Evidence-based medical leadership development: a systematic review

Oscar Lyons 1, Robynne George, Joao R Galante3,4, Alexander Mafi5, Thomas Fordwoh 5, Jan Frich 6, Jaason Matthew Geerts 7,8

ABSTRACT
Health systems invest significant resources in leadership development for physicians and other health professionals. Competent leadership is considered vital for maintaining and improving quality and patient safety. We carried out this systematic review to synthesise new empirical evidence regarding medical leadership development programme factors which are associated with outcomes at the clinical and organisational levels. Using Ovid MEDLINE, we conducted a database search using both free text and Medical Subject Headings. We then conducted an extensive hand-search of references and of citations in known healthcare leadership development reviews. We applied the Medical Education Research Study Quality Indicator (MERSQI) and the Joanna Briggs Institute (JBI) Critical Appraisal Tool to determine study reliability, and synthesised results using a meta-aggregation approach. 117 studies were included in this systematic review. 28 studies met criteria for higher reliability studies. The median critical appraisal score according to the MERSQI was 8.5/18 and the median critical appraisal score according to the JBI was 3/10. There were recurring causes of low study quality scores related to study design, data analysis and reporting. There was considerable heterogeneity in intervention design and evaluation design. Programmes with internal or mixed faculty were significantly more likely to report organisational outcomes than programmes with external faculty only (p=0.049). Project work and mentoring increased the likelihood of organisational outcomes. No leadership development content area was particularly associated with organisational outcomes. In leadership development programmes in healthcare, external faculty should be used to supplement in-house faculty and not be a replacement for in-house expertise. To facilitate organisational outcomes, interventions should include project work and mentoring. Educational methods appear to be more important for organisational outcomes than specific curriculum content. Improving evaluation design will allow educators and evaluators to more effectively understand factors which are reliably associated with organisational outcomes of leadership development.

INTRODUCTION
Health systems invest significant resources in leadership development for physicians and other health professionals. Competent leadership is considered vital for team effectiveness, for clinical and financial performance and for maintaining and improving quality and patient safety.1–5 Clinical leadership development involves activities to promote leadership competencies among clinicians, while medical leadership development refers to activities centred on doctors.

Research suggests that medical leadership development can improve outcomes at individual, organisational and clinical levels.6–11 Evidence backing medical leadership development activities has, however, been variable in quality.1,7–10,12–15 There has been a particular lack of research and evaluation that goes beyond individual learner feedback and subjective outcomes.6–9 One systematic review of 45 studies evaluating leadership development interventions for doctors found that effective interventions were characterised by the use of multiple learning methods, including seminars and group work, alongside action learning projects in multidisciplinary teams.6 These findings were echoed in a recent study by Geerts et al,7 who emphasised that plans need to be in place for transferring learning from the intervention into the working environment.

We undertook this systematic review to synthesise recent empirical evidence regarding medical leadership development programme factors associated with outcomes at the clinical and organisational levels. We specifically investigated links between aspects of programme design, delivery and evaluation and improved outcomes. Given the variable quality of studies highlighted in previous reviews,1,7–9 we applied two validated critical appraisal instruments16–17 to isolate higher reliability findings. This review is the first to apply both instruments in order to identify and synthesise the highest quality empirical evidence in medical leadership development.

METHODS
The design of this review was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses18 and the Best Evidence in Medical Education (BEME) guide for systematic reviews.19 Our methods were based on the review conducted by Frich et al,4 with methodological changes drawn from other reviews.7,9,10,14,15,20 Following the BEME recommendations for systematic reviews,19 we hand-searched references and citations of known reviews extensively to supplement our database search. In line with recommendations from Geerts et al21 and Rosenman et al,22 we assessed study quality using the Medical Education Research Study Quality Indicator (MERSQI), which is designed to measure the methodological quality of quantitative medical education research studies.16 We added
Review

the Joanna Briggs Institute (JBI) Critical Appraisal Checklist, which is designed for meta-aggregation of qualitative research and is well-established in healthcare research.21

Search strategy
We began this review by re-examining the data set identified in the review of leadership development for physicians by Frich et al.8 With assistance from a specialist librarian at the University of Oxford, we then based our search strategy on Frich et al’s review.8 Using the Ovid MEDLINE database, we conducted a search using both free text and Medical Subject Headings. The full search terms are listed in the online supplemental material. This search identified 501 unique publications. We then conducted an extensive hand-search of references and of citations in known healthcare leadership development reviews using Web of Science and Google Scholar. This identified an additional 107 studies for possible inclusion, for a total of 608 records for screening (figure 1).

Inclusion criteria
We included any peer-reviewed study published in English between January 2000 and January 2020 which:
1. Describes a leadership development intervention (programme, workshop, course and so on).
2. Includes physicians as learners (defined here as any practising doctor post-qualification).
3. Evaluates the leadership development intervention.
   Qualitative, quantitative and mixed evaluations were included. We excluded studies where leadership development was a minor focus or where the proportion of physicians was lower than 10% of intervention participants.

Screening process
Two members of the review team (OL and TF) independently screened all study titles and abstracts for eligibility. Articles that were approved by either reviewer progressed to full-text review. Two members of the review team independently reviewed for inclusion the full text of all 207 articles that passed the title and abstract screen (TF and RG reviewed half each, OL reviewed all). Where there was disagreement about inclusion, all three reviewers (OL, TF, RG) reached consensus by discussion, with the third reviewer (TF or RG) arbitrating where required.

Data abstraction
After screening and reviewing for eligibility, 117 unique studies were included for abstraction and analysis. Data were abstracted and coded for educational setting, methods, content, evaluation methods and outcomes. Outcome data were categorised according to an adapted version of Kirkpatrick’s Framework for evaluation of training programmes (see table 1).19 22 One reviewer abstracted and coded all 117 included studies (OL). The second reviewers (RG/JRG/AM/TF) each abstracted and coded at least five studies in full to ensure consistency between reviewers. Data abstraction and coding for all 117 studies was then cross-checked by the second reviewers. Any differences were resolved by consensus, with a third reviewer arbitrating where required. Where possible, statistical tests performed in studies were replicated and checked for accuracy.

Study quality appraisal
Previous reviews have shown marked variation in the quality of studies of medical leadership development.7 9 14 15 20 To isolate the most reliable evidence linking medical leadership programmes to improved outcomes, two researchers independently critically appraised each included study using the MERSQI and JBI Instruments.16 21 Differences in MERSQI and JBI quality score were resolved by consensus, and a third researcher arbitrated where needed.

The MERSQI was applied to all 117 studies. The MERSQI is a validated appraisal tool consisting of 10 items in six domains which relate to design, sampling, type of data collected, validity of evaluation methods, analysis and outcomes.16 Each domain is
Table 1  Kirkpatrick’s Framework for evaluation of training programmes, with adaptations from Frich et al.9

<table>
<thead>
<tr>
<th>Kirkpatrick level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1 Reaction</td>
<td>Participants’ satisfaction with the learning experience, its organisation, presentation, content, teaching methods and quality of instruction</td>
</tr>
<tr>
<td>Level 2A Change in attitudes</td>
<td>Changes in the attitudes or perceptions among participant groups towards leadership, management or administration</td>
</tr>
<tr>
<td>Level 2B Change in knowledge or skills</td>
<td>For knowledge, this relates to the acquisition of concepts, procedures and principles; for skills, this relates to the acquisition of thinking/problem-solving, psychomotor and social skills</td>
</tr>
<tr>
<td>Level 3A Behavioural change (self-reported)</td>
<td>Transfer of learning to the workplace and changes to professional practice, as noted by participants themselves</td>
</tr>
<tr>
<td>Level 3B Behavioural change (observed)</td>
<td>Transfer of learning to the workplace and changes to professional practice, as noted by a third party or by promotions</td>
</tr>
<tr>
<td>Level 4a Results (self-reported)</td>
<td>Organisational outcomes perceived by respondents and group effectiveness perceived by subordinates</td>
</tr>
<tr>
<td>Level 4b Results (observed)</td>
<td>Tangible organisational outcomes, such as reduced costs, improved quality and safety, impact of projects</td>
</tr>
</tbody>
</table>

**Data analysis**

MERSQI and JBI Scores were used to establish which studies presented more reliable evidence of outcomes. Summary statistics were calculated for all 117 studies. In line with Geerts et al.,9 studies with scores of 12 or higher were categorised as higher reliability studies (see the Data analysis section).

The JBI Checklist for Qualitative Studies was also applied where a study used mixed methods (k=53) or qualitative methods (k=10). Fundamental differences in study design, sampling, evaluation instruments and analysis preclude summative comparison of mixed-methods or qualitative studies to quantitative studies using the MERSQI.26 21 23 24 The JBI Checklist is considered the most appropriate qualitative critical appraisal tool for use in pragmatic meta-aggregation of qualitative research.25 It includes 10 items which regard the study’s research questions, methods, analysis and reporting, for a total score of 0–10. Following recommendations from the JBI Reviewers’ Manual,17 a cut-off score for higher reliability studies was predetermined at 6/10. This score was chosen as studies obtaining six or more points included most key elements of high-quality design.

**RESULTS**

**Study reliability (MERSQI and JBI)**

Twenty-eight of 117 studies (25%) were categorised as higher reliability. Two studies were categorised as higher reliability by both the MERSQI and the JBI tool.26 27 14 studies (12%) by the MERSQI only and 12 studies (10%) by the JBI tool only. The median critical appraisal score according to the MERSQI was 8.5 (range 5–16 from possible range of 5–18) and the median critical appraisal score according to the JBI was 3 (range 0–9 from possible range of 0–10). Online supplemental table 1 includes the MERSQI and JBI Scores for all included studies.

Study design showed considerable room for improvement, as shown in online supplemental tables 2 and 3. Nearly half the of studies (46%) relied on post-programme evaluations only, and 92% did not include a control group. Of the nine studies that did include control groups, most had substantial methodological flaws in their selection of control groups. One common method for control group recruitment was to use unsuccessful course applicants.28–30 In terms of evaluation design, the median evaluation instrument score was 0 (range 0–3). The majority of studies (59%) did not fulfil any of the MERSQI requirements for evaluation instruments, including reporting questionnaire design, wording and content. Objective outcome measures were used in only a minority of studies, with 60% relying solely on self-reported measures.

Data analysis and reporting likewise showed considerable limitations. Only one in five studies (20%) met criteria for comprehensive analysis and reporting of data. Few studies analysed their data beyond descriptive statistics to consider the generalisability and implications (13%). In many cases, studies omitted basic statistical significance tests.

Many studies did not contain key reporting elements for qualitative research as outlined in the JBI tool (see online supplemental table 3). There was clear congruency between research methodologies chosen and the research objectives and methods employed in 60% of studies. A minority of studies adequately reported their analysis (28%) and interpretation of data (25%), the potential for the researcher to have influenced data collection and interpretation (23%) and the researcher’s cultural or theoretical orientation (15%). Participant voices were clearly represented through quotes in only 16/53 (30%) of mixed-methods studies and 5/10 (50%) of qualitative studies. There was a statement of ethical approval or ethics exemption in only 26 of 63 studies (40%) which used qualitative methods. No study included a statement of philosophical perspective (normally expected for qualitative research).17

**Programme design**

There was considerable heterogeneity in leadership development intervention design. It was often unclear whether established good practice for development of medical education interventions was followed, as shown in figure 2.9 31 Only 52 studies (44%) reporting having conducted a needs assessment before

![Figure 2](https://example.com/figure2.png)
their intervention, and only 20 studies (17%) explicitly reported using an established capability or competency framework to inform leadership programme goals and objectives. There was, however, a plan for training transfer reported or built into 68 of 117 interventions (59%).

The majority of interventions were carried out in a single hospital department (27%), single hospital (22%) or a single university (12%). Just under a quarter (23%) of interventions were conducted in multiple healthcare centres. A further 15% of studies were conducted within a specialty training programme outside healthcare centres.

Most of the studies took place in the USA (67%) or the UK (16%). The remainder of studies were in other European countries (7%), Canada (4%) or Australia (3%), with a single study each from Africa, India, Israel and Qatar.

Programmes ranged in length from 2 hours to 4 years. The median intervention length was 6 months, and the most common length was 1 year (19%). Only 18 interventions (15%) lasted longer than 1 year. Five interventions (4%) were shorter than 1 day.

Programme faculty
Programmes were predominately delivered by either in-house faculty (36%) or a mix of in-house and external faculty (32%). Programmes delivered by mixed faculty were most likely to show organisational outcomes, as shown in figure 3. The professional backgrounds, qualifications and experience of faculty were generally not reported.

Participants
The majority of programmes included doctors only (76%). Physician learners ranged from residents (60%) to full specialists (30%) and academic medical faculty (19%). Only nine studies of 117 involved doctors from more than one category. Behavioural outcomes were reported in a similar percentage of higher reliability studies for each category (85%–92%), while organisational outcomes were more commonly reported in programmes with academic medical faculty (50%) or full specialists (44%) than in programmes with only residents (20%). The 26 studies (24%) reporting multidisciplinary programmes included a combination of nurses (12%), managers (15%) and allied health professionals (9%). Most studies did not report the gender of participants (74%) or the age of participants (87%).

In terms of participant selection criteria, the majority of interventions included participants who volunteered (27%), were nominated (19%) or who applied to the programme (16%). In some cases the application process was highly competitive. Interventions were mandatory in one-fifth of studies (20%).

A considerable proportion of all studies (23%) did not report the selection process for their learners, including one quarter (25%) of the studies categorised as higher reliability by MERSQI criteria.

Educational methods
A wide range of educational methods were employed in various combinations across the reviewed studies, as shown in figure 4. Most interventions included lectures (68%) and small group work (61%). Project work was included in the majority of studies which did not report organisational outcomes (33%). Individual or team mentoring was also more prevalent where organisational outcomes were reported (47% vs 23%).

Educational content
Educational content varied considerably among interventions. The most consistent content area was leadership theory (reported in 65% of interventions). The other common content areas were performance management (44%), self-management (41%), change management (39%), communication (36%), teamwork (33%), quality improvement (30%), healthcare policy (27%), healthcare finance (26%) and leadership behaviours (20%). There were no notable educational content differences in higher reliability studies or in studies which reported organisational outcomes (Kirkpatrick level 4).

Evaluation methods
A wide range of evaluation methods were employed across the included studies. Nearly half used qualitative methods only for their evaluation (46%). Of the remainder, most studies used mixed methods (45%), with 10 studies (9%) using purely
qualitative methods. These proportions were similar in the higher reliability studies (41% quantitative, 48% mixed methods, 10% qualitative).

Four out of every five studies (82%) used questionnaires in their evaluation. Almost all of these employed Likert Scale items (92%) and one-third included open questions (34%). Only 8% used content or construct validated questionnaires. The proportion of higher reliability studies using validated questionnaires was slightly higher at 20% (MERSQI) and 18% (IBI). An additional six studies (6%) had conducted an expert review of their questionnaire for content validity only.

More than two-thirds of the included studies relied solely on self-ratings (69%). A minority of studies included ratings from subordinates (3%), peers (7%), superiors (12%) or experts (20%). The proportion of higher reliability studies which relied on self-ratings was lower (39%), with increased use of ratings from peers (14%), superiors (25%) or experts (39%).

The majority of studies (72%) included the collection of outcome data regarding behavioural changes (Kirkpatrick level 3, 57%) or organisational outcomes (Kirkpatrick level 4, 24%). Only three studies relied solely on Kirkpatrick level 1 outcomes (reaction).36-38

Nearly half of the studies used single group post-programme only designs (46%), with most of the other half using single group pre-programme and post-programme designs (46%). Most studies included a post-programme evaluation completed immediately at the end of the programme (90%). Only 18 studies (15%) included a longer-term evaluation. In higher reliability studies, longer-term evaluations were associated with increased reporting of organisational outcomes (56%) when compared with immediately-post designs (31%). All 16 higher reliability studies as assessed by the MERSQI used pre and post designs. Six of these included a non-randomised control group (38%), and one study included a randomised control group (6%). This was the only randomised control group used in any of the 117 studies.

**Behavioural and organisational outcomes in higher reliability studies**

A full summary of outcomes from all 117 studies is provided in online supplemental table 1.

There was a range of behavioural (Kirkpatrick level 3) and organisational (Kirkpatrick level 4) outcomes demonstrated in higher reliability studies.

Behavioural changes were objectively demonstrated in higher reliability studies through observed changes in behaviour,26-27 39-43 promotions,44 45 increased responsibilities or titles28 46–49 and project completion.30 31 36-38 Subjective changes in behaviour included improved communication,40 influence,50 delegation,57 collaboration,58 involvement in service improvement42 and application of skills learnt or improved leadership in general.39 40 54-57 These changes were indicated through interviews, free text questionnaire responses and behavioural self-assessments.

Organisational outcomes in higher reliability studies (Kirkpatrick level 4) were defined prospectively and in most cases were objectively demonstrated through leadership project impact evaluations. Projects achieved a range of outcomes, including reduced waiting times,50 improved patient care26 46 50 and cost savings.27 46 47 50 By assessing the financial impact of projects completed during the intervention and relating this to programme costs, one higher reliability study reported a 364% financial return-on-investment (ROI).50 Other objective outcomes included reduced organisational turnover of participants.28

improved departmental working climate,39 reduced sick leave44 and increased promotion of women.45 Organisational outcomes were subjectively indicated through reports of increased staff retention36 and improvement in organisational effectiveness.25 One study reported that ‘intangible benefits’ resulted in a 106% financial ROI.31

Organisational outcomes in higher reliability studies were reported more frequently from programmes delivered by a mix of internal and external faculty than from programmes delivered by only external faculty (83% vs 11%), as shown in figure 2. Organisational outcomes were also more frequently reported from interventions conducted in a whole hospital (57%) or multiple hospitals (40%), compared with interventions conducted in a single specialty (conference or outside-hospital training programme) (33%), single university (25%) or in a single department (0%). There were no notable differences in outcomes related to specific educational content.

Higher reliability studies that reported organisational outcomes were more likely have included project work (70% vs 44%), mentoring (50% vs 22%), coaching (22% vs 11%) and reflective instruments such as personality type assessments (22% vs 6%) than higher reliability studies that did not report organisational outcomes. Organisational outcomes were reported less frequently in higher reliability studies that included simulation or role play (10% vs 33%).

**DISCUSSION**

The aim of this review was to synthesise recent empirical evidence and explore factors associated with higher level outcomes in physician leadership development.

We found a substantial increase in the number of studies which evaluate medical leadership development interventions compared with previous reviews.6-10 14 15 In many studies, it is still not clear whether best practices for design, delivery and evaluation are being followed.31 It is also not clear whether there are sufficient behavioural and organisational outcomes to justify the considerable and increasing investments in medical leadership development.

Compared with previous reviews, we found an increase in the proportion of studies which report the use of active learning methods such as project work, simulation, discussions and reflections, which are widely accepted to be a vital component of leadership development55 and which were associated in our review with increased Kirkpatrick level 4 outcomes.

No single leadership development content area was particularly associated with improved outcomes. With respect to educational methods, however, there was an association between the inclusion of individual or group project work and of mentoring with organisational outcomes. This may support the established position that educational methods are more important than specific curriculum content for leadership development.58 Simulation and role play were less common in higher reliability studies which reported organisational outcomes that those that did not report organisational outcomes. This unexpected finding could result from these studies being situated in a training environment rather than a working environment. Alternatively, it could result from the evaluation process and study designs rather than from a lack of organisational impact. Studies which included simulation and role play tended to focus their evaluations on objective changes in behaviour at the expense of evaluating organisational outcomes (see online supplemental table 1). Interestingly, lacking a leadership development framework did not seem to impede programmes from reporting organisational.
outcomes. This may indicate that programmes which are designed as bespoke solutions to local needs are more likely to achieve organisational impact than pre-packaged approaches to leadership development.

There was an additional association of more senior participant level with organisational outcomes. This may be related to the wider scope of influence or practice of senior physicians compared with resident physicians. It could also indicate that there is a longer post-programme development period before residents are able to have an impact on organisational outcomes. This would align with the finding that programmes which evaluated longer-term outcomes were more likely to report organisational outcomes.

Importantly, our findings indicated that leadership development interventions which used a combination of internal and external faculty were most likely to report organisational outcomes, and those interventions which used external faculty only were least likely. This could have significant implications for procurement and design of leadership development interventions across healthcare, particularly as courses run internally are associated with significantly reduced costs.

As in previous physician leadership development reviews that used critical appraisal instruments, we found that studies frequently did not meet criteria for high reliability. Many studies failed to report important methodological features, which restricts readers’ ability to appraise studies and learn from their findings. This was particularly notable in terms of questionnaire design, with fewer than one in 10 studies using validated questionnaires or reporting their questionnaire content in detail. Most studies also did not report or analyse outcome evaluation data comprehensively. Many study designs were biased towards obtaining positive results, particularly in terms of the absence of control groups, having stringent or undisclosed selection criteria, including leading questions on questionnaires and relying solely on self-ratings. This is likely to have resulted in improved reported outcomes. The lack of evaluation quality seems to indicate perfunctory attention paid to evaluation design and precludes confident conclusions from these studies. Future studies could benefit from consulting study quality appraisal checklists such as the MERSQI and JBI in advance, in order to effectively design their evaluations.

This review does indicate that certain recommendations for improved programme evaluation are beginning to be applied into research. Whereas only 29% of the studies reviewed by Frich et al included qualitative components, 63 (54%) of the 117 studies included in our review used mixed or qualitative methods. In a nascent and complex field such as medical leadership development research, qualitative methods can have value in terms of establishing effective programme design features to achieve desired outcomes, as well as helpful nuances of how, for whom, to what extent or in what circumstances interventions are effective or not.

Additionally, many studies in this systematic review evaluated outcomes at Kirkpatrick level 3 behavioural change (57%) or level 4 organisational outcomes (24%). This is a significant improvement from previous reviews. Changes in behaviour (level 3) and organisational outcomes (level 4) are more closely associated with transfer of learning to the working environment than participant reaction (level 1) and learning (level 2).

**Limitations and strengths**

This review was limited by the reliability of the studies included. We attempted to control for study reliability using critical appraisal tools with cut-off scores for higher reliability studies. To the best of our knowledge, this is the first systematic review of healthcare leadership development interventions to use the JBI critical appraisal tool to critically appraise qualitative studies. The JBI tool enabled us to identify 12 additional higher reliability qualitative and mixed-methods studies which were not identified using the MERSQI. Marked heterogeneity of studies and evaluations precluded a formal meta-analysis, therefore, we adopted a meta-aggregation approach. This enabled us to highlight design components that are correlated with behavioural and organisational outcomes in higher reliability studies.

A substantial majority of studies reported only positive outcomes, which could represent a publication bias, and we limited our review to English language peer-reviewed studies. In line with Frich et al, our database search was limited to MEDLINE, however, we augmented our database search with an extensive hand-search of reference lists and citations using Web of Science and Google Scholar. The hand-search revealed that many relevant empirical studies were absent from recent reviews despite some of those reviews searching a greater range of research databases. This could indicate flaws in healthcare leadership development literature tagging and filing procedures within medical and educational databases.

**CONCLUSION**

Our review has practical implications for those commissioning, designing and evaluating medical leadership development programmes in healthcare. No specific area of curriculum content and no particular leadership development framework were clearly associated with behavioural or organisational outcomes. While relevance and appropriateness of educational content is important, this systematic review has more clear implications for leadership development methods than for specific content. Where possible, interventions should include projects and individual or group mentoring. Transfer of learning from the programme into learners’ daily work and their organisations should be planned into the programme and where possible active learning educational designs should be employed, including opportunities for learners to set their own goals for development. External faculty should be judiciously used to supplement in-house faculty, not as a replacement for in-house expertise.

In terms of evaluation design, efforts should be made to ensure that evaluations are cost-effective and produce data that is useful for both practitioners and researchers. Effective mixed-methods evaluation strategies should be integrated into evaluation designs. Study quality checklists such as the MERSQI and JBI could be consulted in the programme design phase to help build high quality quantitative and qualitative evaluation methods into programmes. At the minimum, evaluation design should include consideration of assessment at multiple time points, inclusion of control groups and collection of objective data, as well as collection of qualitative data from interviews, focus groups, questionnaires or observations. Programme goals and intended organisational outcomes should be explicitly considered during evaluation design so that measures of organisational outcomes (including project outcomes) can be incorporated into the evaluation design. Improving study design and building robust evaluation methods into programmes will allow evaluators and educators to more effectively understand factors which are reliably associated with high level programme outcomes. This could both inform the improvement of individual programmes and
contribute to the medical leadership literature as a whole. It is only through more considered and thorough evaluation of physician leadership development programmes that we will be able to invest the money they represent.

Twitter Oscar Lyons @oscarlyonsnz, Jan Frich @J_Frich and Jaason Matthew Geerts @jaasongeerts

Acknowledgements We would like to thank Tatjana Petrinic, University of Oxford Health Care Healthcare Librarian, for her invaluable assistance and advice in the search process.

Contributors OL, RG, JRG, AM and TF abstracted and coded studies. OL, RG, JRG, AM, TF, JF and JMG contributed to analysis, writing and editing the manuscript.

Funding Oscar Lyons was supported during this work by a Rhodes Scholarship, a Goodger and Schorstein Research Scholarship (University of Oxford) and the Shirtcliffe Fellowship (Universities New Zealand)

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMI Publishing Group Limited (BMI) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMI. BMI disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMI does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID ids Oscar Lyons http://orcid.org/0000-0001-5809-7173 Jan Frich http://orcid.org/0000-0001-9079-7508 Jaason Matthew Geerts http://orcid.org/0000-0001-6672-3859

REFERENCES


7


<table>
<thead>
<tr>
<th>Source/First Author, Year</th>
<th>Setting</th>
<th>Learner Number</th>
<th>Learner Type</th>
<th>Intervention Length</th>
<th>Intervention Description</th>
<th>Teaching Methods</th>
<th>Educational Content</th>
<th>Main Findings by Kirkpatrick Level</th>
<th>JB Score</th>
<th>NRSQI Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolfe, 2008 [1]</td>
<td>Two US KICs</td>
<td>76</td>
<td>Physicians/Teachers</td>
<td>10 months</td>
<td>Six externally provided modules, total 25 hrs</td>
<td>Learning activities, small group skill practice, problem-solving sessions, feedback</td>
<td>Leadership, communication, coordination, problem solving/leadership management, and team culture</td>
<td>3.a. Leaders reported increased satisfaction with their own communication and leadership skills; 3b. Communication skills of KIC nurses and physicians improved significantly in simulation (hours 17 to 45/60). Relationship skills remained high (17/18/40).</td>
<td>0/4</td>
<td>16</td>
</tr>
<tr>
<td>Parson, 2008 [2]</td>
<td>Single US hospital residency</td>
<td>34</td>
<td>Residents</td>
<td>4 days</td>
<td>4 days of simulation scenarios</td>
<td>Crew Resource Management [CRM]</td>
<td>Leadership, behaviours of effective leadership, supervision/mentorship</td>
<td>2b. Each team showed an overall gradual improvement in CRM skills compared to the preceding team, suggesting that observational learning of CRM was effective in this setting.</td>
<td>0/4</td>
<td>14</td>
</tr>
<tr>
<td>Cooper, 2008 [3]</td>
<td>UK advanced life support course</td>
<td>36</td>
<td>mixed medical doctors, nurses and technicians</td>
<td>3 days</td>
<td>3-day resuscitation course</td>
<td>Multi-source feedback</td>
<td>Interpersonal dynamics practice followed by a debriefing and self-assessment</td>
<td>2b. Small group team performance improved in all training courses.</td>
<td>0/4</td>
<td>14</td>
</tr>
<tr>
<td>Malling, 2008 [4]</td>
<td>Single educational region in Denmark</td>
<td>28</td>
<td>Consultants (responsible for education)</td>
<td>6 months</td>
<td>Two-three-day residential modules and follow-up day</td>
<td>Pedagogical knowledge, organization of specialist training, educational evaluation and quality assurance, planning specialist training in the department, supervision of supervisors, implementation strategies, personal development, leadership in specialist training, research in medical education</td>
<td>2b. Technical, administrative and human skills feedback did not improve or differ from the control group</td>
<td>6</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>van Vletter, 2008 [5]</td>
<td>University hospitals in Sweden</td>
<td>52</td>
<td>Specialists, senior physicians, heads of departments</td>
<td>1 year</td>
<td>Three programs, including mentor programs, management networks, and teams, held across 1 year; no details on number, duration, or allocation to programs</td>
<td>Mentoring/Networking/Lectures</td>
<td>2b. Citizenship behaviors did not improve or differ from the control group.</td>
<td>0/4</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Fassiotto, 2008 [6]</td>
<td>Single US hospital</td>
<td>121</td>
<td>Assistants/instructors/full professors</td>
<td>9 months</td>
<td>6 x 1.5 day sessions over 8 months</td>
<td>Interactive teaching methods based on adult learning principles and actual learning projects</td>
<td>Personal development as a leader Managing people and relationships Managing groups and projects</td>
<td>1. Positive qualitative feedback about the course</td>
<td>3</td>
<td>13.5</td>
</tr>
<tr>
<td>Levie, 2008 [7]</td>
<td>Single US academic medical center</td>
<td>47</td>
<td>Residents (Chief residents, medicine and surgery)</td>
<td>1 year</td>
<td>Two-day intensive immersion course, project work</td>
<td>Small group discussions, evidence-based mini-lectures, interactive seminars, overview of project meetings</td>
<td>Management of complex older patients, geriatric principles, giving feedback, approaching the reluctant learner, conflict resolution</td>
<td>2. Increased personal confidence and personal positive feedback</td>
<td>0</td>
<td>12.5</td>
</tr>
<tr>
<td>Hopkins, 2008 [8]</td>
<td>Single US hospital network</td>
<td>113</td>
<td>Senior medical leaders and academic faculty, 18 administrators, 94 doctors</td>
<td>9 months</td>
<td>Six 1.5 day sessions spaced over 8 months</td>
<td>Baseline assessments of their leadership competence, Multi-source feedback, Myers Briggs Type Indicator and the Thomas-Kilmann Conflict-Mode Instrument</td>
<td>Leadership, communication, coordination, problem solving/leadership management, and team culture</td>
<td>3a. Leaders reported increased satisfaction with their own communication and leadership skills; 3b. Communication skills of KIC nurses and physicians improved significantly in simulation (hours 17 to 45/60). Relationship skills remained high (17/18/40).</td>
<td>0/4</td>
<td>16</td>
</tr>
<tr>
<td>Danovs, 2008 [9]</td>
<td>Single US university executive education programme</td>
<td>78</td>
<td>Female academic medical faculty (associate or full professor level)</td>
<td>1 year</td>
<td>Executive leadership development program for senior faculty</td>
<td>Not specified</td>
<td>Leadership, communication, coordination, problem solving/leadership management, and team culture</td>
<td>2a. 4.5% of participants achieved higher leadership position inside an academic health centre; 3b. 56% of participants increased in educational and professional status (from programme to associate professor)</td>
<td>0/4</td>
<td>13</td>
</tr>
<tr>
<td>Orme, 2018 [10]</td>
<td>Single UK hospital trust</td>
<td>625</td>
<td>Consultants and senior healthcare professionals and managers</td>
<td>12 months</td>
<td>12 months with 5 workshop days spread over 6 months</td>
<td>Five face-to-face delivery days, ongoing telephonic coaching and the use of a benchmarked 200 degree profile. Project work, interactive sessions.</td>
<td>Not well reported. Influence, behaviour, team behaviour, finances. From website: Executive people management, evaluation and feedback</td>
<td>4b. Reports of improved communication and efficiency (6/10) were significant</td>
<td>6</td>
<td>12.5</td>
</tr>
</tbody>
</table>
Source (First Author, Year) | Setting | Learner Number | Learner Type | Intervention Length | Intervention Description | Teaching Methods | Educational Content | Main Findings by Kirkpatrick Level | HRQOL Score | MERSQI Score | JBI Score | BMJ Leader
--- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
Codr, 2017 [14] | Single US hospital department | 10 | Residents | (Anesthesiology FOGY/V) | 2 weeks | 2 weeks operating room management and leadership elective rotation | Role-play | Action learning | Literature review | Non-technical skills | Return on investment of BMI/QLI; quality improvements were observed 3a. Increased "re-evaluates and debriefs"; "gathers and actively seeks out information"; "anticipates changing environment" 3b. Increase across a range of metrics measuring teamwork, task management, clinical decision making, situational awareness, as measured by "interpersonal" non-technical skills (INTS) questionnaire 3c. Statistically significant increase in leadership of national committees, production of national workshpap and presentation at national platforms | n/a | 12.5 | n/a | n/a | n/a | Lyons O, et al. BMJ Leader 2020; 8:10. doi: 10.1136/leader-2020-000360
Haftel HM, 2018 [12] | Single US specialty association, 45 sites | 49 | Paediatric academic faculty | | 10 months | 3 sessions focusing on the individual, their training environment and the role of mentorship in interprofessional interaction with physicians | "highly interactive format" | peer mentorship | Professional development, leadership training, administrative skill development | 3b. Participants increased significantly in their performance of 7 of the 10 tasks on the interprofessional ward round assessment tool. Post-test, the participants performed these behaviors significantly more frequently than an experienced control group. | n/a | 12.5 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a | Lyons O, et al. BMJ Leader 2020; 8:10. doi: 10.1136/leader-2020-000360
Ten Haan, 2013 [15] | Four Dutch ICUs (n= Single hospital) | 28 | Exposure intensive care fellows | Control: experienced nurses 1 full day per month & all-day workshop | 23 months | 1 day simulation, group feedback on observed interdisciplinary teamwork; leader who attended before or after post-training added | Multiple learning activities including simulation; small group skill practice and problem solving sessions; performance feedback and reinforcement of newly learned skills; and a planning assignment for on the job applications. | Leading an interdisciplinary ward round | 2b. Statistically significant increase in leadership of national committees, production of national workshpap and presentation at national platforms | n/a | 12.5 | n/a | n/a | n/a | Lyons O, et al. BMJ Leader 2020; 8:10. doi: 10.1136/leader-2020-000360
Wurzel, 2007 [12] | Single US department of surgery | 42 | Surgical fellows | 6 months | Long weekend of didactic study, brainstorm on patient safety-related concepts in 20 modules; didactic group project; monthly conferences; 2 days for didactic sessions and project presentations | Didactic study, group projects with monthly conferences and project presentations and catapults lectures | | | | | | | Lyons O, et al. BMJ Leader 2020; 8:10. doi: 10.1136/leader-2020-000360
Higher Variability Studies (SR Text)

Thaggard, 2016 [18] | US Regional healthcare system | 21 | Physicians across a range of specialties | 10 months | 2-5 meetings/month v. 2-3 of additional learning opportunities | Behavioral style assessment, multi-source feedback, coaching, online discussions, didactic learning modules and project in small groups | Intra/interpersonal effectiveness, resiliency, coaching, communication, teamwork, change management, business acumen, quality focus | | | | | | | Lyons O, et al. BMJ Leader 2020; 8:10. doi: 10.1136/leader-2020-000360
Bergman, 2009 [19] | Single Swedish hospital | 53 | Managers (5 physicians, 10 nurses, and 15 other health personnel) | 7 weeks x 17 sessions | 1. One week intensive course 2. Long-term support group (previously completed intensive); 3. Long-term support group (but not completed intensive) | A one-week course and a long-term support group. The intensive course consisted of modules using reflection and metacognitive reflection supported by theoretical frameworks. The long-term support group met for a half day once a month for 6 months. The long-term support group included didactic lectures that arose in the everyday work. Formal training, debriefing of learning needs, mentoring, project work and on-line peer partner | Group dynamic, communication, leadership theories | | | | | | | Lyons O, et al. BMJ Leader 2020; 8:10. doi: 10.1136/leader-2020-000360
<table>
<thead>
<tr>
<th>Source (First Author, Year)</th>
<th>Setting</th>
<th>Learner Number</th>
<th>Learner Type</th>
<th>Intervention Length</th>
<th>Intervention Description</th>
<th>Teaching Methods</th>
<th>Educational Context</th>
<th>Main Findings by Kirkpatrick level</th>
<th>JBI Score</th>
<th>MEIQ Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beerman, 2002 [22]</td>
<td>Single Australian residency program (surgey)</td>
<td>32</td>
<td>Residents (surgical trainees, interval across all specialties)</td>
<td>2 days</td>
<td>Two day simulation course</td>
<td>Simulation, peer observations, multi-source feedback, reflection, lectures, videos, scenarios</td>
<td>Patient-centred communication, inter-professional communication, teamwork, leadership and professionalism</td>
<td>3b. 60% of graduates reported a new leadership position, 44% of failure graduates, 83% of underrepresented minority graduates. 1. All participants rated the course as good or very good. One third of participants described the communication scenario as “less than useful”. All other aspects of the course were considered useful or highly useful. 2a. Increased awareness of the broader situation and the value of high-quality communication and teamwork. 2b. Self-reported achievement of learning objectives including increased knowledge 1. Participants satisfied 0.5–5.5/7, useful 0.5–5.5/6.5 on 7 topics. 26. 52%-56% report intention to implement</td>
<td>8</td>
<td>7.5</td>
</tr>
<tr>
<td>Carey, 2015 [23]</td>
<td>12 US Primary Care residencies (6 locations)</td>
<td>33</td>
<td>Faculty (Family medicine, internal medicine, family practice)</td>
<td>6 months</td>
<td>2 Day session with follow-up over 6 months</td>
<td>Didactic slide presentations workshops conference calls visits by core faculty</td>
<td>Leadership change management frameworks, personal leadership management, clinical microsystems competency assessment, patient-centredness and patient-centred medical home principles.</td>
<td>3. Changes in relevance and quality of simulations on questionnaire and free text comments rated &gt;4/5. 2a. The simulation helped participants recognize problems with speaking up. 2b. Gained understanding about shifting from blame to learning centered leadership, facilitating communication and teamwork, being welcoming rather than defensive, and other self-reflections.</td>
<td>8</td>
<td>6.5</td>
</tr>
<tr>
<td>Cooper, 2011 [24]</td>
<td>Single UK medical centre</td>
<td>108</td>
<td>Physicians, nurses, allied health professionals, administrative managers</td>
<td>1 day</td>
<td>Workshop</td>
<td>Seminar simulation</td>
<td>Review of data from safety climate surveys team project</td>
<td>3a. Participants were happy with the course 2a. Participants reported improved confidence and judgement skills. 2b. Participants reported improved leadership skills 3b. Participants reported taking new approaches to their roles 3b. Participants reported new roles. 4a. Projects had significant system impact including organisational changes, improved patient experiences, cost savings. 5a. Participants overwhelmingly endorsed the programme and would recommend it to other trainees. 5b. Increased self-confidence as leaders (57% of respondents) and willingness to speak up 6a. Increased understanding of health systems, policy, team working skills with diverse groups, networking. 7a. All graduates report being actively involved in service improvements post-programme. 80% of graduates reported difficulties transferring their learning back to their workplace. 9a. 41% of graduates had gone on to further leadership/management development as a result of the programme 10a. Host organisations describe a range of benefits and examples of the impact of fellows’ work on their organisations, including financial benefit (e.g. income generation, cost savings) and a range of deliverables (e.g. reports, publications, research studies). 11a. All measures above 4.5/5 on latent variable “I recommend this simulation to colleagues at my professional stage” (K-06) and “I would like to take part in similar simulation events in future”. 2a. Self-reported “learning from the simulation will help me deliver better long term care to my patients” (K-01). 2b. Self-reported increase in understanding of healthcare organization (K-02) in all seven knowledge areas (regulation of healthcare providers, role of patient organizations, organizational accountability. role of local authorities, financial climate, role and responsibilities of commissioners and core providers). There was a significant improvement in perceived knowledge scores. 3a. Capability was the only construct that showed a large and significant change post-simulation. There was a moderate but significant change in behavioural intentions, attitudes and subjective norms post-simulation, although there was no significant change detectable in opportunity in the short time between the simulation and feedback. 22 clinicians rated that their experiences in the crucible simulation had directly influenced their leadership practice increase in perceived self-efficacy post-simulation (mean score pre-simulation, 3.87 to post-simulation, 4.08). 4a. All respondents were positive about the value and success of the programme 5a. Participants reported a marked increase in the use of discharge and end of life planning in their practice</td>
<td>7</td>
<td>10.5</td>
</tr>
<tr>
<td>McEldoon, L, 2010 [26]</td>
<td>National UK programme</td>
<td>145</td>
<td>Residents (specialty and GP doctors in training)</td>
<td>1 year</td>
<td>Immersive internship out of practice with the most senior personnel in national and health-related organisations</td>
<td>Immersive internship, visits to other host organisations and training, action learning sets</td>
<td>Policy development, project management, research and analysis, writing and publishing, professional networking skills.</td>
<td>3a. Capability was the only construct that showed a large and significant change post-simulation. There was a moderate but significant change in behavioural intentions, attitudes and subjective norms post-simulation, although there was no significant change detectable in opportunity in the short time between the simulation and feedback. 22 clinicians rated that their experiences in the crucible simulation had directly influenced their leadership practice increase in perceived self-efficacy post-simulation (mean score pre-simulation, 3.87 to post-simulation, 4.08). 4a. All respondents were positive about the value and success of the programme 5a. Participants reported a marked increase in the use of discharge and end of life planning in their practice</td>
<td>7</td>
<td>10.5</td>
</tr>
<tr>
<td>Coyne, 2017 [27]</td>
<td>Multiple UK health care organizations (London)</td>
<td>60</td>
<td>Consultants (52), Registrars (5), Grade not identified</td>
<td>1 day</td>
<td>Workshop</td>
<td>Lecture-style presentations policy leadership simulation</td>
<td>Background of MSc reform healthcare challenges</td>
<td>3b. Statistically significant improvement in leadership strengths and in organizational accountability, role of local authorities, financial climate, role and responsibilities of commissioners and core providers. There was a significant improvement in perceived knowledge scores. 3a. Capability was the only construct that showed a large and significant change post-simulation. There was a moderate but significant change in behavioural intentions, attitudes and subjective norms post-simulation, although there was no significant change detectable in opportunity in the short time between the simulation and feedback. 22 clinicians rated that their experiences in the crucible simulation had directly influenced their leadership practice increase in perceived self-efficacy post-simulation (mean score pre-simulation, 3.87 to post-simulation, 4.08). 4a. All respondents were positive about the value and success of the programme 5a. Participants reported a marked increase in the use of discharge and end of life planning in their practice</td>
<td>7</td>
<td>10.5</td>
</tr>
<tr>
<td>Ruston, 2010 [28]</td>
<td>Single UK diabetes</td>
<td>3</td>
<td>Residents (General Practice Specialty Trainees)</td>
<td>4 months</td>
<td>2 days per week for 4 months</td>
<td>Peer learning sets, meetings, project work, reflective diaries</td>
<td>Strategic and contextual issues, commissioning design and delivery of health care, service redesign, public health agenda, leadership, management and partnership skills</td>
<td>3a. Capability was the only construct that showed a large and significant change post-simulation. There was a moderate but significant change in behavioural intentions, attitudes and subjective norms post-simulation, although there was no significant change detectable in opportunity in the short time between the simulation and feedback. 22 clinicians rated that their experiences in the crucible simulation had directly influenced their leadership practice increase in perceived self-efficacy post-simulation (mean score pre-simulation, 3.87 to post-simulation, 4.08). 4a. All respondents were positive about the value and success of the programme 5a. Participants reported a marked increase in the use of discharge and end of life planning in their practice</td>
<td>6</td>
<td>10.5</td>
</tr>
</tbody>
</table>

Other Included Studies

Hartwick, L, 2019 [29] Single US hospital | 98 | Medical faculty members with supervisory or team leader responsibility | ≥ 10 months | 48 hours of learning over 10 month course. Two one day seminars, one two day off site session, seven monthly four hour meetings, reading assignments | Semiannual, experiential activities, small group discussions, multifaceted feedback, teambuilding activities, reading | Emotional Intelligence, leadership behaviors, leadership foundation skills | 3b. 94% of participants would recommend the programme to others, rating rated as worthwhile investment (K-77). 2a. Statistically significant increase in leadership self-efficacy. 3a. Statistically significant improvement in 2020 leadership measure 4a. Statistically significant increase in setting ways to maximize application of leadership strengths and to overcome leadership limitations 5a. Statistically significant improvement in communication of leadership strengths and in confidence as seen by managers | 3           | 11.5       |
### Source (First Author, Year)

<table>
<thead>
<tr>
<th>Source</th>
<th>Setting</th>
<th>Learner Number</th>
<th>Learner Type</th>
<th>Intervention (Length)</th>
<th>Intervention Description</th>
<th>Teaching Methods</th>
<th>Educational Context</th>
<th>Main Findings by Subject Area</th>
<th>JBI Score</th>
<th>MERSQI Score</th>
<th>BMJ Leader Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATAKA, 2016 (48)</td>
<td>Family resilience programme</td>
<td>150</td>
<td>Residents (family medicine year 1-4)</td>
<td>5 days</td>
<td>In-person management workshop</td>
<td>Didactic teaching, case studies, small group discussions, team exercises</td>
<td>Professional/interpersonal skills practice-based learning and improvement systems-based practice</td>
<td>2a. Increased confidence 2b. Increased self-assessed risk management, conflict management, communication skills, time management, ability to write algorithms</td>
<td>n/a</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Fernandez, 2014 (50)</td>
<td>Single US college (MD)</td>
<td>37</td>
<td>Obstetricians (Senior fellow, junior physicians, senior fellows, specialists)</td>
<td>5 days</td>
<td>5 day national intensive leadership development for OB/GYN physicians</td>
<td>Interactive skills-building workshops, series of leadership and psychological assessment tools, including a 360-degree feedback, free-form discussion, large-group case scenarios, didactic presentations, small group discussions, structured networking, individual leadership reflection, self-assessment, leadership assessment tools</td>
<td>Organizational culture, leadership and empowering, communication, motivation, advocacy, skills, negotiation skills, health policy</td>
<td>1. 100% of respondents indicated that they would recommend the course to colleagues 2. Posttest scores were significantly and meaningfully higher than pretest at the p &lt; 0.001 level in all 10 targeted leadership skills, both immediately following completion of the course and 6 months post completion. Mean difference ranged from 0.8 to 1.5 (5 point scale)</td>
<td>4</td>
<td>11</td>
<td>n/a</td>
</tr>
<tr>
<td>Chang, 2016[32]</td>
<td>Three US healthcare organizations in 26 US states</td>
<td>65</td>
<td>40 Physicians (GI, Gastroenterology) 1 Doctor 3 Nurses 6 Pharmacists 1 Psychologist 1 Social Worker 1 Speech Pathologist</td>
<td>9 months</td>
<td>Silths</td>
<td>Adoption leadership managing complexity and leadership mission, vision, goals/diagnosis assessment, motivation and resilience, managing transnational influence and persuasion, ethics and cultural values, role and influence presentation skills, negotiation and marketing strategy, stakeholders evaluation design, scaling up business case and budget measuring impact project management</td>
<td>Unclear</td>
<td>3a. Respondents reported having used skills learned in their daily work. 3b. GERS respondents had expanded leadership responsibilities in a new role. 1. Measured satisfaction score of 4.8/5 2a. Significant increase in confidence in all subscales. 2b. Self-assessment management, empathy, communication planning, and execution (mean increased from 5.0 to 5.1)</td>
<td>1</td>
<td>1</td>
<td>n/a</td>
</tr>
<tr>
<td>Day, 2010[30]</td>
<td>Single US specialty association (orthopaedics)</td>
<td>100</td>
<td>Orthopaedic surgeons</td>
<td>1 year</td>
<td>Mentoring by established orthopaedic leaders</td>
<td>Mentoring</td>
<td>Unclear</td>
<td>2b. Significant increase in 2b leadership domains: knowledge of theory, tolerance for demands of leadership, and leadership positioning</td>
<td>n/a</td>
<td>11</td>
<td>n/a</td>
</tr>
<tr>
<td>Korschun, 2007[34]</td>
<td>Single US academic medical centre</td>
<td>70</td>
<td>Physicians (25), nurses, and a wide range of administrators</td>
<td>5 months</td>
<td>Five three-day sessions over five months</td>
<td>Lectures, seminars, case studies, experiential exercises, individual assessment, executive coaching, including in 360 assessment, mentoring, team project work</td>
<td>Strategic thinking and personal awareness, leadership qualities, best practices, negotiating tactics and managing conflict, human resources, knowledge, and talent management, building collaborations and influence skills, leadership, change management, team leadership, and patient safety, quality of care and transformation; emerging issues in academic medicine; communication; personal dimensions of medicine</td>
<td>1. Participants reported positive experiences with the program. Attendance was at a rate of 95% or higher. 2a. 34% of respondents disagreed that the mentoring process had helped them with their professional growth. 2b. 98% increased their commitment to and support of the vision and strategies of the organization</td>
<td>n/a</td>
<td>11</td>
<td>n/a</td>
</tr>
<tr>
<td>McCool, 2004[35]</td>
<td>Multiple US and Canada Academic Medical Centers (ACMC)</td>
<td>79</td>
<td>Assistants or full professors</td>
<td>7 months</td>
<td>Three 1-week meetings across 7 months</td>
<td>Leadership skill development, mentoring, and networking</td>
<td>The curriculum focuses on building knowledge and skills in seven domains: paradigms of corporate, government, and academic leadership, financial management; strategic planning and organizational transformation; emerging issues in academic medicine; communication; personal dimensions of leadership; and career advancement strategies. Self-leadership, leading others, collaborative working, delivering excellence through others, fragmented, managing complex change, improving patient experience, political awareness, strategic dexterity, aligning agendas/leading change</td>
<td>2a. Significant increase in knowledge of leadership qualities, leadership best practices, negotiating tactics and managing conflict, human resources, knowledge, and talent management, building collaborations and influence skills, leadership, change management, team leadership, and patient safety, quality of care and transformation; emerging issues in academic medicine; communication; personal dimensions of medicine</td>
<td>n/a</td>
<td>11</td>
<td>n/a</td>
</tr>
<tr>
<td>Edmonstone, 2015[36]</td>
<td>UK National program (Scotland)</td>
<td>117 (from total 1 column)</td>
<td>Senior leaders in clinical leadership, primary care doctors, nurses, GPs, pharmacists, psychologists, dentists, paramedics</td>
<td>12 months</td>
<td>A 3-month residential development centre followed by two-two day residential events held every six weeks</td>
<td>Coaching, Mentoring, Action Learning, Change Management, Mentoring, Mini-Workshops, Workshops, Doctor's Workshops, Shadowing, Development Centre, Leadership Competency Assessment, Research Skills, Personal Development Plan</td>
<td>Successful adoption of most projects into the organisation</td>
<td>4a. Fellows stated that the leadership academy allowed them to establish strong network of links and colleagues from other disciplines, and after the program was over, they found it much easier to seek advice or establish collaborations with peers in other parts of the organization. 4b. Graduates reported being more effective in committees within their school or the university.</td>
<td>5</td>
<td>10.5</td>
<td>5</td>
</tr>
<tr>
<td>Martin, 2015[37]</td>
<td>Single Australian Hospital Department (Gastroenterology, Rehabilitation and Palliative Care)</td>
<td>20</td>
<td>Non-executive middle or senior level medical, nursing and allied health professionals</td>
<td>9-10 months</td>
<td>3-hour monthly sessions, group projects, site visits</td>
<td>Guest speakers and discussions external and clinical, project, supervision</td>
<td>Organisational structure, healthcare context, leadership and political culture, quality of care and clinical errors, complex systems, communication, engagement with patients, conflict, negotiation, change management</td>
<td>1. Participants reported almost unanimously (94/100) for all measures including that the course was relevant and valuable 2. Participants reported increased willingness to take on leadership roles 3. Participants reported changes in comments that they benefited from increased networking and relationships across sites.</td>
<td>5</td>
<td>10.5</td>
<td>5</td>
</tr>
<tr>
<td>Source (First Author, Year)</td>
<td>Setting</td>
<td>Learner Number</td>
<td>Learner Type</td>
<td>Intervention Length</td>
<td>Intervention Description</td>
<td>Teaching Methods</td>
<td>Educational Context</td>
<td>Main finding by Kirkpatrick level</td>
<td>Net Score</td>
<td>NMQOS Score</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td>----------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>-------------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>----------------------------------</td>
<td>-----------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>Correia, 2015 [46]</td>
<td>Single US integrated health care organization</td>
<td>56</td>
<td>Residents (including Chief Residents)</td>
<td>3-4y</td>
<td>85%</td>
<td>3GEM</td>
<td>Enhanced leadership identity, strategies to manage emotions, interpersonal and communication skills, writing, presentation skills</td>
<td>Leadership, managing, and core feedback skills, interpretation of emotional intelligence inventory, interpersonal and communication skills and professionalism</td>
<td>1. Participants reported that the program had a positive impact on their learning needs. Participants reported continued use of the GEM tool</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Patel, 2015 [50]</td>
<td>Single US hospital</td>
<td>30</td>
<td>Residents (PGY-4)</td>
<td>2y</td>
<td>2 years healthcare leadership in quality residency track</td>
<td>Educational intervention, simulated teaching encounters, core curriculum (10 hours over 3 weeks including lectures, readings, videos, small group activities, online modules, facilitated discussion)</td>
<td>Methods and tools of quality improvement and patient safety, human factors engineering and safety culture</td>
<td>1. The core curriculum has been considered highly valuable and necessary (mean 4.61 and 4.68/5 respectively)</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nakajjako, 2015 [46]</td>
<td>Four African and 4 US universities</td>
<td>15</td>
<td>Unclear</td>
<td>1y</td>
<td>1 year fellowship</td>
<td>Educational intervention, 8 weeks of didactic teaching with two 4.5 months of experiential training in health care organizations and 6 online modules, meeting (weekly meetings, monthly mentoring, team meetings, simulation for learning)</td>
<td>Leadership, communication, monitoring and evaluation, health information, research methodology, grant writing, implementation science, and responsible conduct of research</td>
<td>1. Participants reported use of new knowledge and skills in both current and new leadership roles</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuo, 2010 [46]</td>
<td>Single US residency program</td>
<td>16</td>
<td>Residents (Pediatrics PGY1-3)</td>
<td>3y</td>
<td>3 years</td>
<td>Baseline residency, program overview</td>
<td>Themes of leadership, critical thinking, and community engagement. Topics include policy making, project management, decision making and communication</td>
<td>1. Participants reported satisfaction with the programme and outcomes</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brandt, 2013 [42]</td>
<td>Single US residency program (Radiology)</td>
<td>44</td>
<td>Residents (Radiology)</td>
<td>1y</td>
<td>7x60min modules</td>
<td>Faculty teaching</td>
<td>Quality improvement projects, team building, more learning sessions and planning for a six-month action period (following the meetings, Teams from subsequent waves overlapped)</td>
<td>Finance, quality improvement, employment, organizational dynamics, health policy and economics, negotiation and conflict management</td>
<td>1. Improvement of participants’ knowledge and skills (p&lt;.001)</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Green, 2003 [43]</td>
<td>US network of community-owned health care providers and physicians</td>
<td>26</td>
<td>Teams from eight organizational units</td>
<td>2y</td>
<td>Coaching and leadership initiative</td>
<td>Faculty teaching</td>
<td>Quality improvement projects, Team meetings, with small group discussions for learning</td>
<td>Quality improvement projects, team building, project management, coaching and building further organizational capacity</td>
<td>1. Participants reported that they incorporated knowledge and skills into their workshops</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Hamner, 2007 [44]</td>
<td>Single US residency/ fellowship program (pathology)</td>
<td>16</td>
<td>Residents and fellows (Pathology)</td>
<td>1y</td>
<td>6 to 2 day workshops (average 20h per week)</td>
<td>Faculty teaching</td>
<td>Quality improvement projects, team building, project management, small group discussions</td>
<td>Leadership and management basics, managing change and interpersonal skills, emotional intelligence, organisational dynamics, decision-making, accountability, negotiation and conflict management</td>
<td>1. Participants evaluated their positive impact on their learning needs and satisfaction</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>McCurdy, 2008 [45]</td>
<td>Single US academic medical centre</td>
<td>22</td>
<td>Faculty members who were at the time of the course, in a leadership position or likely to move into a leadership position</td>
<td>12 months</td>
<td>Eight 3-hour sessions in two 6 months apart</td>
<td>Coaching and leadership initiative</td>
<td>Faculty teaching</td>
<td>Quality improvement projects, team building, project management, coaching and building further organizational capacity</td>
<td>1. Participants reported that the programme improved their learning needs</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Hadley, 2014 [46]</td>
<td>Single US training discovery</td>
<td>30</td>
<td>Residents (PGY-2 doctors) paired with a management trainee</td>
<td>6-0 months</td>
<td>Project work, mentoring, small group discussions</td>
<td>Project work, mentoring, small group discussions</td>
<td>Leadership and management basics, managing change and interpersonal skills, emotional intelligence, organisational dynamics, decision-making, accountability, negotiation and conflict management</td>
<td>1. Participants provided positive feedback on the programme</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reviewed, 2013 [47]</td>
<td>Single US hospital network</td>
<td>50</td>
<td>Senior Physicians (very good table in the paper with details of each)</td>
<td>6 months</td>
<td>Leadership and management basics, managing change and interpersonal skills, emotional intelligence, organisational dynamics, decision-making, accountability, negotiation and conflict management</td>
<td>Leadership, communication, monitoring and evaluation, health information, research methodology, grant writing, implementation science, and responsible conduct of research</td>
<td>1. Participants reported use of new knowledge and skills in both current and new leadership roles</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osburn, 2004 [44]</td>
<td>Texas US Pediatric Association locations</td>
<td>22 (total of 2 columns)</td>
<td>Pediatricians from AAMC</td>
<td>2-3 years</td>
<td>Three 2-day workshops during first 12 months, then 5 Advanced workshops annually (only one for cohort 2)</td>
<td>Workshops, project work</td>
<td>Quality improvement projects, team building, project management, coaching and building further organizational capacity</td>
<td>1. Participants reported that the programme improved their learning needs</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BMJ Leader

Supplemental material: BMJ Leader disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)
<table>
<thead>
<tr>
<th>Source (First Author, Year)</th>
<th>Setting</th>
<th>Learner Number</th>
<th>Learner Type</th>
<th>Intervention Length</th>
<th>Intervention Description</th>
<th>Teaching Methods</th>
<th>Educational Content</th>
<th>Main Findings by Kirkpatrick level</th>
<th>JBI Score</th>
<th>MERSQI Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weinhrau, 2009 (46)</td>
<td>Single US hospital department</td>
<td>6</td>
<td>Residents (Psychiatry PGY4)</td>
<td>8 months</td>
<td>Weekly seminars (8 modules x 4 sessions each), with projects and mentoring</td>
<td>Lectures, discussions, projects, mentoring</td>
<td>Financial management, human resources management, planning and marketing, information management, risk management, Governance and organizational dynamics, Business and clinical operations, Professional responsibility.</td>
<td>4b. Project resulted in decreased non-attendance by new patients by 50% across 12 months.</td>
<td>r/s</td>
<td>9.5</td>
</tr>
<tr>
<td>Monaghan, 2018 (50)</td>
<td>Single US hospital trust</td>
<td>13 (6 doctors)</td>
<td>Residents (non-clinical training and internal medicine fellows, managers of various departments)</td>
<td>6 months</td>
<td>6-month paired learning: doctors/managers 6-month paired learning matching doctors with mentors, Shadowing, conversation, reflections</td>
<td></td>
<td>2a. Physician participants reported feeling more prepared for a range of leadership requirements, including understanding decisions and working in teams and with managers 2b. Two pairs collaborated to successfully implement a lasting organisational change in the form of a trainee management forum.</td>
<td>4</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Voigt, 2016 (51)</td>
<td>Six Dutch teaching hospitals</td>
<td>98</td>
<td>Residents (radiology)</td>
<td>12 months</td>
<td>Eight sessions offered, Twelve sessions on Fridays at an off-site retreat centre, One hour lightning journal club meeting every fortnight</td>
<td>Facilitated discussions, project work</td>
<td>Quality improvement and leadership, not otherwise specified.</td>
<td>2a. Interviews reported feeling empowered 2b. Interviews reported increased awareness of organisational aspects of healthcare delivery 3a. Self-reported improved communication, interaction with referring physicians, career development team building, group dynamics, and interactions with hospital administration 3b. 95% had assumed leadership roles within three years. Publication of 27 articles in core radiology journals, completion of multiple projects. 4a. New external collaboration for residents.</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Hekimian, 2017 (52)</td>
<td>Single US hospital department</td>
<td>06</td>
<td>Residents (radiology)</td>
<td>NR (variable, journal club)</td>
<td>One hour lightning journal club meeting every fortnight</td>
<td>Journal club, projects, mentoring, leadership role placement</td>
<td>Leadership (topics chosen by the group on an ad hoc basis)</td>
<td>2a. Participants felt better prepared for their careers 2b. Self-reported increased understanding of the business of radiology 3a. Self-reported improved communication, interaction with referring physicians, career development team building, group dynamics, and interactions with hospital administration 3b. 85% had assumed leadership roles within three years. Publication of 27 articles in core radiology journals, completion of multiple projects. 4a. New external collaboration for residents.</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Pearson, 2018 (53)</td>
<td>UK leadership fellowship</td>
<td>13</td>
<td>Residents (junior, medicine, general practice, surgery, obstetrics and gynecology, pediatric, surgery and psychiatry)</td>
<td>1 year</td>
<td>1 year out-of-programme fellowship in a host organization</td>
<td>Symposium and conferences, one-to-one coaching sessions, action learning sets, shadowing opportunities and reflexive practice including completion of a portfolio. Project work for host organisation. Interactive lectures</td>
<td></td>
<td>1. Mentors reported that all components of the course were very or slightly useful. 75% did not access multi-source feedback component. 2a. Most participants reported improved attitudes towards leadership and their ability to make changes in their organisation. Reports of increased confidence. 2b. Reports of increased awareness of other working styles and characteristics of good leadership. 3a. Reports of adopting behavior towards other working styles. 4a. Participants rated the course effectiveness 4.13/5. 4b. The residents’ views towards practice management education in general had mean scores of 4.67/5. 5a. Participants scored significantly higher on self-assessed management skills, from 2.42 to 3.60/5. Average score on knowledge test significantly increased from 74% to 95%. 3b. Three projects undertaken as part of voluntary elective. 4b. One project resulted in a reduction of admission time of 60 minutes.</td>
<td></td>
<td>r/s</td>
</tr>
<tr>
<td>Citrin, 2004 (54)</td>
<td>Single US residency program (internal medicine and pediatrics)</td>
<td>13</td>
<td>Residents (Psychiatry PGY4-5)</td>
<td>1 year</td>
<td>Monthly seminar series</td>
<td></td>
<td>2a. No change in knowledge in exposure or control. No difference in self-assessed change in newly acquired leadership role. 2b. Participants reported increased confidence across a range of key curriculum areas. 3a. Participants reported the programme had helped them to gain new leadership roles, and that they had applied learning from the programme in their new leadership roles. 3b. Participants reported increased confidence across a range of key curriculum areas. 4a. Of those that acquired a new leadership opportunity, 55% of respondents reported participating in the leadership academy program had an impact in their success within the newly acquired leadership role(s). 4b. No change in leadership skills or other measures of interest. 5a. Change in self-assessed leadership skills or other measures of interest (mean scores from 5.1 to 5.5). Average score on knowledge test significantly increased from 76% to 95%. 5b. Three projects undertaken as part of voluntary elective. 4b. One project resulted in a reduction of admission time of 60 minutes.</td>
<td></td>
<td>r/s</td>
<td></td>
</tr>
<tr>
<td>Dickey, 2014 (55)</td>
<td>Single US hospital</td>
<td>Unclear</td>
<td>Residents (Psychiatry PGY4-5)</td>
<td>4 years</td>
<td>Modular leadership programme over 4 years with monthly seminars and monthly clinical skills hour</td>
<td>Seminars, with voluntary simulations, action teams work, eMentoring, mentoring</td>
<td>Leadership of healthcare, healthcare delivery system, quality assurance, risk management, queries of exceptional leaders.</td>
<td>1. Participants rated the course effectiveness 4.13/5. 2a. Participants reported increased confidence across a range of key curriculum areas. 3a. Participants reported the programme had helped them to gain new leadership roles, and that they had applied learning from the programme in their new leadership roles. 3b. Participants reported increased confidence across a range of key curriculum areas. 4a. Of those that acquired a new leadership opportunity, 55% of respondents reported participating in the leadership academy program had an impact in their success within the newly acquired leadership role(s). 4b. No change in leadership skills or other measures of interest. 5a. Change in self-assessed leadership skills or other measures of interest (mean scores from 5.1 to 5.5). Average score on knowledge test significantly increased from 76% to 95%. 5b. Three projects undertaken as part of voluntary elective. 4b. One project resulted in a reduction of admission time of 60 minutes.</td>
<td></td>
<td>r/s</td>
</tr>
<tr>
<td>Fodor, 2008 (56)</td>
<td>Single US medical centre</td>
<td>Unclear</td>
<td>Residents (Psychiatry PGY4-5)</td>
<td>13</td>
<td>Residents</td>
<td>2 years</td>
<td>2-week intensive orientation, 11-month MPH degree and leadership coursework, weekly U2-day didactic sessions, monthly journal club, monthly on-orientation sessions, 2-year cardiology leadership academy</td>
<td>MPH degree, leadership coursework, change project, mentoring</td>
<td>Leadership of small systems in health care 3a. Self-reported improved communication, interaction with referring physicians, career development team building, group dynamics, and interactions with hospital administration. 3b. Participants reported increased confidence across a range of key curriculum areas. 3b. Participants reported increased confidence across a range of key curriculum areas. 4b. One project resulted in a reduction of admission time of 60 minutes.</td>
<td></td>
</tr>
<tr>
<td>Freerme, 2018 (57)</td>
<td>Single US training programme</td>
<td>30</td>
<td>Cardiology fellows-in-training and early career professionals</td>
<td>2 years</td>
<td>2 years cardiology leadership academy</td>
<td></td>
<td>Leadership of change for improvement of quality, value, and safety of health care of individuals and populations</td>
<td>2b. Participants reported increased confidence across a range of key curriculum areas. 3a. Participants reported the programme had helped them to gain new leadership roles, and that they had applied learning from the programme in their new leadership roles. 3b. Participants reported increased confidence across a range of key curriculum areas. 4a. Of those that acquired a new leadership opportunity, 55% of respondents reported participating in the leadership academy program had an impact in their success within the newly acquired leadership role(s). 4b. No change in leadership skills or other measures of interest. 5a. Change in self-assessed leadership skills or other measures of interest (mean scores from 5.1 to 5.5). Average score on knowledge test significantly increased from 76% to 95%. 5b. Three projects undertaken as part of voluntary elective. 4b. One project resulted in a reduction of admission time of 60 minutes.</td>
<td></td>
<td>r/s</td>
</tr>
<tr>
<td>Saums, 2017 (58)</td>
<td>Single German university hospital</td>
<td>50</td>
<td>Residents (PGY4-5 across specialties)</td>
<td>4 weeks</td>
<td>Weekly 2.5hr sessions after clinical duties</td>
<td>Didactic module, standardised simulations, one-on-one feedback on recorded simulations, practicing communication techniques and negotiating</td>
<td>Full Range Leadership Model (Bass); transactional and transformational leadership, simulation of critical incidents, communication techniques</td>
<td>1. Participants rated the programme had helped them to gain new leadership roles, and that they had applied learning from the programme in their new leadership roles. 4b. One project resulted in a reduction of admission time of 60 minutes.</td>
<td></td>
<td>r/s</td>
</tr>
<tr>
<td>Schud, 2013 (59)</td>
<td>US surgical department specialty</td>
<td>9</td>
<td>Residents (DVT)</td>
<td>6 months</td>
<td>Virtue strength assessments (VSA), mentorship meetings, thoughts of the day, internal and external faculty training and development, leadership basic training course</td>
<td></td>
<td>Facilitated discussions, project work</td>
<td>Feedback on recorded simulations, &quot;practicing communication techniques,&quot; group skills training. &quot;thought of the day,&quot; internal and external faculty training and development. Leadership basic training course. Eight sessions offered, Eight sessions offered.</td>
<td>2. No change in knowledge or other measures of interest. 3. Participants reported increased confidence across a range of key curriculum areas. 4b. One project resulted in a reduction of admission time of 60 minutes.</td>
<td></td>
</tr>
<tr>
<td>Steiner, 2007 (60)</td>
<td>Single US hospital network</td>
<td>Unclear</td>
<td>Emerging physician leaders</td>
<td>9 months</td>
<td></td>
<td></td>
<td>Marketing is healthcare, Healthcare finance, Writing a business plan, Emotional intelligence, situational leadership, conflict resolution and negotiation, Mediation issues.</td>
<td>3b. 40% business plans were submitted over 12 courses. 3b. 40% of business plans have been implemented</td>
<td></td>
<td>r/s</td>
</tr>
</tbody>
</table>
**Main findings by Kirkpatrick level**

1. **1-4 years post:**
   - 2a. All trainees who completed the survey reported that development of leadership skills is demonstrated.
   - 2b. Respondents reported increased ability to recognize weaknesses and strengths, and a range of other increased leadership skills.
   - 3a. Reports of improved confidence in addressing and resolving conflict situations. Insight and feedback into “why the game is played.”
   - 2b. Increased perceived ability to locate areas of need. Increased familiarity with workshop content.
   - 2b. Self-reported increased ability to resolve conflict effectively.
   - 3b. 66% of respondents were invited to apply for or offered positions post-programme.
   - 4a. 76% reported that the program lowered the level of stress. Participants reported enhanced job satisfaction, reduced job stress, and an expanded network of educational contacts and resources.

2. **2-3hrs a week for 4 months over a 1 year period:**
   - 2a. Some respondents reported increased insight into others/self/job/leadership.
   - 3b. Observed application of leadership skills in programme improvement.
   - 2b. Increased confidence, feeling more prepared for team challenges.

3. **>100 days:**
   - 2b. For each leadership skill surveyed, the majority of participants agreed that it was improved immediately post.
   - 2b. There was no significant change in the self-rated confidence in leadership skills or team management.
   - 2b. Respondents reported increased ability to recognize weaknesses and strengths, and a range of other increased leadership skills.
   - 3b. Respondents reported increased ability to resolve conflict effectively.

4. **20-40 weeks:**
   - 83% of respondents reported having undertaken a voluntary leadership activity, and 75% reported increased awareness of educational and medical issues and development of strategies to be informed, resolve problems, and advance projects.
   - 3a. Increased confidence, feeling more prepared for team challenges.
   - 4a. 76% reported that the program lowered the level of stress. Participants reported enhanced job satisfaction, reduced job stress, and an expanded network of educational contacts and resources.

5. **1 year:**
   - 2b. Self-reported learning, better understanding of personal strengths and weaknesses as a leader.
   - 4a. 76% reported that the program lowered the level of stress. Participants reported enhanced job satisfaction, reduced job stress, and an expanded network of educational contacts and resources.

6. **>100 days:**
   - 2b. Increased confidence, feeling more prepared for team challenges.
   - 3b. Observed application of leadership skills in programme improvement.

7. **20-40 weeks:**
   - 2b. Increased confidence, feeling more prepared for team challenges.
   - 3b. Observed application of leadership skills in programme improvement.

8. **>100 days:**
   - 2b. Increased confidence, feeling more prepared for team challenges.
   - 3b. Observed application of leadership skills in programme improvement.

**JBI Score**

8.5 8.5 8.5

**BMJ Leader**

<table>
<thead>
<tr>
<th>Source First Author, Year</th>
<th>Setting</th>
<th>Learner Number</th>
<th>Learner Type</th>
<th>Intervention Length</th>
<th>Intervention Description</th>
<th>Teaching Methods</th>
<th>Educational Context</th>
<th>Main Finding(s)</th>
<th>Methodology</th>
<th>JBI Score</th>
<th>BMJ Leader Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pickard, 2006 [48]</td>
<td>Single UK hospital</td>
<td>19</td>
<td>Physicians (new cohort)</td>
<td>2 years</td>
<td>54 monthly training sessions, 1 half day session each 6 months</td>
<td>Adult learning, interactive question, case-based scenario</td>
<td>Decision making, problem solving, strategy, change, situational leadership, communication, negotiation, coaching and mentoring, conflict, finances</td>
<td>1. Mean score 4.7/4 for satisfaction, 4.8 for quality of instruction 2. Participants reported increased awareness of leadership resources (1), motivation to be included in their community (3.3) 3. Participants self-reported increased leadership effectiveness (4.2), teamwork (6.0) and team leading (6.5), as well as new roles (6.0). Participants required using skills learned; 4. Improved organizational collaboration towards strategic initiatives 2. Participants reported increased self-awareness 5. Demonstrated increased knowledge of authentic leadership concepts</td>
<td>5.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>Shah, 2011 [72]</td>
<td>Single UK specialty training</td>
<td>40</td>
<td>Consultant orthopaedic surgeons</td>
<td>2 days</td>
<td>2x 1/2 day interactive sessions</td>
<td>Grouping, team challenges, meta-analysis, role play, and professional roles, interactive presentations, and self-assessment</td>
<td>Admitting vulnerability and uncertainty, taking responsibility for managing risk, doing self and reflective, internalizing authentic leadership</td>
<td>1. Improved self-confidence in leadership, intended changes to leadership style 2. Increased knowledge of desirable leadership characteristics in the organization 3. Several of the participants reported experimenting in their current leadership assignments with concepts discussed during a session, one participant used one of the cases to heighten the leadership awareness of some of his own subordinates. Reported personal benefits of the retreat</td>
<td>2.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>Clapp, 2018 [72]</td>
<td>Single UK hospital department</td>
<td>36</td>
<td>Residents and Faculty (Department of Anesthesiology and Critical Care)</td>
<td>1 year</td>
<td>2x 2 hr evening half-day sessions, 1 half-day session and 2 1/2 hr evening sessions explicated</td>
<td>Case presentations on relevant topic small group discussions presentations back to large group</td>
<td>Feedback supporting colleagues during tough times clinical problem solving</td>
<td>1. Leadership style, conflict management, effective feedback, team building, team leadership, motivation, moving from peer to leader. 2. Improved self-confidence in leadership.</td>
<td>1.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>Pettit, 2011 [74]</td>
<td>Single UK hospital department</td>
<td>119 (response rate)</td>
<td>Residents</td>
<td>1 year (academic)</td>
<td>Monthly 30 sessions</td>
<td>Interactive lecture, self-assessment articles, case studies, self-reflection, discussions, and reading materials</td>
<td>Leadership style, conflict management, effective feedback, team building, team leadership, motivation, moving from peer to leader.</td>
<td>1. &quot;universal primer&quot; from participants for the mandatory first 2 years, 100% opted into the optional years 3-4. 2. If participants have worked on projects as part of the programme, with one having submitted academic manuscript resulting from her project</td>
<td>0.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>Dawson, 2016 [75]</td>
<td>Single US residency programme</td>
<td>Unwritten</td>
<td>Residents (surgery, PGY 1-4)</td>
<td>4 years</td>
<td>4-year residency programme in leadership fundamentals and leadership tracks. Monthly meetings for 2 years, research work with imaging scientists, mentoring, project work</td>
<td>1. &quot;universal primer&quot; from participants for the mandatory first 2 years, 100% opted into the optional years 3-4. 2. If participants have worked on projects as part of the programme, with one having submitted academic manuscript resulting from her project</td>
<td>0.0</td>
<td>8.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gray, 2006 [76]</td>
<td>Single US health system</td>
<td>57</td>
<td>Physicians and non-physicians (finance, IT, nursing, public relations)</td>
<td>8 months</td>
<td>11 residency weeks over dinner (2-3 hrs)</td>
<td>Case-based leadership discussions during two four-hour sessions</td>
<td>Managing vs. leading, forming vision, providing a person's moral compass, risk taking and transactional leadership</td>
<td>1. High ratings for the course 4.4/5 for comparison to other leadership and education programmes experienced. 2. Increased self-confidence in leadership, intended changes to leadership style 3. Increased knowledge of desirable leadership characteristics in the organization 4. Several of the participants reported experimenting in their current leadership assignments with concepts discussed during a session, one participant used one of the cases to heighten the leadership awareness of some of his own subordinates. Reported personal benefits of the retreat</td>
<td>0.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>Rabab, 2006 [77]</td>
<td>Single US residency program (pandemic)</td>
<td>18</td>
<td>Residents (Paediatrics PGY-2, 3)</td>
<td>9 months</td>
<td>Nine sessions</td>
<td>Lectures</td>
<td>A core curriculum focusing on physician compensation, medical economics, healthcare system, leadership and communication, conflict (re), contracts, health law, and customer service</td>
<td>1. Satisfaction scores between 3 and 4 on a four-point scale. 2. Improvement in general comprehension of the subject matter of each lecture, with an average increase of 30% to 40% between tests (5-point scale)</td>
<td>0/2</td>
<td>8/8</td>
<td></td>
</tr>
<tr>
<td>Gade K, 2018 [78]</td>
<td>Indian health care organisation</td>
<td>96</td>
<td>Clinicians and hospital administrators from public and private sector organisations</td>
<td>3 days</td>
<td>2-day leadership residential programme as part of 4-day residential programme</td>
<td>Didactic lecture, small group workshops, focus group discussions, case-based discussions and experience sharing</td>
<td>Leadership style, leadership competencies, team building, strategic management, procurement, operational excellence, legal and ethical issues, budgeting, financial management, conflict management, quality and patient safety, hospital accreditation, communication, human resource development, health care technology, contract management, hospital projects and supply chain management</td>
<td>1. &quot;universal primer&quot; from participants for the mandatory first 2 years, 100% opted into the optional years 3-4. 2. If participants have worked on projects as part of the programme, with one having submitted academic manuscript resulting from her project</td>
<td>0.0</td>
<td>8.0</td>
<td></td>
</tr>
<tr>
<td>Stalker, 2004 [79]</td>
<td>Single US residency program (internal medicine)</td>
<td>32</td>
<td>Residents (PGY-1, 2, 3)</td>
<td>1 day</td>
<td>One-day retreat</td>
<td>Group simulation exercise, group discussions</td>
<td>1. All attendees rated the retreat as valuable. 2. Based on significant changes in residents' responses on the postretreat questionnaire, attended believed that the retreat enhanced their abilities to be better physicians, resident supervisors, and leaders. (p&lt;0.001) 3. Postretreat responses (table 1) indicated significant increase in agreement that good leaders challenge the process, make decisions based on shared vision, allow others to act, recognize individual contributions, and serve as good role models. 4. Participants described with the programme, partly by high expectations not being met. 5. Follow-up attendance 6. Growing sense of identity as clinical leaders, increased confidence</td>
<td>0/2</td>
<td>8/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Edmondstone, 2009 [80]</td>
<td>Multiple UK strategic health authorities</td>
<td>200 (approx.)</td>
<td>Senior medical leaders in primary and secondary care and public health</td>
<td>12 months</td>
<td>Two day residential sessions at the beginning and end of the programme, four 3-hour evening days, three intermittent weekend days</td>
<td>Personal development plan, coaching, mentoring</td>
<td>Leadership for partnership 1. Personal development – through the creation of a personal development plan, provision of coaching, and development of an individualized programme for participants. 2. Workshops: seminars, case-based teaching, role play, and scenario planning</td>
<td>1. All attendees rated the retreat as valuable. 2. Based on significant changes in residents' responses on the postretreat questionnaire, attended believed that the retreat enhanced their abilities to be better physicians, resident supervisors, and leaders. (p&lt;0.001) 3. Postretreat responses (table 1) indicated significant increase in agreement that good leaders challenge the process, make decisions based on shared vision, allow others to act, recognize individual contributions, and serve as good role models. 4. Participants described with the programme, partly by high expectations not being met. 5. Follow-up attendance 6. Growing sense of identity as clinical leaders, increased confidence</td>
<td>5.0</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>Skergjæsloen, 2008 [81]</td>
<td>Single Canadian residency program</td>
<td>52</td>
<td>Residents (PGY-2, 3, 4)</td>
<td>4 half days</td>
<td>Workshop (four half-days)</td>
<td>Interactive teaching as much as possible. Didactic teaching and small group work, interactive techniques: lunch groups, brain-storming, think-pair-share discussion, a debate, and clinical case study</td>
<td>Teamwork, conflict resolution, quality improvement, program planning and evaluation, leadership and change management, system health reform, organizational structure, and self and career development</td>
<td>1. Attendance averaged 58% overall. Workshops averaged around 62.5%. Participants appreciated the reflective and interactive components of the workshops and valued the hands-on exercises and the use of case studies and &quot;real life&quot; examples. They suggested that more time be dedicated to quality improvement and medical error and opportunities to take part in administrative committee and quality improvement projects at their hospital site. Focus requested on current efforts rather than historical overview. 2. Objectives respected skills and knowledge rated the most (7.5 vs. 0.0 as issued)</td>
<td>2.0</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>Berkertosh, 2014 [82]</td>
<td>Single UK university medical centre</td>
<td>14</td>
<td>Residents (OGU, paediatrics, internal medicine)</td>
<td>8 hours</td>
<td>2 x 4-hour sessions 3 weeks apart homework between</td>
<td>Didactic teaching/lectures student presentations simulation</td>
<td>Knowledge of the healthcare system time management</td>
<td>1. Rated 7.4/8.9 (7.5% in our current postsgraduate medical training) 2. Increased interest in leadership development 3. No significant changes (underpowered)</td>
<td>1.0</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>Black, 2017 [82]</td>
<td>11 residency</td>
<td>146</td>
<td>Residents</td>
<td>2 days</td>
<td>2 day programme</td>
<td>Experimental small-</td>
<td>Leadership competencies, self awareness,</td>
<td>1. High satisfaction scores of 6.2 on a scale of 1-7 (6.6)</td>
<td>0.0</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>Source (First Author, Year)</td>
<td>Setting</td>
<td>Learner Number</td>
<td>Learner Type</td>
<td>Intervention Length</td>
<td>Intervention Description</td>
<td>Teaching Methods</td>
<td>Educational Context</td>
<td>Main Findings by Kirkpatrick Level</td>
<td>AB Score</td>
<td>NRSQD Score</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td>----------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>-------------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>-----------------------------------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>Supplemental material</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Awad, 2004 (30)**
- Single residency program (Surgical)
- NR (Residents (Surgical))
- 6 months
- NR (Not specified)

- **Leadership Competency Framework**: Demonstrating personal qualities, working with others, managing services, improving services, setting direction.
- A "Focused program" to train residents to have the balance: work, relationships, and personal health.

**Main Findings by Kirkpatrick Level**:
- 2a. Participants felt the programme helped them in achieving leadership competencies.
- 2b. Respondents reported increased leadership skills, knowledge, and self-awareness.
- 1. Participants believed the course should be offered again and had a positive effect on their professional lives.
- 2b. Respondents reported increased leadership skills, knowledge, and self-awareness.
- 1. 100% of respondents would recommend the programme to peers.

**Score**
- AB Score: 7.5
- NRSQD Score: 7.5

### Bajpai, 2008 [96]
**Source (First Author, Year)**: Single US residency program (family medicine)
**Setting**: NR
**Learner Numbers**: Residents (Family Medicine PGY2-3)
**Intervention Description**: 2 years
**Intervention Duration**: Nine half-day sessions (first year), monthly one-hour sessions (second year)
**Teaching Methods**: Interactive lectures and group assignments
**Educational Context**: Learns through didactic, collaborative, and interactive A education management curriculum. Determining balancing personal and professional goals, practice opportunities, facilities, organization, operation and management, staffing policies, legal issues, marketing, resources, and hospital issues.
**Main Findings by Kirkpatrick level**: 1. Reported that the course was beneficial, a positive experience. 2a. Reported increased interest (and knowledge) in practice management 2b. Self-reported knowledge/skill fit with each of 15 practice management topics increased by roughly 2 points (five point Likert scale)

### Harro, 2012 [97]
**Source (First Author, Year)**: Single-Canadian hospital
**Setting**: NR
**Learner Numbers**: Residents (PGY1 & Surgery)
**Intervention Description**: 1 day
**Intervention Duration**: One-day conference
**Teaching Methods**: Interactive lectures and case-based discussions, role-playing, scenario simulation, small-group problem-solving, series of lectures, and live feedback sessions
**Educational Context**: Giving feedback and delegating duties, building confidence, building hands-on skills, coping with stress, effective learning while on duty, taking feedback and implementing it, and managing conflicts.
**Main Findings by Kirkpatrick level**: 1. 70% felt that management was well-addressed or very well-addressed, of 13 before the course. 2a. Statistically significant improvement on perceived preparedness for all 4 managerial duties (negotiating employees, managing personal finances, handling conflict resolution, managing a practice office). 2b. Participants reported improved skills in giving feedback, delegating, role-playing,呦, coping with stress, effective learning, and effective teaching. These were not significant with confidence corrected applied.

### Patel, 2019 [98]
**Source (First Author, Year)**: Single US hospital residency
**Setting**: NR
**Learner Numbers**: Residents (PGY-1)
**Intervention Description**: 2 years
**Intervention Duration**: Quarterly mentoring sessions for 2 years as part of a healthcare leadership program
**Teaching Methods**: Weekly didactic sessions
**Educational Context**: Monthly one-hour sessions
**Main Findings by Kirkpatrick level**: 1. Participants noted their likelihood of recommending the program at 7.8/10 (10 being extremely likely) 2a. Confidence seemed to increase (limited data reported). 2b. Gained confidence in time management, negotiation, and leadership skills

### Saffie, 2010 [99]
**Source (First Author, Year)**: Single US hospital
**Setting**: NR
**Learner Numbers**: Residents (Senior)
**Intervention Description**: 4 weeks
**Intervention Duration**: Monthly 1hr didactic sessions
**Teaching Methods**: Didactic teaching, team projects
**Educational Context**: Monthly sessions of advanced cardiovascular leadership, giving feedback and delegating duties, building hands-on skills, coping with stress, effective learning while on duty, taking feedback and implementing it, and managing conflicts.
**Main Findings by Kirkpatrick level**: 1. "more than half of the participants considered the program to be beneficial" 2. Overall self-confidence score improved from 2.8 to 3.8

### Gaglani, 2010 [100]
**Source (First Author, Year)**: Single US hospital
**Setting**: NR
**Learner Numbers**: Physicians with leadership responsibilities
**Intervention Description**: 2 years
**Intervention Duration**: Monthly sessions of 2-4 hours
**Teaching Methods**: Lectures and case-based discussion
**Educational Context**: Organizational leadership, financial management, management strategy, applied skills and tools
**Main Findings by Kirkpatrick level**: 1. Participants reported high satisfaction with the course, and most would recommend to colleagues. 2a. Participants reported feeling more interested in and prepared for leadership responsibilities. 2b. 70% of participants reported having changed their approaches to projects or problems as a result of the course.

### Gurney, 2014 [101]
**Source (First Author, Year)**: Single US hospital
**Setting**: NR
**Learner Numbers**: Residents
**Intervention Description**: 5 months
**Intervention Duration**: Didactic teaching, team projects
**Teaching Methods**: (1) Overview of course, exploration of business plan project and expectations; resident interest survey; (2) leadership and work environment model; (3) coping with stress, effective learning, and effective teaching. These were not significant with confidence corrected applied.

### Whata, 2015 [102]
**Source (First Author, Year)**: Single US institution
**Setting**: NR
**Learner Numbers**: Residents (Internal Medicine, Surgery, Emergency, Psychiatry)
**Intervention Description**: 1 week
**Intervention Duration**: 1 week programme
**Teaching Methods**: Case-based learning interactive tasks, small group sessions simulation, Evaluation and feedback
**Educational Context**: Achieved high satisfaction rates and 96%.

### Hadley, 2015 [103]
**Source (First Author, Year)**: Single UK residency
**Setting**: NR
**Learner Numbers**: Residents (FY2 doctor)
**Intervention Description**: Single brief intervention
**Intervention Duration**: Leadership assessment and feedback
**Teaching Methods**: Evaluation and feedback
**Educational Context**: 60% of participants felt that their leadership skills had improved as a result of the feedback received.

### Kaya, 2020 [104]
**Source (First Author, Year)**: Single US residency program (internal medicine)
**Setting**: NR
**Learner Numbers**: Residents (PGY1, PGY2)
**Intervention Description**: 1 day
**Intervention Duration**: 5x one-hour sessions
**Teaching Methods**: Lectures and small group tasks and discussions, scenarios and role play
**Educational Context**: Setting personal visions, leadership, management, building teams, practical application of leadership, problem-solving as a team leader

### Nissen, 2018 [105]
**Source (First Author, Year)**: Single US residency program
**Setting**: NR
**Learner Numbers**: Residents (Anesthesiology)
**Intervention Description**: 1 year
**Intervention Duration**: Monthly 2hr didactic sessions
**Teaching Methods**: Didactic teaching, assignment
**Educational Context**: Personal branding, curriculum vitae, marketing, networking, evaluating and evaluating different types of medical practice, medical staff structure, governance, healthcare reforms, future trends in medicine

### Schwartz, 2019 [106]
**Source (First Author, Year)**: International (US and Canada) Psychiatry Leadership conference
**Setting**: NR
**Learner Numbers**: Residents (All US and Canadian residency programs)
**Intervention Description**: 3 days
**Intervention Duration**: 3-day immersion course
**Teaching Methods**: Large and small group sessions, group tasks, peer and teacher feedback
**Educational Context**: Psychological challenges in leadership situations, personal conflicts, self-reflection and self-awareness, group process, conflict resolution, navigation of challenging leadership roles

### Enola Cole, 2018 [107]
**Source (First Author, Year)**: Single US hospital
**Setting**: NR
**Learner Numbers**: Physicians NOS
**Intervention Description**: 6 months
**Intervention Duration**: 2hrs once a fortnight
**Teaching Methods**: Multi-source feedbacks
**Educational Context**: BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance on the information contained in this document and the use of any information, or of the use of materials referred to in this document.

---

<table>
<thead>
<tr>
<th>Source (First Author, Year)</th>
<th>Setting</th>
<th>Learner Number</th>
<th>Learner Type</th>
<th>Intervention Length</th>
<th>Intervention Description</th>
<th>Teaching Methods</th>
<th>Educational Content</th>
<th>Main Findings by Kirkpatrick level</th>
<th>JB Score</th>
<th>MERSQI Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>O’Donnell, 2011 [116]</td>
<td>Single US hospital</td>
<td>NR</td>
<td>Residents (Geriatric Medicine)</td>
<td>2 to 3 months</td>
<td>8 ½ hour workshop delivered by in-house core leadership to faculty members</td>
<td>Didactic teaching</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased improvement. 2b. Participants reported improved communication skills. 3a. Participants reported increased self-efficacy.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Johnson, 2014 [115]</td>
<td>Single US hospital</td>
<td>NR</td>
<td>Residents (Family Medicine)</td>
<td>2 to 4 months</td>
<td>2 ½ day workshop on diversity and multiculturalism delivered by in-house core leadership</td>
<td>Didactic teaching, small group discussions, case presentations, role play</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased awareness and understanding of cultural competence. 3a. Participants reported improved cultural sensitivity.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Cherry, 2010 [114]</td>
<td>Single US hospital</td>
<td>NR</td>
<td>Residents (Emergency Medicine)</td>
<td>2 to 3 months</td>
<td>3 hour session on leadership and management</td>
<td>Didactic teaching, small group discussions, case presentations, role play</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased awareness and understanding of leadership roles. 3a. Participants reported improved leadership skills.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Bircher, 2013 [111]</td>
<td>Single UK deanery (extension of GP training)</td>
<td>NR</td>
<td>Residents (GP trainees)</td>
<td>2 years</td>
<td>Unspecified number of programme days</td>
<td>Didactic teaching, online learning modules</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased confidence in their leadership roles. 3a. Participants reported improved leadership skills.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Kocher, 2008 [112]</td>
<td>Single US academic medical centre</td>
<td>30</td>
<td>Faculty members</td>
<td>5 months</td>
<td>32 hour course in three modules over five months</td>
<td>Didactic teaching, online learning modules</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased awareness and understanding of leadership roles. 3a. Participants reported improved leadership skills.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Rhee, 2011 [113]</td>
<td>Single US hospital department</td>
<td>NR</td>
<td>Residents (Senior Residents, Emergency Medicine)</td>
<td>1 year</td>
<td>2 hour workshop on leadership and management</td>
<td>Didactic teaching, small group discussions, case presentations, role play</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased awareness and understanding of leadership roles. 3a. Participants reported improved leadership skills.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Cherry, 2010 [114]</td>
<td>Single US university</td>
<td>NR</td>
<td>Junior faculty</td>
<td>9 months</td>
<td>2 hour workshop on leadership and management</td>
<td>Didactic teaching, small group discussions, case presentations, role play</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased awareness and understanding of leadership roles. 3a. Participants reported improved leadership skills.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Johnson, 2014 [112]</td>
<td>Single US hospital department</td>
<td>NR</td>
<td>Residents (Junior Medical Residents)</td>
<td>1/2 day</td>
<td>2 hour workshop on leadership and management</td>
<td>Didactic teaching, small group discussions, case presentations, role play</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased awareness and understanding of leadership roles. 3a. Participants reported improved leadership skills.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>O’Donnell, 2011 [114]</td>
<td>Single US hospital (residency programs)</td>
<td>NR</td>
<td>Residents (PGY1-3)</td>
<td>4 weeks</td>
<td>36 hour programme for leadership and management</td>
<td>Didactic teaching, small group discussions, case presentations</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased awareness and understanding of leadership roles. 3a. Participants reported improved leadership skills.</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Riggsby, 2010 [117]</td>
<td>Single US residency programme</td>
<td>NR</td>
<td>Residents (Family Medicine, PGY1-2)</td>
<td>2 to 4 weeks</td>
<td>5 day workshop on leadership and management</td>
<td>Didactic teaching, small group discussions, case presentations, role play</td>
<td>Principles of leadership, communication, collaborative problem-solving, and team building</td>
<td>2a. Participants reported increased awareness and understanding of leadership roles. 3a. Participants reported improved leadership skills.</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

**Supplementary Table 1: Summary of included studies. NR= Not Reported; NOS=Not Otherwise Specified; n/a=not applicable. MERSQI=Medical Education Research Study Quality Instrument; JB=Joanna Briggs Institute Critical Appraisal Checklist for Qualitative Research.**
References:


Kuo AK, Thyne SM, Chen HC, *et al.* An innovative residency program designed to develop leaders to improve the health of children. *Acad Med* 2010;85:653. doi:10.1097/ACM.0b013e3181b60f6


58 Saravo B, Netzel J, Kiesewetter J. The need for strong clinical leaders - Transformational and transactional leadership as a framework for resident leadership training. *PloS One* 2017;12:e0183019. doi:https://dx.doi.org/10.1371/journal.pone.0183019


105 Ninan D, Patel D. Career and Leadership Education in Anesthesia Residency Training. Cureus 2018;10:e2546. doi:https://dx.doi.org/10.7759/cureus.2546


114 Cherry RA, Davis DC, Thordnyke L. Transforming culture through physician leadership development. Physician Exec 2010;36:38–44.


### Supplementary Table 2: Study characteristics organised by MERSQI heading. Brackets in headings refer to original MERQSI items where headings have been adapted for clarity.

<table>
<thead>
<tr>
<th>MERSQI Component</th>
<th>Classification</th>
<th>All studies (117)</th>
<th>MERSQI&gt;12 (16)</th>
<th>JBI&gt;6 (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Study Design</strong></td>
<td>Single Group Cross-Sectional or Post-programme only</td>
<td>54 (46%)</td>
<td>0 (0%)</td>
<td>8 (57%)</td>
</tr>
<tr>
<td></td>
<td>Single Group Pre and Post Programme</td>
<td>54 (46%)</td>
<td>9 (56%)</td>
<td>5 (36%)</td>
</tr>
<tr>
<td></td>
<td>Non-Randomised Two Group</td>
<td>8 (7%)</td>
<td>6 (38%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Randomised Controlled Trial</td>
<td>1 (1%)</td>
<td>1 (6%)</td>
<td>1 (7%)</td>
</tr>
<tr>
<td><strong>Institution #</strong></td>
<td>Single</td>
<td>81 (69%)</td>
<td>10 (63%)</td>
<td>9 (64%)</td>
</tr>
<tr>
<td></td>
<td>Double</td>
<td>1 (1%)</td>
<td>1 (6%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td></td>
<td>Multi</td>
<td>35 (30%)</td>
<td>5 (31%)</td>
<td>5 (36%)</td>
</tr>
<tr>
<td><strong>Response Rate</strong></td>
<td>&lt;50% or NR</td>
<td>44 (38%)</td>
<td>3 (19%)</td>
<td>1 (7%)</td>
</tr>
<tr>
<td></td>
<td>50-75%</td>
<td>23 (20%)</td>
<td>3 (19%)</td>
<td>6 (43%)</td>
</tr>
<tr>
<td></td>
<td>&gt;75%</td>
<td>48 (41%)</td>
<td>9 (56%)</td>
<td>7 (50%)</td>
</tr>
<tr>
<td><strong>Type of Data</strong></td>
<td>Self-reported</td>
<td>70 (60%)</td>
<td>2 (13%)</td>
<td>9 (64%)</td>
</tr>
<tr>
<td></td>
<td>Observed</td>
<td>47 (40%)</td>
<td>14 (88%)</td>
<td>5 (36%)</td>
</tr>
<tr>
<td><strong>Questionnaire Construct Validity</strong></td>
<td>Reported</td>
<td>9 (8%)</td>
<td>7 (44%)</td>
<td>2 (14%)</td>
</tr>
<tr>
<td></td>
<td>Not Reported</td>
<td>108 (92%)</td>
<td>9 (56%)</td>
<td>12 (86%)</td>
</tr>
<tr>
<td><strong>Questionnaire Content Validity</strong></td>
<td>Reported</td>
<td>45 (38%)</td>
<td>14 (88%)</td>
<td>5 (36%)</td>
</tr>
<tr>
<td></td>
<td>Not Reported</td>
<td>72 (62%)</td>
<td>2 (13%)</td>
<td>9 (64%)</td>
</tr>
<tr>
<td><strong>Relationships to Other Variables</strong></td>
<td>Reported</td>
<td>8 (7%)</td>
<td>5 (31%)</td>
<td>3 (21%)</td>
</tr>
<tr>
<td></td>
<td>Not Reported</td>
<td>108 (92%)</td>
<td>11 (69%)</td>
<td>11 (79%)</td>
</tr>
<tr>
<td><strong>Data Analysis Comprehensiveness</strong></td>
<td>Comprehensive</td>
<td>23 (20%)</td>
<td>14 (88%)</td>
<td>6 (43%)</td>
</tr>
<tr>
<td><strong>Appropriateness</strong></td>
<td>Less Comprehensive</td>
<td>94 (80%)</td>
<td>2 (13%)</td>
<td>8 (57%)</td>
</tr>
<tr>
<td></td>
<td>Descriptive only</td>
<td>102 (87%)</td>
<td>7 (44%)</td>
<td>11 (79%)</td>
</tr>
<tr>
<td><strong>Data Analysis Complexity</strong></td>
<td>Beyond Descriptive</td>
<td>15 (13%)</td>
<td>9 (56%)</td>
<td>3 (21%)</td>
</tr>
<tr>
<td><strong>Outcomes (Kirkpatrick Level)</strong></td>
<td>Level 1</td>
<td>80 (68%)</td>
<td>8 (50%)</td>
<td>14 (100%)</td>
</tr>
<tr>
<td></td>
<td>Level 2a</td>
<td>70 (60%)</td>
<td>7 (44%)</td>
<td>13 (93%)</td>
</tr>
<tr>
<td></td>
<td>Level 2b</td>
<td>79 (68%)</td>
<td>11 (69%)</td>
<td>11 (79%)</td>
</tr>
<tr>
<td></td>
<td>Level 3a</td>
<td>51 (44%)</td>
<td>7 (44%)</td>
<td>10 (71%)</td>
</tr>
<tr>
<td></td>
<td>Level 3b</td>
<td>54 (46%)</td>
<td>14 (88%)</td>
<td>7 (50%)</td>
</tr>
<tr>
<td></td>
<td>Level 4a</td>
<td>9 (8%)</td>
<td>1 (6%)</td>
<td>2 (14%)</td>
</tr>
<tr>
<td></td>
<td>Level 4b</td>
<td>26 (22%)</td>
<td>7 (44%)</td>
<td>4 (29%)</td>
</tr>
<tr>
<td>JBI Component</td>
<td>Description of component</td>
<td>High-reliability studies (n=14)</td>
<td>Mixed-Methods Studies (n=53)</td>
<td>Qualitative Studies (n=10)</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Philosophical Perspective</td>
<td>Congruity between the stated philosophical perspective and the research methodology</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Objectives</td>
<td>Congruity between the research methodology and the research question or objectives</td>
<td>39 (63%)</td>
<td>33 (62%)</td>
<td>6 (60%)</td>
</tr>
<tr>
<td>Methods</td>
<td>Congruity between the research methodology and the methods used to collect data</td>
<td>38 (61%)</td>
<td>32 (60%)</td>
<td>6 (60%)</td>
</tr>
<tr>
<td>Analysis</td>
<td>Congruity between the research methodology and the representation and analysis of data</td>
<td>18 (29%)</td>
<td>15 (28%)</td>
<td>3 (30%)</td>
</tr>
<tr>
<td>Interpretation</td>
<td>Congruity between the research methodology and the interpretation of results</td>
<td>17 (27%)</td>
<td>13 (25%)</td>
<td>4 (40%)</td>
</tr>
<tr>
<td>Researcher Location</td>
<td>Statement locating the researcher culturally or theoretically</td>
<td>10 (16%)</td>
<td>8 (15%)</td>
<td>2 (20%)</td>
</tr>
<tr>
<td>Researcher Influence</td>
<td>Influence of the researcher on the research addressed</td>
<td>13 (21%)</td>
<td>12 (23%)</td>
<td>1 (10%)</td>
</tr>
<tr>
<td>Participant Representation</td>
<td>Participants and their voices adequately represented</td>
<td>21 (34%)</td>
<td>16 (30%)</td>
<td>5 (50%)</td>
</tr>
<tr>
<td>Ethics</td>
<td>Evidence of ethical approval by an appropriate body</td>
<td>26 (42%)</td>
<td>23 (43%)</td>
<td>3 (30%)</td>
</tr>
<tr>
<td>Conclusions Supported</td>
<td>Conclusions drawn in the research report flow from the analysis or interpretation of the data</td>
<td>20 (32%)</td>
<td>16 (30%)</td>
<td>4 (40%)</td>
</tr>
</tbody>
</table>

Supplementary Table 3: Proportion of studies which met Joanna Briggs Institute (JBI) Critical Appraisal Items for Qualitative Studies. Descriptions are adapted from the JBI tool. Higher Reliability Studies scored 6 or more on the JBI tool.
### Supplementary Figure 1: Medline (OVID) Search Strategy, January 2020

<table>
<thead>
<tr>
<th>BMJ Leader, et al. Lyons O</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Searches</strong></td>
<td><strong>Results</strong></td>
<td><strong>Type</strong></td>
<td><strong>Actions</strong></td>
<td><strong>Annotations</strong></td>
<td></td>
</tr>
<tr>
<td>1 PHYSICIAN/</td>
<td>87338</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>2 PHYSICIAN EXECUTIVE/</td>
<td>4211</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>3 CONSULTANTS/</td>
<td>6648</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>4 INTERNSHIP/</td>
<td>49084</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>5 RESIDENCY/</td>
<td>49084</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>6 MEDICAL STAFF/</td>
<td>2059</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>7 (physician* or surgeon* or doctor* or intern* or intern* or resident* or registrar* or consultant* or &quot;house officer&quot;* or &quot;medical staff&quot;)..</td>
<td>230256</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>8 1 or 2 or 3 or 4 or 5 or 6 or 7</td>
<td>304007</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>9 LEADERSHIP</td>
<td>40117</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>10 PRACTICE MANAGEMENT/</td>
<td>1344</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>11 &quot;leader&quot;* or &quot;leader&quot;..</td>
<td>21361</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>12 (practice and manage&quot;)..</td>
<td>9878</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>13 9 or 10 or 11 or 12</td>
<td>58933</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>14 PROGRAM EVALUATION/</td>
<td>61451</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>15 PROGRAM DEVELOPMENT/</td>
<td>28428</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>16 CURRICULUM/</td>
<td>73527</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>17 EDUCATION/MEDICAL CONTINUING/</td>
<td>24644</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>18 EDUCATION/MEDICAL GRADUATE/</td>
<td>28493</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>19 (teach* or train* or educat* or course* or program* or pathway* or curriculum&quot;)..</td>
<td>793306</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>20 14 or 15 or 16 or 17 or 18 or 19</td>
<td>873227</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>21 8 and 13 and 20</td>
<td>1056</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>22 limit 21 to y=2013 -Current*</td>
<td>461</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>23 GENERAL PRACTITIONER/</td>
<td>7485</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>24 FAMILY PHYSICIAN/</td>
<td>16222</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>25 MEDICAL STAFF/HOSPITAL/</td>
<td>22748</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>26 (GP or GPs or &quot;general practitioner&quot; or &quot;family practitioner&quot;)..</td>
<td>22053</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>27 23 or 24 or 25 or 26</td>
<td>852255</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>28 13 and 20 and 27</td>
<td>151</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>29 limit 26 to y=2013 -Current*</td>
<td>63</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
<tr>
<td>30 29 and 22</td>
<td>40</td>
<td>Advanced</td>
<td>Display Results</td>
<td>More</td>
<td></td>
</tr>
</tbody>
</table>