# Continuous Quality Improvement (CQI) for Clinical Teams: A Systematic Review of Reviews

March 2022

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

Prepared by:
Evidence Synthesis Program (ESP) Center
VA Greater Los Angeles Health Care System
Los Angeles, CA
Isomi Miake-Lye, PhD, Co-Director
Paul G. Shekelle, MD, PhD, Co-Director

Authors:
Joel C. Boggan, MD, MPH
Paul G. Shekelle, MD, PhD
Selene Mak, PhD, MPH
Jason Burton, MA
Meron Begashaw, MPH
Isomi Miake-Lye, PhD

VA
U.S. Department of Veterans Affairs
Veterans Health Administration
*Health Services Research & Development Service*

# PREFACE

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

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

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

This topic was developed in response to a nomination from the VA Office of System Redesign and Improvement (10E2F). The scope was further developed with input from Operational Partners (below), the ESP Coordinating Center, the review team, and the technical expert panel (TEP). The ESP consulted several technical and content experts in designing the research questions and review methodology. In seeking broad expertise and perspectives, divergent and conflicting opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Ultimately, however, research questions, design, methodologic approaches, and/or conclusions of the review may not necessarily represent the views of individual technical and content experts. Comments on this evidence report are welcome and can be sent to Nicole Floyd, Deputy Director, ESP Coordinating Center at Nicole.Floyd@va.gov.

**Recommended citation:** Boggan, JC, Shekelle, PG, Mak, SS, Burton, J, Begashaw, MM, Miake-Lye IM. Continuous Quality Improvement (CQI) for Clinical Teams: A Systematic Review of Reviews. Los Angeles: Evidence Synthesis Program, Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs. VA ESP Project #05-226; 2022.

This report is based on research conducted by the Evidence Synthesis Program (ESP) Center located at the **VA Greater Los Angeles Health Care System, Los Angeles, CA,** directed by **Isomi Miake-Lye, PhD and Paul Shekelle, MD, PhD**, and funded by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development.

The findings and conclusions in this document are those of the author(s) who are responsible for its contents and do not necessarily represent the views of the Department of Veterans Affairs or the United States government. Therefore, no statement in this article should be construed as an official position of the Department of Veterans Affairs. No investigators have any affiliations or financial involvement (eg, employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in the report.

# ACKNOWLEDGMENTS


## Operational Partners

Operational partners are system-level stakeholders who help ensure relevance of the review topic to the VA, contribute to the development of and approve final project scope and timeframe for completion, provide feedback on the draft report, and provide consultation on strategies for dissemination of the report to the field and relevant groups.

**Vince Watts, MD**
*Interim Director*
VA Office of System Redesign and Improvement (10E2F)

## Technical Expert Panel (TEP)

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

**David Ganz, MD, PhD**
*Associate Director,* VA HSR&D Center for the Study of Healthcare Innovation, Implementation & Policy, VA Greater Los Angeles Healthcare System
*Professor of Medicine,* UCLA School of Medicine

**Laura Damschroder, MPH, MS**
*Research Investigator,* VA Ann Arbor Center for Clinical Management Research
*Project Principal Investigator,* Personalizing Options through Veteran Engagement Quality Enhancement Research Initiative (QUERI)

**Kyler Godwin, PhD, MPH**
*Director*, Coordinating Center, National VA Quality Scholars Program
*Assistant Professor*, Department of Medicine, Baylor College of Medicine

**Elizabeth Yano, PhD, MSPH**
*Director and Senior Research Career Scientist,* VA HSR&D Center for the Study of Healthcare Innovation & Policy, VA Greater Los Angeles Healthcare System
*Adjunct Professor of Health Policy & Management,* University of California Los Angeles Fielding School of Public Health

## Peer Reviewers

The Coordinating Center sought input from external peer reviewers to review the draft report and provide feedback on the objectives, scope, methods used, perception of bias, and omitted evidence. Peer reviewers must disclose any relevant financial or non-financial conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The Coordinating Center and the ESP Center work to balance, manage, or mitigate any potential nonfinancial conflicts of interest identified.

# TABLE OF CONTENTS

**Preface** ..... **i**
**Acknowledgments** ..... **ii**
**Executive Summary** ..... **1**
Introduction ..... 1
Methods ..... 2
Results ..... 3
Discussion ..... 4
Abbreviations Table ..... 5
**Evidence Report** ..... **6**
**Introduction** ..... **6**
**Methods** ..... **7**
Topic Development ..... 7
Search Strategy ..... 7
Study Selection ..... 7
Data Abstraction ..... 8
Quality Assessment ..... 8
Data Synthesis and Analysis ..... 9
Peer Review ..... 10
**Results** ..... **11**
Literature Flow ..... 11
Key Question 1A: What is the comparative effectiveness of implementing continuous quality improvement frameworks in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change? ..... 21
Key Points (Summary of Findings) ..... 21
Detailed Findings ..... 21
Certainty of Evidence for Key Question 1A ..... 22
Key Question 1B: What is the effectiveness of implementing a continuous quality improvement framework in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change? ..... 22
Key Points (Summary of Findings) ..... 22
Detailed Findings ..... 23
Sustainment of Change in CQI Frameworks ..... 28
Health Care Workers' Reaction, Learning, and Behavior Change to CQI Frameworks ..... 28
Certainty of Evidence for Key Question 1B ..... 29

Key Question 2: What factors (including intervention, inner setting, outer setting, individuals involved, and process by which implementation is accomplished) contribute to the success or failure of these continuous quality improvement frameworks? ..... 29
Key Points (Summary of Findings) ..... 29
Detailed Findings ..... 29
Intervention Characteristics Contributing to Success or Failure of CQI Frameworks ..... 30
Characteristics of Individuals Contributing to Success or Failure of CQI Frameworks ..... 32
Setting and Process Characteristics Contributing to Success or Failure of CQI Frameworks ..... 33
Certainty of Evidence for Key Question 2 ..... 34
**Summary and Discussion ..... 35**
Summary of Evidence by Key Question ..... 35
Limitations ..... 36
Research Gaps/Future Research ..... 37
Conclusions ..... 37
**References ..... 39**

**TABLES**
Table 1. Modified AMSTAR2 Questions ..... 9
Table 2: Total AMSTAR2 Scores, by Criterion ..... 13
Table 3: AMSTAR2 Scoring for Question 1 Describing Reviews Reporting Population/Setting, Intervention, Comparator(s), and/or Outcomes in the Resarch Questions and Inclusion Criteria ..... 15
Table 4: AMSTAR2 Scoring for Question 8 Describing Reviews Reporting Population/Setting, Intervention, Comparator(s), Outcomes, and/or Research Designs in the Results ..... 16

**FIGURES**
Figure 1: Literature Flow Chart ..... 11
Figure 2. Distribution of AMSTAR2 Quality Scores ..... 12
Figure 3. Map of Included Reviews, by Key Questions ..... 18
Figure 4: Number of Included Studies, by Reviews ..... 19
Figure 5: Number of Reviews, by End Year of Search Date ..... 20

**Appendix A. Search Strategies ..... 42**
**Appendix B. Full-Text Screening Form ..... 44**
**Appendix C. Criteria Used in Quality Assessment ..... 46**
**Appendix D. Peer Review Disposition ..... 50**
**Appendix E. evidence Table ..... 60**
**Appendix F. Citations for Excluded publications ..... 70**

# EXECUTIVE SUMMARY

## INTRODUCTION

As part of its mandate to optimize health outcomes for Veterans, the Department of Veterans Affairs (VA) has an incentive to improve the quality and safety of health care. Standardizing a process improvement methodology and training across the entire VA has the potential to expand resources for local improvement activities and improve the quality and efficiency of care delivery.

Continuous quality improvement frameworks are system-level approaches to improving the quality and safety of health care through systematic data-guided activities, iterative development and testing of processes, and designing with local conditions in mind. Lean Management (Lean) has subsequently been applied to a variety of medical and industrial settings and is one of the most popular continuous quality improvement frameworks in health care settings. In December 2019, the Deputy Under Secretary for Health issued a directive outlining the deployment of a new VA-wide program for systems redesign and improvement. As part of this directive, Lean was designated as the primary process improvement methodology to be utilized across the VA.

Despite designation as the preferred continuous quality improvement methodology, there is uncertainty as to whether Lean is superior to other continuous quality improvement strategies, such as Clinical Microsystems or the Institute for Healthcare Improvement (IHI)'s Model for Change. There is also uncertainty as to whether certain intervention-level or health system-level factors affect the success or failure of specific methodologies, such as rigorous training of staff or health system academic affiliation.

Several reviews on continuous quality improvement methodologies exist; however, none identified in a preliminary literature search by the Evidence Synthesis Program (ESP) Coordinating Center currently cover all continuous quality improvement frameworks, settings, and outcomes of interest. Therefore, this current review of reviews was requested by the VA Office of Systems Redesign and Improvement (SRI), which addresses the following Key Questions:

**Key Question 1A:** What is the comparative effectiveness of implementing continuous quality improvement frameworks in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?

**Key Question 1B:** What is the effectiveness of implementing a continuous quality improvement framework in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?

**Key Question 2:** What factors (including intervention characteristics, inner setting, outer setting, individuals involved, and process by which implementation is accomplished) contribute to the success or failure of these continuous quality improvement frameworks?

# METHODS

We developed and followed a standard protocol for this review in collaboration with operational partners and a Technical Expert Panel (PROSPERO registration number CRD42021245263).

## Data Sources and Searches

We conducted broad systematic review searches using terms relating to "quality improvement" or "continuous quality improvement" or "system redesign" in 4 databases: PubMed, CINAHL, DARE, and Cochrane. Search dates for PubMed are from 01/01/2010 through 03/18/21. Search dates for CINAHL and Cochrane are from 01/01/2010 through 03/30/2021. Search dates for DARE are from 01/01/2010 to 03/31/2015.

## Study Selection

Four team members working independently screened the titles of retrieved citations. Full-text review was conducted in duplicate by teams of 2, with any disagreements resolved through discussion.

## Data Abstraction and Quality Assessment

Data extraction was completed in duplicate. All discrepancies were resolved with full-group discussion. We abstracted data on the following: continuous quality improvement framework/strategy discussed, whether the article described context/factors contributing to the success or failure of the framework/strategy, total number of studies included, search dates, health care condition, healthcare setting, and geographical region.

Each systematic review was assessed using a modified version of the Assessing the Methodological Quality of Systematic Reviews 2 (AMSTAR2) criteria. This 16-item tool was designed to assess the methodological quality of systematic reviews. As some AMSTAR2 items concern meta-analysis, we adapted the tool for this review, resulting in a 13-item tool. Assessment of studies using our modified tool was also completed in duplicate, with discrepancies resolved with group discussion. No study was excluded from analysis based on AMSTAR2 score; however, we chose a score of greater than or equal to 8 to represent higher-quality systematic reviews. Studies reporting results relevant to our key questions utilizing an established method of synthesis other than traditional systematic review methods were not assessed with our modified AMSTAR2 tool.

We used the Consolidated Framework for Implementation Research (CFIR) to guide abstraction and synthesis of Key Question 2 around the following domains: intervention characteristics, inner setting, outer setting, individuals involved, and process by which implementation is accomplished. All studies discussed in Key Question 2 were assessed for inclusion in duplicate, with discrepancies resolved with group discussion.

## Data Synthesis and Analysis

Our review is a narrative analysis.

# RESULTS

## Results of Literature Search

We identified 1,795 potentially relevant citations, of which 274 were included at the abstract screening level. A total of 136 publications were reviewed at the full-text stage, and 36 publications were identified at full-text review as meeting initial inclusion criteria. The systematic reviews reported on studies conducted in multiple countries, primarily in North America and Europe. Of the 36 included reviews, 29 reviews were assessed using the modified AMSTAR2 tool. Scores on the AMSTAR2 tool ranged from a high of 11 points, out of a possible 12 points, to a low of 2 points. The median and mode scores on the modified tool were both 5.

## Summary of Results for Key Questions

### ***Key Question 1A**: What is the comparative effectiveness of implementing continuous quality improvement frameworks in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?*

We assessed the literature for evidence regarding the comparative effectiveness of different continuous quality improvement strategies. We were only able to identify a single review,[1] with an AMSTAR2 rating of 2, that met inclusion criteria. This review is over a decade old and found no evidence that any single continuous quality improvement strategy was more effective than others. Instead, the authors found there was significant overlap in defining and implementing different approaches to quality improvement, concluding that the local context should guide which continuous quality improvement framework is ultimately implemented. The authors did identify 7 "necessary, but not sufficient" conditions for successful implementation of any continuous quality improvement strategy: provision of the practical and human resources to enable quality improvement; active engagement of health professionals, especially doctors; sustained managerial focus and attention; use of multi-faceted interventions; coordinated action at all levels of the health care system; substantial investment in training and development; and availability of robust and timely data through supported information technology systems.

### ***Key Question 1B**: What is the effectiveness of implementing a continuous quality improvement framework in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?*

Having found only a single review comparing the effectiveness of different continuous quality improvement strategies, we then assessed the 25 reviews that studied at least 1 framework for evidence of effectiveness for a variety of outcomes. None of the 11 reviews that examined more than 1 methodology reached a conclusion that any strategy was superior to the others in terms of results. However, many frameworks have been successfully implemented in a variety of clinical settings. In some clinical settings, such as in the operating room (OR) and emergency department (ED), multiple different strategies have been used in different geographic locations. Nine of the 22 reviews for which an AMSTAR2 rating was calculated had a modified AMSTAR2 rating of at least 8; however, reviews with higher AMSTAR2 scores did not draw more specific conclusions regarding Key Question 1B outcomes than reviews with lower scores.

Similarly, none of the 7 reviews discussing sustainment of change identified a superior strategy. Additionally, no review directly commented on health care workers' reactions in being involved as part of continuous quality improvement framework implementation, what health care workers learned or retained as part of continuous quality improvement training, nor any health care worker behavior changes noted after implementation of a continuous quality improvement strategy. Five reviews discussed clinician/provider satisfaction as an outcome of implementation of a continuous quality improvement methodology, with mixed results. It is unclear from these 5 reviews whether the clinicians/providers in whom satisfaction was studied were part of the implementation teams or were frontline workers affected by the implementation.

***Key Question 2**: What factors (including intervention, inner setting, outer setting, individuals involved, and process by which implementation is accomplished) contribute to the success or failure of these continuous quality improvement frameworks?*

We then assessed the literature for evidence regarding success or failure factors associated with Consolidated Framework for Implementation Research (CFIR) factors. Ultimately, 20 reviews studied at least 1 of 3 CFIR factors: intervention characteristics (n=15), characteristics of individuals (n=6), and/or inner setting (n=10). None of these 20 reviews compared the success or failure of different continuous quality improvement strategies based on any of these 3 factors, however. Instead, the majority of reviews listed aspects of some or each of these 3 CFIR categories that the authors deemed important for implementation of the studied strategy, with little to no supporting evidence. No publication included in this review discussed whether either outer setting or specific processes during implementation of a continuous quality improvement framework contributed to either the success or failure of implementation for any framework. Four of these 15 reviews for which an AMSTAR2 rating was calculated had a modified AMSTAR2 score of 8 points or higher. However, as none of the conclusions related to the CFIR topics were supported by comparative data, we conclude there is low certainty of evidence for specific intervention characteristics, individuals to be involved, or inner setting aspects leading to success in implementing a continuous quality improvement methodology.

# DISCUSSION

## Applicability of Findings to the VA Population

The vast majority of reviews within these reviews were conducted in Organisation for Economic Co-operation and Development countries, which are a group of 38 countries with mature economies and, often, well-financed healthcare systems. Additionally, some individual studies within these reviews were performed with Veteran populations, both of which improve applicability to VA.

## Research Gaps/Future Research

In brief, comparative research of different continuous quality improvement strategies is needed both in US populations and with Veterans. Additionally, improved reporting of ongoing work would improve the evidence base regarding implementation of continuous quality improvement frameworks. Pursuing such initiatives across large health systems such as the VA has the potential to improve health care for millions of patients.

### Conclusions

Prior systematic reviews of continuous quality improvement strategies have not, with 1 exception, compared the effectiveness of different methodologies. Instead, many published reviews have shown success for 1 or more methodologies within specific contexts. However, these findings are likely subject to significant publication bias from the constituent studies, as it is probable unsuccessful quality improvement work is less likely to receive publication, making the overall certainty of evidence low. Few data are available regarding sustainment of changes made through continuous quality improvement and no systematic reviews we identified discussed health care workers' reactions, learning, or behavior changes related to participating in continuous quality improvement. Similarly, no systematic reviews compared the success or failure of different continuous quality improvement frameworks based on intervention characteristics, characteristics of individuals, or inner setting. Furthermore, no studies discussed whether either the outer setting or specific processes during implementation of a continuous quality improvement framework contributed to either the success or failure of implementation of any methodology. Few systematic reviews included in this review of reviews had high ratings on a modified AMSTAR2 tool, leading us to conclude the overall certainty of evidence related to these topics is low to moderate. Thus, evidence gaps remain regarding whether any continuous quality improvement strategy is superior to others and how any methodology should be implemented at large scale within the VA context.

## ABBREVIATIONS TABLE

| Abbreviation | Meaning |
|---|---|
| CFIR | Consolidated Framework for Implementation Research |
| CMS | Centers for Medicare & Medicaid Services |
| CQI | Continuous Quality Improvement |
| DMAIC | Define, Measure, Analyze, Improve, and Control |
| IHI | Institute for Healthcare Improvement |
| MRSA | Methicillin-resistant Staphylococcus aureus |
| PDCA | Plan-Do-Check-Act |
| PDSA | Plan-Do-Study-Act |
| QI | Quality Improvement |
| RCT | Randomized Control Trial |
| SRI | VA Office of Systems Redesign and Improvement |
| TEP | Technical Expert Panel |
| VA QUERI | VA Quality Enhancement Research Initiative |
| VA | Department of Veterans Affairs |
| VISN | Veterans Integrated Service Network |

# EVIDENCE REPORT

## INTRODUCTION

As part of its mandate to optimize health outcomes for Veterans, the Department of Veterans Affairs (VA) has an incentive to improve the quality and safety of health care. Standardizing a process improvement methodology and training across the entire VA has the potential to expand resources for local improvement activities, particularly in settings such as individual clinics or units that may have fewer currently trained personnel to support project leadership and management, and improve the quality and efficiency of care delivery.

Continuous quality improvement (CQI) frameworks are system-level approaches to improving the quality and safety of health care through systematic data-guided activities, iterative development and testing of processes, and designing with local conditions in mind.[2,3] Lean Management (Lean) – a process adapted from a 1930s manufacturing model by the Toyota Corporation that seeks to increase efficiencies and reduce waste – has subsequently been applied to a variety of medical and industrial settings and is one of the most popular CQI frameworks in health care settings. In December 2019, the Deputy Under Secretary for Health issued a directive outlining the deployment of a new VA-wide program for systems redesign and improvement.[3] As part of this directive, Lean was designated as the primary process improvement methodology to be utilized across the VA.

Despite designation as the preferred CQI methodology, there is uncertainty as to whether Lean is superior to other CQI frameworks, such as the Institute for Healthcare Improvement (IHI)'s Model for Change or Clinical Microsystems. There is also uncertainty as to whether certain intervention-level or health system-level factors affect the success or failure of specific CQI methodologies, such as rigorous training of staff or health system academic affiliation.

Several reviews on CQI methodologies exist; however, none identified in a preliminary literature search by the Evidence Synthesis Program (ESP) Coordinating Center currently cover all interventions, settings, and outcomes of interest. Therefore, this current review of reviews was requested by the VA Office of Systems Redesign and Improvement (SRI), which is charged with partnering with Veterans Integrated Service Networks (VISNs) in implementing the 2019 directive. This report will be used to identify effective CQI frameworks and conditions necessary for their success for dissemination and training across the VA by the SRI.

# METHODS

## TOPIC DEVELOPMENT

This topic was developed in response to a nomination by Vince Watts, MD, Interim Director of VA Office of System Redesign and Improvement (10E2F). Key questions were then developed with input from the topic nominator, the ESP Coordinating Center, the review team, and the technical expert panel (TEP):

**Key Question 1A:** What is the comparative effectiveness of implementing continuous quality improvement frameworks in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?

**Key Question 1B:** What is the effectiveness of implementing a continuous quality improvement framework in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?

**Key Question 2:** What factors (including intervention characteristics, inner setting, outer setting, individuals involved, and process by which implementation is accomplished) contribute to the success or failure of these continuous quality improvement frameworks?

To be eligible for Key Question 1A, a systematic review had to explicitly focus on comparative effectiveness of multiple CQI methodologies as a stated aim. If a review commented on multiple methodologies but did not seek to compare the effectiveness of the strategies within its methods, it was included in Key Question 1B.

The review was registered in PROSPERO: CRD42021245263.

## SEARCH STRATEGY

Our team, which included a medical librarian, developed and conducted broad systematic review searches using terms relating to "quality improvement" or "continuous quality improvement" or "system redesign" in 4 databases: PubMed, CINAHL, DARE, and Cochrane. Search dates for PubMed are from 01/01/2010 through 03/18/2021. Search dates for CINAHL and Cochrane are from 01/01/2010 through 03/30/2021. Search dates for DARE are from 01/01/2010 to 03/31/2015. We restricted our searches to English language publications. See Appendix A for full search strategy.

## STUDY SELECTION

As multiple reviews on individual or subsets of CQI methodologies have been performed previously, this study was designed as a review of these pre-existing reviews. Three team members working independently screened all titles for relevance; any article chosen by any reviewer was included in the abstract screen. Abstracts were then reviewed in duplicate with any discrepancies resolved by group discussion. Full-text review was conducted independently by team members working in pairs, with any disagreements resolved through discussion. In order to be included, a review had to be a systematic review and include CQI as an intervention within any health care setting. An intervention was deemed to be CQI if it was either explicitly referred to as CQI or comprised of 3 essential features of CQI methods: "systematic data guided

activities", "designing with local conditions in mind", and "iterative development and testing."[2] These strategies included Lean, Six Sigma, Lean Six Sigma, Quality Improvement Collaboratives, CQI (hereafter referred to as the 'CQI method' to distinguish from general CQI), Total Quality Management (TQM), Clinical Microsystems, and the Plan-Do-Check-Act (PDCA) or Plan-Do-Study-Act (PDSA) method. Thus, we rejected narrative reviews, scoping reviews, and publications that only reported components of quality improvement (QI) without a full QI framework. Specific definitions for each strategy were not included in our search strategy; rather, we allowed each systematic review to apply its own definitions and labels to the included studies. See Appendix B for the full-text review form.

## DATA ABSTRACTION

Each included systematic review had data abstracted by 2 reviewers independently. Abstracted data included: CQI framework/strategy discussed, whether article described context/factors contributing to the success or failure of the framework/strategy, total number of studies included, search dates, health care condition, health care setting, and geographical region. Any discrepancies in data abstraction were resolved by group discussion.

We used the Consolidated Framework for Implementation Research (CFIR) to guide abstraction and synthesis of Key Question 2 around the following domains: intervention characteristics, inner setting, outer setting, individuals involved, and process by which implementation is accomplished. All studies discussed in Key Question 2 were assessed for inclusion in duplicate, with discrepancies resolved with group discussion.

## QUALITY ASSESSMENT

Each systematic review was assessed using a modified version of the Assessing the Methodological Quality of Systematic Reviews 2 (AMSTAR2) criteria.[4] This 16-item tool was designed to assess the methodological quality of systematic reviews and meta-analysis and is considered the standard for this type of assessment. The tool includes domains such as descriptions of PICO (population, intervention, comparators, and outcomes) in both inclusion criteria and results, PROSPERO registration, use of a comprehensive and timely literature search strategy, duplicate data abstraction, and analysis of risk of bias. As some AMSTAR2 items concern meta-analysis not applicable to our set of studies, we adapted the tool for this review, resulting in a 13-item tool. The criteria in our modified tool are shown in the table below (Table 1), while the full modified tool is available in Appendix C. One item — "If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?" — was found not to be applicable in all included reviews and therefore is excluded from Tables 1 and 2.

Assessment of studies using our modified tool was completed in duplicate, with discrepancies resolved with group discussion. No study was excluded from analysis based on AMSTAR2 score; however, we chose a score of greater than or equal to 8 to represent higher-quality systematic reviews. Reviews with a score of 7 or lower had the potential for multiple methodological flaws (given that they either did not do or did not report doing some key methodological practices), which influenced our interpretation of the completeness and rigor of their findings and conclusions. This cut-off was used in formulating certainty of evidence statements, as described below.

**Table 1. Modified AMSTAR2 Questions**

| | |
|---|---|
| 1. | Did the research questions and inclusion criteria for the review include a) population/setting; b) intervention; c) comparator(s); d) outcomes? |
| 2. | Did the report of the review reference a protocol or PROSPERO registration? |
| 3. | Did the review authors explain their selection of other study designs (non-RCTs) for inclusion in the review? |
| 4. | Did the review authors use a comprehensive literature search strategy? |
| 5. | Did the review authors perform study selection in duplicate? |
| 6. | Did the review authors perform data extraction in duplicate? |
| 7. | Did the review authors provide a list of excluded studies and/or justify the exclusions? |
| 8. | Did the review authors describe the following components in individual included studies in adequate detail: a) populations/settings; b) interventions; c) comparators(s); d) outcomes; e) research designs? |
| 9. | Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? |
| 10. | Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review? |
| 11. | Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review? |
| 12. | Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review? |

Reviews reporting results relevant to our key questions that used an established method of synthesis other than traditional systematic review methods, such as comprehensive reviews and realist reviews, were not assessed with AMSTAR2, given that AMSTAR2 was developed for systematic review and meta-analysis methods.

While quality criteria specifically for quality improvement studies have been developed, like the Quality Improvement Minimum Quality Criteria Set,[5] they are designed for primary research articles, and as such are not appropriate for use with the systematic reviews this report discusses. Nonetheless, we did take into account the Quality Improvement Minimum Quality Criteria Set and our discussion of Key Question 2 is informed by these criteria.

## DATA SYNTHESIS AND ANALYSIS

Our review is a narrative analysis. For KQ1A and KQ1B we narratively synthesized the evidence from relevant included systematic reviews based on CQI frameworks and outcomes reported. For KQ2, we conducted a template analysis using the included systematic reviews as our source material.[6] We looked at factors affecting the success or failure of CQI frameworks according to the 5 main Consolidated Framework for Implementation Research (CFIR) domains.[7] We used the CFIR domains as the basis for our coding structure, capturing any relevant language in the

included systematic reviews that described 1) intervention characteristics, 2) inner setting, 3) outer setting, 4) individuals involved, and/or 5) process by which implementation is accomplished.[7]

Certainty of evidence was determined by use of overall AMSTAR2 scores. We categorized evidence into low, moderate, and high certainty by the frequency with which included reviews for each key question had an AMSTAR2 score greater than or equal to 8. Key questions with fewer than 1/3 of studies having a score ≥ 8 were categorized as low certainty of evidence, those with between 1/3 and 2/3 of studies with scores ≥ 8 as moderate certainty of evidence, and those with > 2/3 of studies with scores ≥ 8 as high certainty of evidence.

## PEER REVIEW

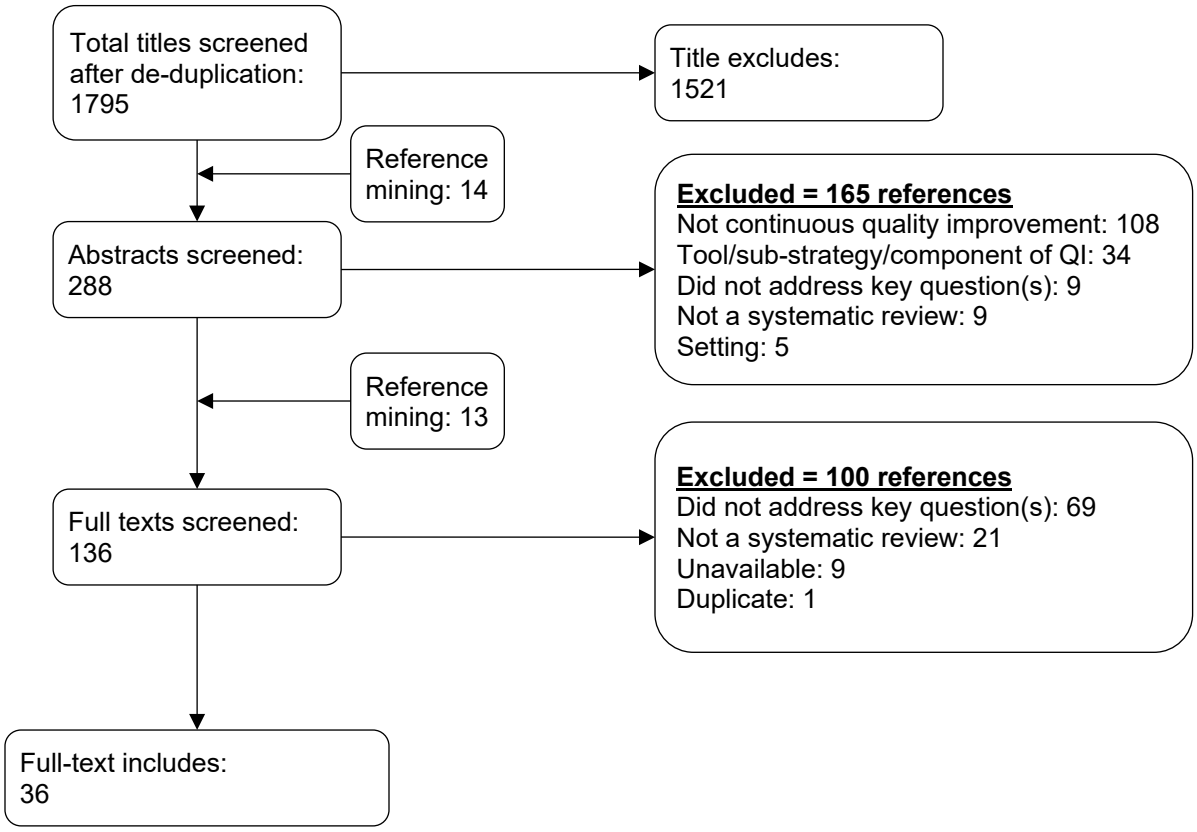
A draft version of the report was reviewed by technical experts and clinical leadership. While Technical Panel Experts are often asked to also serve as peer reviewers, we also invite experts who have not been involved with the current project to serve as peer reviewers. Reviewer comments and our responses are documented in Appendix D.

# RESULTS

## LITERATURE FLOW

The literature search identified 1,795 citations relevant to Key Question 1A/1B and Key Question 2 (Figure 1). After applying the inclusion and exclusion criteria to these 1,795 titles and adding 14 titles obtained from reference mining (*ie*, from citations included in previously identified literature), a total of 288 abstracts were reviewed at abstract stage. From these, a total of 165 abstracts were excluded for the following reasons: not about CQI (n=108), tool/sub-strategy/component of quality improvement (n=34), not a systematic review (n=9), and did not address key question(s) (n=9). After reference mining the cited literature in our screened full-text articles, we identified an additional 13 titles to be reviewed at the full-text stage, resulting in a total of 136 publications. From these, 100 publications were excluded for the following reasons: did not address key question(s) (n=69), not a systematic review (n=21), unavailable (n=9), and duplicate (n=1). A full list of excluded studies from the full-text review is included in Appendix F. Thirty-six studies were retained for abstraction after full-text review.

### Figure 1: Literature Flow Chart




### Quality of Included Systematic Reviews

Of the 36 included reviews, 29 reviews used traditional systematic review methodology and were assessed using the modified AMSTAR2 tool (see Table 1 for the modified AMSTAR2 questions and Appendix C for the full tool). Figure 2 presents the distribution of AMSTAR2 scores for the 29 reviews. The highest score was 11 points out of a possible 12 points, while the lowest score

was 2 points. The median and mode scores are both 5 points, a score which means that just less than half of the quality criteria were met.

**Figure 2. Distribution of AMSTAR2 Quality Scores**

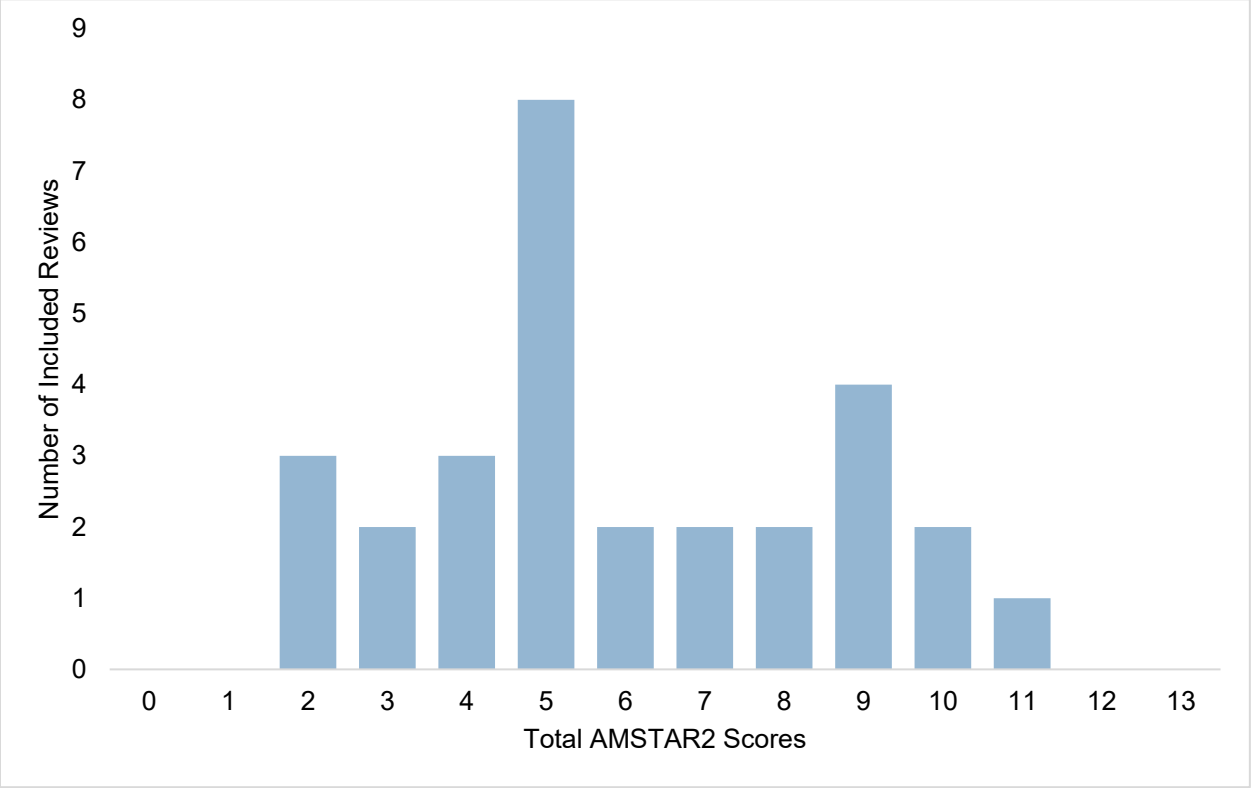



Table 2 shows the breakdown of quality ratings by criterion for each study. Whether the review included a comprehensive search strategy (Criterion 4) was the criterion the AMSTAR2-scored reviews met most frequently, with 22 reviews (76% of the 29 reviews scored by AMSTAR2) meeting this criterion. Criterion 12 was the next most frequently met criterion, which asked whether the review reported conflict of interest, and was met by 21 reviews (72% of 29 reviews with AMSTAR2 scores). All other criteria were met by less than half of the 29 AMSTAR2-scored reviews. The criterion that was met least frequently was Criterion 2, with 4 reviews (14% of 29 AMSTAR2 scored reviews) reporting on their use of a protocol or PROSPERO registration. Two criteria were composite scores with multiple sub-criteria: questions 1 and 8. These are described further below.

## Table 2: Total AMSTAR2 Scores, by Criterion

| Author, Year | 1–Research Question / Inclusion Criteria | 2—Protocol or PROSPERO | 3—Explain selection of Study Design | 4—Use of Search Strategy | 5—Study Selection in Duplicate | 6—Data Extraction in Duplicate | 7—List of Excludes | 8—Description of PICO-Research Design | 9—Assesses Risk of Bias | 10—Account for Risk of Bias in Interpretation | 11—Explanation for Heterogeneity | 12—Report Conflict of Interest | TOTAL SCORE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aij, 2017[8] | X* | | | X | | | | | | | | | 2 |
| Amaratunga, 2016[9] | X* | | X | X | X | X | X | X* | | | | X | 8 |
| Bucci, 2016[10] | X* | | | X | | | X | X* | X | | | | 5 |
| Côté, 2020[11] | X* | | | X | | X | X | X* | X | X | X | X | 9 |
| Dellifraine, 2010[12] | X* | | | X | | | | X* | X | X | | | 5 |
| DelliFraine, 2013[13] | X* | | | X | | | | X* | | | X | X | 5 |
| Dzidowska, 2020[14] | X* | X | X | X | X | X | | X* | | | X | X | 9 |
| Glasgow, 2010[15] | X* | | X | X | | | | X* | | | | X | 5 |
| Hill, 2020[16] | X* | X | X | X | X | X | | X* | X | X | X | X | 11 |
| Hulscher, 2013[17] | X* | | | | X | X | | X* | | | X | X | 6 |
| Isfahani, 2019[18] | X* | | | X | | | | X* | | | | X | 4 |
| Leggat, 2015[19] | X* | | | X | | | | | | | | X | 3 |
| Magalhães, 2016[20] | | | X | X | | | | X* | | | | | 3 |
| Mason, 2015[21] | X* | | X | X | | | X | X* | X | X | | | 7 |
| Moraros, 2015[22] | X* | | | | X | | | X* | X | | | X | 5 |
| Nadeem, 2013[23] | X* | | X | X | X | | | X* | | | | X | 6 |
| Nicolay, 2012[24] | X* | | | X | X | X | X | X* | X | X | X | X | 10 |
| Nunes, 2016[25] | X* | | | X | X | | X | | | | | X | 5 |
| Powell, 2008[1] | X* | | | | | | | | | | | X | 2 |
| Schouten, 2008[26] | X* | | | X | X | X | X | X | X | X | X | X | 10 |
| Talib, 2011[27] | X* | | | | | | | X* | | | | | 2 |

| Author, Year | 1–Research Question / Inclusion Criteria | 2—Protocol or PROSPERO | 3—Explain selection of Study Design | 4—Use of Search Strategy | 5—Study Selection in Duplicate | 6—Data Extraction in Duplicate | 7—List of Excludes | 8—Description of PICO-Research Design | 9—Assesses Risk of Bias | 10—Account for Risk of Bias in Interpretation | 11—Explanation for Heterogeneity | 12—Report Conflict of Interest | TOTAL SCORE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Taylor, 2014 [28] | X* | | | X | X | X | | | | | | X | 5 |
| Tlapa, 2020[29] | X* | X | X | X | X | | | X* | X | X | | | 8 |
| Trakulsunti, 2018[30] | X* | | | | X | X | | X* | | | | X | 5 |
| Tricco, 2012[31] | X* | | | X | X | X | X | X* | X | | X | X | 9 |
| Wackerbarth, 2021[32] | X* | | | | | X | X | | | | | X | 4 |
| Woodnutt, 2018[33] | X* | | | | | | X | X* | X | | | | 4 |
| Zamboni, 2020[34] | X* | | X | X | | | X | X* | X | | | X | 7 |
| Zepeda-Lugo, 2020[35] | X* | X | X | X | X | | | X* | X | X | | X | 9 |
| **Percent included reviews meeting criteria** | | **14%** | **34%** | **76%** | **48%** | **38%** | **38%** | | **45%** | **28%** | **28%** | **72%** | |

*Review noted as meeting this criterion; however, this criterion is only partially met, based on multiple sub-questions. See tables 3 and 4 below for detailed scoring on these 2 criteria.

The criteria for question 1, "Did the research questions and inclusion criteria for the review include a) population/setting; b) intervention; c) comparator(s); d) outcomes?", was scored with sub-criteria for each of the 4 specific components of the question: population/setting, intervention, comparator(s), and outcomes. As such, the 29 reviews scored using AMSTAR2 could receive partial credit for up to 4 components: 2 reviews reported on all 4 sub-criteria, 13 reviews reported on 3 sub-criteria, 8 reviews reported on 2 sub-criteria, and 5 reviews reported on 1 sub-criteria. One review did not report on any sub-criteria. AMSTAR2-scored reviews most often reported population/setting and intervention in their research questions or inclusion criteria (n= 23, 79%). Comparators were reported as a part of the research questions or inclusion criteria in 5 reviews (17%).

**Table 3: AMSTAR2 Scoring for Question 1 Describing Reviews Reporting Population/Setting, Intervention, Comparator(s), and/or Outcomes in the Resarch Questions and Inclusion Criteria**

| Author, Year | 1A— Population/ Setting | 1B— Intervention | 1C— Comparator(s) | 1D— Outcomes | Sub-criteria met |
|---|---|---|---|---|---|
| Côté, 2020[11] | X | X | X | X | 4 |
| Hill, 2020[16] | X | X | X | X | 4 |
| Amaratunga, 2016[9] | X | | X | X | 3 |
| Bucci, 2016[10] | X | X | | X | 3 |
| Dellifraine, 2010[12] | X | X | | X | 3 |
| Dzidowska, 2020[14] | X | X | | X | 3 |
| Hulscher, 2013[17] | X | X | | X | 3 |
| Leggat, 2015[19] | X | X | | X | 3 |
| Nunes, 2016[25] | X | X | | X | 3 |
| Schouten, 2008[26] | X | X | | X | 3 |
| Tlapa, 2020[29] | X | X | | X | 3 |
| Trakulsunti, 2018[30] | X | X | | X | 3 |
| Tricco, 2012[31] | X | X | | X | 3 |
| Zepeda-Lugo, 2020[35] | X | X | | X | 3 |
| Zamboni, 2020[34] | X | X | | X | 3 |
| DelliFraine, 2013[13] | X | X | | | 2 |
| Glasgow, 2010[15] | X | X | | | 2 |
| Isfahani, 2019[18] | X | X | | | 2 |
| Mason, 2015[21] | X | | X | | 2 |
| Nicolay, 2012[24] | X | | X | | 2 |
| Powell, 2008[1] | X | X | | | 2 |
| Talib, 2011[27] | X | | | X | 2 |
| Wackerbarth, 2021[32] | X | | | X | 2 |
| Aij, 2017[8] | | X | | | 1 |
| Moraros, 2015[22] | | X | | | 1 |

| Author, Year | 1A—Population/ Setting | 1B—Intervention | 1C—Comparator(s) | 1D—Outcomes | Sub-criteria met |
|---|---|---|---|---|---|
| Nadeem, 2013[23] | | X | | | 1 |
| Taylor, 2014 [28] | | X | | | 1 |
| Woodnutt, 2018[33] | | X | | | 1 |
| Magalhães, 2016[20] | | | | | 0 |
| Percent included reviews meeting this sub-criterion | 79% | 79% | 17% | 59% | |

The criteria for question 8, "Did the review authors describe the following components in individual included studies in adequate detail: a) populations/settings; b) interventions; c) comparators(s); d) outcomes; e) research designs?", was scored with sub-criteria for each of the 4 specific components of the question: population/setting, intervention, comparator(s), and outcomes. As such, the 29 reviews scored using AMSTAR2 could receive partial credit for up to 4 components: 2 reviews reported on all 5 sub-criteria, 5 reviews reported on 4 sub-criteria, 9 reviews reported on 3 sub-criteria, 5 reviews reported on 2 sub-criteria, and 2 reviews reported on 1 sub-criteria. Six reviews did not report on any sub-criteria. AMSTAR2-scored reviews most often reported population/setting (n=18, 62%) and intervention (n=14, 48%) in their research questions or inclusion criteria. Comparators were reported as a part of the research questions or inclusion criteria in 2 reviews (7%).

### Table 4: AMSTAR2 Scoring for Question 8 Describing Reviews Reporting Population/Setting, Intervention, Comparator(s), Outcomes, and/or Research Designs in the Results

| Author, Year | 8A—Populations/ settings | 8B—Interventions | 8C—Comparators | 8D—Outcomes | 8E—Research Designs | Sub-criteria met |
|---|---|---|---|---|---|---|
| Hill, 2020[16] | X | X | X | X | X | 5 |
| Schouten, 2008[26] | X | X | X | X | X | 5 |
| Bucci, 2016[10] | X | X | | X | X | 4 |
| Moraros, 2015[22] | X | X | | X | X | 4 |
| Tricco, 2012[31] | X | X | | X | X | 4 |
| Woodnutt, 2018[33] | X | X | | X | X | 4 |
| Zamboni, 2020[34] | X | X | | X | X | 4 |
| Amaratunga, 2016[9] | X | | | X | X | 3 |
| Dellifraine, 2010[12] | X | | | X | X | 3 |
| Dzidowska, 2020[14] | X | X | | X | | 3 |
| Hulscher, 2013[17] | X | X | | | X | 3 |
| Mason, 2015[21] | | X | | X | X | 3 |
| Nicolay, 2012[24] | X | | | X | X | 3 |
| Talib, 2011[27] | X | X | | | X | 3 |
| Tlapa, 2020[29] | X | | | X | X | 3 |

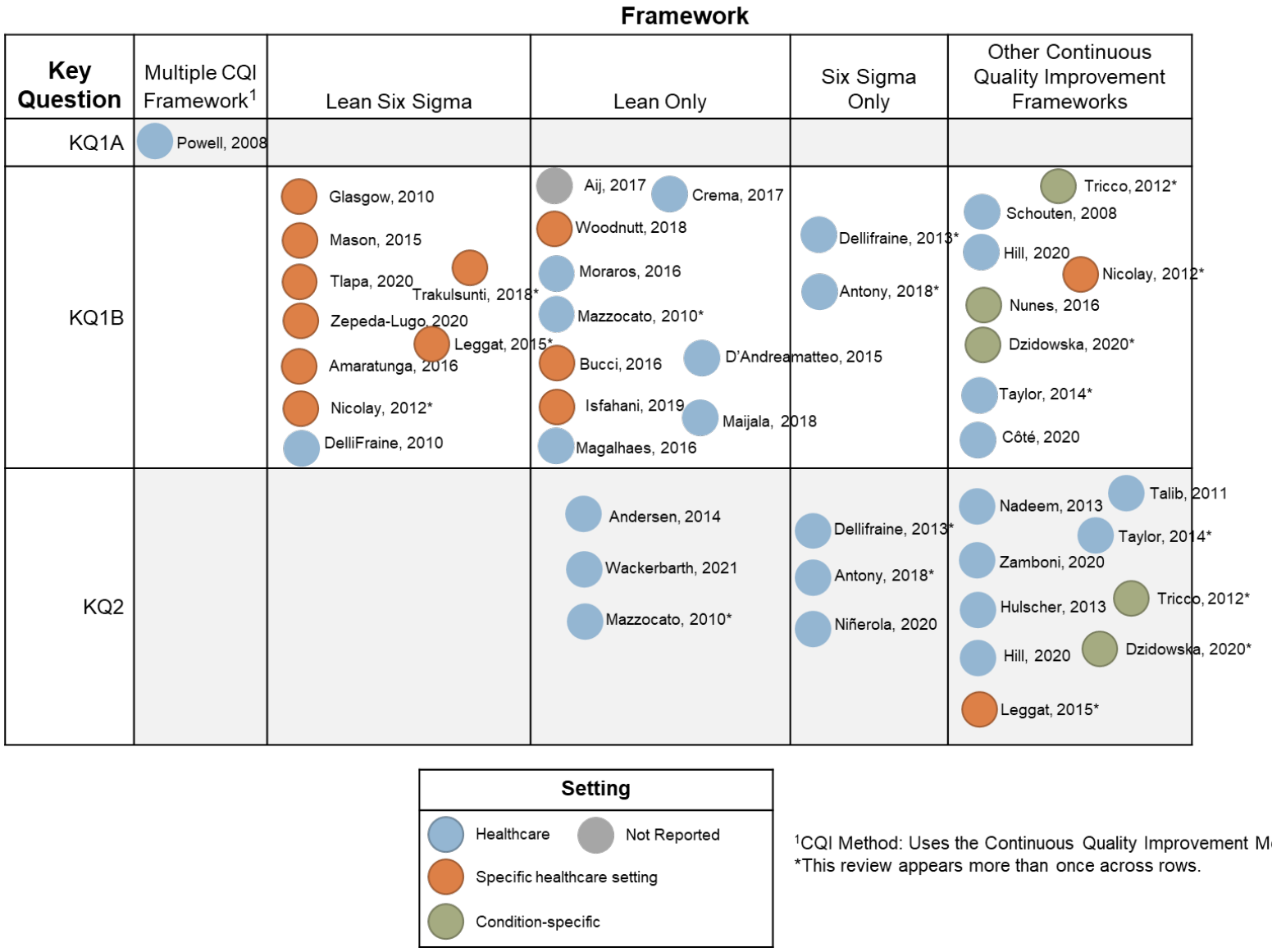
| Author, Year | 8A—Populations/ settings | 8B—Interventions | 8C—Comparators | 8D—Outcomes | 8E—Research Designs | Sub-criteria met |
|---|---|---|---|---|---|---|
| Zepeda-Lugo, 2020[35] | X | | | X | X | 3 |
| Côté, 2020[11] | X | | | X | | 2 |
| DelliFraine, 2013[13] | | X | | X | | 2 |
| Glasgow, 2010[15] | X | | | | X | 2 |
| Isfahani, 2019[18] | | | | X | X | 2 |
| Nadeem, 2013[23] | | X | | | X | 2 |
| Magalhães, 2016[20] | X | | | | | 1 |
| Trakulsunti, 2018[30] | | X | | | | 1 |
| Aij, 2017[8] | | | | | | 0 |
| Leggat, 2015[19] | | | | | | 0 |
| Nunes, 2016[25] | | | | | | 0 |
| Powell, 2008[1] | | | | | | 0 |
| Taylor, 2014 [28] | | | | | | 0 |
| Wackerbarth, 2021[32] | | | | | | 0 |
| Percent included reviews meeting this sub-criterion | 62% | 48% | 7% | 59% | 62% | |

The Map of Included Reviews by Key Questions illustrates each included review, by key question, and their respective frameworks and settings (Figure 3). Reviews could appear more than once in a column if they addressed multiple Key Questions.

For Key Question 1A, there was 1 review that had described comparative effectiveness multiple CQI frameworks (*eg*, Total Quality Management, PDSA, the CQI method, *etc*) in a non-specific health care setting. For Key Question 1B, there were 9 reviews focused on Lean/Six Sigma—8 of these were in specific healthcare settings (*eg*, ambulatory & ED, radiology, *etc*) and 1 in a non-specific health care setting. Seven reviews reported Lean-only interventions for Key Question 1B, with 3 in specific health care settings (*eg*, ED, hospital) and 4 in a non-specific health care setting. Two reviews focused on Six Sigma only for Key Question 1B, both in a non-specific health care setting. Lastly, 8 reviews addressing Key Question 1B focused on other CQI frameworks (*eg*, Quality Improvement Collaboratives, Total Quality Management, clinical microsystems approach, *etc*): 4 reviews in general healthcare settings, 1 review in a hospital setting, and 3 reviews in a condition-specific setting (*eg*, kidney-disease treatment, *etc*).

Three reviews focused on Lean only in Key Question 2 were all in a non-specific health care setting. Two reviews focused on Six Sigma only, again in a non-specific health care setting. Nine studies addressing Key Question 2 focused on other CQI frameworks: 6 reviews in general health care settings, 2 reviews in a condition-specific setting (*eg*, alcohol misuse, diabetes, *etc*), and 1 review in a primary care setting. Reviews may appear more than once across Key Questions.

## Figure 3. Map of Included Reviews, by Key Questions




[1]CQI Method: Uses the Continuous Quality Improvement Method
*This review appears more than once across rows.

The included systematic reviews reported on studies conducted in multiple countries, primarily in North America and Europe. The number of included studies within each review ranged from 9 studies to 295 studies (Figure 4), with search end dates from 2006 to 2019 (Figure 5).

**Figure 4: Number of Included Studies, by Reviews**

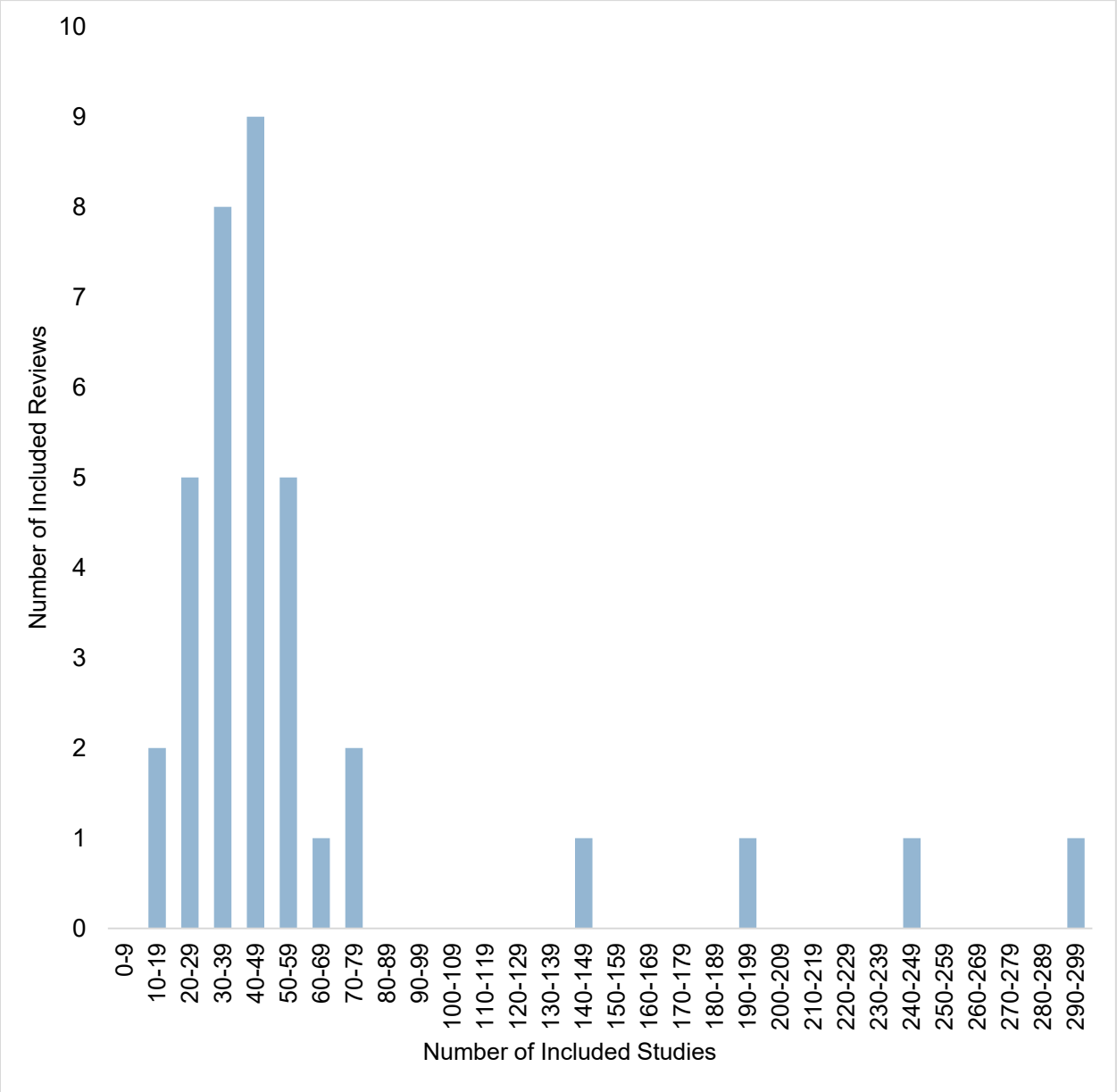
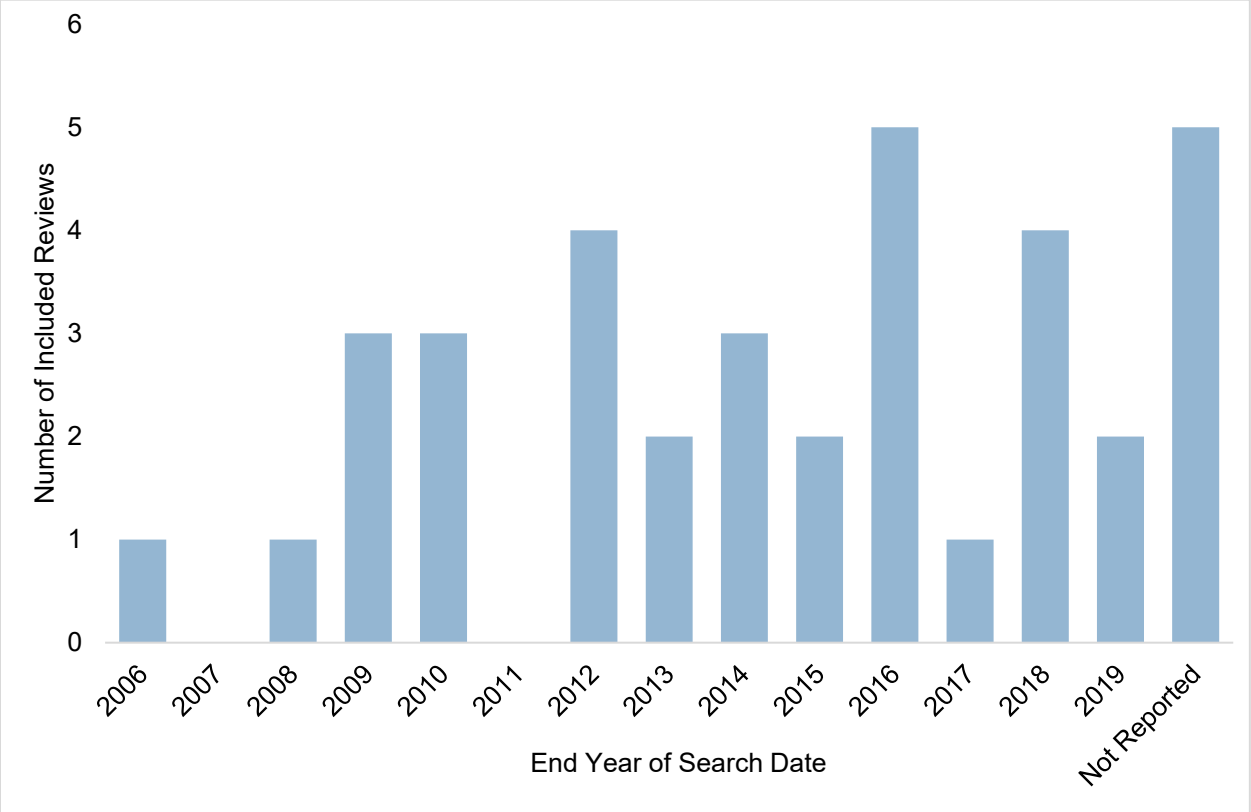


**Figure 5: Number of Reviews, by End Year of Search Date**



## KEY QUESTION 1A: What is the comparative effectiveness of implementing continuous quality improvement frameworks in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?

### Key Points (Summary of Findings)

- Only 1 review of low certainty of evidence from 2008 directly assessed the relative effectiveness of CQI strategies, finding significant overlap in defining and implementing different approaches to CQI.
- The authors concluded that local context should dictate choice of CQI methodology, finding no evidence classifying any strategy as more or most effective.
- The authors identified 7 "necessary, but not sufficient" conditions for successful implementation of any CQI strategy: provision of the practical and human resources to enable quality improvement; active engagement of health professionals, especially doctors; sustained managerial focus and attention; use of multi-faceted interventions; coordinated action at all levels of the health care system; substantial investment in training and development; availability of robust and timely data through supported information technology systems.

### Detailed Findings

#### *Comparative effectiveness of CQI frameworks*

*Multiple CQI frameworks*

Only 1 review[1] by Powell from 2008 identified via reference mining directly assessed the relative effectiveness of a broad number of different QI strategies. In this report performed for NHS Scotland, the authors reviewed 5 organizational-level approaches for quality improvement: Total Quality Management, the CQI method, Lean and Six Sigma, business process reengineering, the IHI's rapid cycle change (Model for Improvement),. The review included case-based analyses of each model, including 1 of the VA QUERI initiative in the United States. The authors note the studied approaches have been variously adopted in different health care settings, but standardization in adoption and implementation has been lacking. Specifically, they state, "there is little uniformity in nomenclature or in the content of programmes, and many organisations have used a combination of tools and approaches eclectically and variably over time". Practically, this has meant that even when a model has been chosen for utilization across a health care system, each model "proceeds by carrying out several secondary activities" that may be outside their historical primary focus. For example, an organization implementing a Lean model may include a focus on human factors in addition to process analysis. Thus, organizations practicing different methodologies often employ the same or similar tools, and most models and organizations focus on similar measurement and data collection strategies. The end result of this evolution is that, from an outside perspective, "for all their differences – the approaches do begin to resemble each other".

Given the variation in implementation of each model and the overlap in tools and practices, the authors' main findings identify "limited evidence available to assess how effective these

approaches are in health care". However, they do note broad lessons may be drawn from these experiences for successful adoption of QI strategies in a variety of settings, including a set of "necessary, but not sufficient" conditions. These conditions include

- "provision of the practical and human resources to enable quality improvement;
- the active engagement of health professionals, especially doctors;
- sustained managerial focus and attention;
- the use of multi-faceted interventions;
- coordinated action at all levels of the health care system;
- substantial investment in training and development;
- and the availability of robust and timely data through supported IT systems."

Ultimately, they conclude, "[i]mportantly, there is no one right method or approach that emerges above the others as the most effective". They then suggest analyzing the local context prior to implementation to determine which strategy "provides the 'best fit' locally (however imperfect)", which then must be applied "in a programmed and sustained way, which may include considerable adaptation of the approach to suit the local circumstances and to respond to emerging developments".

The remainder of the reviews we identified did not assess relative effectiveness of different QI strategies. These reviews dealt with only a single strategy, like Lean or Six Sigma, or a Lean/Six Sigma hybrid strategy and do not provide evidence of differential effectiveness other than among Lean and Six Sigma.

### Certainty of Evidence for Key Question 1A

This systematic review by Powell[1] received a score of 2 on our modified AMSTAR2 tool for inclusion of characteristics of methodologies in the inclusion criteria and reporting of funding source. Details such as a fully defined search strategy, article review, data abstraction, and clear reporting on the characteristics of included studies for the review were not identified. The evidence from this study was, therefore, deemed of low overall certainty.

## KEY QUESTION 1B: What is the effectiveness of implementing a continuous quality improvement framework in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?

### Key Points (Summary of Findings)

- Twenty-five reviews studied at least 1 CQI framework, with 11 studying more than one. These reviews identified successful implementation of CQI frameworks in a variety of clinical settings.
- None of the 11 reviews that included more than 1 CQI strategy reached a strong conclusion that any strategy was superior to any other(s).

- Seven publications commented on sustainment of change using CQI strategies, none of which concluded that any specific methodology had definitive evidence for sustainment of change.
- None of the included reviews concluded any CQI methodology had evidence speaking to health care workers' reaction to use of that general framework, specific learning outcomes related to a CQI framework, nor identifiable behavioral changes made based on CQI training. Five studies discussed clinician/provider satisfaction as an outcome of implementation of a CQI methodology, with mixed results.

## Detailed Findings

### *Effectiveness of CQI frameworks*

*Lean, Six Sigma, and/or Lean Six Sigma*

Overall, 21 reviews studied the effectiveness of Lean, Six Sigma, and/or Lean Six Sigma frameworks in health care. Of these, 10 publications reviewed only Lean strategies,[8,10,18,20,22,33,36-39] 2 only Six Sigma strategies,[13,40] and 9 a combination of the three strategies[9,12,15,19,21,24,29,30,35]. One review of the 3 strategies also reviewed other frameworks addressed in the 'Other CQI Frameworks' section and is discussed in both sections.[24] None of the reviews that assessed some combination of the QI strategies reached a strong conclusion that one strategy was superior to the others. Of the 9 reviews of a combination of the 3 strategies, 6 included a discussion of differential effectiveness.[15,21,24,29,30,35] Of these 6 reviews, 4 explicitly stated they could not conclude 1 strategy was superior to the others,[15,21,24,30] while the remaining two[29,35] concluded without supporting comparative data that a combination of Lean and Six Sigma outperformed either one by itself. Several other articles addressing aspects other than effectiveness of these strategies are discussed elsewhere in this document under Key Question 2.

Each strategy has been studied in a variety of clinical settings. Collectively, these include outpatient medical and surgical clinics, emergency care settings, inpatient wards, operating theatres, radiology units, hospital and UK National Health Service trust pharmacies, and entire hospital or UK National Health Service trust systems. Often, more than 1 strategy has been used in the same general clinical setting in different locations (such as both Lean and Six Sigma used in increasing emergency department patient volume throughput, improving outpatient waiting times, or improving Methicillin-resistant Staphylococcus aureus [MRSA] infection rates in different studies), though they have not been directly compared in any study.

Six reviews directly commented on when to utilize 1 of the 3 strategies or improvement systems.[15,21,24,29,30,35] Glasgow[15] reviewed 47 studies and noted, "Lean and Six Sigma have been effectively applied to all components, from admission to discharge, of a patient's hospital experience." Additionally, the authors identify, "there is sufficient flexibility [in these methods] to improving quality in the acute care setting". However, given the variability in how these methods were defined, implemented, and reported across the included literature, they conclude "there is not sufficient evidence to recommend broad adoption of Lean, Six Sigma, or Lean Sigma".

Two reviews by the same author group looked at Lean strategies in the inpatient setting[35] and Lean strategies generally related to "patient flow"[29] across multiple settings. The former review

identified 40 studies, while the latter identified 39, with 3 overlapping studies. Both reviews found improvements in multiple outcomes with Lean strategies, including 6 of 7 outcomes in the inpatient setting and shorter length of stay and reduced outpatient waiting times in the work focusing on patient flow. Both papers suggested combining Lean and Six Sigma methodologies into Lean Six Sigma outperformed both Lean and Six Sigma, although these statements seem based on the frequency of reporting included in their review. In neither case were direct comparisons offered as evidence of the superiority of Lean Six Sigma. Specifically, Tlapa noted Lean Six Sigma "outperforms the use of only one other methodology; however, this combination tends to be composed of larger, private hospitals with more resources for quality improvement". Zepeda-Lugo identified "25 out of 39 studies combining lean with tools and principles of the six sigma methodology, suggesting that such integration offers a more robust approach to improving speed, quality and costs, increasing customer satisfaction, and maximizing shareholder value".

Similarly, the Nicolay[24] and Mason[21] reviews had some overlapping authors and shared 10 references for Lean, Six Sigma, and Lean Six Sigma involving the operating theatre. In the later work,[21] Mason also included additional 11 Lean, Six Sigma, and Lean Six Sigma works from surgical outpatient clinics and wards, while Nicolay also reviewed other methodologies such as the CQI method and PDCA in the operating room that are discussed later in this report. The Mason review identified a "role for Lean and Six Sigma QI methodologies within surgery, with significant improvements demonstrated across a variety of outcomes within the pre-operative, operative and in-patient settings". These outcomes included outpatient clinic efficiency, operating room turnaround time, use of prophylactic perioperative antibiotics, glycemic control postoperatively while hospitalized, and nosocomial infections. Nicolay further detailed, "these methodologies have been applied successfully to many different aspects of care, but in particular those that are repetitive and can be standardized". Mason highlighted the flexibility of both methods to meet "unique challenges in a particular place at a particular time". In both reviews, the authors were unable to make recommendations favoring 1 strategy over the others, with Mason concluding, "This review is unable to make formal recommendations on the use of Lean and Six Sigma methodologies in improving specified outcomes in surgical practice" and Nicolay, "it is not possible to make evidence-based recommendations for different indications, as different studies implemented different aspects of various methodologies to varying extents, and in different contexts".

Trakulsunti[30] et al identified 24 articles across the 3 methodologies focused on reducing medication errors. In this review, 22 Lean and 22 Six Sigma tools were "used across the medication process including prescribing, transcribing, dispensing and administration". The tools were generally reported to be successful, with their results determining "Lean, Six Sigma and Lean Six Sigma implementation can reduce errors in the medication delivery process". In specifying tools by framework, the authors conclude that "very few studies use pure Lean, Six Sigma and Lean Six Sigma to reduce medication errors". Similar to Tlapa[29] and Zepeda-Lugo,[35] Trakulsunti suggests, "the integration of Lean and Six Sigma may lead to better results", specifically as "practitioners can use tools from both philosophies". However, this statement was not based on comparative data, as they note, "the current literature does not provide a Lean, Six Sigma or Lean Six Sigma road map for practitioners to follow in order to reduce medication errors in their hospitals".

Three other publications[9,12,19] reviewed more than 1 of the 3 methodologies but did not comment on utilizing one approach over another. Amaratunga[9] reviewed 23 articles across the 3

methodologies involving Radiology modalities and units. The authors identified 7 different themes to which the strategies had been applied, including wait times, patient volumes, costs, and satisfaction and found "benefits can be derived through the application of Lean and Six Sigma methodologies within the field of radiology". In their appraisal, however, they determined the evidence was insufficient to determine the overall effectiveness of the methodologies, "consequently preventing one from recommending their widespread implementation" when considering possible upfront costs relative to other methodologies such as audit-and-feedback. Dellifraine[12] reviewed 34 articles across different health care settings, with 26 following a Six Sigma strategy and 20 included in a later review[13] on just Six Sigma methodologies. In this review, the individual strategy often was not identified during specific discussion points, making it difficult to draw detailed conclusions. However, the authors note, "the level of evidence supporting a positive relationship between the use of [Six Sigma and Lean] and performance improvement was weak" and therefore the limited evidence may only identify explicit potential applications. Finally, Leggatt[19] reviewed multiple strategies for process redesign in health care across 41 articles, including Lean and Six Sigma. The authors included 4 reviews, including those from Dellifraine[12] and Nicolay.[24] Of the 41 articles, 17 correlated practice changes with outcomes, although the structure of the article prevents identification of specific strategies for the stated improved outcomes.

*Lean only*

The 10 publications of only Lean strategies reviewed implementation in emergency settings,[10,18] general health care settings,[20,22,36,38,39] and within the UK National Health Service.[33] Two additional reviews did not report setting.[8,37] Four publications were not appropriate for use of the AMSTAR2 tool,[36-39] and none of the other 6 publications had AMSTAR2 ratings $\geq$8.

The first of the 2 reviews of the effectiveness of Lean in emergency settings by Bucci[10] identified 9 studies on patient flow and found specific improvements reported on outcomes that included increased patient volumes, decreased rates of length of stay and patients who left without being seen, cost reduction, and patient satisfaction. The Lean strategies adopted across these 9 studies were not standardized, though "[a]lmost all studies considered showed that Lean interventions contributed to the EDs performance improvement". Importantly, no study included by Bucci expanded their patient flow efforts to other health care settings, such as primary care, that have direct or indirect effects on emergency care volumes or access. The second review by Isfahani[18] was a descriptive overview of the literature and identified 26 articles documenting 23 studies. Very little was stated regarding effectiveness of Lean strategies, though the authors did identify an average duration of intervention of 10 months as reported across 16 of those studies.

Five publications reviewed the effectiveness of Lean in any health care setting, with 3 finding positive results in general categories of care, such as "productivity" and "clinical quality".[20,38,39] In the most recent of these articles, Magalhaes[20] identified 47 articles but did not provide specific examples or citations of improvement. Instead, the authors noted multiple general areas in which improvements have been documented, such as waiting times and improved teamwork. Similarly, Mazzocato[38] reviewed 33 articles and commented on effectiveness in general improvement categories such as "time-savings" and "several quality aspects including reduction in errors or mistakes" in their results with individual citations provided. Most of Mazzocato's work focused on a realist review of contextual aspects related to 4 change mechanisms more appropriately discussed under Key Question 2.

D'Andreamatteo[39] and Moraros[22] attempted to discuss Lean strategies in any health care setting at a more systemic level, with the former identifying 243 articles, more than 90% of which were noted to be from the acute care setting. The authors identified 7 studies addressing a systemic organizational Lean approach, stating, "[w]hen Lean was implemented within a plan of actions aimed to improve the whole organization performance, the organizations appeared to become more process-oriented, reduce costs and increase quality". However, as relatively few articles reviewed Lean at a high level, they also commented on studies with narrower scopes, finding "similar [positive] results were achieved with projects implemented in a single ward, in other specific units or addressing only one organizational process". Moraros,[22] on the other hand, identified 22 articles, 4 of which considered health outcomes, 3 a combination of health and process outcomes, and 15 process outcomes only. In reviewing the health outcomes, only 1 of the 4 studies found a statistically significant result. This finding was on reduced MRSA infection incidence, with another 1 of the 4 studies finding a null result on the same outcome, and the other 2 studies finding null results on adverse events and 30-day mortality. Similarly, only 2 studies on patient visits and surgical consults out of the 15 process outcome studies identified a statistically significant positive effect of Lean implementation. The overall conclusion by the authors was "[w]hile some may strongly believe that Lean interventions lead to quality improvements in healthcare, the evidence to date simply does not support this claim. It is far more likely that Lean is but one of many strategies that might or might not have an impact on healthcare delivery".

Woodnutt[33] reviewed Lean processes across the UK National Health Service, finding 12 articles with different methodologies, including 1 systematic review. Lean adoption differed between the studies, with 22 different outcomes measured. The most commonly studied outcome was waiting times, which had 5 positive results and 3 null results across 8 studies. In reviewing the literature across time, the authors noted a general "evolution from pragmatic (quasi-scientific and experiential) research to more academic and scientific designs". However, even in the UK National Health Service, the authors identified different definitions of value in Lean that may influence different implementation strategies and that there is "no blueprint to guide the introduction of QI in complex organisations". As such, they conclude "Lean has ostensible value but it is difficult to draw a conclusion on efficacy or sustainability".

*Six Sigma only*

The 2 publications of only Six Sigma strategies both reviewed implementation across multiple health care settings.[13,40] The Antony[40] study focuses mainly on success factors and challenges with Six Sigma, although it briefly reviews outcomes for which some effectiveness has been shown. These outcomes include 16 sub-categories, including the 5 most frequently reported of patient satisfaction, process speed (reduction of process cycle time), revenue enhancement, cost savings, and defect reduction. They conclude, "the most common benefits of Six Sigma implementation in healthcare are improvement in patient safety, improvement in process speed (i.e. increased productivity) and revenue enhancement (i.e. bottom-line savings)." Dellifraine[13] narrowed their earlier work from 2010[12] to only Six Sigma methodologies, including 20 studies from the earlier work and an additional 35 works for inclusion. Of the 55 total articles, only 30 identified each of the Define, Measure, Analyze, Improve, and Control (DMAIC) steps in full and only 16 reported a sigma level achieved. Most often, the 'Control' step was not included (20 of the 25 studies that did not report all steps). The studies included multiple outcomes, such as medication administration errors, operating room throughput, pain management, hospital hand hygiene, falls, ventilator-associated pneumonia, cost savings, and patient safety. Most articles

noted some improvement, though only 7 included statistical comparisons. The authors note, "articles that focused more narrowly on more targeted areas were able to present stronger in-depth evidence that Six Sigma can improve processes of care" and that a greater number of more narrowly focused studies reported statistical testing and/or sigma level achieved. They conclude, "the level of evidence supporting a positive relationship between the use of Six Sigma and statistically significant improvement in quality was weak." Additionally, the authors note the difficulty in achieving Six Sigma levels of standardization, as having no more than 1 defect per 3.4 million opportunities "may not be realistic or applicable to many QI issues faced by health care organizations". Even with these concerns, however, the authors state, "Six Sigma is a good tool to identify the key procedures or problems in the process of care."

*Other CQI frameworks*

In addition to the reviews focused on the use of Lean and/or Six Sigma strategies in health care, we also identified a number of reviews on other topics: 1 review on the "evidence for the impact of quality improvement collaboratives"[26]; 3 reviews on the effectiveness of the CQI method, 1 of which was the CQI method "for developing professional practice and improving health care outcomes",[16] another 1 of which was the use of the CQI method "in nephrology",[25] and the third of which was the role of the CQI method "to improve practice, detection and treatment of unhealthy alcohol use in primary health care"[14]; 1 review on the use of PDSA "to improve quality in health care"[28]; 1 review about the Clinical Microsystems (CMS) approach[11]; and 2 reviews that included many different QI strategies in specific clinical situations, 1 of which was the management of diabetes,[31] and other of which was about "surgical healthcare" (also discussed in the 'Lean, Six Sigma, and/or Lean Six Sigma' section for their results studying those frameworks).[24] We discuss these only briefly, as none had head-to-head comparisons of QI approaches and almost all had included either Lean or Six Sigma as part of their eligibility criteria.

These last 2 publications[24,31] reviewed all identified QI articles meeting eligibility criteria within their respective clinical foci. In the review by Tricco and colleagues on diabetes (search end date of 2010), 142 original studies (including 48 cluster randomized trials) were included.[31] Studies were characterized by the components of the intervention used, such as audit-and-feedback, clinician education, clinician reminders, etcetera. The CQI method was 1 such component, for which 4 studies were included. A meta-analysis and meta-regression analysis found that almost all QI components were effective at improving outcomes such as reduction in hemoglobin A1c and reduction in LDL, etcetera. However, only 1 of the CQI method studies entered into the meta-analysis, and none entered into the meta-regression analysis, meaning conclusions about use of the CQI method were not possible. The review of QI strategies in "surgical healthcare"[24] (search end date of 2010) included 4 original studies using PDSA, 3 studies using statistical process control, 9 studies using the CQI method, 5 studies using Total Quality Management, 5 studies using Six Sigma, 4 studies using Lean, and 1 study using Lean Six Sigma; however, there were no head-to-head studies comparing different QI methods. All studies improved some aspect of surgical health care (such as infection control, operating room turnaround time, or use of pre-operative beta-blockade, *etc*).

The 3 reviews focused on the CQI method[14,16,25] found mixed results. In the review about "improving health care outcomes"[16] (search end date of 2019), 28 RCTs were identified that assessed the effectiveness of the CQI method. Over half did not report statistically significant

improvements on clinical, patient, or other outcomes. In the review focused on the CQI method "to improve practice, detection, and treatment of unhealth alcohol use"[14] (search end date of 2018), 56 publications met eligibility criteria, and more than 90% of these reported some improvement. However, only 5 studies reported clinical outcomes, and of these none reported statistically significant differences. The third review, about use of the CQI method "in nephrology"[25] (search end date of 2014), identified 76 studies meeting eligibility criteria. Only 1 included study was a randomized trial, and it found better blood pressure control among dialysis patients in the intervention group. All 3 reviews concluded more and better-conducted research of the CQI method was needed. Although not the same as the CQI method, the use of PDSA is common to many general CQI frameworks, and so we include here a discussion of the review focused on PDSA.[28] This review (search end date of 2012) identified 73 articles meeting eligibility criteria. The review reported details about the design, execution, and reporting of the studies (such as the country of origin, clinical focus, how many studies used iterative cycles, *etc*), but it did not report any data on effectiveness. This review concluded that better reporting was needed.

The last 2 non-Lean, non-Six Sigma reviews were about Quality Improvement Collaboratives[26] and "the CMS approach".[11] The review of Quality Improvement Collaboratives (search end date of 2006) identified 72 studies meeting eligibility criteria, of which 12 publications (describing 9 studies) found "moderate positive effects", with 7 studies reporting some statistically significant effects on outcomes and 2 finding no significant effect. The review of CMS[11] (search end date of 2018) identified 35 studies that could be of any research design (including case studies and cross-sectional studies) which contained in the title or abstract of the article "keywords related to clinical microsystems" and could concern any health care provider on any clinical topic. The authors stated that "all the included studies underlined the positive aspects regarding the achievement of the targeted objectives, namely a higher quality of care and of better patient safety".

### Sustainment of Change in CQI Frameworks

No publication included in our review concluded that any specific methodology had definitive evidence for sustainment of change. Several, in fact, specifically noted the literature is lacking on this aspect of QI in their analyses.[10,13,15,33,35,39,40] Only 1 review calculated an average duration of study implementation, finding an average of 10 months reported across 16 of 26 studies on Lean management in emergency settings,[18] while another stated only 15% of 39 studies reported a follow-up duration greater than 1 year.[35] Another review suggested reporting data for at least 2 years post-implementation would be "necessary to determine whether process changes are accepted and become part of the permanent culture".[15]

### Health Care Workers' Reaction, Learning, and Behavior Change to CQI Frameworks

No publication included in this review concluded that any specific methodology had evidence speaking to health care workers' reaction to that general framework in perceptions of the usefulness of the CQI training. Additionally, no study commented on specific learning outcomes related to any CQI framework nor tested participant knowledge or skill acquisition with tools such as the Quality Improvement Knowledge Application Tool. Furthermore, no study analyzed whether staff made specific behavioral changes based on training. Several reviews included

provider/clinician satisfaction as a possible outcome,[9,20,22,38,39] although these outcomes may represent staff working in locations with implemented CQI strategies, rather than those who received training as part of each study. In studies reporting results for employee satisfaction outcomes, Moraros[22] concluded implementation of Lean had "a negative association" with worker satisfaction, while 3 other studies drew the opposite conclusion.[20,38,39]

### Certainty of Evidence for Key Question 1B

Of the 25 studies speaking to Key Question 1B, 22 were rated using the modified AMSTAR2.[9-16,18-22,24-26,28-31,33,35] The AMSTAR2 scores for these 22 reviews ranged from 3 to 11 points, with a mean of 6.5. Nine[9,11,14,16,24,26,29,31,35] of these 22 studies had a modified AMSTAR2 score of 8 points or higher. None of the articles with higher AMSTAR2 ratings had supported conclusions widely diverging from the rest of the articles. Therefore, we conclude there is moderate certainty of evidence that CQI frameworks may be successfully implemented in a variety of clinical settings.

## KEY QUESTION 2: What factors (including intervention, inner setting, outer setting, individuals involved, and process by which implementation is accomplished) contribute to the success or failure of these continuous quality improvement frameworks?

### Key Points (Summary of Findings)

- Twenty reviews studied at least 1 CFIR factor related to success or failure of CQI frameworks. Fifteen discussed intervention characteristics, 6 discussed individuals involved, and 10 discussed inner setting.
- No reviews compared the success or failure of different CQI strategies based on intervention characteristics, characteristics of individuals, or inner setting. Instead, the majority of studies listed aspects of some or each of these 3 CFIR categories deemed important for implementation of that methodology with little to no supporting evidence or discussion of the influence of the category on ultimate project success or failure.
- No publication included in this review discussed whether either outer setting or specific processes during implementation of a CQI framework contributed to either the success or failure of implementation for any framework.

### Detailed Findings

#### *Factors contributing to success or failure of CQI frameworks*

We looked at factors affecting the success or failure of CQI according to the 5 main Consolidated Framework for Implementation Research (CFIR) domains[7]. Of these, we found relevant data on intervention characteristics, characteristics of individuals, and inner setting. However, no studies compared the success or failure of different frameworks based on intervention characteristics, characteristics of individuals, or inner setting. Additionally, no studies discussed whether either the outer setting or specific processes during implementation of a CQI framework contributed to either the success or failure of implementation of any methodology.

Overall, few studies addressed intervention characteristics, characteristics of individuals, or inner setting in detail. Rather, the majority of studies listed aspects of some or each of these categories deemed important with little to no supporting evidence or discussion of the influence of the category on ultimate project success or failure. When discussed, the factors were often paired with an example of how the aspect in question might be helpful.

## Intervention Characteristics Contributing to Success or Failure of CQI Frameworks

### *Multiple CQI frameworks*

In all, 15 reviews[13,14,16,17,19,23,27,28,31,32,34,38,40-42] mentioned intervention characteristics contributing to either the success of failure of CQI frameworks. These articles reviewed Lean[32,38,42] and Six Sigma frameworks,[13,40,41] Total Quality Management,[27] the CQI method,[14,31] Quality Improvement Collaboratives,[17,23,34] PDCA,[28] or multiple strategies.[16,19]

Three reviews specifically discussed characteristics of Lean strategies in terms of success or failure. Two only superficially analyzed these characteristics, however. Andersen[42] identified 23 facilitators associated with successful Lean interventions, including administrative support, IT systems, physicians, and teamwork. None of the 23 facilitators were specifically discussed in terms of how they might be best fostered or included within a Lean framework, and many of the 23 factors might be considered related to inner setting (such as IT systems and administrative support). Wackerbarth[32] defined Lean interventions as having 8 specific steps, of which articles in their review documented an average of 2.77 steps. The authors felt adhering to more of these steps would likely increase overall project success; however, their categorizations of these steps and tools within them were unvalidated.

Mazzocato[38] performed a realist review of Lean studies and identified several common characteristics across various studies, but did not specify how these common characteristics serve to facilitate or hinder quality improvement work. They also identified 4 general components of Lean that led to different mechanisms of improvement: "methods to understand processes in order to identify and analyse problems; methods to organise more effective and/or efficient processes; methods to improve error detection, relay information to problem solvers, and prevent errors from causing harm; [and] methods to manage change and solve problems with a scientific approach". The authors then gave some general examples of how these mechanisms may lead to success. Using tools to understand a process as in the first component creates shared understanding, for example, which "helps members of different professions to communicate and see how their roles and their work relate to the bigger picture". Detecting errors and sharing data on their occurrence with leaders quickly as in the third component leads to clarity for those in different roles and "makes workarounds or a lack of routines more noticeable and enables stakeholders to promptly address deviations". Finally, creating a team-based and collaborative approach for problem-solving can "strengthen the belief among staff that errors are preventable and change a culture of blame into one of safety and continual improvement".

As noted in the Key Question 1 discussion, DelliFraine[13] found that "articles that focused more narrowly on more targeted areas were able to present stronger in-depth evidence that Six Sigma can improve processes of care", specifically examples around operating room or ED throughput and turnaround time. Additionally, they identified which DMAIC stages each article included,

finding many did not mention 'Control', although the presence or absence of discussing 'Control' within their reviewed articles was not correlated with success nor failure in their discussion. Antony[40] identified 19 challenges and 16 success factors for Six Sigma methodologies in the literature, discussing that 8 challenges account for 80% of the total reported: "availability of data, cultural issues, resistance to change, sustainability of results, insufficient resources, inadequate knowledge of Six Sigma, complexity of current practice and lack of leadership commitment". Five authors specifically noted availability of quality data as being the single greatest challenge faced. Similarly, 7 success factors accounted for 80% of the total reported: "understanding of Six Sigma tools and techniques, management involvement and commitment, communication, organization infrastructure and culture, training, patient focus and cultural change". No evidence was available for ways to foster success factors or avoid challenges.

In their Table 2, Talib[27] reviewed the included articles for 8 best practices they identified in Total Quality Management: top-management commitment, teamwork and participation, process management, customer focus and satisfaction, resource management, organization behavior and culture, continuous improvement, and training and education. Although 4 managerial implications of this work are listed in the discussion, none of these implications provided specific guidance on designing a Total Quality Management intervention.

Quality Improvement Collaboratives were explored in 3 studies.[17,23,34] Nadeem[23] reviewed both quasi-experimental and randomized studies of 3 different named models, the IHI Breakthrough Series, Chronic care model, and Vermont Oxford Network, as well as an "other model" category. Their Table 2 details the presence or absence of 13 components within each study, and they state more recent collaboratives more often reported on the Breakthrough Series model. Furthermore, they found, "[f]ormal pre-collaborative preparation was rarely reported and was described in only five studies" and "[o]n average, each study implemented an average of six or seven Quality Improvement Collaboratives components." All 20 studies reported in-person learning sessions, three-quarters reported use of PDSA cycles and new data collection, and 14 reported use of multidisciplinary QI teams. Very few details were shared about the way these components were delivered, however, preventing comment on aspects related to ultimate success or failure. Both Zamboni[34] and Hulscher[17] list some characteristics of their studied interventions within their reviews, but focused more on setting and individuals as noted below. Hulscher[17] did include a brief section on "the collaborative process", finding no specific evidence for success based on the "intensity of intervention" nor "exchange and sharing information". A very small sample of studies found positive influence from "being on preconference calls" and "timeliness of submission of reports".

Both reviews including the CQI method focused on narrower clinical problems than otherwise represented in our systematic review. Tricco[31] included the CQI method as an analyzed strategy while using HbA1c data to provide the only statistical evidence related to specific intervention characteristics leading to success. However, this work focused more on individual QI tools than a specific framework. Specifically, the authors note, "HbA1c was further lowered when the QI strategy included team changes (0.33%), case management (0.21%), promotion of self-management (0.21%), clinician education (0.19%), patient education (0.16%), facilitated relay (0.12%), an electronic patient registry (0.08%), and patient reminders (0.02%)". Furthermore, "[d]ecreases in HbA1c of more than 0.5% were noted for four QI strategies (team changes, case management, patients" education, and promotion of self-management) in trials enrolling patients

with HbA1c greater than 8.0%, and one QI strategy (facilitated relay) in trials enrolling patients with HbA1c of 8.0% or less". Finally, "[a]ll QI strategies were associated with significant changes in HbA1c, except for clinician education" and "[w]e noted greater improvements in HbA1c control for QI strategies targeting health systems and patients". Dzidowska[14] reported the 3 CQI method elements described by Rubenstein[2] of 1) using "systematic data guided activities" to identify problems and achieve improvement, 2) "designing with local conditions in mind", and 3) using an "iterative development and testing process" in their review. They found any of the elements in 22 and all 3 elements in 12 of their 56 reports. Additionally, they determined studies with all 3 elements "had implementation and follow-up durations above the median; utilised multifaceted designs; targeted both practice and health system levels; [and] improved screening and brief intervention [for unhealthy alcohol use]" relative to studies without all 3 CQI method elements.

The PDCA strategy was reviewed by Taylor,[28] with PDCA intervention principles defined as iterative cycles, prediction-based test of change, small-scale testing, use of data over time, and documentation. Although more than 45 of their 73 identified articles reported many of these characteristics, only 2 articles reported all of them. Only 7 articles reported data over time and only 4 documented any predictions. PDSA was the most frequent strategy reviewed by Hill,[16] with 19 of the 28 studies using some variant of the model. In their review, the authors noted that "[i]mportant characteristics of approaches to CQI were infrequently reported" across the various frameworks, including the frequency and total number of team meetings, duration of these meetings, and training type and duration. Finally, Leggatt[19] identified 20 examples where the "process redesign programme included the injection of resources (financial, physical and human)" but did not state further if there was evidence as to timing or type of additional resources that best led to success.

## Characteristics of Individuals Contributing to Success or Failure of CQI Frameworks

### *Multiple CQI frameworks*

Six articles[10,15-17,23,42] mentioned the individuals involved in the CQI frameworks in their reviews. These articles reviewed Lean[10,15,30,42] and Six Sigma,[15,30] Quality Improvement Collaboratives,[17,23] or multiple strategies[16].

Bucci[10] generally noted in their review of 9 studies on Lean in ED flow that "the staff involved was generally comprised of not only clinicians and nurses, but also by assistants and engineers", and "[i]n almost all the studies a quality improvement facilitator, often a Lean consultant, led the team". They included 2 comments regarding personnel and success or failure: in 1 study, "best results were obtained when Lean intervention was owned by the frontline workers who worked in the ED and the commitment of the leadership was principally involved sustaining the improvement". Additionally, they found, "[w]hen both leadership commitment and frontline workers" involvement were missing, "lack of improvement or even a worsening in LOS [length of stay] and patient satisfaction were observed." Andersen[42] lists personnel in an "Application" section, specifically "collaborating and multiskilled teams" and "physicians and management". These teams and members are mentioned to "facilitate local applications of Lean" and "encourage change", respectively. Glasgow[15] stated that few details regarding team members were available in the 47 articles reviewed, with only 8 specifically mentioning that physicians

were directly involved in the project. They do mention that "[t]eams are typically multidisciplinary, including representatives from most groups that provide any direct patient care potentially impacted by the QI project". Trakulsunti[30] found only 5 of their included studies on medication errors "mentioned project team members, without explaining the criteria to select such members and their responsibilities".

The other methodologies reviewed included similar information about participants. Nadeem[23] found that Quality Improvement Collaboratives began with an in-person learning session "attended by the multidisciplinary quality improvement teams (QI teams) and led by Quality Improvement Collaborative expert faculty". Follow-up was typically done with a combination of in-person learning sessions and phone meetings, both of which involved multiple sites. These multidisciplinary teams were named as such within their analysis, but the authors found the individual studies "did not always specify whether [the multidisciplinary teams] represented a range of positions within the organization's hierarchy". Additionally, these studies usually did not "describe the team members' roles within the organization". Finally, "[n]ine studies reported that the organization's leadership was involved in the Quality Improvement Collaboratives, but it was unclear whether the organizational leadership was included on the QI team or was engaged through other means." Only 6 total studies "reported that the QI team members trained additional staff in the organization." Hulscher[17] included some analysis of nursing engagement; however, this seemed more related to the inner setting of the Quality Improvement Collaboratives studied rather than nursing staff inclusion in the intervention team. Hill[16] very generally documented whether or not included studies utilized multidisciplinary teams, finding "[m]ulti-disciplinary teams were used in 19 RCTs, with 8 RCTs not adequately describing membership of their teams. One RCT explicitly stated that they did not use an multi-disciplinary team approach."

## Setting and Process Characteristics Contributing to Success or Failure of CQI Frameworks

### *Multiple CQI frameworks*

Ten articles[11,17,19,30,31,34,38-40,42] discussed the inner setting of CQI frameworks. These articles reviewed Lean[30,38,39,42] and Six Sigma frameworks,[30,40] the CQI method,[31] Quality Improvement Collaboratives,[17,34] the CMS Approach,[11] or multiple strategies.[19]

Four groups[30,39,40,42] named some inner setting aspects in their sections on facilitators and challenges for Lean and/or Six Sigma strategies, while 1 reviewing Lean Six Sigma and other strategies[19] and 1 reviewing the CMS approach[11] also listed similar facilitators. Across these 6 papers, the general inner setting areas mentioned included IT systems, administrative support, staff engagement, organizational knowledge, physicians, knowledge of the methodology, training, management/leadership involvement, teamwork, and cultural issues. None of these papers, however, discussed how these facilitators could be cultivated within or across organizations or provided strong evidence of their influence on overall success or failure.

A few authors provided more discussion and context of inner setting. Mazzocato[38] found "similar methods were used in different settings to address problems and that they yielded concrete and easily implementable suggestions". This finding suggests that systems have "either the existence of similar needs (*eg*, reduce excessive inventory, delays and waiting times) irrespective of setting" and that there may be "flexibility of the lean methods". Notably, an organization's

general approach and frontline/managerial structure matters, although no specific guidance on how to construct these theories and hierarchy were provided — "A team-based (collaborative) approach for systematic problem solving reinforces the understanding and values which can transform an error into a learning opportunity. When problem resolution required authority and information beyond that of the front-line staff, stable structures involving managers were effective". Trakulsunti,[30] on the other hand, reviewed use of Lean and Six Sigma in medication errors and identified "lack of top management support and availability of data" as 2 of the biggest challenges to successful implementation of either strategy. The authors also identified 7 success factors — "understanding of Lean Six Sigma tools and techniques and its philosophy, top management support, training, staff engagement, leadership capability, appropriate team formation, or implementation infrastructure and cultural change" — without further specifying how they might be cultivated.

Non-Lean, non-Six Sigma strategies generally provided more in-depth analysis of inner setting. Tricco[31] looked at specific strategies that were shown to statistically significantly improve HbA1c, some of which are facets of inner setting. The strategies included team changes, which were defined as changes to the structure or organization of the primary health care team such as adding a team member or multidisciplinary group, creation of patient registries, and case management. Unlike many other reviews, Zamboni[34] commented directly on inner setting and outcomes, finding "no conclusive evidence that facility size, voluntary or compulsory participation in the Quality Improvement Collaboratives programme, and baseline performance influence Quality Improvement Collaboratives outcomes". They also found inconclusive evidence for the positive impact of "health facility readiness", which included various factors such as health information systems and senior level commitment mentioned above. Mixed evidence was found for project-specific factors of external support and functionality of quality improvement teams. Hulscher[17] examined 2 Quality Improvement Collaboratives papers discussing "engagement of nurses", finding "four comparisons showed positive effects and seven comparisons showed no relationship". Additionally, they looked at 3 papers invoking "previous quality-improvement experience" and found no relationship to success. Finally, they listed "Essential Features" similar to the 3 Lean Six Sigma studies on facilitators and barriers, finding mixed results or no effects on success in 4 papers examining "organisational readiness and commitment", 6 on "leadership support", 4 on "team climate", 2 on "shared vision", and 2 on alignment of goals. Three of their papers were performed across VA facilities on adverse drug events, patient safety, and falls, respectively. Two papers on VA Quality Improvement Collaboratives work showed, specifically, that "'frontline staff support' did not influence success".

### Certainty of Evidence for Key Question 2

Of the 20 studies included in the analysis for Key Question 2, 15 were rated using the modified AMSTAR2.[10,11,13-17,19,23,27,28,30-32,34] The AMSTAR2 scores for these 15 reviews ranged from 3 to 11 points; 4[11,14,16,31] of these studies had a modified AMSTAR2 score of 8 points or higher. However, as none of the conclusions related to the CFIR topics were supported by comparative data, we conclude there is low certainty of evidence for specific intervention characteristics, individuals to be involved, or inner setting aspects leading to success in implementing a CQI methodology.

# SUMMARY AND DISCUSSION

We sought to identify effective CQI interventions (Key Question 1) and conditions necessary for their success (Key Question 2) in a review of reviews.

Our systematic review of reviews is innovative in that it used a rigorous search and review methodology to include a breadth of systematic reviews on each of these Key Questions. Additionally, we evaluated the certainty of evidence for each Key Question using a modification of the AMSTAR2 tool for systematic reviews. No included reviews specifically addressed care of Veterans as a study objective nor were any conducted fully in VA settings, although several individual articles within the included systematic reviews were based at VA.

## SUMMARY OF EVIDENCE BY KEY QUESTION

***Key Question 1A****: What is the comparative effectiveness of implementing continuous quality improvement frameworks in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?*

We assessed the literature for evidence regarding the comparative effectiveness of different CQI strategies. However, we were only able to identify a single study[1] with an AMSTAR2 rating of 2 that met inclusion criteria. This study is over a decade old and found no evidence that any single CQI strategy was more effective than others. Instead, the authors concluded that local context should guide which framework is ultimately implemented.

***Key Question 1B****: What is the effectiveness of implementing a continuous quality improvement framework in terms of health care workers' reaction, learning, behavior change, results, and sustainment of change?*

Having found only a single study comparing the effectiveness of different CQI strategies, we then turned to assess the 25 reviews that studied at least 1 framework for evidence of effectiveness for a variety of outcomes. None of the 11 reviews that examined more than 1 CQI methodology reached a conclusion that any strategy was superior to the others in terms of results. However, many frameworks have been successfully implemented in a variety of clinical settings. In some clinical settings, such as in the OR and ED, multiple different strategies have been used in different geographic locations. Several studies had AMSTAR2 ratings of at least 8; however, studies with higher AMSTAR2 scores did not draw more specific conclusions than studies with lower scores.

Similarly, none of the 7 reviews discussing sustainment of change identified a superior strategy. Additionally, no study directly commented on health care workers' reactions in being involved as part of CQI framework implementation, what health care workers learned or retained as part of CQI training, nor behavior changes noted after implementation of a CQI strategy.

***Key Question 2****: What factors (including intervention, inner setting, outer setting, individuals involved, and process by which implementation is accomplished) contribute to the success or failure of these continuous quality improvement frameworks?*

We then assessed the literature for evidence regarding success or failure factors associated with CFIR factors. Ultimately, 20 reviews studied at least 1 CFIR factor; however, none compared the success or failure of different CQI strategies based on intervention characteristics, characteristics of individuals, or inner setting. Instead, the majority of studies listed aspects of some or each of these 3 CFIR categories deemed important for implementation of that CQI methodology with little to no supporting evidence. No study included in this review discussed whether either outer setting or specific processes during implementation of a CQI framework contributed to either the success or failure of implementation for any framework.

## LIMITATIONS

Our review has a number of strengths, including a protocol-driven design, a comprehensive search, and careful quality assessment. Both our review and the literature, however, have limitations. Our review was limited to English-language publications, but the likelihood of identifying relevant data unavailable from English-language sources is low. Identified studies, with 1 exception, did not discuss comparative effectiveness, and none compared different methodologies for the other outcomes of interest. The single study explicitly discussing comparative effectiveness was identified via reference mining given its age. We chose to include this article, despite not making our publication date cut-off, as it was the only study we identified directly address the first Key Question. Similarly, one limitation of the review of reviews approach is an inability to closely align the inclusion criteria or scopes of the individual systematic reviews. Other limitations are detailed below.

### Publication Bias

Multiple included systematic reviews noted the possibility of publication bias across the quality improvement methodologies, with positive results greatly outnumbering null or negative findings. Given our focus on systematic reviews for this study, quantitative publication bias analysis was not appropriate.

### Quality of Included Reviews

We were also limited by the existing literature. We utilized a modified version of the AMSTAR2 tool for articles deemed appropriate for its use, finding 9 of 29 studies for which we were able to use the tool had a score of at least 8. These nine studies did not differ substantially in their conclusions from other articles included with lower AMSTAR2 scores. These low ratings on the AMSTAR2 tool leads us to conclude the overall certainty of evidence related to these reviews is low to moderate.

Reasons for the low AMSTAR2 scores of identified studies may include underlying issues with study design and implementation in the primary QI literature, as well as with study reporting. Most QI initiatives are designed and implemented in pragmatic clinical or operational settings, which may limit overall study quality for inclusion into a systematic review. Even when sufficient elements are included in QI projects, however, the elements may not be fully explained in subsequent manuscripts. Such issues may then be compounded in performing and reporting systematic reviews. In general, fewer than 40% of our included systematic reviews reported the following: specific comparators as an inclusion criteria (19%); PROSPERO registration (15%); explanation for inclusion of non-randomized controlled trial methodologies within the review (37%); duplicate data extraction (37%); justifications in a list or figure for exclusion of studies

(37%); results of comparators found in individual trials (4%); review for risk of bias of included studies (26%); and discussion of heterogeneity found (26%).

### Applicability of Findings to the VA Population

Several studies included work done within VA as part of their larger reviews. Additionally, most reviews focused on studies performed in Organisation for Economic Co-operation and Development countries, which generally have well-funded health care systems and which improves applicability to VA. The only review mentioning comparative effectiveness was conducted in the United Kingdom and included reflections on the VA QUERI program. Across included reviews, there were limited data on intervention and setting characteristics to compare to the overall VA population. However, the findings presented here likely have applicability to any large health care system seeking to implement a multi-site improvement methodology.

Within VA, there are several ongoing initiatives that use CQI methodologies or frameworks with CQI elements, such as the Evidence-Based Quality Improvement (EBQI) Training Hub and The Learn. Engage. Act. Process. (LEAP) Program, in addition to the focus on Lean in systems redesign. These current efforts may not be represented universally in the published literature, nor were any of these initiatives specifically identified in the systematic reviews that comprised this review of reviews, limiting our ability to discuss comparisons with other strategies. However, these initiatives may provide opportunities for future comparative evaluations of CQI methods.

## RESEARCH GAPS/FUTURE RESEARCH

This comprehensive review of the literature identified several gaps in the current evidence that warrant future investigation, even across studies with higher AMSTAR2 ratings. The most notable of these gaps is a lack of comparative studies analyzing multiple CQI methodologies within health care settings. Although such comparative examples may be identifiable in other fields, such as manufacturing, the complexity of healthcare processes, the differences in outcomes pertaining to health, and the variability in local contexts for implementation of QI methodologies within health care warrants creation of such studies. As noted above, current VA efforts from resources such as the EBQI or LEAP programs may provide opportunities to promote such comparative investigations. An alternative approach would be to attempt meta-analysis of multiple CQI methodologies with narrow questions in specific settings from the existing primary literature, such as waiting times in emergency departments. However, given the limitations of the systematic reviews in our review of reviews, such an initiative would likely require sharing and re-analysis of the primary data from the original studies.

The overall evidence base would be improved if future reported individual QI studies more closely adhered to SQUIRE 2.0 reporting standards.[5,43] Similarly, future systematic reviews of these works would be stronger by considering the AMSTAR2 categories we found to be often lacking, such as specifying comparators as an inclusion criteria and sharing the results of comparators from individual trials, PROSPERO registration, and explicitly discussing risk of bias.

## CONCLUSIONS

Prior systematic reviews of CQI strategies have, with 1 exception, not compared the effectiveness of different methodologies. Instead, many published reviews have shown success

for 1 or more methodologies within specific contexts. However, these findings are likely subject to significant publication bias from the constituent studies, making the overall certainty of evidence low. Additionally, few data are available regarding sustainment of changes made through CQI and no systematic reviews we identified discussed health care workers' reactions, learning, or behavior changes related to participating in CQI. Similarly, no systematic reviews compared the success or failure of different CQI frameworks based on intervention characteristics, characteristics of individuals, or inner setting. Furthermore, no studies discussed whether either the outer setting or specific processes during implementation of a CQI framework contributed to either the success or failure of implementation of any methodology. Few systematic reviews included in this review of reviews had high ratings on a modified AMSTAR2 tool, leading us to conclude the overall certainty of evidence related to these topics is low to moderate. Thus, evidence gaps remain regarding whether any CQI strategy is superior to others or how any such methodology should be implemented at large scale within the VA context. Future work should emphasize comparative designs for CQI methodologies. Available resources at VA may help facilitate such work in the future.

# REFERENCES


1. Powell A, Rushmer R, Davies H. A systematic review of quality improvement models in health care. *NHS Quality Improvement Scotland.* 2008.
2. Rubenstein L, Khodyakov D, Hempel S, et al. How can we recognize continuous quality improvement? *Int J Qual Health Care.* 2014;26(1):6-15.
3. Hill JE, Stephani AM, Sapple P, Clegg AJ. The effectiveness of continuous quality improvement for developing professional practice and improving health care outcomes: a systematic review. *Implement Sci.* 2020;15(1):23.
4. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008.
5. Rubenstein LV, Hempel, S., Liu, J.L., Danz, M.J., Foy, R., Lim, Y.W., Motala, A. and Shekelle, P.G. The Minimum Quality Criteria Set (QI-MQCS) for critical appraisal: advancing the science of quality improvement. *Implementation Science: IS.* 2015;10:A19.
6. King N. Template analysis. 1998.
7. Damschroder LJ, Aron, D.C., Keith, R.E. et al. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science.* 2009;4(50).
8. Aij KH, Teunissen M. Lean leadership attributes: a systematic review of the literature. *J Health Organ Manag.* 2017;31(7-8):713-729.
9. Amaratunga T, Dobranowski J. Systematic Review of the Application of Lean and Six Sigma Quality Improvement Methodologies in Radiology. *J Am Coll Radiol.* 2016;13(9):1088-1095.e1087.
10. Bucci S, de Belvis AG, Marventano S, et al. Emergency Department crowding and hospital bed shortage: is Lean a smart answer? A systematic review. *Eur Rev Med Pharmacol Sci.* 2016;20(20):4209-4219.
11. Côté A, Beogo I, Abasse KS, Laberge M, Dogba MJ, Dallaire C. The clinical microsystems approach: Does it really work? A systematic review of organizational theories of health care practices. *J Am Pharm Assoc (2003).* 2020;60(6):e388-e410.
12. DelliFraine JL, Langabeer JR, II, Nembhard IM. Assessing the evidence of Six Sigma and Lean in the health care industry. *Qual Manag Health Care.* 2010;19(3):211-225.
13. DelliFraine JL, Zheng W, McCaughey D, Langabeer Ii JR, Erwin C. The Use of Six Sigma in Health Care Management: Are We Using It to Its Full Potential? *Qual Manag Health Care.* 2013;22(3):210-223.
14. Dzidowska M, Lee KSK, Wylie C, et al. A systematic review of approaches to improve practice, detection and treatment of unhealthy alcohol use in primary health care: a role for continuous quality improvement. *BMC Fam Pract.* 2020;21(1):33.
15. Glasgow JM, Scott-Caziewell JR, Kaboli PJ. Guiding inpatient quality improvement: a systematic review of Lean and Six Sigma. *Jt Comm J Qual Patient Saf.* 2010;36(12):533-540.
16. Hill JE, Stephani A-M, Sapple P, Clegg AJ. The effectiveness of continuous quality improvement for developing professional practice and improving health care outcomes: a systematic review. *Implement Sci.* 2020;15(1):23.
17. Hulscher MEJL, Schouten LMT, Grol RPTM, Buchan H. Determinants of success of quality improvement collaboratives: what does the literature show? *BMJ Qual Saf.* 2013;22(1):19-31.

18. Mousavi Isfahani H, Tourani S, Seyedin H. Features and Results of Conducted Studies Using a Lean Management Approach in Emergency Department in Hospital: A Systematic Review. *Bull Emerg Trauma.* 2019;7(1):9-20.
19. Leggat SG, Bartram T, Stanton P, Bamber GJ, Sohal AS. Have process redesign methods, such as Lean, been successful in changing care delivery in hospitals? A systematic review. *Public Money & Management.* 2015;35(2):161-168.
20. Pestana Magalhães AL, Erdmann AL, Lima da Silva E, Guedes dos Santos JL. Lean thinking in health and nursing: an integrative literature review. *Revista Latino-Americana de Enfermagem (RLAE).* 2016;24:1-13.
21. Mason SE, Nicolay CR, Darzi A. The use of Lean and Six Sigma methodologies in surgery: a systematic review. *Surgeon.* 2015;13(2):91-100.
22. Moraros J, Lemstra M, Nwankwo C. Lean interventions in healthcare: do they actually work? A systematic literature review. *Int J Qual Health Care.* 2016;28(2):150-165.
23. Nadeem E, Olin SS, Hill LC, Hoagwood KE, Horwitz SM. Understanding the components of quality improvement collaboratives: a systematic literature review. *Milbank Q.* 2013;91(2):354-394.
24. Nicolay CR, Purkayastha S, Greenhalgh A, et al. Systematic review of the application of quality improvement methodologies from the manufacturing industry to surgical healthcare. *Br J Surg.* 2012;99(3):324-335.
25. Nunes JW, Seagull FJ, Rao P, Segal JH, Mani NS, Heung M. Continuous quality improvement in nephrology: a systematic review. *BMC Nephrol.* 2016;17(1):190.
26. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ.* 2008;336(7659):1491-1494.
27. Talib F, Rahman Z, Azam M. Best practices of total quality management implementation in health care settings. *Health Mark Q.* 2011;28(3):232-252.
28. Taylor MJ, McNicholas C, Nicolay C, Darzi A, Bell D, Reed JE. Systematic review of the application of the plan-do-study-act method to improve quality in healthcare. *BMJ Qual Saf.* 2014;23(4):290-298.
29. Tlapa D, Zepeda-Lugo CA, Tortorella GL, et al. Effects of Lean Healthcare on Patient Flow: A Systematic Review. *Value Health.* 2020;23(2):260-273.
30. Trakulsunti Y, Antony J, Ghadge A, Gupta S. Reducing medication errors using LSS methodology: A systematic literature review and key findings. *Total Quality Management and Business Excellence.* 2018:1-19.
31. Tricco AC, Ivers NM, Grimshaw JM, et al. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. *Lancet.* 2012;379(9833):2252-2261.
32. Wackerbarth SB, Bishop SS, Aroh AC. Lean in Healthcare: Time for Evolution or Revolution? *Journal for Healthcare Quality: Promoting Excellence in Healthcare.* 2021;43(1):32-38.
33. Woodnutt S. Is Lean sustainable in today's NHS hospitals? A systematic literature review using the meta-narrative and integrative methods. *Int J Qual Health Care.* 2018;30(8):578-586.
34. Zamboni K, Baker U, Tyagi M, Schellenberg J, Hill Z, Hanson C. How and under what circumstances do quality improvement collaboratives lead to better outcomes? A systematic review. *Implement Sci.* 2020;15(1):27.

35. Zepeda-Lugo C, Tlapa D, Baez-Lopez Y, et al. Assessing the Impact of Lean Healthcare on Inpatient Care: A Systematic Review. *Int J Environ Res Public Health.* 2020;17(15).
36. Crema M, Verbano C. Lean Management to support Choosing Wisely in healthcare: the first evidence from a systematic literature review. *Int J Qual Health Care.* 2017;29(7):889-895.
37. Maijala R, Eloranta S, Reunanen T, Ikonen TS. SUCCESSFUL IMPLEMENTATION OF LEAN AS A MANAGERIAL PRINCIPLE IN HEALTH CARE: A CONCEPTUAL ANALYSIS FROM SYSTEMATIC LITERATURE REVIEW. *Int J Technol Assess Health Care.* 2018;34(2):134-146.
38. Mazzocato P, Savage C, Brommels M, Aronsson H, Thor J. Lean thinking in healthcare: a realist review of the literature. *Qual Saf Health Care.* 2010;19(5):376-382.
39. D'Andreamatteo A, Ianni L, Lega F, Sargiacomo M. Lean in healthcare: A comprehensive review. *Health Policy.* 2015;119(9):1197-1209.
40. Antony J, Palsuk P, Gupta S, Mishra D, Barach P. Six Sigma in healthcare: a systematic review of the literature. *International Journal of Quality & Reliability Management.* 2018.
41. Niñerola A, Sánchez-Rebull M-V, Hernández-Lara A-B. Quality improvement in healthcare: Six Sigma systematic review. *Health Policy.* 2020;124(4):438-445.
42. Andersen H, Røvik KA, Ingebrigtsen T. Lean thinking in hospitals: is there a cure for the absence of evidence? A systematic review of reviews. *BMJ Open.* 2014;4(1):e003873.
43. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf.* 2016;25(12):986-992.

# APPENDIX A. SEARCH STRATEGIES

DATABASE SEARCHED & TIME PERIOD COVERED:

PubMed – From 01/01/2010 to 03/18/2021

SEARCH STRATEGY: 879 results

(Total Quality Management[MESH] OR Quality Improvement[Mesh]) OR (CQI[Title/Abstract] OR "Continuous Quality Improvement"[Title/Abstract] OR "Quality Improvement" [Title/Abstract] OR "Lean Six Sigma"[Title/Abstract] OR "Clinical MicroSystems"[Title/Abstract] OR "Model for Improvement"[Title/Abstract] OR "lean management"[Title/Abstract] OR "lean thinking"[Title/Abstract] OR "lean healthcare"[Title/Abstract] OR "lean principles"[Title/Abstract] OR Toyota[Title/Abstract] OR IHI[Title/Abstract])

---

DATABASE SEARCHED & TIME PERIOD COVERED:

CINAHL – From 01/01/2010 to 03/30/2021

SEARCH STRATEGY: 1068 results

MH Quality Improvement OR TI (CQI OR "Continuous Quality Improvement" OR "Quality Improvement" OR "Lean Six Sigma" OR "Clinical MicroSystems" OR "Model for Improvement" OR "lean management" OR "lean thinking" OR "lean healthcare" OR "lean principles" OR Toyota OR IHI OR "system redesign" OR Total Quality Management”) OR AB (CQI OR "Continuous Quality Improvement" OR "Quality Improvement" OR "Lean Six Sigma" OR "Clinical MicroSystems" OR "Model for Improvement" OR "lean management" OR "lean thinking" OR "lean healthcare" OR "lean principles" OR Toyota OR IHI OR "system redesign" OR “Total Quality Management”)

---

DATABASE SEARCHED & TIME PERIOD COVERED:

DARE – From 01/01/2010 to 03/31/2015

SEARCH STRATEGY: 78 results

MeSH descriptor: [Quality Improvement]

OR

(CQI OR "Continuous Quality Improvement" OR "Quality Improvement" OR "Lean Six Sigma" OR "Clinical MicroSystems" OR "Model for Improvement" OR "lean management" OR "lean thinking" OR "lean healthcare" OR "lean principles" OR Toyota OR IHI OR "system redesign" OR “Total Quality Management”)

---

DATABASE SEARCHED & TIME PERIOD COVERED:

Cochrane – From 01/01/2010 to 03/30/2021

SEARCH STRATEGY: 54 results

MeSH descriptor: [Quality Improvement]

OR

(CQI OR "Continuous Quality Improvement" OR "Quality Improvement" OR "Lean Six Sigma" OR "Clinical MicroSystems" OR "Model for Improvement" OR "lean management" OR "lean thinking" OR "lean healthcare" OR "lean principles" OR Toyota OR IHI OR "system redesign" OR “Total Quality Management”)

# APPENDIX B. FULL-TEXT SCREENING FORM

1. Is this an include? *Select all that apply.*

- ☐ a. Include: systematic review describes comparative effectiveness of CQI frameworks/methods
- ☐ b. Include: systematic review describes effectiveness of CQI framework(s)/method(s) SPECIFY HERE:
- ☐ c. Include: SR describes context/factors contributing to the success or failure of CQI framework(s)/method(s)
- ☐ d. Exclude (second tier): subset of included articles address key questions
- ☐ e. Exclude: this SR does not answer a KQ nor does it present data from included studies to address KQs STOP HERE
- ☐ f. Exclude: not a systematic review STOP HERE
- ☐ g. Unclear/to discuss STOP HERE
- ☐ h. DUPLICATE

2. Total number of studies included in the review

3. Number of included studies identified as CQI/QI in each study design category:

- ☐ Case study/case series with no formal analysis (fill in number)
- ☐ Pre/post (fill in number)
- ☐ Time series or concurrent comparator design (fill in number)
- ☐ RCT (fill in number)
- ☐ Qualitative analysis addressing KQ (fill in number)
- ☐ Other/not reported (fill in number)

**IF YES TO 1A, 1B, 1C above, continue:**

4. What were the **search dates** for this review (NR if not reported):

5. What **health conditions** were included:

- a. Specific population as inclusion/exclusion criteria for review:
- b. Varies/not reported/unclear

6. What **healthcare settings** were included:

- a. Specific setting as inclusion/exclusion criteria for review:
- b. Varies/not reported/unclear

7. What **regions/countries/geographical areas** were included:

- a. Specific areas/countries as inclusion/exclusion criteria for review:
- b. Varies/not reported/unclear

8. What **outcomes** were reported? *Select all that apply*

- a. Clinical/patient level
- b. Provider level (satisfaction, experience, etc)
- c. System/efficiency (e.g. wait times, length of stay, etc)
- d. Patient safety outcome (HAIs, harms, Costs)
- e. Various/other
- f. Not reported/unclear

9. Main findings (can copy/paste from abstract results/findings)

10. Did review have a definition or citation for CQI? *Select all that apply*

- a. Yes; a definition or citation for CQI was described as background/conceptual frame
- b. Yes, a definition or citation for CQI was used as inclusion/exclusion criteria
- c. No
- d. Unclear/to discuss

# APPENDIX C. CRITERIA USED IN QUALITY ASSESSMENT

1A. Did the research questions and inclusion criteria for the review include the population/setting?

- Yes
- No
- To Discuss

1B. Did the research questions and inclusion criteria for the review include the intervention?

- Yes
- No
- To Discuss

1C. Did the research questions and inclusion criteria for the review include the comparator(s)?

- Yes
- No
- To Discuss

1D. Did the research questions and inclusion criteria for the review include the outcomes?

- Yes
- No
- To Discuss

2. Did the report of the review reference a protocol or PROSPERO registration?

- Yes
- No

3. Did the review authors explain their selection of other study designs (non-RCTs) for inclusion in the review?

- Yes, justification given
- No

4. Did the review authors use a comprehensive literature search strategy?

*searched at least 2 databases (relevant to research question) provided key word and/or search strategy conducted search within 24 months of publication of the review*

- Yes
- No

5. Did the review authors perform study selection in duplicate?
For yes, either ONE of the following:
*at least two reviewers independently agreed on selection of eligible studies and achieved consensus on which studies to include*
*OR two reviewers selected a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder selected by one reviewer*

- Yes
- No

6. Did the review authors perform data extraction in duplicate?
For yes, either ONE of the following:
*at least two reviewers achieved consensus on which data to extract from included studies*
*OR two reviewers extracted data from a sample of eligible studies and achieved good agreement (at least 80 percent), with the remainder extracted by one reviewer*

- Yes
- No

7. Did the review authors provide a list of excluded studies and/or justify the exclusions? (list of available excludes with reasons or flow with count of excludes by individual exclusion criteria categories for abstract and full-text excludes)

- Yes
- No

8A. Did the review authors describe the described populations/settings for individual included studies in adequate detail? (e.g. an evidence table)

- Yes
- No
- To Discuss

8B. Did the review authors describe the described interventions for individual included studies in adequate detail? (e.g. an evidence table) - *need components, not just brand name of intervention.*

- Yes
- No

- To Discuss

8C. Did the review authors describe the described comparators for individual included studies in adequate detail? (e.g. an evidence table)

- Yes
- No
- To Discuss

8D. Did the review authors describe the described outcomes for individual included studies in adequate detail? (e.g. an evidence table) - *not looking for numbers but need clarity (not "improve patient care", for example)*

- Yes
- No
- To Discuss

8E. Did the review authors describe the described research designs for individual included studies in adequate detail? (e.g. an evidence table)

- Yes
- No
- To Discuss

9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? (because of topic/nature of interventions, criteria aren't relevant)

- Yes, do use and report tool used
- No

10. Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?
For Yes:
included only low risk of bias RCTs
*OR, if RCTs with moderate or high RoB, or NRSI were included the review provided a discussion of the likely impact of RoB on the results*

- Yes
- No

11. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?
For Yes:
*There was no significant heterogeneity in the results*
*OR if heterogeneity was present the authors performed an investigation of sources of any heterogeneity in the results and discussed the impact of this on the results of the review (can be narratively discussed)*

- Yes
- No

12. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?

For Yes:
*performed graphical or statistical tests for publication bias or discussed the likelihood and magnitude of impact of publication bias*

- Yes
- No
- N/A, not a quantitative synthesis

***We have excluded question 12 from our AMSTAR table because all reviews were marked "N/A, not a quantitative synthesis".***

13. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?

- Yes
- No

# APPENDIX D. PEER REVIEW DISPOSITION

| Reviewer | Reviewer Comments | Author Responses |
|---|---|---|
| **Executive Summary** | | |
| 2 | page 7, lines 8-9, not sure how being the largest integrated system is an incentive in and of itself to improve quality/safety. Not getting what aspect here is the incentive. | Edited to read, '***As part of its mandate to optimize health outcomes for Veterans***, the Department of Veterans Affairs (VA) has an incentive to improve the quality and safety of health care'. This was also changed in the Introduction. |
| 2 | Lines 28-29, are both the Model for Change and Clinical Microsystems IHI's? Thought clinical microsystems were another approach or model, but not IHI's. Would be good to clarify. Actually, if both are IHI, no change would be needed. | Switched order of Model for Change and Clinical Microsystems to identify Model for Change as only IHI approach - '***Clinical Microsystems*** or the Institute for Healthcare Improvement (IHI)'s Model for Change' |
| 2 | page 8, why are DARE search dates truncated to 2015? | Bibliographic records published on DARE ceased at the end of March 2015. These databases can still be accessed but have not been and will not be updated. |
| 2 | page 9, line 5, should add "and" before the final condition | Added as suggested |
| 2 | Line 21-22, grammatical glitch, need to be either "had a modified AMSTAR2 rating" or "had modified AMSTAR2 ratings." | Removed as suggested |
| 2 | Line 34, should be "satisfaction was studied." | Replaced as suggested |
| 2 | page 10, line 19, need to decide on use of VA vs. VHA for consistency throughout (was VHA just a few lines above on line 11). | Changed to 'VA' throughout. |
| 2 | Lines 23-24, would move "not" before the "with one exception" for ease in interpretation (have not, with one exception,...). | Moved as suggested |
| 7 | Abbreviations Table: This review uses so many abbreviations – could you limit their use if word count is not an issue? Check the abbreviations table for "L = Lean" does not seem correct – pg. 4, line 58 | Thank you for your comment. We have reduced the abbreviations throughout. |
| 7 | Could you explain an AMSTART2 rating of 2 (pg. 2, line 51) and the AMSTAR2 rating of 8 (pg. 3, line 23) (page 3, line 54) for I had to go looking for information in the text. | The following was added to the Executive Summary under Data Abstraction and Quality Assessment, ***'Each systematic review was assessed using a modified version of the Assessing the Methodological Quality of Systematic Reviews 2 (AMSTAR2) criteria [4]. This 16-item tool was designed to assess the*** |

| | | |
|---|---|---|
| | | ***methodological quality of systematic reviews. As some AMSTAR2 items concern meta-analysis, we adapted the tool for this review, resulting in a 13-item tool. Assessment of studies using our modified tool was also completed in duplicate, with discrepancies resolved with group discussion. No study was excluded from analysis based on AMSTAR2 score; however, we chose a score of greater than or equal to 8 to represent higher quality systematic reviews. Studies based solely on qualitative analysis were not scored with the modified tool.'*** |
| 7 | Please spell out/explain OECD countries and why this was pertinent for the review (page 4, line 10) | Added the following, 'conducted in ***Organisation for Economic Co-operation and Development*** countries, ***which are a group of 38 countries with mature economies and, often, well-financed healthcare systems***,' |
| 7 | How many were performed with veteran populations (pg. 4, line 11)? | No SR was performed solely in veterans; however, individual studies often were performed in veteran populations. We did not track the full number of these included individual studies as part of this work. We did add 'some ***individual studies*** were performed with Veteran populations, which improves applicability to VA' earlier in the Discussion to clarify. |
| 7 | You don't abbreviate to CQI in first line of conclusion (pg. 4, line 23) or on lines 32, 35. I prefer it spelled out in the executive summary. | Removed CQI abbreviations from Executive Summary. |
| 7 | Could you explain the publication bias (e.g., only successful studies were ultimately published (pg. 4, line 27)? | Added '***as it is probable unsuccessful quality improvement work is less likely to receive publication'***. |
| 7 | Could you explain what a high AMSTAR 2 rating (e.g., 8 or higher) on pg. 4 line 38 | Please see comment above. |
| **Introduction** | | |
| 2 | Page 12, lines 12-14, not sure of intent or meaning of "particularly in settings with fewer currently trained personnel..." As opposed to what? | Modified to the following, 'particularly in settings ***such as individual clinics or units that may have*** fewer currently trained personnel to support project leadership and management' |
| 7 | Could you expand on why Lean was selected by the VA in 2019? Was this because it is the most popular, or for other reasons (pg. 6, lines 27-28). Would help the reader appreciate why this review was requested after the decision was made (versus before). | Unfortunately, we do not have any information about the history of why Lean was selected by the VA beyond what was shared with us in the topic nomination brief. |
| **Methods** | | |

| | | |
|---|---|---|
| 2 | Page 14 (of document, appears to be page 8 of report, so other page references above are of the document itself), "rejected ... publications that only reported components of QI without a full QI framework." Could the authors clarify the volume of publications that fell in this category and consider the implications of not having a "full QI framework"? This phrase leaves me wondering what was left out and how much I should be concerned about it. | Since the scope of the review was to discuss CQI frameworks/methods, any publications that only mentioned components of QI (such as checklists, or audit and feedback) would not address either of the key questions and therefore were excluded from our review. Out of 165 publications excluded at abstract phase, 34 publications were excluded for not discussing a CQI framework/method. Some of these focused on specific QI components without an overarching framework or strategy, while others focused on interventions that did not describe themselves as CQI or match the description of CQI described by Rubenstein and colleagues (e.g., in-hospital interventions to reduce diabetes readmissions). |
| 2 | Lines 27-28, grammatical glitch "were not get assessed..." (take out word "get"). There are just enough of these kind of glitches that I recommend a thorough re-read by the authors to ensure they address any I missed. | Deleted this specific example, as well as further editing on full document |
| 7 | Did you work with a librarian to validate the search method? | Yes, a librarian at the UCLA Library was one of our co-authors and team members. We added the following under the Search Strategy section, "***Our team, which included a medical librarian, developed and*** conducted broad systematic review searches using terms…" |
| 7 | What is the DARE database and why only searched until 2015? | Bibliographic records published on DARE ceased at the end of March 2015. These databases can still be accessed but have not been and will not be updated. |
| 7 | Pg. 11 – line 11: You don't abbreviate key questions – which is much better. Can you change the KQ in the figure 1 lit review to key questions and throughout the document? | We changed the abbreviations throughout. |
| **Findings** | | |
| 2 | Literature flow, page 15, recommend past tense for "does not address key questions" etc. (line 11) since this section is otherwise in past tense. Same for line 13. | Changed as suggested |
| 2 | Figure 1 typo lines 10-12 "de-depulication" ?? (assume they mean de-duplication). | Changed as noted |
| 2 | What does RefMine mean? | RefMine stands for "reference mining." It involves using bibliographies of a source to find more related sources. We have added language in text to clarify the addition of articles by reference mining, or "RefMine". |
| 2 | Would be helpful if Figure 2 was self-documenting (e.g., what is SS only? line 12). | Edited figure to reflect this. |

| | | |
|---|---|---|
| 2 | Page 19, line 50, typo "this has meant the even..." should be "that even". Time spent on this level of a read is going to exceed time available, so urge authors to do a very careful read-through to remedy these grammatical and typographical errors throughout. | Changed as noted |
| 2 | page 23, line 35, should be "reviewed implementation [in] emergency setting..." | Added as noted |
| 2 | Line 55 "general categories of care" an e.g., would be helpful. | Added the following, 'general categories of care, ***such as 'productivity' and 'clinical quality'*.** |
| 2 | page 25, lines 25-36 appears to be a single sentence and it is very hard to read through and stay on top of. | Regrettably, this may be a computer glitch. We apologize for the inconvenience! |
| 3 | I would include as an appendix a table of all of the included studies with their respective quality ratings, broken down by criterion in other words, "show us the math" when it comes to the calculation of the quality rating for each study. The benefit of this is that, if necessary, a can be used in follow on analyses to more granularly assess the effect of these interventions (for example, via network meta-analysis; see comment #4 below)). Along related lines, p. 11 lists in text how many reviews received a given AMSTAR2 rating. This information is better conveyed through a histogram. | We have added a table with AMSTAR2 ratings, as well as more discussion of the scores. |
| 4 | Minor issue on page 19: in the paragraph at the top of the page, a review is quoted as "having no more than 3.4 million defects per opportunity…" There is something wrong here, because it should be the other way around – not having more than 1 defect per 3.4 million opportunities | Thank you for noting this, as this was a direct quote from the referenced article. We have removed the quotations from this error and corrected the statement as follows, 'as having ***no more than one defect per 3.4 million opportunities '***may not be realistic or applicable to many QI issues faced by health care organizations'. |
| 5 | Page 10, Figure 1 - RefMine is included in the figure but it not referenced in the text. Please explain the addition of articles by RefMine in the literature flow. | We have added language in text to clarify the addition of articles by reference mining, or "RefMine". |
| 7 | Pg. 13, line 37: Can you explain reference mining and why this article was not captured using the review methods? | We have added language in text to clarify the addition of articles by reference mining, or "RefMine". The article in question was identified by reference mining and not our searches because it was published prior to the dates specified in our search. Given that our search did not identify any literature within the desired time frame addressing KQ1A, this publication merited inclusion despite its age. |
| 7 | Pg. 17, line 26 – which National Health Service? Same on Pg. 18., lines 34-45 | The UK National Health Service; we have inserted this in the report. |

| | | |
|---|---|---|
| 7 | Pg. 19, line 37 – double negative is hard to read | We have revised this sentence. |
| 7 | Pg. 20, line 27 – what is a QIC? At this point I'm so overwhelmed by the abbreviations… | QIC is Quality Improvement Collaborative. We have streamlined the number of abbreviations throughout for readability. |
| 7 | Pg 23, line 26: typo "review" not reviewed | Edited to read as follows, 'reviewed ***the*** included articles' |
| 7 | Pg 25, line 45 – typo – check font on multiple strategies | Regrettably, this may be a computer glitch. We apologize for the inconvenience! |
| 7 | Page 26, line 51 – mention why the five articles were not rated using the AMSTART2 | We provided explanation in text. |
| 7 | Page 27, line 21-22. The sentence "no systematic reviews addressed Veterans" is confusing. I'm assuming you mean there was not a specific review that focused exclusively on Veterans or VHA, but there were a couple studies that were VA based included. Rewording would help comprehension. | Reworded to the following, 'No systematic reviews specifically addressed ***care of*** Veterans ***as a study objective nor*** were ***any*** conducted fully in VA ***settings***, although several ***individual*** articles within ***the included*** systematic reviews were based at VA |
| **Discussion** | | |
| 3 | This may be an issue of systematic review writing style, but I could have done with a bit more discussion, reflection, interpretation, and recommendations for future directions. For example, one of the most important points I think that the authors make is the idea that there seems to be no one best way to conduct quality improvement in an organization. That is certainly consistent with current frameworks and implementation science, as well as general theries of organizational development and change. And this is a conclusion that really stands beyond the Powell et al. 2008 study; this conclusion is appropriately reached by the preponderance of evidence of a single studies of the studies that tested the effectiveness of single interventions as well. | We have expanded the following sections: Limitations, Study Quality, Applicability to VA, Research Gaps/Future Research, and Conclusions. |
| 3 | In another part of the conclusions the authors state: "Few systematic reviews included in this review of reviews had high ratings on a modified AMSTAR2 tool, leading us to conclude the overall quality of evidence related to these topics is low-to-moderate." Why is that? Why are so many QI studies of low to moderate quality? The cynic might answer, "because there was no IRB holding these QI teams to the rigorous standards of research." A more pragmatic answer, however, is likely that change is messy. In other words quality improvement must be tested in situ, not a laboratory, | We have renamed the 'Study Quality' section **'Quality of Included Reviews'** to clarify that we are not referring to quality of individual studies but rather, overall quality of systematic reviews. We have expanded this point, which now reads, ***'These low ratings on the AMSTAR2 tool leads us to conclude the overall certainty of evidence related to these reviews is low-to-moderate.'***<br>***Reasons for the low AMSTAR2 scores of identified reviews may include underlying issues with study design and implementation in the primary QI literature, as well as with study reporting. Most QI initiatives are designed and implemented in pragmatic clinical*** |

| | | |
|---|---|---|
| | in order to fully understand its effectiveness in this setting question. Certainly the aforementioned suggestion to include the breakdown of quality ratings by criterion for each study (see comment #1, above) would be helpful in better understanding what is going on with respect to study quality. | ***or operational settings, which may limit overall study quality for inclusion into a systematic review. Even when sufficient elements are included in QI projects, however, the elements may not be fully explained in subsequent manuscripts. These issues may then be compounded in performing and reporting systematic reviews.'***<br><br>We have also added a table re: AMSTAR2 information (Table 2). |
| 3 | Related to the study quality and future directions points, perhaps an appropriate future direction is to adopt a more meta-analytic approach to answering these evidence synthesis questions. In other words, (to be clear: not for the systematic review, but rather for the follow on study that should come next) the suggestion is to be analyze the included studies (and since this is a review of reviews, I would go so far as to say to use the original empirical studies included in the reviews examined) according to their individual study quality characteristics, and try to come up with some effect sizes with study quality accounted for. This approach is in many ways a analogous to the Hunter and Schmidt style of meta-analysis, which in psychological studies accounts for the reliability and validity of the criterion and predictor variables of each study as moderators and the overall meta-analysis, thereby yielding a more nuanced assessment of effect size. | This section has been expanded to now include, '***The most notable of these gaps is a lack of comparative studies analyzing multiple continuous quality improvement methodologies within healthcare settings. Although such comparative examples may be identifiable in other fields, such as manufacturing, the complexity of healthcare processes and the variability in local contexts for implementation of QI methodologies within healthcare warrants creation of such studies. As noted above, current VA efforts from resources such as the EBQI or LEAP programs may provide opportunities to promote such comparative investigations. An alternative approach would be to attempt meta-analysis of multiple continuous quality improvement methodologies with narrow questions in specific settings from the existing primary literature, such as waiting times in emergency departments. However, given the limitations of the systematic reviews in our review of reviews, such an initiative would likely require sharing and re-analysis of the primary data from the original studies.'*** |
| 4 | On page 29, the section on "Research Gaps/Future Research" seems underdeveloped. It would be helpful to go on to provide detail about the "several gaps" noted in the one sentence in that section, which seems like it should be a topic sentence of a full paragraph. | Please see the above response. |
| 7 | Since the only paper describing comparative effectiveness was found using reference mining because it was a report and not indexed, you should address how this may have impacted your overall findings | This study was identified via reference mining as it was excluded from original searches given its age. We have added the following to highlight this, '***The single study explicitly discussing comparative effectiveness was identified via reference mining given its age. We chose to include this article, despite not making our publication date cut-off, as it was the only study we identified directly address the first Key Question.***' |

| | | |
|---|---|---|
| 7 | Pg. 28, line 5-6: can you explain why OECD countries improves the applicability to VHA? | We have added the following, 'Organisation for Economic Co-operation and Development (OECD) countries, ***which generally have well-funded healthcare systems and*** improves applicability to VA'. |
| 7 | Research gaps/future research: Pg 29, lines 17-19 -This is where work from outside healthcare would be beneficial. Could you briefly look to see if other industries have compared CQI methods. If yes, which came out on top. If no, can you include possible reasons why this work is challenging and research methods that could address these questions. | This is an important note, although we feel work from outside healthcare is likely ungeneralizable into a healthcare context given the difference in contexts and outcomes and complexity of healthcare workflows. We did attempt to address this point as follows, '***The most notable of these gaps is a lack of comparative studies analyzing multiple continuous quality improvement methodologies within healthcare settings. Although such comparative examples may be identifiable in other fields, such as manufacturing, the complexity of healthcare processes and the variability in local contexts for implementation of QI methodologies within healthcare warrants creation of such studies.*** |
| 5 | However, the conclusion is unsatisfying as it leaves the reader without a map for next steps. What are the next steps from here? What is a reader to do if one wants to determine the most effective improvement methodology? Consider adding to the discussion if there are comparative effectiveness studies of continuous quality improvement outside of healthcare. Can we learn something from another industry about reporting on CQI that we can apply in future effectiveness studies of CQI in healthcare? | Please see the amended section and comment above. |
| **General/Misc.** | | |
| 1 | Are there any published or unpublished studies that we may have overlooked?: Yes - Work by Michael Parchman on Practice Facilitation REview QUERI Learning Hubs to determine if any of them also provide training in QI methodologies- notably LEAP, EBQI, Quality Scholars, Facilitation, LOCI, etc- see full list here: https://www.queri.research.va.gov/training_hubs/default.cfm Conceptual work by Friedman on the Learning Health System cycle | One article of Dr. Parchman's was included in a systematic review we included in this study (Hill, 2020). However, as he has not performed any systematic reviews, these other works would not have been eligible for inclusion in this project. Such issues with literature inclusion/scope are certainly limitations of the review of reviews approach, as we have emphasized in the Limitations as follows, ***'Similarly, one limitation of the review of reviews approach is an inability to closely align the inclusion criteria or scopes of the individual systematic reviews***'.<br>We have added the following to the 'Applicability of Findings to the VA Population' section, '***Within VA, there are several ongoing initiatives that use continuous quality improvement methodologies or frameworks with continuous quality improvement elements, such as the Evidence-Based Quality Improvement (EBQI) Training Hub and The Learn. Engage. Act. Process. (LEAP) Program, in addition*** |

| | | |
|---|---|---|
| | | ***to the focus on Lean in systems redesign. These current efforts may not be represented universally in the published literature, nor were any of these initiatives specifically identified in the systematic reviews that comprised this review of reviews, limiting our ability to discuss comparisons with other strategies. However, these initiatives may provide opportunities for comparative evaluations of continuous quality improvements methods.'***<br>We also added the following to the Conclusions, ***'Future work should emphasize comparative designs for continuous quality improvement methodologies. Available resources at VA may help facilitate such work in the future.***' |
| 1 | This is a very thorough review. The authors should also consider framing their findings within the context of current VHA initiatives around knowledge translation and HRO, which are part of the VHA Long-range goal of innovating as a learning and teaching organization. For example, it might be good to see which program offices (e.g., QUERI training programs, NCPS, NCOD, etc) are training others or using some of the 'necessary, but not sufficient' conditions for successful implementation of CQI: human resources, provider (MD) involvement), sustained managerial focus, etc. etc | Please see the latter half of the above comment. |
| 2 | Minor comment on TEP: Dr. Godwin is not just at Baylor -- she is a VA researcher as well, and leads the National VA Quality Scholars Coordinating Center. Unintentionally may misrepresent her as unattached to VA. May want to consider adding her VA title and adding academic titles to other TEP members as well. Small thing is that you refer to some folks as being from the VA Greater LA Healthcare System but describe the ESP as not. May want to consider adding GLA to the WLA reference. | We have edited this section to reflect the appropriate titles, thank you! |
| 2 | Minor comment on Peer Reviewers: When I have been a TEP member on this and other ESP reports, it is not uncommon for me to be asked to peer review the report as well. So when I see the section on peer reviewers, it makes me wonder if those are indeed the same as TEP or a mix of TEP and non-TEP. Might be useful to clarify. | Thank you for your comment. ***We have added, 'Technical Panel Experts are often asked to also serve as peer reviewers, but we also invite experts who have not been involved with the current project to serve as peer reviewers.'*** |

| | | |
|---|---|---|
| 2 | A brief summary of what the improvements would be might be helpful to make it easier for other professionals in this space to refer to when designing, conducting, and writing up their work -- actually, given that such guidelines exist, the authors may want to add a sentence about re-emphasizing their use in the field. The report would benefit from a very careful re-read by one of the authors for typos/grammatical glitches. I worked to include all I saw but may not have caught them all. | We have added the following to the Research Gaps / Future Research section, ***'The overall evidence base would be improved if future reported individual QI studies more closely adhered to SQUIRE 2.0 reporting standards. Similarly, future systematic reviews of these works would be stronger by considering the AMSTAR2 categories we found to be often lacking, such as specifying comparators as an inclusion criteria and sharing the results of comparators from individual trials, PROSPERO registration, and explicitly discussing risk of bias.'*** |
| 2 | page 21, lines 46, reference to "trust pharmacies" and "trust systems" -- suspect these are from NHS type system but for a US readership, it would be helpful to add a definition or e.g., or something. Similarly, it would be helpful if the authors did a read-through/search to make sure that each abbreviation is spelled out on first use. There is inconsistency in how this is handled (e.g., ED vs. emergency department) -- many of these are not in the table of abbreviations, which is fine because you do not want readers to have to continue to go back to the table frequently. There are also variations in tense that warrant attention, e.g., page 22, line 7, two studies..."look at Lean strategies," while the next section is past tense "both review found..." | Your intuition was correct and we have added '***National Health Service'*** to clarify the language around trusts. Thank you for the other careful edits – we have clarified the abbreviations and abbreviations table and attempted to clear the remainder of the typos with this draft. |
| 2 | Ref #1 is incomplete. I did not go through the rest but omission on the 1st reference suggests all should be re-reviewed for completeness. | Thank you for your comment. We have reviewed all references. |
| 3 | I would also try to model the behavior the authors lamented being absent in some of the studies that they themselves review: put together an AMSTAR2 checklist and use it as a reporting standards checklist for this review of reviews (the same way one would include a PRISMA checklist for a meta-analysis | Thank you for your suggestion. We are unaware of the existence of such a tool for a review of reviews. We have included in our report the areas that may apply from the AMSTAR2 tool, such as registration of our protocol and duplicate data collection. However, as many aspects of AMSTAR2, such as the appropriateness of meta-analytic techniques and risk of bias for included primary studies, do not apply to a review of reviews, we have created a list. |
| 3 | If resources and the structure of the reporting mechanism allow, it might be nice to create a more interactive version of this Appendix E table by which users could sort and filter that could be made available in addition. | Thank you for your comment. While a very interesting idea, we are unable to add interactive features to the reports at this time. |
| 4 | When I first read Key Questions 1A and 1B, I was unclear on how they differed (and may still be a bit unclear now). It | This is correct. We attempted to clarify under Topic Development, as follows, ***'To be eligible for Key Question 1A, a systematic review*** |

| | | |
|---|---|---|
| | seems like Key Question 1B still involved looking for indirect comparisons of effectiveness (at least, conclusions by authors of the studied systematic reviews regarding differential evidence of effectiveness between Lean and Six Sigma). For example, on page 15 (part of question 1B), it is noted that "None of the 11 reviews that included more than one CQI strategy reached a strong conclusion that any strategy was superior to any other(s)." If I understand Key Question 1A correctly, eligible reviews had to have to actually have comparative effectiveness as their stated focus? If the authors could do more to clarify the criteria for deciding how a study became eligible for Key Question 1A as opposed to being in Key Question 1B, that would be helpful. | ***had to explicitly focus on comparative effectiveness of multiple continuous quality improvement methodologies as a stated aim. If a review commented on multiple methodologies but did not seek to compare the effectiveness of the strategies within its methods, it was included in Key Question 1B.***' |
| 4 | Given how frequently Lean and Six Sigma are discussed in the report, it would be helpful to provide a brief description of what constitutes a Lean approach or a Six Sigma approach. The directive to use Lean makes it important to define (in general terms) what constitutes a Lean approach, and since Six Sigma is one of the most popular alternatives to Lean, a brief definition of Six Sigma would be helpful as well. | We did not explicitly define each method for our search strategy, as we allowed each systematic review to categorize the included individual studies as the authors had chosen originally. This is now emphasized in the Study Selection section, ***'Specific definitions for each strategy were not included in our search strategy; rather, we allowed each systematic review to apply its own definitions to the included studies.***' |
| 7 | Pg 7, lines 56-59: A table listing the CQI strategies with a brief definition of each approach would enhance the readers understanding of each method. Appendix B does not include this information. | Please see above comment and response. We did not define strategies for this review of reviews and individual studies had considerable variability in their own definitions/discussion of strategies. |
| 7 | The big question is if any work has been done outside healthcare? The review was comprehensive but inconclusive. Is there knowledge from other fields that could inform healthcare? If yes, it should be briefly included in the review. If no, perhaps address why there is not work comparing CQI methods?<br><br>Omitted evidence<br>• Work outside of healthcare<br>• Non-indexed program reports | Please see our earlier comment about work outside healthcare having poor generalizability to the healthcare context. We have expanded comments about other VA work that may help to answer these questions in the future, in addition to the expanded Research Gaps previously mentioned. |

# APPENDIX E. EVIDENCE TABLE

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| Aij, 2017[8] | **Search end date:** 2016<br>**Number of included studies:** 32<br>**Quality score:** 2 | **Intervention:** Lean<br>**Setting**: Not reported<br>**Condition:** Not reported<br>**Geographical region:** Not reported | This work contributes to the development of a new framework for describing leadership attributes within lean management of health care. Originality/value – The summary of attributes can provide a model for health care leaders to apply lean in their organizations. |
| Amaratunga, 2016[9] | **Search end date:** June 9, 2015<br>**Number of included studies:** 23<br>**Quality score:** 8 | **Intervention:** Lean, Six Sigma, Lean/Six Sigma<br>**Setting**: Radiology<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | Of the 278 articles returned, 23 studies were suitable for inclusion. Of these, 10 assessed Six Sigma, 7 assessed Lean, and 6 assessed Lean Six Sigma. The diverse range of measured outcomes can be organized into 7 common aims: cost savings, reducing appointment wait time, reducing in-department wait time, increasing patient volume, reducing cycle time, reducing defects, and increasing staff and patient safety and satisfaction. All of the included studies demonstrated improvements across a variety of outcomes. However, there were high rates of systematic bias and imprecision as per the Grading of Recommendations Assessment, Development and Evaluation guidelines. |
| Andersen, 2014[42] | **Search end date:** 2012<br>**Number of included studies:** 18<br>**Quality score:** N/A | **Intervention:** Lean<br>**Setting**: Not reported<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | We provide a framework emphasising context by relating facilitators to domains and dimensions of capability. 23 factors enabling a successful lean intervention in hospitals were identified in the systematic review, where management and a supportive culture, training, accurate data, physicians and team involvement were most frequent. |
| Antony, 2018[40] | **Search end date:** 2016<br>**Number of included studies:** 68<br>**Quality score:** N/A | **Intervention:** Six Sigma<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | The findings of the systematic review reveal a growing interest in research on Six Sigma adoption in healthcare. The findings indicate that Six Sigma applications in healthcare have been focused on the entire hospital with no real focus on a particular department or function. The key findings on benefits, success factors, challenges and common tools of Six Sigma from the existing literature are also presented in the paper. |
| Bucci, 2016[10] | **Search end date:** NR<br>**Number of included studies:** 9<br>**Quality score:** 5 | **Intervention:** Lean<br>**Setting**: Emergency department<br>**Condition:** Not reported<br>**Geographical region:** | Nine before-and-after studies met these eligibility criteria. Management of patient flow was the main intervention. Almost all studies showed EDs performance improvement: increased patient volume, decreased length of stay and number of patients left without |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | | Varies across included studies | being seen, reduced costs, and increased patient satisfaction. Only 1 case reported worse results after Lean intervention implementation. |
| Côté, 2020[11] | **Search end date:** 2018<br>**Number of included studies:** 35<br>**Quality score:** 9 | **Intervention:** Clinical Microsystems<br>**Setting**: Varies across included studies<br>**Condition:** Varies across included studies<br>**Geographical region:** Varies across included studies | The impact of the project was most often measured using a single metric (59.1%) that was operational (*eg*, waiting time). Although most Lean project publications reported the use of tools to "break down the problem" (84.4%, Step 2) and "see countermeasures through" (70.0%, Step 6), fewer than half described using tools associated with each of the other steps. Projects completed an average of 2.77 steps and none of the projects completed all steps. Although some may perceive low adherence to the tenets of Lean as a deficiency, it may be that Lean approaches are evolving to better meet the needs of healthcare. |
| Crema, 2017[36] | **Search end date:** March 2014<br>**Number of included studies:** 16<br>**Quality score:** N/A | **Intervention:** Lean<br>**Setting**: Varies across included studies<br>**Condition:** Varies across included studies<br>**Geographical region:** Varies across included studies | Sixteen articles were included in the analysis. Links between the identified LHM [Lean Healthcare Management] purposes of adoption and CW objectives were discovered: through process understanding, optimization, evaluation and control, LHM contributes to the reduction of overuses in healthcare, but also to the delivery of a more effective and evidence-based care (EBC). Moreover, it provides an objective approach useful for choosing the most cost-effective solution among different alternatives. |
| D'Andreamatteo, 2015[39] | **Search end date:** September 2013<br>**Number of included studies:** 243<br>**Quality score:** N/A | **Intervention:** Lean<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | 243 articles were selected for analysis. Lean is best understood as a means to increase productivity. Hospital is the more explored setting, with emergency and surgery as the pioneer departments. The USA appears to be the leading country for number of applications. The theoretical works have been focused mainly on barriers, challenges, and success factors. Sustainability, framework for measurement and critical appraisal remain underestimated themes. Evaluations of "system wide approach" are still low in number. |
| Dellifraine, 2010[12] | **Search end date:** June 2009<br>**Number of included studies:** 34<br>**Quality score:** 5 | **Intervention:** Lean, Six Sigma, and Lean/Six Sigma<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Not reported | The authors identified 177 articles on SS/L published in the last 10 years. However, only 34 of them reported any outcomes of the SS/L projects studied, and less than one-third of these articles included statistical analyses to test for significant changes in outcomes. |
| DelliFraine, 2013[13] | **Search end date:** May 2012 | **Intervention:** Six Sigma<br>**Setting**: Varies across | The authors identified 310 articles on SS published in the last 15 years. However, only 55 were empirical peer-reviewed articles, 16 of |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | **Number of included studies:** 55<br>**Quality score:** 5 | included studies<br>**Condition:** Varies across included studies<br>**Geographical region:** Not reported | which reported the correct use of SS. Only 7 of these articles included statistical analyses to test for significant changes in quality of care, and only 16 calculated defects per million opportunities or sigma level. This review demonstrates that there are significant gaps in the Six Sigma health care quality improvement literature and very weak evidence that Six Sigma is being used correctly to improve health care quality. |
| Dzidowska, 2020[14] | **Search end date:** September 2018<br>**Number of included studies:** 56<br>**Quality score:** 9 | **Intervention:** CQI method<br>**Setting**: Primary care<br>**Condition:** Alcohol misuse<br>**Geographical region:** Varies across included studies | Fifty-six papers representing 45 projects were included. Of these, 24 papers were randomized controlled trials, 12 controlled studies and 20 before/after and other designs. Most reported on strategies for improving implementation of screening and brief intervention. Only 6 addressed relapse prevention pharmacotherapies. Only 5 reported on patient outcomes and none showed significant improvement. The 3 essential CQI method elements were clearly identifiable in 12 reports. More studies with 3 essential CQI method elements had implementation and follow-up durations above the median; utilised multifaceted designs; targeted both practice and health system levels; improved screening and brief intervention than studies without the CQI method elements. |
| Glasgow, 2010[15] | **Search end date:** July 15, 2010<br>**Number of included studies:** 37<br>**Quality score:** 5 | **Intervention:** Lean, Six Sigma, Lean Sigma<br>**Setting**: Varies across included studies<br>**Condition:** Varies across included studies<br>**Geographical region:** Not reported | Database searches identified 539 potential articles. After review of titles, abstracts, and full text, 47 articles met inclusion criteria: 10 articles summarized multiple projects, 12 reported Lean projects, 20 reported Six Sigma projects, and 5 reported Lean Sigma projects. Generally, the studies provided limited data, with only 15 articles providing any sort of follow-up data; of the 15, only 3 report a follow-up period greater than 2 years. |
| Hill, 2020[16] | **Search end date:** February 23, 2019<br>**Number of included studies:** 28<br>**Quality score:** 11 | **Intervention:** Continuous quality improvement vs non-continuous quality improvement<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | Twenty-eight RCTs assessed the effectiveness of different approaches to continuous quality improvement with a non-continuous quality improvement comparator in various settings, with interventions differing in terms of the approaches used, their duration, meetings held, people involved, and training provided. All RCTs were considered at risk of bias, undermining their results. Findings suggested that the benefits of continuous quality improvement compared to a comparator on clinical process, patient, and other outcomes were limited, with less than half of RCTs showing any effect. Where benefits were evident, it was usually on clinical process measures, with the model used (*ie*, Plan-Do-Study-Act, Model of |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | | | Improvement), the meeting type (*ie*, involving leaders discussing implementation) and their frequency (*ie*, weekly) having an effect. None considered socio-economic health inequalities. |
| Hulscher, 2013[17] | **Search end date:** Week 2 of June 2009<br>**Number of included studies:** 23<br>**Quality score:** 6 | **Intervention:** Quality Improvement Collaboratives<br>**Setting**: Varies across included studies<br>**Condition:** Varies across included studies<br>**Geographical region:** Not reported | Of 1367 abstracts identified, 23 papers (reporting on 26 collaboratives) provided information on potential determinants and their relationship with effectiveness. We categorised potential determinants of success using the definition for collaboratives as a template. Numerous potential determinants were tested, but only a few related to empirical effectiveness.Some aspects of teamwork and participation in specific collaborative activities enhanced short-term success. If teams remained intact and continued to gather data, chances of long-term success were higher. There is no empirical evidence of positive effects of leadership support, time, and resources. |
| Isfahani, 2019[18] | **Search end date:** 2016<br>**Number of included studies:** 26<br>**Quality score:** 4 | **Intervention:** Lean<br>**Setting**: Emergency department<br>**Condition:** Varies across included studies<br>**Geographical region:** Varies across included studies | According to the studies, responsibility of organization's senior management and his/her supports; increasing the knowledge of the characteristics and dimension of lean among the providers of health service; and decreasing the resistance and consulting with external counselors can have great effect on the success of lean management |
| Leggat, 2015[19] | **Search end date:** April 2013<br>**Number of included studies:** 41<br>**Quality score:** 3 | **Intervention:** process redesign methodologies<br>**Setting**: Hospital<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | Success factors for the changes included mechanisms to facilitate participation throughout the process, clearly documented protocols and expectations for the health professionals which were supported by education, mechanisms to audit and provide feedback on behaviours and performance, as well as being able to hold the participating health professionals accountable. The success of process redesign methodologies is found to be highly dependent on these performance-based human resource management (HRM) practices. |
| Maijala, 2018[37] | **Search end date:** February 2016<br>**Number of included studies:** 12<br>**Quality score:** N/A | **Intervention:** Lean<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** | Considering the results using the windshield concept emphasizes the philosophy, principles, and tools of Lean thinking. Lean leadership and management factors in health care were mainly conceptualized as skills and capabilities such as problem solving, making changes occur, empowering, communicating, coaching, supporting, facilitating, being democratic, organizational learning, and organizational |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | | Varies across included studies | success, all of which represented middle-stage or advanced managerial skills and capabilities. |
| Magalhães, 2016[20] | **Search end date:** March 2015<br>**Number of included studies:** 47<br>**Quality score:** 3 | **Intervention:** Lean<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | This study enabled us to show that Lean thinking in health is a management model that improves the structure, process and outcome, from the care and management actions. The principles of Lean thinking are widespread in various contexts of health, such as emergency, oncology, pharmacy, intensive care unit, radiology, orthopedics, mental health clinics, and cardiology services. The main impacts from the application of this thinking in health are increasing productivity and team efficiency; reduction in waiting time for patient care; standardization of care process, reducing costs, improved teamwork, reduction in the patient's hospital length; increasing the quality of service provided; increased patient satisfaction; increasing patient safety and health professionals; and employee satisfaction. |
| Mason, 2015[21] | **Search end date:** January 1, 2014<br>**Number of included studies:** 23<br>**Quality score:** 7 | **Intervention:** Lean, Six Sigma, Lean/Six Sigma<br>**Setting**: Surgical inpatients, ORs, or outpatient surgery clinics<br>**Condition:** Surgical patients<br>**Geographical region:** Varies across included studies | Of the 124 studies returned, 23 were suitable for inclusion, with 11 assessing Lean, 6 Six Sigma, and 6 Lean Six Sigma. The broad range of outcomes can be collated into 6 common aims: to optimise outpatient efficiency, to improve operating theatre efficiency, to decrease operative complications, to reduce ward-based harms, to reduce mortality, and to limit unnecessary cost and length of stay. The majority of studies (88%) demonstrate improvement; however, high levels of systematic bias and imprecision were evident. |
| Mazzocato, 2010[38] | **Search end date:** February 2008<br>**Number of included studies:** 33<br>**Quality score:** N/A | **Intervention:** Lean<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Not reported | The authors reviewed 33 articles and found a wide range of lean applications. The articles describe initial implementation stages and emphasise technical aspects. All articles report positive results. The authors found common contextual aspects which interact with different components of the lean interventions and trigger 4 different change mechanisms: understand processes to generate shared understanding; organise and design for effectiveness and efficiency; improve error detection to increase awareness and process reliability; and collaborate to systematically solve problems to enhance continual improvement. |
| Moraros, 2015[22] | **Search end date:** NR<br>**Number of included studies:** 22 | **Intervention:** Lean<br>**Setting**: Varies across included studies | Our electronic search identified 22 articles that passed methodological quality review. Among the accepted studies, 4 were exclusively concerned with health outcomes, 3 included both health |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | **Quality score:** 5 | **Condition:** Varies across included studies<br>**Geographical region:** Varies across included studies | and process outcomes, and 15 included process outcomes. Our study found that Lean interventions have: (i) no statistically significant association with patient satisfaction and health outcomes; (ii) a negative association with financial costs and worker satisfaction, and (iii) potential, yet inconsistent, benefits on process outcomes like patient flow and safety. |
| Nadeem, 2013[23] | **Search end date:** April 2012<br>**Number of included studies:** 24<br>**Quality score:** 6 | **Intervention:** Quality Improvement Collaboratives<br>**Setting**: Varies across included studies<br>**Condition:** Varies across included studies<br>**Geographical region:** Varies across included studies | We found 14 crosscutting components as common ingredients in health care QI collaboratives (*eg*, in-person learning sessions, phone meetings, data reporting, leadership involvement, and training in QI methods). The collaboratives reported included, on average, 6-7 of these components. The most common were in-person learning sessions, plan-do-study-act (PDSA) cycles, multidisciplinary QI teams, and data collection for QI. The outcomes data from these studies indicate the greatest impact of QI collaboratives at the provider level; patient-level findings were less robust. |
| Nicolay, 2012[24] | **Search end date:** November 24, 2010<br>**Number of included studies:** 34<br>**Quality score:** 10 | **Intervention:** Lean, Lean Six Sigma, Six Sigma, Total Quality Management, CQI method, Statistical Process Control or Statistical Quality Control, PDSA/PDCA<br>**Setting**: Surgical settings<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | Some 34 of 1595 articles identified met the inclusion criteria after consensus from 2 independent investigators. Nine studies described continuous quality improvement (CQI method), 5 Six Sigma, 5 total quality management (TQM), 5 plan-do-study-act (PDSA) or plan-do-check-act (PDCA) cycles, 5 statistical process control (SPC) or statistical quality control (SQC), 4 Lean, and 1 Lean Six Sigma; 20 of the studies were undertaken in the USA. The most common aims were to reduce complications or improve outcomes (11), to reduce infection (7), and to reduce theatre delay (7). There was 1 randomized controlled trial. |
| Niñerola, 2019[41] | **Search end date:** 2017<br>**Number of included studies:** 196<br>**Quality score:** N/A | **Intervention:** Six Sigma<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Not reported | (1) Six Sigma publications in healthcare sector have been carried out mostly in the USA, (2) multiple specialities and services have used this tool; among them, we can emphasize the operating room and radiology service, (3) the case study has been the most used methodology and, (4) the objectives are focused mainly on achieving reductions of time, costs and errors, for the improvement of the quality and the satisfaction of the patients. |
| Nunes, 2016[25] | **Search end date:** October 13, 2014<br>**Number of included studies:** 76 | **Intervention:** CQI method<br>**Setting**: Nephrology<br>**Condition:** Kidney disease<br>**Geographical region:** | We initially identified 468 publications; 40 were excluded as duplicates or not available/not in English. An additional 352 did not meet criteria for full review due to: 1. Not meeting criteria for inclusion = 196 (*eg*, reviews, news articles, editorials) 2. Not nephrology- |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | **Quality score:** 5 | Varies across included studies | specific = 153, 3. Only available as abstracts = 3. Of 76 publications meeting criteria for full review, the majority [45 (61%)] focused on ESRD care. 74% explicitly stated use of specific continuous quality improvement tools in their methods. The highest number of publications in a given year occurred in 2011 with 12 (16%) articles. 89% of studies were found in biomedical and allied health journals and most studies were performed in North America (52%). Only 1 was randomized and controlled although not blinded. |
| Powell, 2008[1] | **Search end date:** NR<br>**Number of included studies:** 59<br>**Quality score:** 2 | **Intervention:** TQM/continuous quality improvement, IHI/PDSA, Lean, SS<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | The 5 models are described and evaluated separately in order to bring out some important conceptual differences between them. In practice, however, distinctions between the models are not always clear-cut: there are many areas of overlap, with many of the approaches employing very similar tools and techniques. In implementation, health care organisations have tended to apply a combination of tools and approaches in a piecemeal and eclectic way. What the models do have in common is that they all require the same broad set of 'necessary, but not sufficient' conditions for successful implementation. These conditions emerge strongly from the studies reviewed in this report as well as from the broader literature on health service change. They include: the active engagement of health professionals, especially doctors; the active participation of middle and senior managers, and the support of board members; the use of multifaceted interventions and sustained action at different levels of the health care system; the alignment of quality improvement activities with the strategic goals of the organisation; and the embedding of quality improvement as an integral part of the everyday work of all staff (rather than as the responsibility of a separate directorate or team). Effective quality improvement work – whatever the model that structures this work – also needs to be supported by robust IT providing timely local data, and requires significant investment in staff training and development. |
| Schouten, 2008[26] | **Search end date:** June 2006<br>**Number of included studies:** 9<br>**Quality score:** 10 | **Intervention:** Quality Improvement Collaboratives<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Not reported | Of 1104 articles identified, 72 were included in the study. Twelve reports representing 9 studies (including 2 randomised controlled trials) used a controlled design to measure the effects of the quality improvement collaborative intervention on care processes or outcomes of care. Systematic review of these 9 studies showed moderate positive results. Seven studies (including 1 randomised controlled trial) reported an effect on some of the selected outcome |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | | | measures. Two studies (including 1 randomised controlled trial) did not show any significant effect. |
| Talib, 2011[27] | **Search end date:** 2009<br>**Number of included studies:** 15<br>**Quality score:** 2 | **Intervention:** Total Quality Management (TQM)<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | Eight supporting TQM practices, such as top-management commitment, teamwork and participation, process management, customer focus and satisfaction, resource management, organization behavior and culture, continuous improvement, and training and education were identified as best practices for TQM implementation in any health care setting. |
| Tlapa, 2020[29] | **Search end date:** December 2018<br>**Number of included studies:** 40<br>**Quality score:** 8 | **Intervention:** Lean, Lean Six Sigma<br>**Setting**: Ambulatory & ED settings<br>**Condition:** Not reported<br>**Geographical region:** Varies across included studies | According to our findings, LH [lean healthcare] helped to reduce waiting time and LOS in ambulatory care, mainly owing to its focus on identifying and minimizing non-value added (NVA) activities. Nevertheless, evidence of the impact of LH on patient/ staff satisfaction and the translation of the obtained benefits into savings is scarce among studies. |
| Trakulsunti, 2018[30] | **Search end date:** December 2016<br>**Number of included studies:** 24<br>**Quality score:** 5 | **Intervention:** Lean, Six Sigma, Lean Six Sigma<br>**Setting**: Hospital<br>**Condition:** Medication errors<br>**Geographical region:** Varies across included studies | A total of 24 studies were identified from the search meeting the criteria for the systematic literature review. Increased interest in the application of process excellence methodologies such as Lean, Six Sigma, and Lean Six Sigma to reduce medication errors, especially from the developed countries, was found. Several themes have emerged in this paper including: tools and techniques of Lean and Six Sigma in the context of medication errors, Lean and Six Sigma methodology, types of medication errors, Lean Six Sigma project selection, benefits, challenges, and success factors. The study is expected to benefit health care practitioners in implementing the Lean Six Sigma methodology to reduce medication errors. |
| Tricco, 2012[31] | **Search end date:** July 2010<br>**Number of included studies:** 142<br>**Quality score:** 9 | **Intervention:** Compares CQI method to other methods<br>**Setting**: Varies across included studies<br>**Condition:** Diabetes<br>**Geographical region:** | We reviewed 48 cluster randomised controlled trials, including 2538 clusters and 84 865 patients, and 94 patient randomised controlled trials, including 38 664 patients. In random effects meta-analysis, the QI strategies reduced HbA1c by a mean difference of 0·37% (95% CI 0·28–0·45; 120 trials), LDL cholesterol by 0·10 mmol/L (0·05–0.14; 47 trials), systolic blood pressure by 3·13 mm Hg (2·19–4·06, 65 trials), and diastolic blood pressure by 1·55 mm Hg (0·95–2·15, 61 trials) versus usual care. We noted larger effects when baseline |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | | Varies across included studies | concentrations were greater than 8·0% for HbA1c, 2·59 mmol/L for LDL cholesterol, and 80 mm Hg for diastolic and 140 mm Hg for systolic blood pressure. The effectiveness of QI strategies varied depending on baseline HbA1c control. QI strategies increased the likelihood that patients received aspirin (11 trials; relative risk [RR] 1·33, 95% CI 1·21–1·45), antihypertensive drugs (10 trials; RR 1·17, 1·01–1·37), and screening for retinopathy (23 trials; RR 1·22, 1·13–1·32), renal function (14 trials; RR 128, 1·13–1·44), and foot abnormalities (22 trials; RR 1·27, 1·16–1·39). However, statin use (10 trials; RR 1·12, 0·99–1·28), hypertension control (18 trials; RR 1·01, 0·96–1·07), and smoking cessation (13 trials; RR 1·13, 0·99–1·29) were not significantly increased. |
| Wackerbarth, 2021[32] | **Search end date:** NR<br>**Number of included studies:** 295<br>**Quality score:** 4 | **Intervention:** Lean<br>**Setting**: Varies across included studies<br>**Condition:** Not reported<br>**Geographical region:** Not reported | The most frequent type of publication reported empirical research (48.6%) and most of these (80.3%) shared the results of the Lean projects. Of the 237 publications reporting Lean projects, more than half (71.3%) used an experimental, one-site, pre/postdesign. The impact of the project was most often measured using a single metric (59.1%) that was operational (*eg*, waiting time). Although most Lean project publications reported the use of tools to "break down the problem" (84.4%, Step 2) and "see countermeasures through" (70.0%, Step 6), fewer than half described using tools associated with each of the other steps. Projects completed an average of 2.77 steps and none of the projects completed all steps. Although some may perceive low adherence to the tenets of Lean as a deficiency, it may be that Lean approaches are evolving to better meet the needs of healthcare. |
| Woodnutt, 2018[33] | **Search end date:** NR<br>**Number of included studies:** 12<br>**Quality score:** 4 | **Intervention:** Lean<br>**Setting**: Hospital<br>**Condition:** Varies across included studies<br>**Geographical region:** England | Lean has ostensible value but it is difficult to draw a conclusion on efficacy or sustainability. Higher-quality scientific research into Lean and the effect of staffing cultures on initiatives are needed to ascertain the extent that Lean can affect health care quality and subsequently be sustained. |
| Zamboni, 2020[34] | **Search end date:** June 2018<br>**Number of included studies:** 32<br>**Quality score:** 7 | **Intervention:** Quality Improvement Collaboratives<br>**Setting**: Varies across included studies<br>**Condition:** Varies across included studies | We screened 962 abstracts of which 88 met the inclusion criteria, and we retained 32 for analysis. Adequacy and appropriateness of external support, functionality of quality improvement teams, leadership characteristics and alignment with national systems and priorities may influence outcomes of quality improvement collaboratives, but the strength and quality of the evidence is weak. |

| Author, year | Description of systematic review | Description of intervention* | Findings relevant to continuous quality improvement |
|---|---|---|---|
| | | **Geographical region:** Varies across included studies | Participation in quality improvement collaborative activities may improve health professionals' knowledge, problem-solving skills and attitude; teamwork; shared leadership and habits for improvement. Interaction across quality improvement teams may generate normative pressure and opportunities for capacity building and peer recognition. |
| Zepeda-Lugo, 2020[35] | **Search end date:** 2019<br>**Number of included studies:** 39<br>**Quality score:** 9 | **Intervention:** Lean, Lean Six Sigma<br>**Setting**: Varies across included studies<br>**Condition:** Varies across included studies<br>**Geographical region:** Varies across included studies | Our findings show that by focusing on reducing non-value-added activities, LH contributed to improving patient flow and efficiency within inpatient care. |

*Only the study by Hill et al sufficiently described comparators to receive credit in our modified AMSTAR review.

# APPENDIX F. CITATIONS FOR EXCLUDED PUBLICATIONS

Does Not Discuss Key Questions, n=69


1. Agarwal, A., et al., Hospital-based quality improvement interventions for patients with heart failure: a systematic review. Heart (British Cardiac Society), 2019. 105(6): p. 431-438.
2. Alagoz, E., et al., The use of external change agents to promote quality improvement and organizational change in healthcare organizations: a systematic review. BMC health services research, 2018. 18(1): p. 42.
3. Antierens, A., et al., How much of Toyota's philosophy is embedded in health care at the organisational level? A review. Journal of Nursing Management (John Wiley & Sons, Inc.), 2018. 26(4): p. 348-357.
4. Atassi, K., Strategies to increase colorectal cancer screening. Nurse Practitioner, 2012. 37(7): p. 21-26.
5. Bahiru, E., et al., Hospital-Based Quality Improvement Interventions for Patients With Acute Coronary Syndrome: A Systematic Review. Circulation. Cardiovascular quality and outcomes, 2019. 12(9): p. e005513.
6. Bakker, F.C., S.H.M. Robben, and M.G.M. Olde Rikkert, Effects of hospital-wide interventions to improve care for frail older inpatients: a systematic review. BMJ quality & safety, 2011. 20(8): p. 680-691.
7. Bastemeijer, C.M., et al., Patient experiences: a systematic review of quality improvement interventions in a hospital setting. Patient related outcome measures, 2019. 10: p. 157-169.
8. Bergerum, C., et al., How might patient involvement in healthcare quality improvement efforts work—A realist literature review. Health Expectations, 2019. 22(5): p. 952-964.
9. Beyranvand, T., A. Aryankhesa, and A.A. Hashjin, Quality improvement in hospitals' surgery-related processes: A systematic review. Journal of Medical Council of Islamic Republic of Iran, 2019. 33(1): p. 1-9.
10. Chakraborty, A., et al., The effectiveness of in-hospital interventions on reducing hospital length of stay and readmission of patients with Type 2 Diabetes Mellitus: a systematic review. Diabetes research and clinical practice, 2020: p. 108363.
11. Chaudhuri, D., et al., Effectiveness of Quality Improvement Interventions at Reducing Inappropriate Cardiac Imaging: A Systematic Review and Meta-Analysis. Circulation. Cardiovascular quality and outcomes, 2016. 9(1): p. 7-13.
12. Clay-Williams, R., et al., Do large-scale hospital- and system-wide interventions improve patient outcomes: a systematic review. BMC health services research, 2014. 14: p. 369.
13. Cleveringa, F.G.W., et al., Computerized decision support systems in primary care for type 2 diabetes patients only improve patients' outcomes when combined with feedback on performance and case management: a systematic review. Diabetes Technology & Therapeutics, 2013. 15(2): p. 180-192.
14. Conry, M.C., et al., A 10 year (2000-2010) systematic review of interventions to improve quality of care in hospitals. BMC health services research, 2012. 12: p. 275.

15. Coory, M., et al., Systematic review of quality improvement interventions directed at cancer specialists. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 2013. 31(12): p. 1583-1591.
16. Crema, M. and C. Verbano, Investigating the connections between health lean management and clinical risk management. International Journal of Health Care Quality Assurance (09526862), 2015. 28(8): p. 791-811.
17. Curatolo, N., et al., A critical analysis of Lean approach structuring in hospitals. Business Process Management Journal, 2014.
18. Davidson, K., A. Perry, and L. Bell, Would continuous feedback of patient's clinical outcomes to practitioners improve NHS psychological therapy services? Critical analysis and assessment of quality of existing studies. Psychology & Psychotherapy: Theory, Research & Practice, 2015. 88(1): p. 21-37.
19. Deblois, S. and L. Lepanto, Lean and Six Sigma in acute care: a systematic review of reviews. International journal of health care quality assurance, 2016. 29(2): p. 192-208.
20. Dilley, J.A., B. Bekemeier, and J.R. Harris, Quality improvement interventions in public health systems: a systematic review. American journal of preventive medicine, 2012. 42(5 Suppl 1): p. S58-71.
21. Donnellan, D., et al., The effect of thermoregulation quality improvement initiatives on the admission temperature of premature/very low birth-weight infants in neonatal intensive care units: A systematic review. Journal for Specialists in Pediatric Nursing, 2020. 25(2): p. 1-13.
22. Fry, T.J., S. Marfurt, and S. Wengier, Systematic Review of Quality Improvement Initiatives Related to Cue-Based Feeding in Preterm Infants. Nursing for women's health, 2018. 22(5): p. 401-410.
23. Gallagher, H., et al., Quality-improvement strategies for the management of hypertension in chronic kidney disease in primary care: a systematic review. The British journal of general practice : the journal of the Royal College of General Practitioners, 2010. 60(575): p. e258-265.
24. Gao, T. and B. Gurd, Organizational issues for the lean success in China: exploring a change strategy for lean success. BMC health services research, 2019. 19(1): p. 66.
25. Gettel, C.J., N. Pertsch, and E.M. Goldberg, A Systematic Review of Interventions to Improve Nursing Home to Emergency Department Care Transitions. Annals of Long Term Care, 2020. 28(2): p. e12-s4.
26. Green, T., et al., Use and reporting of experience-based codesign studies in the healthcare setting: a systematic review. BMJ Quality & Safety, 2020. 29(1): p. 64-76.
27. Hansen, L.O., et al., Interventions to reduce 30-day rehospitalization: a systematic review. Annals of internal medicine, 2011. 155(8): p. 520-528.
28. Healy, H., et al., A systematic review of reports of quality improvement for bronchopulmonary dysplasia. Seminars in fetal & neonatal medicine, 2021: p. 101201.
29. Henrique, D.B., & Godinho Filho, M., A systematic literature review of empirical research in Lean and Six Sigma in healthcare. Total Quality Management & Business Excellence, 2018: p. 1-21.

30. Irwin, R., T. Stokes, and T. Marshall, Practice-level quality improvement interventions in primary care: a review of systematic reviews. Primary Health Care Research & Development (Cambridge University Press / UK), 2015. 16(6): p. 556-577.

31. Ivers, N., et al., Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database of Systematic Reviews, 2012(6).

32. Jabin, M.S.R., et al., A Mixed-Methods Systematic Review of the Effectiveness and Experiences of Quality Improvement Interventions in Radiology. Journal of patient safety, 2020.

33. Jesus, T.S., et al., Key Characteristics of Rehabilitation Quality Improvement Publications: Scoping Review From 2010 to 2016. Archives of Physical Medicine & Rehabilitation, 2018. 99(6): p. 1141-1148.e4.

34. Kampstra, N.A., et al., Health outcomes measurement and organizational readiness support quality improvement: a systematic review. BMC health services research, 2018. 18(1): p. 1005.

35. Kelendar, H., et al., The use of lean methodology in healthcare settings in developing countries: a narrative review. British Journal of Healthcare Management, 2020. 26(6): p. 1-13.

36. Keller, S.C., et al., Ambulatory Antibiotic Stewardship through a Human Factors Engineering Approach: A Systematic Review. Journal of the American Board of Family Medicine : JABFM, 2018. 31(3): p. 417-430.

37. Knudsen, S.V., et al., Can quality improvement improve the quality of care? A systematic review of reported effects and methodological rigor in plan-do-study-act projects. BMC health services research, 2019. 19(1): p. 683.

38. Kourtis, S.A. and J.P. Burns, Quality improvement in pediatric intensive care: A systematic review of the literature. Pediatric investigation, 2019. 3(2): p. 110-116.

39. Lamontagne, M.-E., et al., A Scoping Review of Clinical Practice Improvement Methodology Use in Rehabilitation. Rehabilitation Process & Outcome, 2016(5): p. 1-11.

40. Marang-van de Mheen, P.J. and L. van Bodegom-Vos, Meta-analysis of the central line bundle for preventing catheterrelated infections: a case study in appraising the evidence in quality improvement. BMJ Quality & Safety, 2016. 25(2): p. 118-129.

41. Maurer, N.R., T.H. Hogan, and D.M. Walker, Hospital- and System-Wide Interventions for Health Care-Associated Infections: A Systematic Review. Medical care research and review : MCRR, 2020: p. 1077558720952921.

42. McPheeters, M.L., et al., Quality improvement interventions to address health disparities: closing the quality gap - revisiting the state of the science. 2012: Centre for Reviews and Dissemination (UK).

43. Mery, G., et al., Evaluating investment in quality improvement capacity building: a systematic review. BMJ open, 2017. 7(2): p. e012431.

44. Mileski, M., et al., An investigation of quality improvement initiatives in decreasing the rate of avoidable 30-day, skilled nursing facility-to-hospital readmissions: a systematic review. Clinical interventions in aging, 2017. 12: p. 213-222.

45. Mohandas, A., et al., Role of Quality Improvement in Prevention of Inappropriate Transfusions. Quality Management in Health Care, 2011. 20(4): p. 298-310.

46. Nayak, S. and S.L. Greenspan, How Can We Improve Osteoporosis Care? A Systematic Review and Meta-Analysis of the Efficacy of Quality Improvement Strategies for Osteoporosis. Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research, 2018. 33(9): p. 1585-1594.

47. Neo, J.R.J., et al., Evidence-Based Strategies in Using Persuasive Interventions to Optimize Antimicrobial Use in Healthcare: a Narrative Review. Journal of Medical Systems, 2020. 44(3): p. 1-13.

48. Ng, G.K.B., et al., Factors affecting implementation of accreditation programmes and the impact of the accreditation process on quality improvement in hospitals: a SWOT analysis. Hong Kong Medical Journal = Xianggang Yi Xue Za Zhi, 2013. 19(5): p. 434-446.

49. Nuckols, T.K., et al., Economic Evaluation of Quality Improvement Interventions Designed to Improve Glycemic Control in Diabetes: A Systematic Review and Weighted Regression Analysis. Diabetes care, 2018. 41(5): p. 985-993.

50. Nuckols, T.K., et al., Economic Evaluation of Quality Improvement Interventions for Bloodstream Infections Related to Central Catheters: A Systematic Review. JAMA internal medicine, 2016. 176(12): p. 1843-1854.

51. Olisemeke, B., et al., The Effectiveness of Service Delivery Initiatives at Improving Patients' Waiting Times in Clinical Radiology Departments: A Systematic Review. Journal of Digital Imaging, 2014. 27(6): p. 751-778.

52. Parikh, K., S. Keller, and S. Ralston, Inpatient Quality Improvement Interventions for Asthma: A Meta-analysis. Pediatrics, 2018. 141(5): p. 1-15.

53. Patel, P. and G. Mitera, A Systematic Scoping Literature Review of Incorporating a Total Quality Culture Within Radiotherapy Staffing Models: A Management Strategy to Improve Patient Safety and Quality of Care in Radiation Therapy Departments. Journal of Medical Imaging & Radiation Sciences, 2011. 42(2): p. 81-85.

54. Peimbert-García, R.E., Analysis and Evaluation of Reviews on Lean and Six Sigma in Health Care. Quality Management in Health Care, 2019. 28(4): p. 229-236.

55. Perla, R.J., E. Bradbury, and C. Gunther-Murphy, Large-Scale Improvement Initiatives in Healthcare: A Scan of the Literature. Journal for Healthcare Quality: Promoting Excellence in Healthcare, 2013. 35(1): p. 30-40.

56. Reynolds, C., E. Esrailian, and D. Hommes, Quality Improvement in Gastroenterology: A Systematic Review of Practical Interventions for Clinicians. Digestive diseases and sciences, 2018. 63(10): p. 2507-2518.

57. Sales, A.E., et al., The use of data for process and quality improvement in long term care and home care: a systematic review of the literature. Journal of the American Medical Directors Association, 2012. 13(2): p. 103-113.

58. Schouten, L.M., R.P. Grol, and M.E. Hulscher, Factors influencing success in quality-improvement collaboratives: development and psychometric testing of an instrument. Implement Sci, 2010. 5: p. 84.

59. Silver, S.A., et al., Effectiveness of Quality Improvement Strategies for the Management of CKD: A Meta-Analysis. Clinical journal of the American Society of Nephrology : CJASN, 2017. 12(10): p. 1601-1614.
60. Soban, L.M., et al., Preventing pressure ulcers in hospitals: a systematic review of nurse-focused quality improvements interventions. Joint Commission Journal on Quality & Patient Safety, 2011. 37(6): p. 245-252.
61. Terens, N., et al., Quality improvement strategies at primary care level to reduce inequalities in diabetes care: an equity-oriented systematic review. BMC endocrine disorders, 2018. 18(1): p. 31.
62. Tricco, A.C., et al., Effectiveness of quality improvement strategies for coordination of care to reduce use of health care services: a systematic review and meta-analysis. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne, 2014. 186(15): p. E568-578.
63. Vos, L., et al., Towards an organisation-wide process-oriented organisation of care: a literature review. Implement Sci, 2011. 6: p. 8.
64. Walker, C., Strategies for Improving Patient Throughput in an Acute Care Setting Resulting in Improved Outcomes: A Systematic Review. Nursing Economic$, 2016. 34(6): p. 277-288.
65. Ward, K., et al., Strategies to improve vaccination uptake in Australia, a systematic review of types and effectiveness...[corrected] [published erratum appears in AUST NZ J PUBLIC HEALTH 2012; 36(5):490]. Australian & New Zealand Journal of Public Health, 2012. 36(4): p. 369-377.
66. Wells, S., et al., Are quality improvement collaboratives effective? A systematic review. BMJ quality & safety, 2018. 27(3): p. 226-240.
67. Williams, S., S. Keogh, and C. Douglas, Improving paediatric pain management in the emergency department: An integrative literature review. International Journal of Nursing Studies, 2019. 94: p. 9-20.
68. Worswick, J., et al., Improving quality of care for persons with diabetes: an overview of systematic reviews - what does the evidence tell us? Systematic reviews, 2013. 2: p. 26.
69. Xie, A. and P. Carayon, A systematic review of human factors and ergonomics (HFE)-based healthcare system redesign for quality of care and patient safety. Ergonomics, 2015. 58(1): p. 33-49.

Not a Systematic Review, n=21

1. Beyranvand, T., A. Aryankhesal, and A. Aghaei Hashjin, Quality improvement in hospitals' surgery-related processes: A systematic review. Medical journal of the Islamic Republic of Iran, 2019. 33: p. 129.
2. Brackett, T., L. Comer, and R. Whichello, Do Lean Practices Lead to More Time at the Bedside? Journal for Healthcare Quality: Promoting Excellence in Healthcare, 2013. 35(2): p. 7-14.
3. Brennan, S., et al., Continuous quality improvement: effects on professional practice and healthcare outcomes. Cochrane Database of Systematic Reviews, 2009(4).

4. Costa, L.B.M. and M. Godinho Filho, Lean healthcare: review, classification and analysis of literature. Production Planning & Control, 2016. 27(10): p. 823-836.
5. De Souza, L.B., Trends and approaches in lean healthcare. Leadership in health services, 2009.
6. de Souza, L.B. and M. Pidd, Exploring the barriers to lean health care implementation. Public Money & Management, 2011. 31(1): p. 59-66.
7. Hewson-Conroy, K.M., D. Elliott, and A.R. Burrell, Quality and safety in intensive care--A means to an end is critical. Australian Critical Care, 2010. 23(3): p. 109-129.
8. Holden, R.J., Lean Thinking in emergency departments: a critical review. Ann Emerg Med, 2011. 57(3): p. 265-78.
9. Honeycutt, L.C. and S.D. Keller, Effectiveness of the Lean process compared to other quality improvement initiatives on length of stay and wait times in healthcare organizations: a systematic review protocol. JBI database of systematic reviews and implementation reports, 2018. 16(1): p. 12-20.
10. Kaplan, H.C., et al., The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. BMJ Qual Saf, 2012. 21(1): p. 13-20.
11. Kim, C.S., et al., Implementation of lean thinking: one health system's journey. Jt Comm J Qual Patient Saf, 2009. 35(8): p. 406-13.
12. Kollberg, B., J.J. Dahlgaard, and P.-O. Brehmer, Measuring lean initiatives in health care services: issues and findings. International Journal of Productivity and Performance Management, 2006. 56(1): p. 7-24.
13. Lukas, C.V., et al., Transformational change in health care systems: an organizational model. Health Care Manage Rev, 2007. 32(4): p. 309-20.
14. McIntosh, B. and G. Cookson, Lean management in the NHS: fad or panacea. British Journal of Healthcare Management, 2012. 18(3): p. 130-135.
15. Morrow, E., et al., Emerald article: implementing large-scale quality improvement: lessons from The Productive Ward: Releasing Time to Care. 2012.
16. Poksinska, B., The current state of Lean implementation in health care: literature review. Qual Manag Health Care, 2010. 19(4): p. 319-29.
17. Radnor, Z.J., M. Holweg, and J. Waring, Lean in healthcare: the unfilled promise? Soc Sci Med, 2012. 74(3): p. 364-371.
18. Vest, J.R. and L.D. Gamm, A critical review of the research literature on Six Sigma, Lean and StuderGroup's Hardwiring Excellence in the United States: the need to demonstrate and communicate the effectiveness of transformation strategies in healthcare. Implement Sci, 2009. 4: p. 35.
19. Vituri, D.W. and Y.D. Martinez Évora, Total Quality Management and hospital nursing: an integrative literature review. Revista Brasileira de Enfermagem, 2015. 68(5): p. 660-667.
20. Walshe, K., Pseudoinnovation: the development and spread of healthcare quality improvement methodologies. Int J Qual Health Care, 2009. 21(3): p. 153-9.

21. Walshe, K. and T. Freeman, Effectiveness of quality improvement: learning from evaluations. Qual Saf Health Care, 2002. 11(1): p. 85-7.

Unavailable, n=9

1. Abdallah, A.B. and R.Z. Alkhaldi, Lean bundles in health care: a scoping review. Journal of Health Organization & Management, 2019. 33(4): p. 488-510.
2. Cacciatore, P., et al., Lean thinking in the hospital setting - Results of a systematic review of literature. Igiene e sanita pubblica, 2019. 75(1): p. 29-50.
3. DelliFraine, J.L., et al., The use of six sigma in health care management: are we using it to its full potential? Quality Management in Health Care, 2013. 22(3): p. 210-223.
4. DelliFraine, J.L., et al., The Use of Six Sigma in Health Care Management: Are We Using It to Its Full Potential? Quality Management in Health Care, 2014. 23(4): p. 240-253.
5. Hayward, L.M., How applicable is lean in mental health? A critical appraisal. International Journal of Clinical Leadership, 2012. 17(3): p. 166-173.
6. Khan, C. and G. Ollenschläger, [Effectiveness of quality assurance programmes for inpatient care in Germany: a literature review]. Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen, 2014. 108(10): p. 576-586.
7. Ralston, S., et al., Effectiveness of quality improvement in hospitalization for bronchiolitis: a systematic review. Pediatrics, 2014. 134(3): p. 571-581.
8. Taylor, M.J., et al., Systematic review of the application of the plan--do--study--act method to improve quality in healthcare. BMJ Quality & Safety, 2014. 23(4): p. 290-298.
9. Vecchi, S., et al., [Audit and feedback, and continuous quality improvement strategies to improve the quality of care for type 2 diabetes: a systematic review of literature]. Epidemiologia e prevenzione, 2016. 40(3-4): p. 215-223.

Duplicate, n=1

1. Wright Nunes, Julie, F. Jacob Seagull, Rao Panduranga, Jonathan H. Segal, Nandita S. Mani, Michael Heung, Julie Wright Nunes, and Panduranga Rao. 2016.'Continuous quality improvement in nephrology: a systematic review', BMC nephrology, 17: 190-90.