Key characteristics of successful quality improvement curricula in physician education: a realist review

Anne C Jones,1,2,3,4 Scott A Shipman,2,3,5 Greg Ogrinc1,2

ABSTRACT

Purpose Quality improvement (QI) is a common competency that must be taught in all physician training programmes, yet, there is no clear best approach to teach this content in clinical settings. We conducted a realist systematic review of the existing literature in QI curricula within the clinical setting, highlighting examples of trainees learning QI by doing QI.

Method Candidate theories describing successful QI curricula were articulated a priori. We searched MEDLINE (1 January 2000 to 12 March 2013), the Cochrane Library (2013) and Web of Science (15 March 2013) and reviewed references of prior systematic reviews. Inclusion criteria included study design, setting, population, interventions, clinical and educational outcomes. The data abstraction tool included categories for setting, population, intervention, outcomes and qualitative comments. Themes were iteratively developed and synthesised using realist review methodology. A methodological quality tool assessed the biases, confounders, secular trends, reporting and study quality.

Results Among 39 studies, most were before-after design with resident physicians as the primary population. Twenty-one described clinical interventions and 18 described educational interventions with a mean intervention length of 6.58 (SD=9.16) months. Twenty-eight reported successful clinical improvements; no studies reported clinical outcomes but little effect on patient outcomes.6–8 Wong et al9,10 identified three categories of QI education, most of which fall into the first category: (1) formal curricula that teach concepts or methods intended to facilitate trainees’ participation in QI activities; (2) educational activities that impart specific related skills and (3) QI initiatives that involve trainees as active or passive participants. Many others worldwide have developed clinical teaching of QI, aiming to engage physician trainees to improve the care of the patients they serve and the function of the system in which they practice.10

Conclusions This realist review identified success characteristics to guide training programmes, medical schools, faculty, trainees, accrediting organisations and funders to further develop educational and improvement resources in QI educational programmes.
question of whether or not QI educational interventions had an impact on physician trainees’ ability to gain knowledge and sought to identify themes associated with successful QI curricula, but did not assess the specific mechanistic and contextual factors that predicted success, especially for improvement of patient care and system performance outcomes. The realist review approach offers one approach to deconstruct such complex interventions to assess key success characteristics and develop recommendations.11 12

In this study, we define key characteristics of successful QI curricula in medical education. This realist review determines how the teaching of QI in the clinical setting enhances patient care and system performance while increasing trainee knowledge and skills.

METHODS

Review framework

A realist review is based on the premise that complex interventions are successful when certain characteristics facilitate the optimal functioning of a system to produce a particular outcome when a complex intervention is applied.13 In a realist review, an iterative approach is used to identify the characteristics of complex interventions in the following categories: ‘what works,’ ‘for whom,’ ‘under what circumstances’ and ‘to achieve what outcomes’.11 The realist review begins with the articulation of candidate theories that may explain the characteristics required for interventions to be successful.11 Next, identification and selection of studies is achieved through standard systematic review approach.11 Once relevant studies are chosen for inclusion, data are systematically abstracted from the studies and the studies are read and reread to identify themes.11 An iterative approach is used to identify data, quotations, tables and figures that either support or refute the candidate theories articulated at the outset. Theories are refined as more data are gathered from the articles.11

We began by searching the literature for existing theories which explained teaching of QI in the clinical setting. We evaluated the prior systematic reviews on the topic of QI medical education,7–9 13 spoken with experts in the field,7 9 and prepared a candidate conceptual framework (see online supplement 1) and accompanying theory for review.7 14 Our candidate theory hypothesised that the process of educating physicians begins with a curriculum and is impacted by characteristics of the learner, teacher, community and others, all encompassed within the educational context. From within the educational context emerges engaged learners and teachers, who produce improved educational and clinical outcomes. The combination of these successes produces physicians who are capable of and believe that it is their job to do their work and improve their work.

The second step of the realist review is development of inclusion criteria, search strategies, a data abstraction tool and methodological quality assessment for review of the literature and analysis of included studies. Throughout the above process, the candidate theories were tested and refined and new theories added. As the studies were evaluated, themes emerged that were based on the predetermined theories. Each theme was assigned a code and linked to a quotation in the study. As a new theme emerged it was assigned a new code; we then searched for this theme in all the included studies in the review. New and revised candidate theories were synthesised into the set of candidate theories to test in our realist review.

Study eligibility criteria

Included studies had the following criteria:

- Study design—original journal articles (no commentaries, letters to the editor, editorials or position pieces).
- Setting—medical schools, residency and fellowship programmes worldwide.
- Population—physician trainees (medical students, residents and/or fellows).
- Interventions—whether clinical or educational—that engage trainees in QI work, where they are involved in changes to the delivery of care to patients within the clinical setting.
- Reporting of clinical outcomes (patient care outcomes and system performance improvements) as the primary outcome measure.

Search methods

In collaboration with a professional librarian, one reviewer (ACJ) developed search strategies for the following databases: MEDLINE (2000 to 12 March 2013), Cochrane Library (2013) and Web of Science (15 March 2013). To locate potentially relevant studies in MEDLINE, we used exploded Medical Subject Headings terms and key words to generate sets for the themes of QI and medical education. We then used the Boolean term ‘AND’ to find their intersection. This basic approach was modified as necessary to search each electronic database. No language restriction was applied. Time limit was applied to obtain articles published after 2000, which corresponded with the publication of the Institute of Medicine reports To Err Is Human and Crossing the Quality Chasm. We excluded commentaries, editorials and letters. The full search strategy is available upon request. Reference reviews of the four earlier systematic reviews7–9 13 were performed by obtaining all references cited and searching forward using Web of Science to find all papers which cited these reviews and including them in title and abstract review.

Study selection

One reviewer (ACJ) independently screened each title and abstract for eligibility. Then, two non-blinded reviewers (ACJ and GO) independently assessed the
eligibility of each full text record. Discrepancies were resolved by consensus between both reviewers after full text review.

Data collection
One reviewer (ACJ) abstracted data from the full text articles. A standardised data collection tool was used to capture identifying information, intervention summaries, details of study protocol, all primary and secondary outcome data and a section to extract quotes from the articles for the realist review (see online supplement 2).

Analysis
Analysis of interventions and outcomes
Anticipating that the QI interventions and outcomes would be complex and different depending on the training programme, we used an iterative approach to categorise the different types of interventions and outcomes described in each study. We focused on any qualitative or quantitative reports of change in clinical outcomes. If outcomes were reported quantitatively, we determined whether statistical analysis was performed, either in the form of enumerative statistics or analytical statistics using statistical process control, a method of time-ordered analysis for QI.15 For the secondary outcome, we noted the findings of the educational outcomes such as knowledge, skills and attitudes.

Assessment of methodological quality
The Medical Education Research Quality Instrument16 is a validated instrument for methodological quality assessment of the medical education literature, but does not allow assessment of the methodological quality of the QI education literature specifically. Thus, we developed a set of criteria based upon the Cochrane Risk of Bias Tool and the Standards for Quality Improvement Reporting Excellence publication guidelines.17 18 Criteria were based on attention to quality factors in three major categories: the population, intervention and outcome reporting (see online supplements 3 and 4). Each study was assessed on all the criteria for quality, taking into account bias, confounding and study quality.

A final methodological quality score was given to each study, based on a scale ranging from ‘fair,’ if almost none of the criteria were met, to ‘good,’ when minor flaws were found, to ‘very good,’ when high-quality reporting was achieved in at least one criterion each for population, intervention and outcome reporting, and ‘excellent,’ when high quality was achieved for all criteria (see online supplement 5).

The realist review
After completing the methodology of a systematic review, we used the realist review methodology to rigorously test the conceptual framework. We iteratively identified relevant themes, and through continuous data collection and rereading the articles, we tested the conceptual framework. Specifically, we looked for examples of when the curriculum was developed at the outset of the intervention to support the educational context of the candidate theory. We also looked for examples of successful completion of QI curricula and examples of physicians who are capable of and believe that it is their job to do their work and improve it.

RESULTS
Results of search
We included 39 studies for our final review (figure 1; excluded studies after full text review is available upon request), most of which were before–after studies or case reports in internal medicine or family medicine residency programmes (table 1). Less common were controlled trials, studies with medical students or reports from subspecialty resident programmes. Of the 29 studies that reported a sample size, the mean sample size was 5.6 trainees (SD=102), the median sample size was 24 trainees and the range was from 3 to 510 trainees.

Among the interventions, 21 were primarily clinical interventions in which the goal was clearly to improve patient care or system performance in the clinical setting with education of the trainee not the primary focus of the intervention; 18 studies primarily had educational interventions, in which the goal was to deliver a curriculum to trainees focused on learning about improvement of which a component was to improve patient care and system performance in the clinical setting. The mean intervention length was 6.6 months (SD=9.2).

Twenty studies reported system performance outcomes (e.g., improved documentation) while three reported only patient care outcomes (e.g., haemoglobin a1c, blood pressure) and 16 reported both patient care and system performance outcomes. Among the clinical outcomes, 28 studies reported successful improvements, two of which were not sustained. Ten studies demonstrated improvement in some measures and two clinical outcome reports were equivocal. No studies reported clinical outcomes that worsened.

Nineteen studies reported educational outcomes, using various measures of knowledge, skills and attitudes such as the Quality Improvement Knowledge Application Tool (QIKAT),19 satisfaction surveys or objective structured clinical examinations.

Description of studies
The most common types of interventions were team projects and involvement in an existing clinical QI team (table 2). Among the team project interventions, trainees worked together to make improvements in the clinical setting; these studies had varied methodological quality scores. Seven of these studies demonstrated significant
improvement in documentation, critical care measures and medication adherence; six studies showed no significant clinical outcomes.

Of the interventions that involved trainees taking a role within an existing clinical interprofessional QI team, most were of ‘good’ methodological quality. Most showed no significant clinical outcomes, but some showed statistically significant improvement in diabetes measures, vaccination rates, chronic care measures and critical care measures. The third most common type of intervention used a chart audit. Most of these had ‘good’ methodological quality. About half these studies showed statistically significant outcomes, specifically in improved diabetes care, preventive care measures, documentation and critical care measures; and the other half did not show statistically significant outcomes.

The least common intervention used individual projects in which the trainee worked independently. Most of these studies were rated ‘good’, and one received a ‘fair’ rating. Most of these studies showed no significant clinical outcomes, while one study showed significant improvement in care for heart failure patients.

Methodological quality
A majority of studies (30) had ‘good’ methodological quality. Six were ‘fair’ studies and 3 were ‘very good’. There were no ‘excellent’ studies.

Among the included studies, all but three studies described the intervention in sufficient detail so that it could be replicated. Although all the studies had, by our inclusion criteria, trainees participating in QI work within the clinical setting, only 14 studies articulated educational objectives for trainees. There were five studies which described a clinical intervention rather than an educational intervention and reported no educational outcomes (table 2).
It is notable that three studies took steps to minimise bias and confounding. Holmboe et al.\textsuperscript{35} matched each second year resident in the intervention track to a third year control. Dysinger and Pappas\textsuperscript{21} enrolled the entire fourth year medical school class in a required monthly clinical QI rotation over 3 years, resulting in 510 students completing the curriculum, and allowing for comparison over 3 years of data collection to observe and account for secular trends. Asao et al.\textsuperscript{49} in addition to enrolling only second year residents in the chart audit curriculum intervention, completed a multivariate analysis to account for the trainee’s experience as an auditor, the duration of exposure to the curriculum independent of training level and number of comorbidities in the resident’s patient sample.

### Synthesis of results and realist review

We identified several major themes through realist review. These themes are organised by ‘what works,’ ‘for whom,’ ‘under what circumstances’ and ‘to achieve what outcomes’ (table 3). After synthesising the range of interventions, clinical and educational outcomes, the methodological quality and realist review of the 39 studies, we tested the candidate theory and conceptual framework by iteratively analysing the major themes which emerged from the data. Specifically, we looked for evidence of a predetermined curriculum and educational context, as we hypothesised at the outset. However, we failed to find evidence of these mechanistic and contextual factors as important determinants in producing physicians who are lifelong learners and improvers. Thus, we revised the candidate theories and developed a conceptual framework (figure 2).

Many different types of curricula are described in the included studies, some of which are distinct educational interventions and some of which involve trainees in existing clinical QI in their practices. Among the included studies, these interventions fell into four categories, and we found examples of statistically significant improvements in clinical outcomes in each category. Those studies with ‘fair’ methodological quality scores did not report significant results or failed to report numerical results with statistical analysis. Therefore, the higher-quality studies suggest that, with certain contexts and mechanisms, significant improvement in clinical outcomes is achievable when trainees are exposed to QI within the clinical setting.

Several success characteristics were common to different contexts of clinical QI education (table 3; for illustrative quotes see online supplement 6). As noted by Sockalingham et al.\textsuperscript{36} ‘residents identified workload as a major barrier to (doing QI work).’ Successful QI teaching programmes were consistently clear about the time required for trainee work-hour rules, competing demands and for faculty involvement. Success of specific programmes will also depend on whether it makes more sense to train all faculty members in QI principles or to have a dedicated select faculty group in charge of the QI curriculum. The availability of data through information systems is also a facilitator to trainee satisfaction and engagement. The sustained improvement reported by Halverson et al.\textsuperscript{36} was achieved through the use of timely regular data feedback to all providers, including trainees, about the care of patients with diabetes in the practice. Two studies also highlighted the challenges that occur when trainees must abstract their own data, such as is the case in a practice which has not implemented an electronic medical record, or when data feedback is not timely enough for continuous QI.\textsuperscript{34 44} Choice of the project topic is also important for trainees. QI educators need to consider the needs of the clinical setting as well as level of trainee; however, no consensus emerged as to the best approach.
<table>
<thead>
<tr>
<th>Author and year</th>
<th>Population (n)</th>
<th>Intervention description</th>
<th>Clinical setting</th>
<th>Clinical outcomes</th>
<th>Statistically significant*</th>
<th>Educational outcomes</th>
<th>Quality assessment†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chart audit</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gould et al 2002</td>
<td>2nd year medical students (77)</td>
<td>Learners audited random sample of diabetes charts using Qualidigm, before and after ‘Project in a Box’</td>
<td>Outpatient primary care clinic</td>
<td>Improved rates of Hba1c, foot exam and eye exam documentation</td>
<td>Yes</td>
<td>Improved knowledge, skills, attitudes; poor learner satisfaction</td>
<td>Good</td>
</tr>
<tr>
<td>Paukert et al 2003</td>
<td>FM residents (36)</td>
<td>Self and peer audits done at several time points</td>
<td>Outpatient primary care clinic</td>
<td>Improved, but unsustained, composite ‘preventive’ score</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Holmboe et al 2005</td>
<td>2nd year IM residents (26)</td>
<td>Learners audited sample of their own diabetes patients using Qualidigm</td>
<td>Outpatient primary care clinic</td>
<td>Improved patient care (Hba1c, LDL) and system performance (monofilament testing documentation)</td>
<td>Yes</td>
<td>Improved perceived value and motivation</td>
<td>Very good</td>
</tr>
<tr>
<td>Kadden et al 2006</td>
<td>Residents and medical students (NR)</td>
<td>Structured session every 24 hrs of all X-ray and culture reports among ED patients</td>
<td>Paediatric emergency department</td>
<td>Decrease in chart requiring attending comment and change in treatment course</td>
<td>Yes</td>
<td>High learner satisfaction</td>
<td>Good</td>
</tr>
<tr>
<td>Krajewski et al 2007</td>
<td>Residents (NR)</td>
<td>Performed audit on self-identified problem in radiology reporting; presented at M&amp;M conference</td>
<td>Radiology department</td>
<td>Multiple system improvements implemented</td>
<td>No</td>
<td>NR</td>
<td>Fair</td>
</tr>
<tr>
<td>Asao et al 2009</td>
<td>2nd and 3rd IM year residents (80)</td>
<td>Peer audit performed; feedback and motivational plan proposed by learner</td>
<td>Outpatient primary care clinic</td>
<td>Improved composite measures for COPD, CHD, DM, HTN, LVF</td>
<td>No</td>
<td>NR</td>
<td>Very good</td>
</tr>
<tr>
<td>Carek et al 2009</td>
<td>FM residents (20)</td>
<td>Chart audit before and after educational session</td>
<td>Outpatient primary care clinic</td>
<td>Improvement in oral and mental health, care of underserved populations, antibiotic use, elderly care, preventive services</td>
<td>No</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Kirschenbaum et al 2010</td>
<td>CC fellows and residents (NR)</td>
<td>Team audit and analysis of all transfers from GMU to MICU</td>
<td>MICU</td>
<td>Improvement in number of cardiac arrests and deaths, number of times MET called</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Smith et al 2012</td>
<td>IM residents (20)</td>
<td>Peer RCA on near-miss cases, presented at M&amp;M conference</td>
<td>Various departments</td>
<td>Multiple system improvements implemented, some successful</td>
<td>No</td>
<td>High learner satisfaction</td>
<td>Fair</td>
</tr>
<tr>
<td>Participant on clinical QI team</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coleman et al 2003</td>
<td>FM residents (NR)</td>
<td>Participation at training site QI team; learner chose projects and improvement approaches</td>
<td>Outpatient primary care clinic</td>
<td>Improved microalbuminuria screening, medication list completion, data summary sheets</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Mohr et al 2003</td>
<td>Peds residents (8)</td>
<td>Resident team chose project for interprofessional practice team improvement</td>
<td>Outpatient paediatric clinic</td>
<td>Improved vaccination rates for the practice</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Landis et al 2006</td>
<td>FM residents (126)</td>
<td>Residents involved in statewide learning collaborative to improve diabetes care</td>
<td>Outpatient primary care clinic</td>
<td>Improved ACIC scores</td>
<td>NR</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Author and year</td>
<td>Population (n)</td>
<td>Intervention description</td>
<td>Clinical setting</td>
<td>Clinical outcomes</td>
<td>Statistically significant*</td>
<td>Educational outcomes</td>
<td>Quality assessment†</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>---------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Halverson et al 2007&lt;sup&gt;98&lt;/sup&gt;</td>
<td>FM residents (NR)</td>
<td>Resident served on committee to improve diabetes care</td>
<td>Outpatient primary care clinic</td>
<td>Improved percentage of patients at goal for Hba1c, LDL, BP</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Stapleton et al 2009&lt;sup&gt;42&lt;/sup&gt;</td>
<td>Peds senior residents (NR)</td>
<td>RPFW on improving senior resident rotation</td>
<td>Paediatric inpatient service MICU</td>
<td>Improved system performance measures</td>
<td>NR</td>
<td>NR</td>
<td>Fair</td>
</tr>
<tr>
<td>Buckley et al 2010&lt;sup&gt;32&lt;/sup&gt;</td>
<td>CC fellows and residents (NR)</td>
<td>Mandatory participation on MICU improvement team</td>
<td>Outpatient primary care clinic</td>
<td>Improved iatrogenic pneumothorax rates, sepsis-specific mortality, sepsis bundle compliance</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Fischman et al 2010&lt;sup&gt;15&lt;/sup&gt;</td>
<td>IM residents (4)</td>
<td>Controlled trial with involvement on clinical QI team</td>
<td>Outpatient primary care clinic</td>
<td>Improvement in no-show, continuity measures, doctor-patient relationships</td>
<td>NR</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Stevens et al 2010&lt;sup&gt;39&lt;/sup&gt;</td>
<td>IM, FM, peds residents (NR)</td>
<td>Involvement in statewide collaborative for diabetes care</td>
<td>Outpatient primary care clinic</td>
<td>Improved ACIC scores</td>
<td>Yes</td>
<td>High learner participation</td>
<td>Good</td>
</tr>
<tr>
<td>Yu et al 2010&lt;sup&gt;41&lt;/sup&gt;</td>
<td>FM residents (6)</td>
<td>Resident participation in statewide collaborative to improve diabetes care</td>
<td>Outpatient primary care clinic</td>
<td>Improved ACIC scores</td>
<td>NR</td>
<td>Improved % of residents reviewing performance reports, demonstrating improved behaviours</td>
<td>Good</td>
</tr>
<tr>
<td>Vidyarthi et al 2011&lt;sup&gt;43&lt;/sup&gt;</td>
<td>Fellows and residents (NR)</td>
<td>Financial incentive (US$1200 per trainee) for involvement in improvement work at medical centre</td>
<td>Various departments</td>
<td>Several system performance measures improved</td>
<td>NR</td>
<td>NR</td>
<td>Fair</td>
</tr>
<tr>
<td>Stueven et al 2012&lt;sup&gt;20&lt;/sup&gt;</td>
<td>Residents and medical students (249)</td>
<td>Residents surveyed for relevant patient safety and QI improvement issues, attended retreats</td>
<td>Various departments</td>
<td>Several system performance measures improved</td>
<td>NR</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Carey et al 2013&lt;sup&gt;33&lt;/sup&gt;</td>
<td>Neonatology fellows (3)</td>
<td>Fellows participate on Q and S committee, chair working group when issues arise</td>
<td>NICU</td>
<td>Improvement in broncho-pulmonary disease, catheter-associated bloodstream infections</td>
<td>NR</td>
<td>Learner self-reflections about QI, presentations at conferences</td>
<td>Good</td>
</tr>
<tr>
<td>Team project</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varkey et al 2006&lt;sup&gt;31&lt;/sup&gt;</td>
<td>Fellows, residents, medical students (7)</td>
<td>Multidisciplinary trainee teams in training hospital</td>
<td>Various departments</td>
<td>Improvement in documentation of medication reconciliation</td>
<td>NR</td>
<td>Improved QIKAT scores</td>
<td>Good</td>
</tr>
<tr>
<td>Oyler et al 2008&lt;sup&gt;23&lt;/sup&gt;</td>
<td>2nd year IM residents (34)</td>
<td>Mandatory project while on ambulatory rotation</td>
<td>Outpatient primary care clinic</td>
<td>Improvement in documentation of height, weights, BMI</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Varkey et al 2008&lt;sup&gt;30&lt;/sup&gt;</td>
<td>Preventive medicine, endocrinology fellows (9)</td>
<td>QI project taught jointly by medicine and engineering faculty</td>
<td>Various departments</td>
<td>Improvement in patient understanding on treatment</td>
<td>NR</td>
<td>Improved QIKAT scores, learner satisfaction</td>
<td>Fair</td>
</tr>
<tr>
<td>Author and year</td>
<td>Population (n)</td>
<td>Intervention description</td>
<td>Clinical setting</td>
<td>Clinical outcomes</td>
<td>Statistically significant*</td>
<td>Educational outcomes</td>
<td>Quality assessment†</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>---------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Tomolo et al 2009</td>
<td>IM residents (42)</td>
<td>QI project while rotating on inpatient medicine service</td>
<td>Various departments</td>
<td>Multiple sustained system performance measures improved (i.e., missing lab values)</td>
<td>NR</td>
<td>High learner satisfaction</td>
<td>Good</td>
</tr>
<tr>
<td>Varkey et al 2009</td>
<td>Preventive medicine fellows (19)</td>
<td>Fellows develop and implement projects over 4-week rotation</td>
<td>Various departments</td>
<td>Multiple system performance and patient care measures improved</td>
<td>Yes</td>
<td>Improved QIAT scores, OSCE performance, learner satisfaction</td>
<td>Good</td>
</tr>
<tr>
<td>Diaz et al 2010</td>
<td>2nd and 3rd year FM residents (61)</td>
<td>Clinical scholars curriculum delivered over 1 year of residency</td>
<td>Outpatient primary care clinic</td>
<td>Multiple system performance and patient care measures improved</td>
<td>Yes</td>
<td>Increase in number of publications and presentations</td>
<td>Good</td>
</tr>
<tr>
<td>Shiner et al 2010</td>
<td>Psychiatry residents (12)</td>
<td>Aiming to improve care for major depressive disorder</td>
<td>Outpatient psychiatry clinic</td>
<td>Improved percentage of patients seen within 6 weeks of starting MDD treatment</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Clark et al 2011</td>
<td>General surgery residents (33)</td>
<td>Team completed needs assessment and improvement in signout process by template</td>
<td>Inpatient surgery service</td>
<td>Improvement in signout processes</td>
<td>NR</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Dysinger et al 2011</td>
<td>4th year medical students (510)</td>
<td>Required rotation, putting students into practices focused on improvement</td>
<td>Various departments</td>
<td>Improved documentation</td>
<td>Yes</td>
<td>Some improved learner satisfaction</td>
<td>Very good</td>
</tr>
<tr>
<td>Laiteerapong et al 2011</td>
<td>IM residents (10)</td>
<td>Team project to improve documentation</td>
<td>Outpatient primary care clinic</td>
<td>Improvement in height, weight, BMI documentation</td>
<td>Yes</td>
<td>Improved learner satisfaction</td>
<td>Good</td>
</tr>
<tr>
<td>Ogrinc et al 2011</td>
<td>2nd year medical students (22)</td>
<td>Health Leadership Practicum Elective brings student teams into clinical settings in need of improvement</td>
<td>Various departments</td>
<td>Multiple reported improvements (i.e., urine samples screened in pregnant women)</td>
<td>NR</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Oyler 2011</td>
<td>IM residents (64)</td>
<td>Required rotation encouraging teams to develop projects</td>
<td>Outpatient primary care clinic</td>
<td>Improvement in ASA use, documentation on BMI and smoking history</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
</tr>
<tr>
<td>Arbuckle et al 2013</td>
<td>3rd year psychiatry residents (12)</td>
<td>Longitudinal QI curriculum over 37 weeks, including longitudinal project</td>
<td>Outpatient psychiatry clinic</td>
<td>Improvement in monitoring of depression symptoms, screening</td>
<td>NR</td>
<td>Improved QIAT scores</td>
<td>Good</td>
</tr>
<tr>
<td>Individual project</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weingart et al 2004</td>
<td>2nd and 3rd year IM residents (26)</td>
<td>RCA/QI project on voluntary elective</td>
<td>Various departments</td>
<td>Improvement in system performance and patient care measures</td>
<td>NR</td>
<td>Improved learner satisfaction</td>
<td>Fair</td>
</tr>
<tr>
<td>Canal et al 2007</td>
<td>3rd year general surgery residents (15)</td>
<td>Mandatory QI project on research time</td>
<td>Various departments</td>
<td>Improvement in several system performance measures</td>
<td>NR</td>
<td>Improved QI curriculum pre/post test</td>
<td>Good</td>
</tr>
<tr>
<td>Author and year</td>
<td>Population (n)</td>
<td>Intervention description</td>
<td>Clinical outcomes</td>
<td>Statistically significant*</td>
<td>Educational outcomes</td>
<td>Quality assessment†</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>-------------------</td>
<td>---------------------------</td>
<td>---------------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>Sockalingham et al 2010&lt;sup&gt;16&lt;/sup&gt;</td>
<td>2nd and 3rd year psychiatry residents (40)</td>
<td>Mandatory QI project while on rotation</td>
<td>Improvement in consult/referral services</td>
<td>NR</td>
<td>Information by focus group and questionnaire</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Oujiri et al 2011&lt;sup&gt;54&lt;/sup&gt;</td>
<td>3rd year IM residents</td>
<td>Learners chose QI project as part of elective rotation</td>
<td>Improvement in several patient care and system performance measures for heart failure patients</td>
<td>Yes</td>
<td>NR</td>
<td>Good</td>
<td></td>
</tr>
<tr>
<td>Reardon et al 2011&lt;sup&gt;35&lt;/sup&gt;</td>
<td>Psychiatry residents (16)</td>
<td>Learner assigned to unit and completed mandatory project</td>
<td>Improvement in consult processes</td>
<td>NR</td>
<td>Improved QIKAT scores</td>
<td>Good</td>
<td></td>
</tr>
</tbody>
</table>

* If 'Yes' the majority of findings in the results of the study had to be shown to be statistically significant, either by p value or statistical process control (SPC) analysis.

† We modified methodological quality assessment of these studies to take into account the important factors of educational literature: clear accounting for systematic bias, clear description of the intervention, learning objectives articulated for educational interventions, minimisation of other exposures which could have impact on the outcome, minimisation of bias and confounding.

QI, quality improvement; ACIC, assessment of chronic illness care; ASA, aspirin; BMI, Body Mass Index; BP, blood pressure; CC, critical care; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; DM, diabetes mellitus; ED, emergency department; FM, family medicine; GMU, general medical unit; Hba1c, haemoglobin a1c; HTN, hypertension; IM, internal medicine; LDL, low density lipoprotein; LVF, left ventricular failure; M&M, morbidity and mortality; MET, medical emergency team; MICU, medical intensive care unit; N, no; NICU, neonatal intensive care unit; NR, not reported; QIKAT, quality improvement knowledge assessment test; OSCE, objective structured clinical examination; RCA, root cause analysis; RPIW, rapid process improvement workshop; Y, yes.
around education and acquisition of improvement knowledge and skills. As Batalden and Davidoff wrote, ‘Learning how to do quality improvement and actually carrying out quality improvement are essentially one and the same; both are special forms of experiential learning.’ In fact, medical education is currently embracing this culture in teaching clinical skills using Adult Learning Theory, which reminds us that professionals learn best when they see a need to acquire the knowledge and skills for fulfilment of goals. If trainees see that faculty are asking questions of the process and needing to learn more to improve the system, then they have the opportunity to engage with them. Interestingly, Asch et al have demonstrated that within obstetrical residencies, it is possible and, perhaps beneficial, to rank obstetrical programmes based on overall performance on clinical rather than educational outcomes, representing yet another innovation for the future of the medical profession. The ACGME has acknowledged the importance of the clinical learning environment as an essential component of resident education and, accordingly, adopted the Clinical Learning Environment Review, ‘to generate national data on programme and institutional attributes that have a salutary effect on quality and safety in settings where residents learn and on the quality of care rendered after graduation.’ Through the knowledge, skills and attitudes that trainees achieve during their educational programme while taking part in the improvement of the clinical care, we expect to nurture lifelong learners and improvers who will advance clinical improvements in patient care and system performance.

This realist review has limitations that begin with the known publication bias in this field. This bias was corroborated in our review, as none of the published studies described a clinical process that was worsened. Not sharing failed pilots and curriculum limits the learning that can occur across programmes. Also, although our search strategy allowed us to analyse any circumstance of trainees being involved in QI in the clinical setting, this may have caused us to falsely criticise studies with primarily clinical interventions, because they did not aim to prove that involving trainees made a difference in their clinical outcomes. The realist review process, however, helps to differentiate these studies and their important qualitative information about ‘what works,’ ‘for whom,’ ‘under what circumstances’ and ‘to achieve what outcomes’.

Because a methodological quality tool does not exist for assessment of the QI education literature, we created a tool by combining elements of existing validated tools. Although we did not use a validated instrument, the tool we developed contains the specificity for the QI education literature and thus made the quality assessment more rigorous. We did not, however, identify any ‘excellent’ quality studies (see online supplement 5). Although we found that all the lowest-quality studies did not demonstrate significant results, there were no strong studies to show improvement in both clinical and educational outcomes. We identified many studies with minor weaknesses, and the realist review process helps to glean the notable characteristics from these data. More high-quality studies would take steps to minimise bias in the study population, clearly describe the intervention and educational objectives, minimise other exposures or secular trends that could have accounted for the results, analyse results with enumerative or analytic statistics, explain all biases and confounders and report funding sources. Among the studies in this review, many would have been improved by particular attention to minimising bias and confounding in the study population, and a clear articulation of educational objectives.
CONCLUSION
The studies in this review reported on many more clinical outcomes than had been described in previous reviews, in large part due to the development of clinically oriented QI programmes since the prior reviews. Using the realist approach allowed us the advantage of synthesising these data to not just update, but reconceptualise (figure 2) the current landscape of QI teaching. Advances in teaching and doing QI has made tremendous strides in the past decade, but further work is needed to determine the factors that reliably facilitate the development of physicians who will believe and are capable of doing their work and improving their work—ultimately, physicians who are lifelong learners and improvers.

Acknowledgements The authors thank Tom Mead, MLS, for assistance with the search strategy; Aurora Leute Matzkin, PhD and Martha Reagan-Smith, MD, EdD (Professor Emerita, Geisel School of Medicine) for their assistance in reading drafts of the work and preparing the conceptual framework.

Contributors The conception or design of the work and interpretation of data was performed by ACJ, SAS and GO. The acquisition and analysis of data was performed by ACJ and GO. The manuscript was drafted by ACJ, and revised critically for important intellectual content by ACJ, SAS and GO. All authors approved the final version of the manuscript. ACJ, SAS and GO agree to be accountable for all aspects of this work and will ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This work was supported by the VA National Quality Scholars Fellowship Program, the VA Office of Academic Affiliations, and the Geisel School of Medicine at Dartmouth Office of Health Systems and Clinical Improvement with the use of facilities and materials from the White River Junction VA in White River Junction, VT.

Competing interests ACJ and SAS have no competing interests to report. GSO is an Associate Editor of BMJ Quality and Safety but otherwise has no competing interests to report.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement ACJ had full access to all the data in the study, and takes responsibility for the integrity of the data and the accuracy of the data analysis.

REFERENCES
14. Scott Shipman M, MPH; Course Director, Capstone Series at The Dartmouth Institute for Health Policy and Clinical Practice; Director for Primary Care Affairs and Workplace Analysis, American Association for Medical Colleges.


