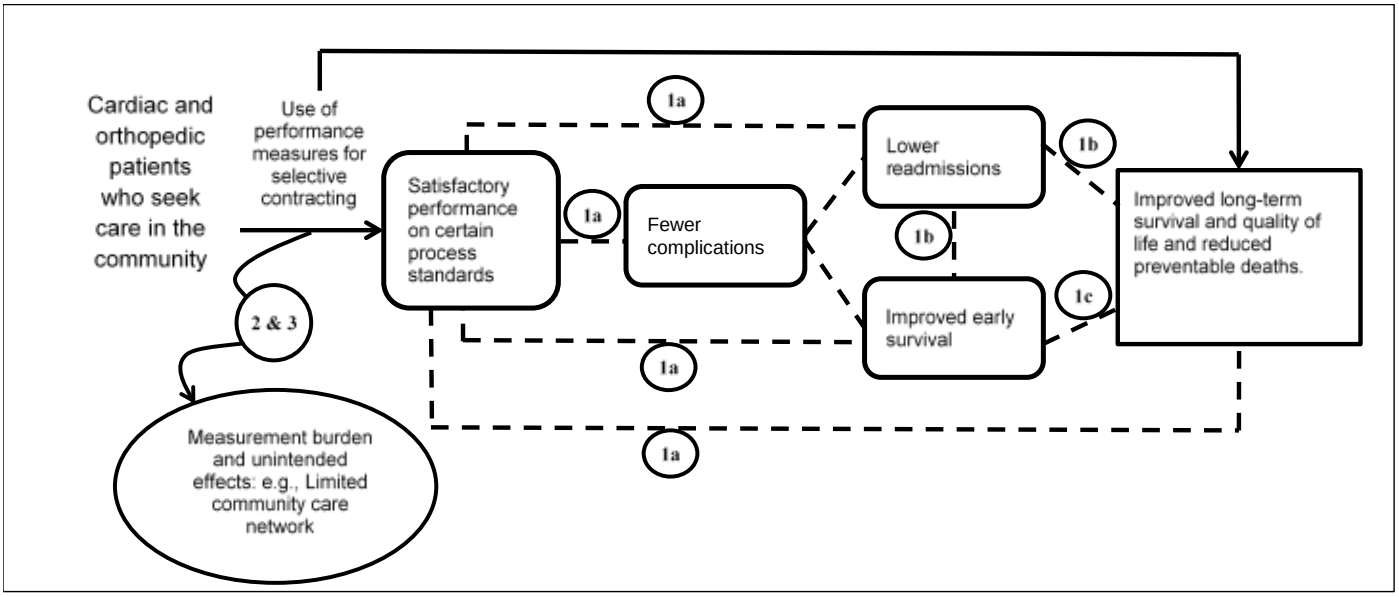
However, several information gaps make it difficult to adapt performance measurement to the purposes of selecting and monitoring community providers. Although there are a large number of measures meant as indirect indicators of health outcomes (*eg*, readmissions, process measures, *etc*), there is uncertainty about their actual association with health outcomes (*eg*, mortality, quality of life, or function). For example, although a 1997 review by Ashton et al found 31-day readmission to be a valid indicator of quality of care based on its association with process of care measures, it did not evaluate association with patient-centered outcomes.[52] Also, although there is general agreement that in order to fairly compare hospital performance based on health outcomes, measures must be risk-adjusted based on patient-level factors to account for differences in case mix, further development may be needed to improve their ability to not discriminate against high-risk patients and to answer questions about whether and how to adjust for broader social and environmental factors.[27,53-56] Finally, although performance measurement has been key in increasing utilization of evidence-based practices, concern has been expressed that it has created an array of unintended consequences including overtreatment, diminished attention to patient needs and preferences, providers' feelings of loss of autonomy, and gaming. However, recent systematic reviews have found that few studies have evaluated unintended effects at least in the context of value-based payment implementations.[46,57]

The goal of this rapid evidence brief is to determine whether certain performance measures meant as indirect indicators of health outcomes are associated with actual health outcomes and evaluate their measurement burden, and their potential unintended consequences.

SCOPE/ANALYTIC FRAMEWORK

The underlying logic that guided our review is that the impact of using performance measures as criteria for selecting providers depends on whether the measures are indicative of lower complication and readmission rates, improved long-term survival and quality of life, and have acceptable measurement burden and unintended effects. The analytic framework (Figure 1) visually depicts this logic. The Key Questions and Inclusion Criteria define our specific focus within this framework.

Figure 1: Analytic Framework



Key Question 1: When used as criteria for selecting providers, are certain performance measures valid indicators of surgical quality?

a) Is satisfactory performance on certain process standards associated with improved complications, lower readmissions, early survival, and long-term outcomes?

b) Are reduced readmission rates associated with improved early survival and long-term outcomes?

c) Is improved early survival associated with improved long-term outcomes?

Key Question 2: What is the comparative measurement burden of different performance measures?

Key Question 3: What are the unintended effects of using performance measures for selective contracting of community providers?

ELIGIBILITY CRITERIA

The ESP included studies that met the following criteria:

- **Surgical services:** Orthopedic and cardiac surgery services. We prioritized non-emergent total hip replacement, total knee replacement, and coronary artery bypass surgeries (CABG) as those are the most common VASQIP-assessed procedures.
- **Health outcome performance measures:** Early readmission rates, mortality rates
- **Process performance measures:** Post-operative plan of continuing care (*eg*, rehabilitation plan), wait times, adherence to surgical standards (*eg*, guideline compliance, infection-related outcome sets (*eg*, SCIP))
- **Measurement burden:** IT environment characteristics (*eg*, automated methods for collecting and auditing data, certification, interoperability), public reporting, audit and feedback, and electronic decision-support tools, chart review
- **Unintended effects**: Gaming, risk-based patient selection, appropriateness of care, changes in disparities, spillover effects, et cetera
- **Timing**: Any study follow-up durations
- **Setting**: Any
- **Study design**: Any, but may prioritize to accommodate timeline using a best-evidence approach

METHODS

To identify articles relevant to the key questions, our research librarian searched MEDLINE®, CINAHL, CCRT, CDSR, and NHS Economic Evaluation up to 06/19/2017 using terms for quality indicators, cardiac or orthopedic surgery, and specific measures of interest (see Appendix B in the supplemental materials for complete search strategies). Additional citations were identified from hand-searching reference lists and consultation with content experts. We limited the search to published and indexed articles involving human subjects available in the English language. Study selection was based on the eligibility criteria described above. Titles, abstracts, and full texts were reviewed by one investigator and checked by another. All disagreements were resolved by consensus.

We used predefined criteria to rate the internal validity of all controlled cohort studies.[58] We abstracted data on setting, population, measurement methods, and results from all controlled cohort studies using a standardized form. All data abstraction and internal validity ratings were first completed by one reviewer and then checked by another. All disagreements were resolved by consensus.

We informally graded the strength of the evidence based on the AHRQ Methods Guide for Comparative Effectiveness Reviews, by considering risk of bias (includes study design and aggregate quality), consistency, directness, and precision of the evidence.[59] Ratings typically range from high to insufficient, reflecting our confidence that the evidence reflects the true effect. For this review, we applied the following general algorithm: evidence comprised primarily of uncontrolled case series or retrospective cohort studies with high risk of bias received ratings of ‘insufficient’; evidence consisting of a single retrospective good-quality study received a rating of ‘low strength’; and evidence consisting of multiple retrospective and/or prospective, consistent, precise, fair-to-good quality studies received a rating of ‘moderate strength’. We found no ‘high-strength’ evidence, but this generally would have been comprised of multiple, prospective, good-quality, precise cohort studies. We synthesized the evidence qualitatively, as meta-analysis was not suitable due to limited data or heterogeneity.

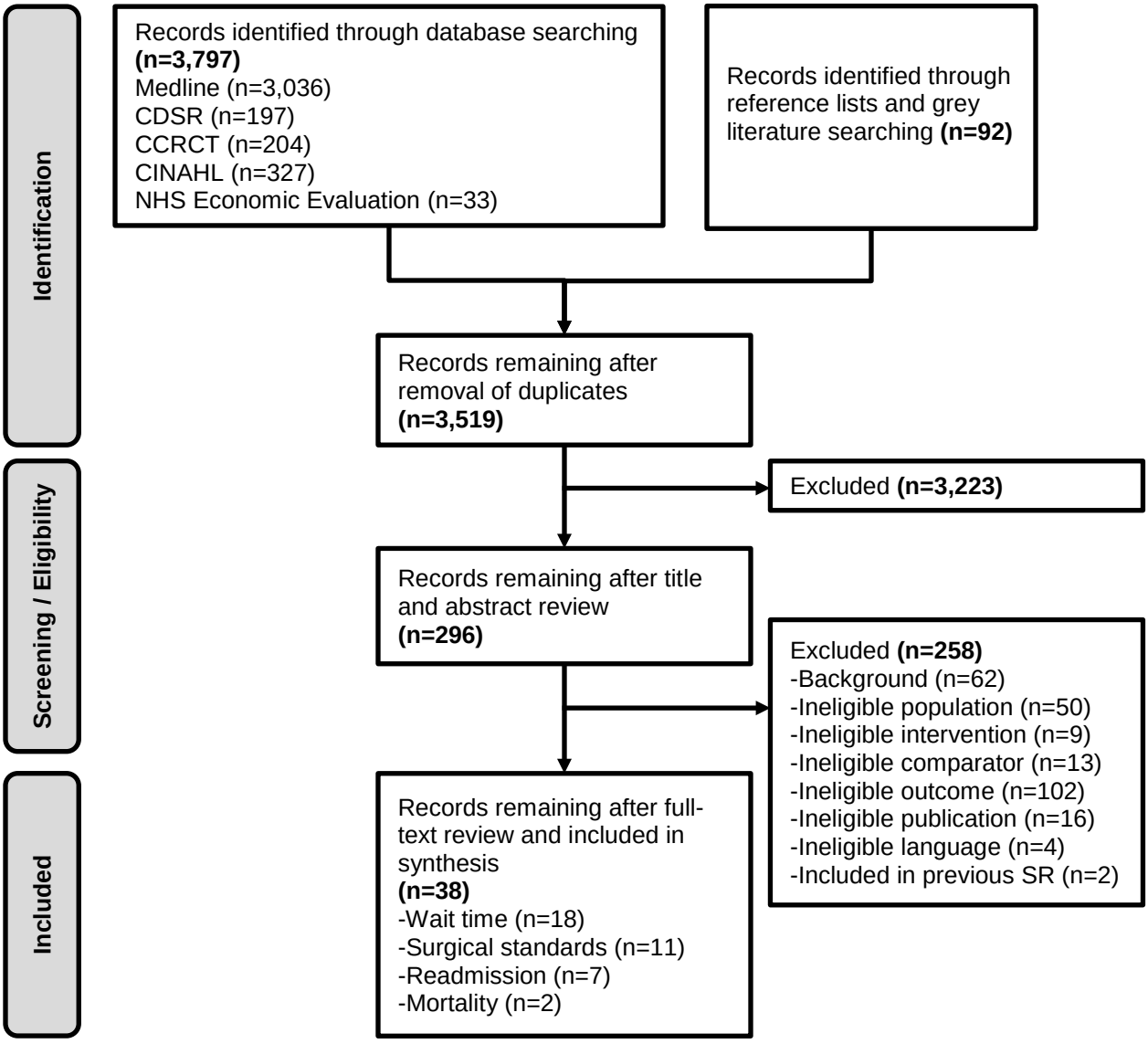
A draft version of this report was reviewed by technical experts selected to represent relevant specialties including surgery, value-based purchasing, and systematic review methodology. Their comments and our responses are available in Appendix E in the supplemental materials.

RESULTS

LITERATURE FLOW

The literature flow diagram (Figure 2) summarizes the results of the search and study selection processes.

Figure 2: Literature Flowchart



Searches resulted in 3,519 unique potentially relevant articles. Overall, we included 38 studies: 6 systematic reviews[60-65] one randomized controlled trial,[66] and 31 observational studies.[1-12,14-18,67-80] Only 4 of the observational studies were prospective cohorts,[8,14,17,66] with the rest being retrospective. Eleven studies addressed adherence to a surgical standard,[10-12,14-16,68,71,72,74,76] 7 addressed readmission,[3-9] 18 addressed wait times,[17,18,60-67,69,70,73,75,77,79-81] and 2 addressed mortality.[1,2] Detailed reasons for study exclusion are provided in Appendix C in the

supplemental materials. More than half of the studies were multicenter (66%) and within the US (53%), but only 44% used data collected in the last decade. Median sample size was 2,218 (range 47 to 2,121,215). Of the studies we formally quality assessed, most were good[2-4,7-12,72] or fair quality.[1,5,6,14-18] We rated 4 studies as poor quality for lack of adjustment for potential confounders, lack of information on missing data, and/or potential risk of selection bias.[68,71,74,76] Sixteen studies had high levels of adjustment for patient-, procedure-, and provider-level factors,[2-5,7-12,14,16,18,66,72,81] while 7 studies adjusted for patient-level factors only,[1,6,15,17,67,74,76] and 2 studies did not adjust for any confounding factors.[68,71] We did not formally assess the quality of the systematic reviews or 7 studies in orthopedic fracture populations,[69,70,73,75,77,79,80] as these emergency populations were of less direct relevance to this review (full details of data abstraction and quality assessment in Appendix D in the supplemental materials).

KEY QUESTION 1A: Is satisfactory performance on certain process standards associated with improved complications, lower readmissions, early survival, and long-term outcomes?

Wait Time

Adherence to standardized CABG wait time protocols shows a decrease in the likelihood of mortality.[17,18] We found no studies on wait times for total knee replacements or total hip replacements. For emergent hip fracture repair in the elderly, delays beyond 24-48 hours are associated with increased mortality and complications, but this evidence likely has limited applicability to the non-emergent, elective surgery populations of greatest interest in this review.

Cardiac Surgery

Delays exceeding British Columbia provincial wait list guidelines (6 weeks for semiurgent and 12 weeks for nonurgent) were associated with increased probability of in-hospital death based on findings from a large, nationally representative retrospective cohort studiy in a cardiac surgery population awaiting CABG (Table 1).[18] This study evaluated the association of adherence to 2 separate established wait time guidelines – British Columbia provincial and the stricter Canadian Cardiac Society (CCS) (2 weeks for semiurgent and 6 weeks for nonurgent) – and in-hospital mortality by comparing outcomes across 3 categories of delays: short (within CCS timeframes), prolonged (longer than CCS but within provincial targets), and excessive (longer than both guidelines). Adherence to at least the provincial guidelines could potentially decrease risk of in-hospital mortality, as the short CCS timeframes had a protective effect compared with nonadherence with both guidelines, but not compared to the provincial guideline adherence. The validity of adherence to a strict wait time protocol as an indicator of quality of care is also supported by the finding that use of a wait-time triage system to assign standard wait times of 0 days for emergent cases, 7 days for in-hospital urgent, 21 days for out-of-hospital semi-urgent A, and 56 days for semi-urgent B resulted in equitable mortality and morbidity across waiting groups.[17] Although the supporting study likely had limited statistical power and applicability because it focused on a small sample of patients with left main coronary artery disease from a single hospital in Halifax, its approach to evaluating the safety of using a standardized triage system for determining CABG wait times is an important mechanism for evaluating the validity of wait time protocols as quality indicators.[17] No studies reported on readmission, complication or other outcomes of interest. Both studies[17,18] adjusted for patient, clinical, and surgical factors (see appendix D), but were conducted outside the US, so their applicability to the US health care system is limited.

Table 1. Cardiovascular Surgery Wait Times and Patient Outcomes

Author, Year N	Setting	Population	Mortality
Légaré 2005[17] N=561	1 hospital in Halifax, Nova Scotia 1999-2003	CABG among patients with stenosis of the left main coronary artery	Composite measure* including in-hospital mortality rate for: Wait longer than standard waiting time OR 0.7 (95% CI 0.4 to 1.2) Queue assignment compared to out-of-hospital semi-urgent B for composite score: Emergent OR 2.5 (95% CI 0.95 to 6.5) In-house urgent OR 0.9 (95% CI 0.4 to 1.9) Out-of-hospital semi-urgent A OR 0.7 (95% CI 0.3 to 1.6)
Sobolev 2012[18] N=9,593	4 cardiac centers in British Columbia, population-based patient registry 1992-2006	CABG	In-hospital mortality rate for: Excessive delay OR 1.00 (Reference) Prolonged delay OR 0.78 (95% CI 0.38 to 1.63) Short delay OR 0.32 (95% CI 0.20 to 0.51)

*Comprised in-hospital death from any cause, mechanical ventilation for more than 24 hours postoperatively, and postoperative length of hospital stay of more 9 days
Abbreviations: CABG=Coronary Artery Bypass Grafting; OR=Odds Ratio

Emergent Fracture Repair

Early surgery (usually within 24-48 hours) was associated with reduced mortality (OR 0.74, 95% CI 0.67 to 0.81) in elderly patients with hip fractures needing emergent repair according on the best evidence from a 2012 good-quality systematic review of 35 studies.[62] However, it is important to note that adjusted and unadjusted data were combined, and the cut-off time for wait time and mortality varied by study. When stratified by cut-off time for surgical delay (<12, <24, <48, <96 hours), results showed a significant association between wait time and mortality for wait times within 24-48 hours, but no association betweeen other cut-off times. Among systematic reviews that looked at complications in this population,[61,64,65] the most recent good-quality systematic review[65] found early surgery (within 24-48 hours) to be associated with decreased rates of pneumonia (RR 0.59; 95% CI 0.37 to 0.93; 2 studies; N=2793) and pressure sores (RR 0.48; 95% CI 0.34 to 0.69; 3 studies; N=3023), but not deep vein thrombosis (RR 0.97; 95% CI 0.56 to 1.68; 2 studies; N=4679) or pulmonary embolism (RR 0.66; 95% CI 0.17 to 2.58; 2 studies; N=2822). The study authors caution that their confidence in the finding is limited because the included primary studies did not adjust for potential confounding factors.[65]

Findings from subsequently published primary literature neither confirm nor refute these findings due to their serious methodological limitations or use of a prolonged follow-up period.[67,75,81] Wait times exceeding 24 hours were not associated with 1-year mortality in a sample of 567 people admitted to a single orthogeriatric unit in Oslo.[67] This study was underpowered to find a difference; post-hoc power for the study was 42.2% and revealed that a future study to would require 1,249 patients to reach adequate power.[82] In a sample of 828

patients in Italy, wait times exceeding 48 hours were not associated with 2-year mortality.[75] A strength of this study was its adjustment for comorbidities. But it is unclear how comparable findings from a 2-year follow-up period are to the most recent systematic review[62] in which the mean follow-up period was unspecified, but most likely shorter. Finally, the most recent study[81] found no association between wait time and mortality in the elderly hip fracture population. However, we have limted confidence in its findings due to the lack of adjustment for any potential confounders and small sample size.

In hip fracture repair populations not restricted to the elderly, one good-quality systematic review found that the majority of studies (14/24) that adequately controlled for confounding did not find a correlation between wait time and mortality, but found that delays increased risk of complications.[60] Results were consistent regardless of variation in adjustment, cut-off time for wait time, and cut-off time for mortality. Among 4 subsequently published primary studies,[66,70,73,77] results from the single study[77] that adequately controlled for potential confounding (adjusted for age, gender, race, comorbidity burden, insurance status, day of admission, hospital size, teaching status, and region) were consistent with and strengthen the findings of the systematic review (Table 2).

In patients undergoing emergent ankle fracture repair, delayed surgery (>24 hours, up to 14 days) significantly increased risk of wound complications in 6 of 11 studies.[63,79] Although we did not formally rate the quality of these studies in this lower-priority population, the strength of evidence is likely very low because they were small – sample sizes < 100 – and did not control for potential confounding.[63,79]

In patients with multiple fractures, one retrospective cohort of 1005 patients at a single Level 1 trauma center provides low-strength evidence that performing fixation of unstable axial fractures within 24 hours of the trauma significantly reduced risk of complications.[80] The applicability of this evidence to the non-emergent surgeries of interest is likely low.

Table 2. Wait Time and Patient Outcomes in Orthopedic Surgery (Studies not captured in systematic reviews)

Author, Year N	Population/ Setting	Mortality	Complications/Readmissions
Hip fracture (elderly)			
Holvik 2010[67] N=567	Oslo, Norway 2007-2008	1-year all-cause mortality rate for ≤24 hours vs >24 hours RR= 0.48 (95% CI 0.21 to 1.10)	NR
Karademir 2015[81] N=115	Istanbul, Turkey Timeframe NR	1-year all-cause mortality rate for ≤5 days vs >5 days P=0.5	NR
Meessen, 2014[75] N=828	Varese, Italy 2009	2y all-cause mortality rate for ≤48h vs >48h*: (data not reported) P>0.05	NR
Hip fracture (non-elderly)			
Clague, 2002[69] N=462	1 UK hospital, 1996-1999	In-hospital mortality rate for ≤24h vs >24h (data not reported) P>0.05	NR

		90d mortality rate for ≤24h vs >24h (data not reported) P>0.05	
Griffiths, 2013[70] N=60	1 UK hospital	30d mortality rate for ≤72h vs >72h (data not reported) P=0.2	30d complication rate for ≤72h vs >72h (Data not reported) P=0.008
Lund, 2014[73] N=6,143	Danish National Registry of Patients	1-year all-cause mortality rate 0-12h: HR 1.00 (reference) 12-24h: HR 0.89 (95% CI 0.78 to 1.01) 24-48h: HR 1.03 (95% CI 0.91 to 1.17) 48-72h: HR 1.02 (95% CI 0.84 to 1.24) 72-96h: HR 1.10 (95% CI 0.83 to 1.44) >96h: HR 1.05 (95% CI 0.81 to 1.36)	NR
Lurati-Buse 2014[66] N=60	2 hospitals in Canada and 1 in India 2011-2012	30-day mortality rate for accelerated care (shorter wait time) and standard care (longer wait time) OR 0.22 (95% CI 0.02 to 2.14)**	NR
Ryan, 2015[77] N=2,121,215	US National inpatient sample 2000-2009	In-hospital mortality rate* 0-day: OR 1 (reference) 1-day: OR 1.02 (95% CI 0.96-1.09) 2-day: OR 1.14 (95% CI 1.06-1.23) ≥3-days: OR 1.34 (95% CI 1.23-1.46)	In-hospital complication rate* 0-day: OR 1 (reference) 1-day: OR 1.09 (95% CI 1.06-1.12) 2-day: OR 1.33 (95% CI 1.29-1.39) ≥3-days: OR 2.08 (95% CI 2.00-2.16)

Abbreviations: OR odds ratio; HR=hazard ratio; RR=relative risk; h=hours; y=year; NR=not reported
*=adjusted for confounding variables; **ESP-calculated

Surgical Standards

Low 'total care quality' scores moderately increased risk of 30-day mortality and complications following CABG (OR range, 1.51 to 1.91; low strength of evidence; 2 studies).[10,12] The 'total care quality' score ranges from 0 to 5-6, where 0 means no process measures were met and 5-6 is the total number of process measures used. Adherence to individual cardiac surgery process measures was generally not associated with reduced mortality and complications, except that violation of an antibiotic prophylaxis guideline-based protocol increased risk of mortality and surgical site infections (OR range 7.03 to 10.16; low strength of evidence; one study).[14] For orthopedic surgery, among 4 studies the best studies provided low-strength evidence that a composite quality score of CMS NQF measures, and adherence to individual SCIP measures were not valid indicators of either mortality or complications.[15,16]

Cardiac Surgery

In cardiac surgery populations, composite scores of 'total care quality' (*ie*, total number of individual process measures missed) may better predict patient health outcomes than adherence to an individual process measure (Table 3). Two good-quality studies examined adherence to surgical quality measures in cardiac surgery using a composite quality score based on the total number of recommended NQF or SCIP measures achieved.[10,12] One study of 81,289 CABG patients found increased 30-day mortality with increasing proportion of patients who failed to receive recommended Surgical Care Improvement Project (SCIP) measures.[10] This study found no association between the number of missed measures and readmission.[10] Another study of 2,218 CABG patients found higher odds of several complications for patients with a low quality

score based on the number of National Quality Forum (NQF) process measures achieved (range 0 to 5, 5 = high quality care, all other scores = low quality care).[12]

Among individual cardiac surgery process measures studied (Table 3),[11,13,14,74] only adherence to an antibiotic prophylaxis guideline-based protocol improved outcomes.[14] Initiating the first skin incision for elective cardiac surgery *before* the end of vancomycin infusion was associated with large increased risk of surgical site infection rates and mortality in-hospital or within 30 days in a prospective study of 741 adults from a single center in a 1200-bed tertiary care university hospital in Italy.[14] Three other studies investigated adherence or lack of adherence to a single measure, pre-operative beta-blocker use[11] and blood glucose maintenance.[13,74] These studies found no association between adherence to surgical standards and 30-day mortality or complications.[11,13,74] An additional pre-post study found decreased infection rates after implementation of an enhanced targeted infection control program, including education, improved hygeiene, MRSA screening, care pathways, and antimicrobial surgical prophlaxis. However, this study was limited by no control for potential confounders and did not specifically assess adherence to the antimicrobial surgical prophylaxis.[68] These results indicate that composite measures or 'all or none' measures of surgical standard adherence may better measure overall quality of care, as they evaluate the overall system of care for several processes. In contrast, adherence to single surgical standards may be influenced by the practices of single team members.[83]

These studies were generally of good quality, using rigorous methods to adjust for potential confounding, and 3 out of the 7 studies were multicenter.[10-12] The exceptions were 2 studies that inadequately addressed potential confounding factors, either only controlling for a single factor or not controlling for any potential confounders.[68,74] As most of these studies were in CABG patients – which is a well-developed and frequently performed surgery – these results might not be applicable to other types of cardiac surgery.[12]

Table 3. Adherence to Surgical Standards and Patient Outcomes in Cardiac Surgery

Author, Year N	Setting	Measure Details	Findings
Composite Measure			
Auerbach, 2009[10] N=81,289	164 US hospitals in Perspective database, 2003-2005	Proportion of patients who failed to receive recommended SCIP measures (# of missed measures).	*Mortality:* 3 missed measures vs none missed OR 1.54 (95 % CI 1.20 to 1.98), 4 or more missed measures vs none missed OR 1.63 (95% CI 1.24 to 2.15) *Readmissions:* 4 or more missed vs none OR 1.02 (95% CI 0.93 to 1.13)
Kurlansky, 2012[12] N=2,218	5 cardiac surgery programs associated with Columbia University, 2007-2009	Total quality score (0 to 5) based on number of NQF process measures achieved. 5=high-quality care, all other scores=low-quality care	*Complications:* Low vs high quality score: stroke OR 1.51 (95% CI 1.18 to 1.93), reoperation OR 1.65 (95% CI 1.25 to 2.16), prolonged ventilation OR 1.54 (95% CI 1.21 to 1.96), and renal failure OR 1.91 (95% CI 1.09 to 3.35). No association with sternal infection.
Individual Measure			
Brinkman, 2014[11] N=506,110	1,107 centers in STS database, 2008-2012	Adherence to use of preoperative beta-blocker within 24 hours preceding surgery (NQF).	*Mortality:* Beta-blocker use vs no use: OR 0.96 (95% CI 0.88 to 1.04) *Complications:* Beta-blocker use vs no use*:* atrial fibrillation OR 1.09 (95 % CI 1.06 to 1.12). No association with other complications (stroke, prolonged ventilation, reoperation, renal failure).
Cotogoni, 2017[14]	1 hospital in Turin, Italy, Timeframe NR	Adherence to prophylactic vancomycin administration timing protocol	*Mortality:* Association of increased mortality with protocol violation: OR 10.16 (95% CI 2.48 to 41.58) *Complications:* Association of increased SSI with protocol violation: OR 7.03 (95% CI 3.41 to 14.52)
LaPar, 2014[72] N=1,703	University of Virginia Hospital, 2010-2012	Adherence to SCIP measure of maintenance of 6am blood glucose levels on post-operative days 1 and 2	*Mortality:* SCIP measure failure: OR 1.49 (95% CI 0.54 to 4.09) *Complications:* SCIP measure failure: major morbidity OR 1.51 (95% CI 0.86 to 2.67), major sternal complications OR 1.58 (95% CI 0.18 to 13.7).
McDonnell, 2013[74] N=832	Boston University Medical Center, 2008-2011	Adherence to SCIP measure of maintenance of 6am blood glucose levels on post-operative days 1 and 2	*Mortality:* SCIP measure failure 1.8% vs SCIP compliant 1.7%, P=0.55 *Complications:* No association with complications for MI, stroke, deep sternal infection, multisystem failure, or atrial fibrillation

Abbreviations: STS=Society of Thoracic Surgeons; SCIP=Surgical Care Improvement Project; NQF=National Quality Forum; MI=myocardial infarction; OR=odds ratio; SSI=Surgical Site Infection

Orthopedic Surgery

Adherence to orthopedic surgical standards was consistently not associated with health outcomes in all of 4 included studies.[15,16,71,76] One study examined hospital quality in hip or knee replacement surgery in 4 performance tiers based on a composite quality score of CMS NQF measures.[15] This composite score included process measures (adherence to antibiotic prophylaxis standards, *etc*), as well as outcome measures (readmission avoidance, *etc*), but the differences in the top 20% of hospitals based on performance and the remaining hospitals were driven by the process measures. There was no significant difference in mortality across hospital tiers, but a trend toward a higher rate of mortality in tier 4 hospitals, classified as the bottom 50% of performance ($r = 0.116$, $P = 0.088$). Additionally, complication rates did not differ significantly by hospital tier, and readmission avoidance did not differ significantly between the top 20% of hospitals and all other hospitals.[15] Another study examined level of compliance to SCIP measures (*eg,* 'highly compliant' indicated greater than median level of compliance) and complications among hip arthroplasties at 128 New York state hospitals.[16] This study found no association of adherence to the individual SCIP measures and post-operative infection rates or hospital-level SSI rates, with the exception of *increased* infection rates with higher adherence to the SCIP "VTE-2" prevention measure, which measures the percentage of hospital patients who received appropriate venous thromboembolism prophylaxis within 24 hours prior to surgery to 24 hours after surgery (post-operative infection: OR 1.50, 95% CI 1.07 to 2.12; hospital-level SSI: OR 1.91, 95% CI 1.31 to 2.79).

Two studies examined rates of complications (SSI) before and after implementation of surgical standards.[71,76] One poor-quality study with no adjustment for potential confounders examined adherence to surgical antibiotic prophylaxis (SAP) among 6 hospitals in Korea before and after implementation of a national hospital evaluation program.[71] Although adherence to SAP improved ($P < 0.01$), there were no significant changes in SSI rate for hip or knee arthroplasty or spine surgery. Another poor-quality study with minimal adjustment for confounders examined SSI rates before and after implementation of SCIP in a single institute in the US.[76] After implementation, the SCIP rate adherence was more than 98%. Multivariate analysis (adjusting for surgery and SCIP factors) showed that SCIP adherence was associated with increased risk of superficial infection (estimate: -4.45, $P < 0.001$), and pulmonary embolism (estimate: 0.34, $P < 0.001$), but there were no significant associations with deep infection ($P = 0.46$) or deep vein thrombosis ($P = 0.51$). These findings suggest that adherence to surgical standards may not predict surgical complication rates. However, these studies are limited by lack of a concurrent control, minimal or no adjustment for potential confounding variables, and limited information on processes prior to implementation of the surgical standard programs.

Postoperative Plan of Care

In some cases, community providers will offer episode-of-care services ('bundles of care') that include outpatient preoperative care, inpatient care, and postoperative rehabilitation services. We did not find any studies of performance measures collected in the inpatient setting that directly assessed the quality of postoperative bundles of care.

We did not evaluate the effectiveness of strategies to reduce readmission rates through better coordination and planning of post-discharge care. Nevertheless, if an indicator such as readmission rate is used, it may be important to assess whether the provider applies strategies to prevent or detect post-discharge complications early.[84] Some believe that scheduling a timely

post-hospital visit while the patient is still in the hospital may be helpful because readmissions for complications often occur before the first follow-up visit.[85] The clarity of discharge instructions, better communication at the time of discharge, follow-up nursing calls, and communication between inpatient and outpatient rehabilitation services are also considered potentially useful.[86]

We are aware of additional studies that investigate the use of "clinical or care pathways" on patient health outcomes.[87] However, these studies are complex system-level implementation interventions designed to encourage use of evidence-based processes (*eg*, flowcharts, patient education materials, checklists, mapping out of sequences) that are used for quality improvement. Although these studies measure use of these support systems, they typically don't measure adherence to the actual bundles of postoperative care standards, and this makes it difficult to determine which factors influence outcomes. As our goal is to assess measures collected in the inpatient setting that directly assessed the quality of postoperative bundles of care, we did not evaluate this broader "clinical or care pathway" literature.

KEY QUESTION 1B: Are reduced readmission rates associated with improved early survival and long-term outcomes?

For CABG surgery, the 3 most rigorous studies[3,4,7] found a modest but consistent significant association between 30-day readmissions and 30-day mortality (*r* range, 0.32 to 0.38; OR 1.14; moderate strength of evidence), but not with process measures. For hip replacement, there was a correlation between 30-day readmission and mortality (one study, low strength of evidence),[7] but the association between readmission rates and process measures was inconsistent. Among 2 studies,[7,9] readmission rates were only correlated with aggregate process compliance measures in the study with a broader orthopedic population that had wider variation in readmission rates.[9]

Coronary Artery Bypass Grafting (CABG)

30-day Readmission Association with 30-day Mortality

Among 4 retrospective cohort studies, the 3 most rigorous studies[3,4,7] found a modest but consistent significant association between 30-day readmissions and 30-day mortality (*r* range, 0.32 to 0.38; OR 1.14) (Table 4).[3-5,7] The most broadly applicable data comes from the only study using a national data source.[7] Others used data from the states of New York,[3,4] or California.[5] All used risk-adjusted readmission and mortality variables, but each used slightly different analytic approaches. For example, the national Medicare data study used a standardized readmission rate whereby they calculated the composite readmission rate for each hospital by averaging the observed-to-expected (O/E) readmission ratios for the 6 procedures for the hospital and weighting each ratio on the basis of the number of cases for that procedure. Then they compared the standardized risk-adjusted readmission rates between hospitals in the highest quartile for mortality to the lowest quartile.[7] In contrast, the California study observed *trends* in discordance among hospitals that were O/E outliers (95% CI excluded 1.0) for mortality *or* readmission and not both ("discordant").[5] Regardless, 3 of the 4 studies found a significant association between 30-day readmission and 30-day mortality.[3,4,7] We have lower confidence in the validity of the outlier study, however, due to their use of weaker statistical methods.[5] Authors of the outlier study noted that most hospitals had concordant O/E readmission and mortality rates, which "points to the likely existence of some association between readmission and mortality rates." But, they based their conclusions on the *trends* they observed among the

outliers. Among outliers, 85% were "discordant" for readmission and mortality (for example, hospitals with high readmission and low mortality). This approach may exaggerate the significance of the outliers as it neglects the statistical effects among the data as a whole, which was predominantly comprised of concordant hospitals.

Table 4. 30-day Readmission in Coronary Artery Bypass Grafting (CABG) Patients

Author, Year N	Setting	Measure Details	Findings
Hannan, 2011[4] N=33,936	NY State Cardiac Surgery Reporting System, 2005-2007	Risk-adjusted hospital 30-day readmissions	*Mortality:* Association with risk-adjusted 30-day hospital mortality rates (r=0.32, P=0.047), and with hospital risk-adjusted mortality rate in highest tertile (r=0.38, P=0.03)
Hannan, 2003[3] N=16,325	NY State Cardiac Surgery Reporting System, 1999	Risk-adjusted hospital 30-day readmissions	*Mortality:* No statistically significant association with overall hospital risk-adjusted mortality rate (r=0.09, P=0.64), but association with hospital RAMR in highest tertile, OR 1.14 (95% CI 1.03 to 1.25)
Parina, 2015[5] N=296,063	299 hospitals in CA, 1995-2009	Risk-adjusted 30-day readmission. Outliers based on 95% CIs of O/E ratio excluding 1; classified 'discordant' if readmission and mortality rates were not both high or both low.	*Mortality:* No association: among outliers 85% were discordant, CABG discordance rate: 78.3%
Tsai, 2013[7] N=153,496	National Medicare data, 2009-2010	Hospital-level composite of procedure-specific risk-adjusted 30-day readmission rates	*Mortality:* Readmission rate in highest mortality quartile=18.1% vs in lowest to third quartile=17.3%-17.4%; P=0.013. *Process Measures:* HQA surgical score. No statistically significant difference in readmission with HQA surgical score quartile (P=0.751)

Abbreviations: OR=odds ratio; O/E=observed/expected; CABG=coronary artery bypass grafting; HQA=Hospital Quality Alliance

Readmission Associated with 100-day Mortality

One retrospective cohort study evaluated the association between longer-term readmission rates (within 100 days) and 30-day mortality among 14 hospitals in Israel in 1994.[8] Although this study found that high mortality-ranked hospitals had higher rates of readmission (OR 1.34, P = 0.003), this finding likely has low relevance to contemporary care in the US.

Readmission Associated with Process Measures

Among 3 retrospective cohort studies that used diverse process measure assessment,[6,7,9] none found a significant association with readmission. Process measure compliance assessment ranged widely and included (1) peer-reviewing charts based on a set of Health Care Financing Administration (HCFA)-specified generic quality screens to evaluate care provided as acceptable

or problematic,[6] (2) evaluating compliance with surgery-specific SCIP process measures (overall compliance with all 9 process measures, or number of compliant events divided by the number of opportunities to provide recommended care (*ie*, non-smoker not eligible for smoking sessation counseling),[9] and (3) comparing hospitals in the highest and lowest decile of the Hospital Quality Alliance (HQA) surgical score (composite of 9 process measures).[7] Some authors speculated that the lack of association may be because these process measures represent just a fraction of the entire spectrum of care, and do not capture other important features such as care coordination at the time of discharge.[9]

Readmission associated with composite process/outcome measures

Lower readmission rates were weakly correlated (Spearman rank correlation= –0.154) with higher composite scores (including mortality, major morbidity, internal mammary artery graft, and NQF-endorsed perioperative medications) in a secondary subgroup analysis of 827 CMS CABG providers from the 2010 STS database.[88] However, we have insufficient information to determine the strength of this evidence as this finding was only very briefly noted in the discussion section of the main study, which was devoted to the development of the readmission measure. No other information about the methodology were provided in the publication or via author request.

Orthopedic

30-day Readmission Association with 30-day Mortality

One Medicare study (described above)[7] evaluated the association between 30-day readmission and 30-day mortality in an orthopedic population. Among 206,175 hip replacements from National Medicare data from 2009 to 2010, there were significantly higher readmissions among hospitals in the highest quartile for mortality compared to the lowest quartile (11.7% vs 10.2%; $P < 0.001$).

30-day Readmission Association with Process Measures

An evaluation of a broad population of fee-for-service Medicare patients from the Hospital Inpatient Quality Reporting Program who underwent *any* orthopedic procedure in 2007 found significant correlations between process performance scores and risk-standardized all-cause 30-day readmission rates.[9] Correlations were -0.06 ($P = 0.003$) for the OM and -0.05 ($P = 0.03$) for the ACM for orthopedic surgery overall. Median hospital performance was 92.9% (InterQuartile Range [IQR], 88.3% to 95.9%) for the OM and 71.1% (IQR, 55.8% to 82.6%) for the ACM and median risk-standardized readmission rates were 9.1% (6.9% to 12.7% across 10th to 90% percentiles). The 2 aggregate compliance measures are described above. In a narrower population of hip replacements with less variability in readmission rates across hospitals, there was no significant difference in risk-adjusted 30-day readmission across quartiles of Hospital Quality Alliance (HQA) surgical score (lowest = 11.2%, 2nd quartile = 11.0%, 3rd quartile = 10.7%, and highest = 10.8%).[7] Both studies were of similar size and methodological quality, with only few and minor limitations. Therefore, it is likely that the significant variation in their population and process measures led to their different findings.

KEY QUESTION 1C: Is improved early survival associated with improved long-term outcomes?

Correlation Between In-Hospital All-Cause Mortality and Preventable In-Hospital Mortality

No significant correlation was found between all-cause, risk-adjusted in-hospital mortality rates and *preventable* in-hospital deaths based on a retrospective analysis of 347 randomly selected deaths following CABG at 9 institutions in Ontario between 1998 and 2003 (Spearman coefficient, -0.42, P = 0.26).[1] Preventability was rated by 2 experienced, blinded surgeons who used a standardized tool to identify preventable deaths from nurse-abstracted chart summaries. Despite that the deaths came from high-volume hospitals with a long-standing public reporting, 32% were judged preventable. Deviations in perioperative management standards were attributed as the reason for a majority of the preventable deaths. The preventability criteria, low level of inter-rater agreement (*k* range, 0.16 to 0.26), and reliance on retrospective chart review are major deficiencies that reduce our confidence in these findings, however. Preventability was scored using a subjective 7-point Likert scale, with ratings of "none", "slight", "modest", "<50 to 50", "<50 to 50 but close call", "strong" and "certain". Although an advantage of a Likert scale is that it can be quick and efficient, detection of preventability could have been strengthened by use of a more objective set of criteria representing specific standards of care.

Correlation Between 30-day Mortality and 1-year Survival

Deciles of 30-day mortality were associated with significantly different rates of survival at 365 days following VASQIP-assessed cardiac (N = 10,042) and orthopedic (N = 60,515) surgeries performed within the VHA from October 1, 2011 to September 30, 2013.[2] For cardiac surgery overall, the highest 30-day mortality risk decile had a significantly lower risk of 365-day survival than other risk deciles (approximately, 81% vs 91%-98%). Findings were similar for orthopedic surgery overall, with a 365-day survival probability of 77% for the highest 30-day risk decile compared with 96% to 99% for the other risk deciles. Although the methodology of this study was strong, as this study was conducted *within* the VHA, it is unclear how applicable these findings are to community hospitals with varying levels of quality improvement programs in place.

KEY QUESTION 2: MEASUREMENT BURDEN

Performance measures vary in their burden of data collection and reporting. Measurement burden is an important consideration, as successful implementation depends on realistic attainability.[89] However, we found no studies that formally evaluated comparative measurement burden. In general, measures such as mortality and readmission are often more appealing because they are the most feasible to collect as they are already being regularly collected and reported for internal and public quality improvement initiatives. However, they have the potential disadvantages of insufficiently discriminating quality-sensitive, preventable events. However, determining preventability would require committee or chart audits and would likely be resource-prohibitive.[1] Also, risk-adjustment is necessary, but administrative databases may not capture all relevant variables[6] and there is little consensus on which methods to use.[32] Surgical standard adherence, postoperative plan of care, and wait times are likely associated with greater measurement burden. This is because (1) there is more variability in whether any are measured and, if so, which ones; (2) if tracked electronically, sources may have low reliability;[10]

and (3) they may require additional staff to set up and implement,[74] possibly through chart review.[85] Additionally, wait times may present additional barriers to collection. Different wait time standards may be needed for different subgroups, as safe wait times may vary by patient subgroups and surgery types.[69] There is a risk that clinically-justified longer wait times may be penalized if medical records and/or databases lack a mechanism for recording supporting documentation. For example, frailer patients may be more likely to be delayed due to need for stabilization prior to surgery.[60,62]

KEY QUESTION 3: UNINTENDED EFFECTS

Concern has been expressed that focus on specific performance measures may result in unintended effects such as gaming, risk-based patient selection, appropriateness of care, changes in disparities, and spillover effects. For example, 30-day readmission may be gamed by delaying readmission to 31 days instead of 30 days.[90] Or for 30-day mortality, intensive care management and end-of-life care may be modified against the patient's best interest to extend death beyond 30 days.[2] Also, adherence to some surgical standards is debated for some patients (*ie*, beta-blockers for specific cases of off-pump revascularization) and adherence may harm those patients.[11] For wait times, as demonstrated in the Phoenix VA Medical Center in 2014, use of a 2-week wait time performance measure – regardless of clinical need, facility capacity, and other accommodations such as secure messaging or telehealth – resulted in the manipulation of scheduling practices in order to meet performance goals.[48] Finally, an unintended effect of using any of these performance measures for selective contracting with Choice providers is that it may result in an undersupply of providers, which could diminish Choice's intended effect on reducing Veterans' wait times.

Evaluation of such unintended consequences has largely been lacking. Recent systematic reviews of using various performance measures in value-based payment programs found very limited evidence assessing the extent of gaming, inconsistent evidence of effects on health disparities, and only some evidence of both positive and negative effects on unincentivized measures.[91] The only additional evidence we found was from a recent VHA study that examined surgical mortality between postoperative days 25 and 35 and found no evidence of delayed mortality within the highest readmission decile.[2]

SUMMARY AND DISCUSSION

Health care performance measurement is the regular collection of data on health care processes, efficiency, experiences, and/or patient outcomes to determine the degree to which providers and health systems are providing health services that "increase the likelihood of desired health outcomes and are consistent with current professional knowledge."[26] Development of performance measures has proliferated over the past 2 decades, resulting in a wide range of individual measures of both direct and indirect indications of health outcomes, as well as comprehensive, multidimensional composite measures meant to broadly encapsulate the overall quality of care by combining information from multiple individual outcome and process performance measures. No gold standard exists and different organizations use different sets for different purposes (*eg*, quality improvement, competency assessment, criteria for network selection, *etc*). One challenge in selecting performance measures for determining Choice community provider eligibility is that there are a large number of measures meant as indirect indicators of health outcomes (*eg*, readmissions, process measures, *etc*) but uncertainty about their actual association with health outcomes (*eg*, mortality, quality of life, or function). The purpose of our review was to determine whether certain performance measures that are indirect indicators of health outcomes are associated with health outcomes and compare their measurement burden and unintended consequences.

The key findings of this review are:

1. Although 30-day mortality is the most commonly reported direct measure of a health outcome, its singular use has been recently criticized as unintentionally failing to accommodate patients who refuse aggressive treatments, neglecting the impact of postoperative complications on longer-term survival, and promoting manipulation of patient care to achieve better statistics.[30] However, it is encouraging that the usefulness of 30-day mortality as a surrogate for longer-term survival was reinforced in a recent VHA study, which also found no evidence of gaming to meet a 30-day metric. It may also be feasible to collect from hospitals who are already mandated reporters through their participation in the Hospital Inpatient Quality Reporting Program.

2. Thirty-day readmission meets several of the criteria for performance measures for selecting community providers. Thirty-day readmission rates would likely be feasible to collect as all community Medicare providers are already mandated to collect this information. For CABG, it is consistently (although modestly) associated with 30-day mortality (*eg*, *R* range, 0.32 to 0.38; OR 1.14). For hip replacement, evidence is weaker but also points towards 30-day readmission as being a valid indicator of 30-day mortality. However, 30-day readmission has several potential limitations that are discussed in detail below.

3. Adherence to standardized CABG wait time protocols, a set of cardiac surgery process measures, and a specific cardiac surgery antibiotic prophylaxis guideline-based protocol, but not other studied individual process measures, has decreased the likelihood of mortality and complications in cardiac surgery. For orthopedic surgery, a composite quality score of CMS NQF measures was not a valid indicator of either

mortality or complications. We did not find any studies of performance measures collected in the inpatient setting that directly assessed the quality of postoperative bundles of care.

4. Public reporting is an implementation factor that may mediate or moderate the association of performance measures with health outcomes. Discussions with experienced pay-for-performance researchers[89] indicated that public reporting itself may be a strong motivator for hospital administrators and individual providers alike to improve quality. Decision-makers may consider determining eligibility of Choice community providers based on their participation in a public reporting program that involves periodic auditing. This could have the added benefit of demonstrating that providers could produce reliable measures.

Limitations of 30-day Readmission Rates

Although evidence suggests 30-day readmission is a moderately valid indicator of surgical quality and is likely reasonably feasible to assess, its limitations and potential consequences must be considered. A systematic overview of methodological aspects of readmission rates[32] and several editorials have provided detailed evaluations of the limitations of 30-day readmissions.[28,31,43,44,92,93] The most common criticism about all-cause readmission is that it is insensitive to true quality of care issues because only a minority are avoidable (~27%),[94] with some being planned, and the majority being outside of the hospitals' control and likely driven by factors such as mental illness, poor social support, and poor community outpatient resources.[28] A proposed solution is to limit analyses to preventable/avoidable readmission.[32] However, as doing so would require additional resources in the form of independent expert judgment, this is likely technically infeasible to implement. Also, our review identified evidence supporting all-cause readmission as a valid indicator of quality in the form of 30-day mortality. A second common criticism of readmission rates is that their validity is limited by variability in and lack of consensus on the methods used to adjust for case-mix. Although risk prediction models consistently adjust for some variety of patient-, surgical-, and hospital-level variables, current evidence provides limited guidance on which specific variables should be included because studies have been highly heterogenous in their methodology, patient groups, and considered variables, and have found different factors increased risk of readmission. One of the key controversies in the ongoing dialogue about risk adjustment methods is over whether to adjust for socioeconomic status (SES). CMS and NQF have argued not to include adjustment for SES in readmission risk adjustment, because factors such as SES, race, and gender may be associated with inequalities in care and these inequalities may be obscured if these factors are included in risk models. Others argue that differences in outcomes by SES may in fact be due to poorer living conditions, lack of support, and/or poorer nutrition, for which hospitals should not be accountable.[95]Although there is much debate around this issue, it is unclear if adding SES to risk adjustment models would change actual hospital rankings.[96] Third, although not necessarily unique to readmissions, the lack of widely acceptable thresholds for minimum performance may be a barrier to widespread use.[44] Although thresholds must be set, there is an inherent risk of misclassifying providers or otherwise unfairly excluding providers with lower performance for idiosyncratic reasons. Finally, concern has been expressed that focus on readmission may result in neglect of other important aspects of care quality and/or gaming to achieve lower rates at the expense of patient care (*ie*, admitted at 31 days instead of 30 days). Although evaluation of such unintended consequences is largely lacking, a recent VHA study that examined surgical

mortality between postoperative days 25 and 35 by risk decile showed no evidence of delayed mortality within the highest readmission decile.[2] Also, as participation in the Choice network would be voluntary and not universally imposed on the community at large, this may minimize potential unintended consequences of readmission. However, adding performance measures to Choice network participation eligibility requirements may be seen as an additional barrier for community providers and may perpetuate the current network inadequacy challenges.[21]

General Limitations

The main general limitations of the evidence included understudied measures, methodological features, and applicability. With few exceptions,[1,2] evidence on the validity of the surgical performance measures evaluated in this review has largely neglected evaluating their association with the most patient-centered outcomes of long-term health outcomes. Also, although care coordination and follow-up care interventions have been shown to impact readmissions and there has been a call for development of innovative care transitions and postoperative care plan measures, other than one study of time to follow-up appointment timing (before or after the recommended 3 weeks),[85] evidence is lacking on potential postoperative plan of care measures such as presence of pre-discharge assessment and ensuring every patient is scheduled for a follow-up visit. Also, as mentioned above, although there is much concern about the unintended consequences of using performance measures for various purposes, little research has been done in this area. Methodologically, the main strengths of these studies are that the reliability of the data sources was generally high, in that they were from established and regularly audited quality improvement databases, adequacy of adjustment for potential confounding was at least moderate, in that most at least adjusted for patient-level factors, and they likely had adequate statistical power with a median sample size of 2,218 (range 47 to 2,121,215). The main methodological weaknesses were that all but 4 were retrospective, none was able to capture events that may have occurred at hospitals other than the index hospital, and most were based on administrative databases that did not include information not captured by billing codes, such as disease severity. Although a majority of studies were within the US and multicenter, applicability is still limited because most focused on CABG and over half involved samples more than a decade old.

The main limitations of our review methods include (1) our literature search, (2) our focused scope, and (3) our use of sequential instead of independent dual assessment. For our literature search, although we searched 5 databases and other sources and used a comprehensive search strategy, the inconsistent terminology used in literature on performance measures and quality of care may have increased our risk of missing relevant studies. Second, to meet our condensed timeframe we focused our scope to the highest-priority populations and measures of the Office of Community Care. However, this limits the applicability of our findings to other populations and measures of interest. Third, although sequential dual review is a widely used method, its comparison to independent dual review has not yet been empirically studied and may have increased the risk of error and bias.

Future Research

To improve evidence about the validity of performance measures as indicators of surgical quality and increase its applicability to broader populations, high priorities for future research include (1) prospective evaluation of a broader array of modern surgical populations; (2) improving consensus on case-mix adjustment methods, which could be in the forms of an updated systematic review, as well as a new primary study that compares the use of various adjustment

methods; (3) evaluation of associations with longer-term patient-important outcomes; (4) evaluation of unintended consequences including gaming and risk-based patient selection, evaluation of appropriateness of care, changes in disparities, and spillover effects;[97] and (5) evaluation of postoperative plan of care measures such as presence of pre-discharge assessment and ensuring every patient is scheduled for a follow-up visit.

As the reason for this review was to identify performance measures that are valid indicators of quality, for implementation in selective contracting with Choice community providers to improve quality of care for Veterans, implementation should be accompanied by plans not only to extend the evidence on validity, but also to evaluate the effectiveness of the implementation, including how selective contracting is affecting community providers' experience. For the purposes of this review, although the primary interest in the validity of performance measures is in their use for selective contracting, likely they will continue to be used to monitor Choice patient outcomes as well for general quality improvement. Currently, data is routinely collected from community providers using a one-page form that is scanned into the Veterans' charts. However, this process does not permit systematic electronic querying for data analysis. To improve capabilities to monitor Veterans' outcomes in the community and compare to care within VHA, VA Secretary Shulkin and others have called for creation of data standards and standardized electronic data systems that would better permit aggregation of data across sites.[19,44]

Implications for Policy and Implementation

Possible minimum requirements for Choice providers include acceptable performance on national rankings, compatible operational infrastructure, and ability to comply with an agreed-upon wait time threshold. In addition to these minimum requirements, the Office of Community Care could consider the added value of the performance measures that this review has identified as being indicators of desirable health outcomes. These include the single measures of 30-day mortality, a direct health outcome measure, and/or 30-day readmission, the indirect measure with the strongest association with mortality – both of which are commonly measured by surgery programs. Another option is to consider use of a composite performance measure that includes mortality, readmission, and other process measures – such as STS's composite CABG measure. We also recommend use of public reporting program participation and measures of efficiency as additional considerations for Choice community provider minimum standards. However, an unintended effect of stricter performance measure-based criteria for selective contracting with Choice providers may be an undersupply of providers, which could diminish Choice's effect on reducing Veterans' wait times.

An alternative to using individual performance measures that are direct or indirect indicators of health outcomes is to use a composite performance measure[36] that includes mortality, readmission, and other process measures – such as STS's composite CABG measure. Composite performance measures are meant to broadly encapsulate the overall quality of care by combining information from multiple individual outcome and process performance measures into a single, comprehensive, multidimensional measure. An advantage of composite measures is that they can reduce data burden by translating information on a broad range of indicators that may not otherwise be possible to track. However, as with any composite performance measure, including intelligence quotients, there are a few challenges to consider. Use of a single composite measure may lose important detail on variation in performance across individual component measures. For example, an intermediate score on a composite measure may reflect excellent performance

on some measures and below-average on others. Also, all users' unique interests, values, and preferences may not be reflected by the subjective relative weighting of different components in scoring composite measures. Ideally, choice of a composite performance measure should consider the strength of the association between its components and health outcomes, the transparency and comprehensibility of its inputs and rules, the regularity of updating to maintain clinical relevance as new information becomes available, and its reliability and validity. The Office of Community Care could consider use of a composite measure as a highly feasible and comprehensive approach to determine eligibility of community providers if they determined that the potential advantages of a composite measure outweighed its challenges and identified a rigorously developed and validated composite measure that is widely accepted and used.

We examined whether performance measures meant as indirect indicators of health outcomes are related to health outcomes and whether they are difficult (feasible) to measure, but not whether potential Choice community providers collect and report valid data. To apply performance measurement-based selective contracting, VA must be able to determine whether the providers' data are reliably measured. Accordingly, in addition to selecting providers based on meeting thresholds for acceptable 30-day readmission and 30-day mortality rates, decision-makers may also require Choice community providers to participate in a public reporting program that involves periodic auditing, including publically available claims databases such as CMS, commercial comprehensive clinical databases such as Vizient, Crimson, Premier, etcetera, or specialty registries such as STS, ACC, or NSQUIP. This would ensure the reliability of Choice community providers' performance measures, and the participation in public reporting itself may also be a strong motivator for quality improvement. Preferable characteristics of performance reporting programs include (1) use of defined populations, (2) longer time in operation, (3) transparency through providing data that identifies specific providers/surgical centers, and (4) evaluation across the continuum of care.

Also, although we did not evaluate measures such as length of stay and patient visits per physician per month, we suggest decision-makers consider the usefulness of such efficiency measures as a supplement to indicators of patient outcomes. For providers that meet minimum standards on direct or indirect health outcome measures, choosing more efficient providers could benefit national efforts to improve health care value.

CONCLUSIONS

Among performance measures meant as indirect indicators of health outcomes, 30-day readmission is the strongest indicator of 30-day mortality for CABG and for hip replacement and is feasible to measure. Its use for selecting Choice providers is reasonable, but its potential limitations must be considered. Contrary to recent criticism, the 30-day mortality measure is likely a valid surrogate for long-term survival, with a lower than expected risk of gaming. Use of a robust and widely used composite measure of direct and indirect indicators of health outcomes may also be a highly feasible and comprehensive approach to determining eligibility of Choice providers. Also, we recommend use of public reporting program participation and measures of efficiency as additional considerations for Choice community provider minimum standards. An unintended effect of using performance measures as criteria for selecting providers in general is that it may result in an undersupply of Choice providers that could diminish Choice's intended effect of reducing Veterans' wait times.

Acknowledgements: We would like to thank Julia Haskin, MA for editorial support and Steve Meurer, PhD for providing technical expertise.

REFERENCES

1. Guru V, Tu JV, Etchells E, et al. Relationship between preventability of death after coronary artery bypass graft surgery and all-cause risk-adjusted mortality rates. *Circulation.* 2008;117(23):2969-2976.
2. Smith T, Li X, Nylander W, Gunnar W. Thirty-day postoperative mortality risk estimates and 1-year survival in Veterans Health Administration surgery patients. *JAMA Surgery.* 2016;151(5):417-422.
3. Hannan EL, Racz MJ, Walford G, et al. Predictors of readmission for complications of coronary artery bypass graft surgery. *JAMA.* 2003;290(6):773-780.
4. Hannan EL, Zhong Y, Lahey SJ, et al. 30-day readmissions after coronary artery bypass graft surgery in New York State. *Jacc: Cardiovascular Interventions.* 2011;4(5):569-576.
5. Parina RP, Chang DC, Rose JA, Talamini MA. Is a low readmission rate indicative of a good hospital? *Journal of the American College of Surgeons.* 2015;220(2):169-176.
6. Thomas JW. Does risk-adjusted readmission rate provide valid information on hospital quality? *Inquiry.* Fall 1996;33(3):258-270.
7. Tsai TC, Joynt KE, Orav EJ, Gawande AA, Jha AK. Variation in surgical-readmission rates and quality of hospital care. *New England Journal of Medicine.* 2013;369(12):1134-1142.
8. Zitser-Gurevich Y, Simchen E, Galai N, Braun D. Prediction of readmissions after CABG using detailed follow-up data: the Israeli CABG Study (ISCAB). *Medical Care.* 1999;37(7):625-636.
9. Stefan MS, Pekow PS, Nsa W, et al. Hospital performance measures and 30-day readmission rates. *Journal of General Internal Medicine.* 2013;28(3):377-385.
10. Auerbach AD, Hilton JF, Maselli J, Pekow PS, Rothberg MB, Lindenauer PK. Follow the crowd or shop for the best? Volume, quality, and outcomes of coronary artery bypass surgery. *Annals of Internal Medicine.* 2009;150(10):696-704.
11. Brinkman W, Herbert MA, O'Brien S, et al. Preoperative beta-blocker use in coronary artery bypass grafting surgery: national database analysis. *JAMA Intern Med.* 2014;174(8):1320-1327.
12. Kurlansky PA, Argenziano M, Dunton R, et al. Quality, not volume, determines outcome of coronary artery bypass surgery in a university-based community hospital network. *The Journal of Thoracic and Cardiovascular Surgery.* 2012;143(2):287-293. e281.
13. LaPar DJ, Speir AM, Crosby IK, et al. Postoperative atrial fibrillation significantly increases mortality, hospital readmission, and hospital costs. *Annals of Thoracic Surgery.* 2014;98(2):527-533; discussion 533.
14. Cotogni P, Barbero C, Passera R, Fossati L, Olivero G, Rinaldi M. Violation of prophylactic vancomycin administration timing is a potential risk factor for rate of surgical site infections in cardiac surgery patients: a prospective cohort study. *BMC Cardiovascular Disorders.* 2017;17(1). doi: 10.1186/s12872-017-0506-5.
15. Bhattacharyya T, Freiberg AA, Mehta P, Katz JN, Ferris T. Measuring the report card: the validity of pay-for-performance metrics in orthopedic surgery. *Health Aff (Millwood).* 2009;28(2):526-532.
16. Wang Z, Chen F, Ward M, Bhattacharyya T. Compliance with Surgical Care Improvement Project measures and hospital-associated infections following hip arthroplasty. *Journal of Bone & Joint Surgery - American Volume.* 2012;94(15):1359-1366.

17. Légaré JF, MacLean A, Buth KJ, Sullivan JA. Assessing the risk of waiting for coronary artery bypass graft surgery among patients with stenosis of the left main coronary artery. *CMAJ.* 2005;173(4):371-375.

18. Sobolev BG, Fradet G, Kuramoto L, Rogula B. An observational study to evaluate 2 target times for elective coronary bypass surgery. *Medical Care.* 2012;50(7):611-619.

19. Shulkin D. Understanding Veteran Wait Times. *Annals of Internal Medicine.* 2017. doi: 10.7326/M17-0900.

20. Expanded Access to Non-VA Care Through the Veterans Choice Program. Final Rule. *Federal Register.* Oct 29 2015;80(209):66419-66429.

21. VA Office of Inspector General. Veterans Health Administration: Review of the Implementation of the Veterans Choice Program. Washington, DC: Veterans Affairs; 2017.

22. O'Hanlon C, Huang C, Sloss E, et al. Comparing VA and non-VA quality of care: a systematic review. *Journal of General Internal Medicine.* 2017;32(1):105-121.

23. Barnett P. Differences in Quality, Cost, and Access between VA and Fee Basis CABG and PCI (IIR 11-049). 2016; Abstract. Available at: https://www.hsrd.research.va.gov/research/abstracts.cfm?Project_ID=2141702066. Accessed May 3, 2017.

24. VA Office of Public Affairs. Top VA Health Care Official Announces Initiatives and Progress Made to Improve Access to Care: *New Initiatives Chart Roadmap to Access-to-Care Improvements*. Washington DC: Veterans Affairs; 2016.

25. Donabedian A. The quality of medical care methods for assessing and monitoring the quality of care for research and for quality assurance programs. *Science.*1978;200:856-864.

26. Institute of Medicine Committee on Quality of Health Care in America. *Crossing the quality chasm: a new health system for the 21st century.* Washington, DC: National Academy Press; 2001.

27. Chee TT, Ryan AM, Wasfy JH, Borden WB. Current State of Value-Based Purchasing Programs. *Circulation.* 2016;133(22):2197-2205.

28. Joynt KE, Jha AK. Thirty-day readmissions—truth and consequences. *New England Journal of Medicine.* 2012;366(15):1366-1369.

29. Health Resources and Services Administration. Performance Management & Measurement. https://www.hrsa.gov/quality/toolbox/methodology/performancemanagement/. Accessed May 9, 2017.

30. Schwarze ML, Brasel KJ, Mosenthal AC. Beyond 30-day mortality: aligning surgical quality with outcomes that patients value. *JAMA Surgery.* 2014;149(7):631-632.

31. Krumholz HM. Seeking better outcomes in coronary artery bypass grafting: lessons from past experience. *Circulation.* Jun 10 2008;117(23):2963-2965.

32. Fischer C, Lingsma HF, Marang-van de Mheen PJ, Kringos DS, Klazinga NS, Steyerberg EW. Is the readmission rate a valid quality indicator? A review of the evidence. *PLoS ONE [Electronic Resource].* 2014;9(11):e112282.

33. Fry DE. Surgical site infections and the surgical care improvement project (SCIP): evolution of national quality measures. *Surgical Infections.* 2008;9(6):579-584.

34. Shiloach M, Frencher SK, Steeger JE, et al. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. *Journal of the American College of Surgeons.* 2010;210(1):6-16.

35. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Annals of Surgery.* 1998;228(4):491.

36. Peterson ED, DeLong ER, Masoudi FA, et al. ACCF/AHA 2010 Position Statement on Composite Measures for Healthcare Performance Assessment: a report of American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures (Writing Committee to Develop a Position Statement on Composite Measures). *J Am Coll Cardiol.* 2010;55(16):1755-1766.

37. D'Agostino RS, Jacobs JP, Badhwar V, et al. The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2017 Update on Outcomes and Quality. *Ann Thorac Surg.* 2017;103(1):18-24.

38. Jacobs JP, Shahian DM, He X, et al. Penetration, Completeness, and Representativeness of The Society of Thoracic Surgeons Adult Cardiac Surgery Database. *Ann Thorac Surg.* Jan 2016;101(1):33-41; discussion 41.

39. Shahian DM, Jacobs JP, Edwards FH, et al. The Society of Thoracic Surgeons National Database. *Heart.* 2013;99(20):1494-1501.

40. The Society of Thoracic Surgeons. The Society of Thoracic Surgeons Quality Performance Measures. 2017; https://www.sts.org/quality-research-patient-safety/quality/quality-performance-measures. Accessed June 28, 2017.

41. Centers for Medicare & Medicaid Services. Hospital Compare. https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html. Accessed June 28, 2017.

42. Olmsted MG, Geisen E, Murphy J, et al. *Methodology: US News & World Report 2016-17 Best Hospitals: Specialty Rankings.* 2016.

43. Girotti ME, Shih T, Dimick JB. Health policy update: rethinking hospital readmission as a surgical quality measure. *JAMA Surgery. 2*014;149(8):757-758.

44. Landon BE, Normand S-LT, Blumenthal D, Daley J. Physician clinical performance assessment: prospects and barriers. *JAMA.* 2003;290(9):1183-1189.

45. Centers for Medicare & Medicaid Services. Hospital Readmissions Reduction Program: Fiscal Year (FY) 2017 Fact Sheet. 2016. http://www.qualitynet.org/dcs/ContentServer?cid=1228772412458&pagename=QnetPublic%2FPage%2FQnetTier4&c=Page. Accessed May 9, 2017.

46. Damberg CL, Sorbero ME, Lovejoy SL, Martsolf GR, Raaen L, Mandel D. *Measuring success in health care value-based purchasing programs: Summary and Recommendations.* Santa Monica, CA: 2014.

47. Berwick DM. Era 3 for medicine and health care. *JAMA.* 2016;315(13):1329-1330.

48. Atkins D, Kilbourne AM, Shulkin D. Moving From Discovery to System-Wide Change: The Role of Research in a Learning Health Care System: Experience from Three Decades of Health Systems Research in the Veterans Health Administration. *Annual Review of Public Health.* 2017.

49. Dimick JB, Ghaferi AA. Hospital readmission as a quality measure in surgery. *JAMA.* 2015;313(5):512-513.

50. Croft AM, Lynch P, Smellie JS, Dickinson CJ. Outpatient waiting times: indicators of hospital performance? *J R Army Med Corps.* 1998;144(3):131-137.

51. Press MJ, Scanlon DP, Ryan AM, et al. Limits of readmission rates in measuring hospital quality suggest the need for added metrics. *Health Aff (Millwood).* 2013;32(6):1083-1091.

52. Ashton CM, Del Junco DJ, Souchek J, Wray NP, Mansyur CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. *Medical Care.* 1997;35(10):1044-1059.

53. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA.* 2011;306(15):1688-1698.

54. Shahian DM, Blackstone EH, Edwards FH, et al. Cardiac surgery risk models: a position article. *Ann Thorac Surg.* 2004;78(5):1868-1877.

55. Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 3--valve plus coronary artery bypass grafting surgery. *Ann Thorac Surg.* 2009;88(1 Suppl):S43-62.

56. Shahian DM, O'Brien SM, Filardo G, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1--coronary artery bypass grafting surgery. *Ann Thorac Surg.* 2009;88(1 Suppl):S2-22.

57. Kondo K, Damberg C, Mendelsen A, et al. Understanding the intervention and implementation factors associated with benefits and harms of pay for performance programs in healthcare. VA-ESP Project #05-225; 2015.

58. Sterne J, Higgins J, Reeves B. A Cochrane risk of bias assessment tool: For non-randomized studies of interventions (ACROBAT-NRSI). 2014.

59. Berkman ND, Lohr KN, 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 Methods Guide for Effectiveness and Comparative Effectiveness Reviews.* Rockville MD: Agency for Healthcare Research and Quality; 2013.

60. Khan SK, Kalra S, Khanna A, Thiruvengada MM, Parker MJ. Timing of surgery for hip fractures: a systematic review of 52 published studies involving 291,413 patients. *Injury.* 2009;40(7):692-697.

61. Leung F, Lau TW, Kwan K, Chow SP, Kung AW. Does timing of surgery matter in fragility hip fractures? *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA.* 2010;21(Suppl 4):S529-534.

62. Moja L, Piatti A, Pecoraro V, et al. Timing matters in hip fracture surgery: patients operated within 48 hours have better outcomes. A meta-analysis and meta-regression of over 190,000 patients. *PloS One.* 2012;7(10):e46175.

63. Schepers T, De Vries MR, Van Lieshout EM, Van der Elst M. The timing of ankle fracture surgery and the effect on infectious complications; a case series and systematic review of the literature. *International Orthopaedics.* 2013;37(3):489-494.

64. Shiga T, Wajima Z, Ohe Y. Is operative delay associated with increased mortality of hip fracture patients? Systematic review, meta-analysis, and meta-regression. *Canadian Journal of Anaesthesia.* 2008;55(3):146-154.

65. Simunovic N, Devereaux PJ, Sprague S, et al. Effect of early surgery after hip fracture on mortality and complications: systematic review and meta-analysis. *CMAJ.* 2010;182(15):1609-1616.

66. Lurati Buse GA, Bhandari M, Devereaux P, Inverstigators T. Accelerated care versus standard care among patients with hip fracture: the HIP ATTACK pilot trial. *CMAJ.* 2014;186(1):E52-60.
67. Holvik K, Ranhoff AH, Martinsen MI, Solheim LF. Predictors of mortality in older hip fracture inpatients admitted to an orthogeriatric unit in oslo, norway. *J Aging Health.* 2010;22(8):1114-1131.
68. Schelenz S, Tucker D, Georgeu C, et al. Significant reduction of endemic MRSA acquisition and infection in cardiothoracic patients by means of an enhanced targeted infection control programme. *Journal of Hospital Infection.* 2005;60(2):104-110.
69. Clague JE, Craddock E, Andrew G, Horan MA, Pendleton N. Predictors of outcome following hip fracture. Admission time predicts length of stay and in-hospital mortality. *Injury.* 2002;33(1):1-6.
70. Griffiths EJ, Cash DJ, Kalra S, Hopgood PJ. Time to surgery and 30-day morbidity and mortality of periprosthetic hip fractures. *Injury.* 2013;44(12):1949-1952.
71. Kim ES, Park SW, Lee CS, Gyung Kwak Y, Moon C, Kim BN. Impact of a national hospital evaluation program using clinical performance indicators on the use of surgical antibiotic prophylaxis in Korea. *Int J Infect Dis.* 2012;16(3):e187-192.
72. LaPar DJ, Isbell JM, Kern JA, Ailawadi G, Kron IL. Surgical Care Improvement Project measure for postoperative glucose control should not be used as a measure of quality after cardiac surgery. *Journal of Thoracic & Cardiovascular Surgery.* 2014;147(3):1041-1048.
73. Lund CA, Moller AM, Wetterslev J, Lundstrom LH. Organizational factors and long-term mortality after hip fracture surgery. A cohort study of 6143 consecutive patients undergoing hip fracture surgery. *PLoS ONE [Electronic Resource]*. 2014;9(6):e99308.
74. McDonnell ME, Alexanian SM, Junqueira A, Cabral H, Lazar HL. Relevance of the Surgical Care Improvement Project on glycemic control in patients undergoing cardiac surgery who receive continuous insulin infusions. *Journal of Thoracic & Cardiovascular Surgery.* 2013;145(2):590-594; discussion 595-597.
75. Meessen JM, Pisani S, Gambino ML, et al. Assessment of mortality risk in elderly patients after proximal femoral fracture. *Orthopedics.* 2014;37(2):e194-200.
76. Rasouli MR, Jaberi MM, Hozack WJ, Parvizi J, Rothman RH. Surgical care improvement project (SCIP): has its mission succeeded? *Journal of Arthroplasty.* 2013;28(7):1072-1075.
77. Ryan DJ, Yoshihara H, Yoneoka D, Egol KA, Zuckerman JD. Delay in Hip Fracture Surgery: An Analysis of Patient-Specific and Hospital-Specific Risk Factors. *Journal of Orthopaedic Trauma.* 2015;29(8):343-348.
78. Sari I, Acar Z, Nurkalem Z, et al. Preoperative clinical status but not waiting time predicts in-hospital outcomes of surgery in patients with left main coronary artery stenosis. *Tohoku J Exp Med.* 2007;213(2):173-180.
79. Tennent TD, Calder PR, Salisbury RD, Allen PW, Eastwood DM. The operative management of displaced intra-articular fractures of the calcaneum: a two-centre study using a defined protocol. *Injury.* 2001;32(6):491-496.
80. Vallier HA, Super DM, Moore TA, Wilber JH. Do patients with multiple system injury benefit from early fixation of unstable axial fractures? The effects of timing of surgery on initial hospital course. *Journal of Orthopaedic Trauma.* 2013;27(7):405-412.

81. Karademir G, Bilgin Y, Ersen A, et al. Hip fractures in patients older than 75 years old: Retrospective analysis for prognostic factors. *International Journal Of Surgery.* 2015;24(Pt A):101-104.

82. Kane S. Post-hoc Power Calculator. *Evaluate statistical power of an existing study* 2017; http://clincalc.com/stats/Power.aspx. Accessed 6/20/2017, 2017.

83. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA.* 2006;295(10):1168-1170.

84. Merkow RP, Ju MH, Chung JW, et al. Underlying reasons associated with hospital readmission following surgery in the United States. *JAMA.* 2015;313(5):483-495.

85. Maniar HS, Bell JM, Moon MR, et al. Prospective evaluation of patients readmitted after cardiac surgery: analysis of outcomes and identification of risk factors. *Journal of Thoracic & Cardiovascular Surgery.* 2014;147(3):1013-1018.

86. Nomellini V, Tevis S, O'Leary M, et al. Utilizing Patient Perspectives to Help Reduce Postoperative Readmission Rates. *International Journal of Surgery Research and Practice.* 2015;2(3).

87. Neuman MD, Archan S, Karlawish JH, Schwartz JS, Fleisher LA. The relationship between short-term mortality and quality of care for hip fracture: a meta-analysis of clinical pathways for hip fracture. *Journal of the American Geriatrics Society.* 2009;57(11):2046-2054.

88. Shahian DM, He X, O'Brien SM, et al. Development of a clinical registry-based 30-day readmission measure for coronary artery bypass grafting surgery. *Circulation.* 2014;130(5):399-409.

89. 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.

90. Hannan EL, Sarrazin MS, Doran DR, Rosenthal GE. Provider profiling and quality improvement efforts in coronary artery bypass graft surgery: the effect on short-term mortality among Medicare beneficiaries. *Medical Care.* 2003;41(10):1164-1172.

91. 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.* 07 2017;166(5):341-353.

92. Axon RN, Williams MV. Hospital readmission as an accountability measure. *JAMA.* 2011;305(5):504-505.

93. Benbassat J, Taragin M. Hospital readmissions as a measure of quality of health care: advantages and limitations. *Archives of Internal Medicine.* 2000;160(8):1074-1081.

94. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med.* 2009;360(14):1418-1428.

95. Atkinson G, Giovanis T. Conceptual errors in the CMS refusal to make socioeconomic adjustments in readmission and other quality measures. *J Ambulatory Care Manage.* 2014;37(3):269-272.

96. Blum K, Gottlieb SS. The effect of a randomized trial of home telemonitoring on medical costs, 30-day readmissions, mortality, and health-related quality of life in a cohort of community-dwelling heart failure patients. *J Card Fail.* 2014;20(7):513-521.

97. Portland Evidence-based Synthesis Program. Pay-for-Performance and Veteran Care: Effects, Implementation, and Unintended Consequences. *ESP Reports in Progress* https://www.hsrd.research.va.gov/publications/esp/in_progress.cfm#pay-perf. Accessed May 9, 2017.