Abstract

Background

Academic institutions benefit from researchers adopting leadership positions and, subsequently, leadership development programmes are of increasing importance. Despite this, no evaluation of the evidence basis for leadership development programmes for healthcare researchers has been conducted. In the present study, the authors reviewed leadership development programmes for healthcare researchers and aimed to identify their impact and the factors which influenced this impact.

Methods

The authors searched MEDLINE, EMBASE, CINAHL and PsycINFO between January 2000 and January 2023 for evaluations of leadership development programmes with healthcare researchers. The authors synthesised results through exploratory meta-analysis and meta-aggregation, and used the Medical Education Research Study Quality Instrument (MERSQI) and Joanna Briggs Institute (JBI) Checklist for Qualitative Studies to identify higher-reliability studies.

Results

Forty-eight studies met inclusion criteria, of which approximately half (22) met criteria for higher reliability. The median critical appraisal score was 10.5/18 for the MERSQI and 3.5/10 for the JBI. Common causes of low study quality appraisal related to study design, data analysis

and reporting. Evaluations principally consisted of questionnaires measuring self-assessed outcomes. Interventions were primarily focused on junior academics. Overall, 163/168 categorised programme outcomes were positive. Coaching, experiential learning/project work and mentoring were associated with increased organisational outcomes.

Conclusion

Educational methods appeared to be more important for organisational outcomes than specific educational content. To facilitate organisational outcomes, educational methods should include coaching, project work and mentoring. Programmes delivered by external faculty were less likely to be associated with organisational outcomes than those with internal or mixed faculty, but this needs further investigation. Finally, improving evaluation design will allow educators and evaluators to more effectively understand factors which are reliably associated with organisational outcomes of leadership development.

What is already known on this topic

Many leadership programmes have been developed for healthcare researchers. While there have been several publications related to individual programmes, no systematic review of these publications has been conducted to date.

What this study adds

This is the first systematic review of the literature evaluating leadership development programmes for healthcare researchers.

We have identified a lack of high-quality qualitative evaluations, with limited use of validated evaluation instruments and randomised control studies.

We have demonstrated meta-analysis can be applied to outcomes reported within the field, provided standardised instruments are used.

We found the literature to be skewed towards the reporting of positive programme outcomes.

This study reinforces that organisational outcomes are more aligned with educational methods, such as experiential learning, than any specific educational content.

How this study might affect research, practice or policy

Future research describing leadership programmes need to be more transparent in reporting key elements and negative findings. We recommend authors are guided by critical appraisal tools/reporting checklists.

Future research would benefit from a relevant critical appraisal tool which is designed to specifically evaluate mixed-methods research.

Future programmes seeking organisational outcomes should be led by internal or mixed faculty and include experiential learning components such as coaching, project work or mentoring.

Introduction

Healthcare research has steadily been organised into larger, multidisciplinary groups.[1] There are indications that university performance improves after established researchers are appointed to leadership roles[2] and that active biomedical research centres have a net positive impact on local patient care.[3] Leadership development programmes for healthcare researchers have become more popular in recent years.[4,5] While there is limited data on expenditure in healthcare research, the UK National Health Service spends an estimated quarter of its organisational development budget on leadership development, approximately £1 billion, every year.[6] A survey of Academic Health Centres in North America by Lucas and colleagues suggests this investment is mirrored or exceeded in the United States.[4]

There have been multiple systematic reviews of leadership development programmes for participants in related fields, including physicians,[7–13] medical students[14–16] and academic medical centres.[4,5] Their authors have particularly recommended the use of experiential learning methods,[8,11,13] (for more effective transfer of learning[11]) and the inclusion of faculty from both inside and outside and organisation rather than solely external faculty.[13] These recommendations have been limited by a general lack of high-quality evidence[4,7,10,12] and it is not clear how transferrable these conclusions are to a healthcare researcher population.

To the best of our knowledge there has not yet been a systematic review of leadership development programmes for healthcare researchers. Because of inherent differences between research and clinical practice environments, best practices for leadership development for

healthcare scientists may differ from best practices for leadership development for clinical physicians or medical educators.[5]

We conducted this systematic review to identify and synthesise evaluations of leadership development programmes for healthcare researchers. Our review was guided by the research question: "What is the impact of leadership development interventions for healthcare researchers, and what factors influence the impact of these interventions?"

Methods

This systematic review is reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement.[17] Our review protocol was prospectively made available on the Open Science Framework.[18] Any deviations from the protocol are noted below. We followed a modified "2-week systematic review" (2weekSR) process (see Appendix 1).[19]

Inclusion and exclusion criteria

We included studies which met the following criteria.

1. *Described a leadership development intervention*. We included all interventions described as leadership interventions regardless of the definition of leadership, alongside those which we judged to be describing leadership development.
2. *Included participants who were healthcare researchers*. We defined healthcare researchers as individuals conducting research related to human healthcare. Researchers

could be of any seniority. Programmes for clinicians met this criterion only if these clinicians had an explicit research role or if the programme also included healthcare researchers.

3. *Evaluated the leadership development intervention*. All forms of intervention evaluation were included, provided they investigated the impact of the leadership development intervention.

There was no restriction on study methodology/design. We included published, peer-reviewed studies, in any language.

We excluded studies that met the following criteria.

1. *Leadership development was not the main focus*; the focus was on improving clinical or educational leadership rather than research leadership; or where there was no intervention to improve research leadership.
2. *Publications available as titles/abstracts only*.
3. *Protocols, editorials or commentaries without published results.*
4. *Review publications.*

Outcomes

We organised outcomes according to Kirkpatrick's Framework for Training Programmes, adapted from previous reviews of medical leadership development programmes (Table 1).[13,20,21] After summarising and organising study outcomes, one reviewer (HKS) categorised study outcomes as positive, neutral or negative. Another reviewer (OL) checked these categorisations.

[Table 1 to follow]

Kirkpatrick Level	Description
Level 1 Reaction	Participants' satisfaction with the learning experience, its organisation, presentation, content, teaching methods, and quality of instruction.
Level 2A Change in Attitudes	Changes in the attitudes or perceptions among participant groups towards leadership, management, and/or administration.
Level 2B Change in Knowledge or Skills	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.
Level 3A Behavioural change (self-reported)	Documents the transfer of learning to the workplace and changes to professional practice, as noted by participants themselves.
Level 3B Behavioural change (observed)	Documents the transfer of learning to the workplace and changes to professional practice, as noted by a third party or by promotions.
Level 4 Results (self-reported or observed)	Organisational results/group effectiveness either perceived by respondents/subordinates, or tangible outcomes such as reduced costs, improved quality and safety, and impact of projects.

Search strategies

We searched Medline (PubMed), Embase (Elsevier), CINAHL and PsycINFO from 1 January 2000 until 25 January 2023. The search was limited to studies from the year 2000 onwards based on previous systematic reviews in related topics.[5,8,11,13]

We designed a search string in Medline (PubMed) that included the following concepts: Research AND Leadership AND Education/Programme AND Evaluation. The complete search strings for all databases are provided in Appendix 2.

On 1 February 2023, we conducted a forward and backward citation search on 36 includable

references identified from the screen of database search results, using SpiderCite.

Study screening and selection

Titles and abstracts, and then full texts, were independently screened, in duplicate, against the inclusion criteria. At both stages, disagreements were resolved by discussion or consensus with a third author. Figure 1 outlines the screening process. In Appendix 3 we have listed studies excluded at the full-text screening stage, with reasons for their exclusion.

Data extraction

We used a data extraction form for study characteristics and outcome data, which we piloted on four studies in the review. Where available, we extracted the following data:

1. Participant Characteristics:
    a) Number of participants
    b) Gender
    c) Age
    d) Participant occupation(s)
2. Methods:
    a) Details of the study site
    b) Details of intervention methods
    c) Details of data collection methods
3. Outcomes:
    a) Main findings

Critical appraisal of included studies

To isolate the most reliable evidence, two researchers independently appraised each study using the Medical Education Research Study Quality Instrument (MERSQI) and Joanna Briggs Institute (JBI) Checklist for Qualitative Studies.[22,23] Differences in MERSQI and JBI score were resolved by consensus with another researcher.

We applied the MERSQI to all studies which included a quantitative component. The MERSQI consists of ten items in six domains.[22] Total scores can range from five to 18. In line with Lyons and colleagues,[13] studies with scores of 12 or higher were categorised as higher reliability studies.

We applied the JBI Checklist for Qualitative Studies to all studies which included a qualitative component. It includes ten items, for a total score of 0-10. Following recommendations from the JBI Reviewers' Manual[24] and Lyons and colleagues,[13] a cut-off score for higher reliability studies was predetermined at 6/10.

Measurement of effect and data synthesis

Where ≥2 studies reported on the same quantitative outcome, we conducted a meta-analysis, using risk ratios for dichotomous outcomes and mean differences or standardised mean differences for continuous outcomes. Anticipating considerable heterogeneity, we used a random effects model. We used the I^2 statistic to measure heterogeneity among the included studies. Where data was presented only as a mix of text and figures, we used WebPlotDigitizer to interpret the values. We used STATA version 16.1 to conduct the meta-analyses.

Where qualitative data were reported, we followed a meta-aggregation approach.[23] We grouped qualitative findings according to meaning using MaxQDA 2022 (VERBI Software, 2021). Findings were synthesised and described iteratively in a collaborative process between reviewers (KK, HKS, OL).

Unit of analysis

We used the individual as the unit of analysis, where possible. Where individualised data was not available, we extracted the information as it was presented, e.g. programme/team/project outcomes.

Missing data

In an adjustment to our protocol,[18] we contacted the authors of two studies[25,26] for missing data used in our meta-analysis.

Subgroup and sensitivity analyses

We calculated descriptive statistics for the design and evaluation of subgroups which achieved Kirkpatrick level four outcomes and those categorised as higher reliability using the MERSQI and JBI tools. We did not conduct sensitivity analyses.

Results

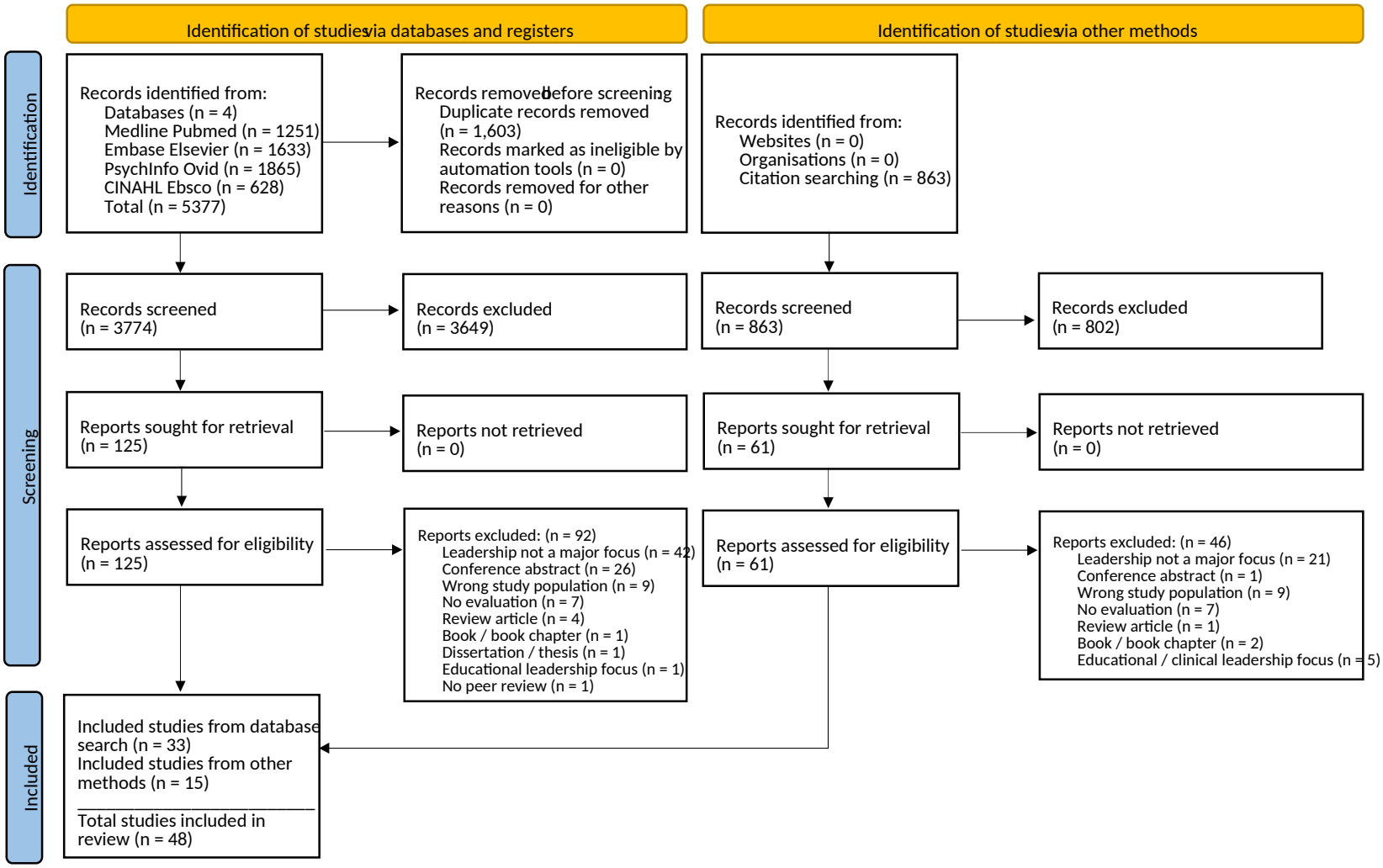
Search results

Database searching and the citation analysis yielded 3774 and 863 records, respectively, after de-duplication (see Appendix 4). After title and abstract screening, we retrieved 125 studies from the database search and 61 studies from the citation analysis.

During full-text screening, we identified 33 studies from the database search and 15 studies from the citation analysis which met our inclusion criteria.

The included 48 studies represented an estimated 3999 individuals (3772 intervention, 227 controls). This is an estimate because two studies[51,61] omitted participant numbers and others likely had a crossover of participants from previous studies.

[Figure 1 to follow]


Identification of studies via databases and registers
Identification of studies via other methods
Identification
Screening
Included
Records identified from:
Databases (n = 4)
Medline Pubmed (n = 1251)
Embase Elsevier (n = 1633)
PsychInfo Ovid (n = 1865)
CINAHL Ebsco (n = 628)
Total (n = 5377)
Records removed before screening:
Duplicate records removed (n = 1,603)
Records marked as ineligible by automation tools (n = 0)
Records removed for other reasons (n = 0)
Records identified from:
Websites (n = 0)
Organisations (n = 0)
Citation searching (n = 863)
Records screened (n = 3774)
Records excluded (n = 3649)
Records screened (n = 863)
Records excluded (n = 802)
Reports sought for retrieval (n = 125)
Reports not retrieved (n = 0)
Reports sought for retrieval (n = 61)
Reports not retrieved (n = 0)
Reports assessed for eligibility (n = 125)
Reports excluded: (n = 92)
Leadership not a major focus (n = 42)
Conference abstract (n = 26)
Wrong study population (n = 9)
No evaluation (n = 7)
Review article (n = 4)
Book / book chapter (n = 1)
Dissertation / thesis (n = 1)
Educational leadership focus (n = 1)
No peer review (n = 1)
Reports assessed for eligibility (n = 61)
Reports excluded: (n = 46)
Leadership not a major focus (n = 21)
Conference abstract (n = 1)
Wrong study population (n = 9)
No evaluation (n = 7)
Review article (n = 1)
Book / book chapter (n = 2)
Educational / clinical leadership focus (n = 5)
Included studies from database search (n = 33)
Included studies from other methods (n = 15)
Total studies included in review (n = 48)

Study reliability (MERSQI and JBI)

Twenty-two studies met our threshold for higher reliability. Seven studies met the criteria using both critical appraisal tools, 11 studies by the MERSQI only and four studies by the JBI only. The median score according to the MERSQI was 10.5 (range 5.5-17). The median score according to the JBI was 3.5 (range 0-8). Methodological approaches were similar in the higher reliability group compared to overall. Appendix 5 is a summary table, including the MERSQI and JBI scores for all included studies.

Five studies (10%) included a control group, three of which were randomized.[25–27] One study used data from a previously published sample in the same population,[28] and one study used individuals who did not apply or were not selected as a control.[29]

Median MERSQI scores for each of questionnaire content, internal structure and relationships to other variables was zero. The data analysis in twenty studies (43%) appropriately answered their research questions. Twenty-one studies (47%) conducted an analysis beyond descriptive statistics. Seven studies (15%) conducted an appropriate analysis beyond descriptive statistics. Evaluations often applied statistical tests which were unsuitable for the data (e.g. applying *t*-tests to non-parametric data) or presented changes as significantly different without describing statistical tests.

Most studies omitted key qualitative reporting elements according to the JBI (see Appendix 6). No study reported a philosophical perspective, and few explicitly reported research

methodology. One study (3%) located the researchers culturally/theoretically.[30] Three studies[31–33] (9%) addressed the influence of the researchers on the research. Across all included studies, nearly a third (31%) did not declare if there were any conflicting interests. No study clearly utilised the TIDieR checklist for intervention reporting.[34]

Evaluation methods

Thirty studies (63%) used mixed methods. Sixteen (33%) used only quantitative methods and two (4%) used only qualitative methods. The most used data collection methods were questionnaires (n=40, 83%), analysis of documents (n=14, 29%) and interview/focus groups (n=8, 17%).

The majority of questionnaires used Likert scales (n=38, 95%) and open/free-text questions (n=29, 73%). Most questionnaires were not validated (n=24, 60%). A quarter of studies (n=11, 28%) used externally validated instruments such as the Mentoring Competency Assessment (MCA),[35] or conducted an expert review for content validity (n=5, 10%).

Two studies used examination questions.[36,37] Data used in evaluations typically consisted of participants' self-assessed subjective outcomes (n=40, 83%), self-reported objective outcomes such as promotions/publications (n=17, 35%) or findings from organisational documents (n=9, 19%). Eight studies (17%) collected data from subordinates, peers, superiors or experts.

Thirty studies (62%) used pre-post designs. Some studies (n=14, 29%) included long-term follow-up between three months[38] to thirteen years[39] post-intervention. Follow-up consisted of questionnaires in all instances.

Study Outcomes by Kirkpatrick Level						
L4	**L3b**	**L3a**	**L2b**	**L2a**	**L1**	**Study: Author and Year**
●	●	●	●	●	●	van Dongen, 2021[32]
●	●	●	●	●	●	Hartvigsen, 2021[40]
●	●	●	●	●	●	Pillai, 2018[41]
●	●	●	●	●	●	Hickey, 2014[42]
●	●	●	●	●	●	Bragg, 2021[43]
●	●	●		●	●	Gillespie, 2017[31]
●	●	●		●	●	Denicola, 2018[38]
●	●		●		●	Abraham, 2020[44]
●	●			●		Hafsteinsdóttir, 2020[45]
●	●				●	Kingsley-Smith, 2022[46]
●		●	●	●	●	Trejo, 2021[47]
●		●	●	●	●	Desai, 2021[48]
●		●	●	●	●	Durbin, 2019[49]
●		●	●	●	●	Feldman, 2012[50]
●		●	●	●	●	McBride, 2006[51]
●		●		●	●	Galaviz, 2019[52]
●		●		●	●	Vassallo, 2021[33]
●			●	●	●	Jaffe, 2009[36]
●			●		●	Stacciarini, 2019[53]

●					●	Butler, 2017[54]
●						Gianfredi, 2019[37]
●						Byington, 2016[55]
●						Deiner, 2017[39]
●		●	●		●	Walsh, 2022[56]
	●	●	●	●	●	Göç, 2022[57]
	●	●	●	●		Williams, 2020[58]
	●		●	●	●	Pfund, 2014[26]
	●			●		Gandhi, 2019[59]
	●				◐	McBride, 2017[60]
		●	●	●	●	Pfund, 2013[27]
		●	●	●	●	Johnson, 2021[61]
		●	●	●	●	Weber-Main, 2019[25]
		●	●	●	●	Aliyu, 2022[62]
		●	●	●	●	Norman, 2021[63]
		●	●	●		Wides, 2014[64]
		●	●	●		Nearing, 2020[65]
		●	●		●	Libby, 2018[66]
		●	●			McMahon, 2019[67]
		●	●			Adams, 2018[68]
		●				Mayowski, 2019[69]

			●	●	●	DiFrances, 2019[70]
			●	●	●	Feldman, 2009[71]
			●	●	●	Steen, 2021[29]
			●		●	Johnson, 2014[72]
			●		●	Behar-Horenstein, 2019[30]
			●			Gandhi, 2016[28]
				●	●	Martina, 2014[73]
				●	●	McBride, 2018[74]

Participants

Most programmes (n=39, 81%) exclusively included healthcare researchers. Most studies included junior academics (e.g. post-doctoral fellows) (n=35, 73%) and/or senior academics (e.g. professors) (n=25, 52%). A minority included doctoral (n=10, 21%), masters (n=3, 6%), or undergraduate students (n=1, 2%). Sixteen studies included participants from more than one experience level (33%).

Most programmes recruited participants using a competitive application (n=27, 56%) or nomination process (n=6, 13%). The remainder included volunteers (n=8, 17%) or did not report their recruitment process (n=7, 15%).

Programme faculty

Twenty programmes (42%) did not describe their faculty. Of the remaining, eight used in-house faculty only (29%). Four used external faculty only (15%). Sixteen studies used a mix of in-house and external faculty (57%).

Programme design

Programmes included components in 26 countries. Appendix 7 contains a list of these countries. Most programmes included a component in the United States (n=35, 73%).

The median amount of time spent on the intervention was 2.5 days. The median period over which the intervention was delivered was eight months.

Most of the included studies explicitly focused on leadership (n=27, 56%). The remainder focused on mentoring development (n=21, 44%) and described leadership development, without necessarily using the term. Only four studies (8%) clearly defined leadership.

Educational methods

Most studies included small-group work, lectures/large-group sessions and/or experiential learning (Figure 2). Organisational outcomes were present in most studies which used coaching (6/8), experiential learning/project work (16/24) and mentoring (15/23).

Few programmes explicitly conducted a needs assessment (31%), incorporated participant goal setting (29%) or had a plan for transfer of learning (29%).

Only 19 studies (40%) reported applying learning theory during their intervention. Reported theories included cognitive theory,[28,59,72] adult learning theory,[73] self-determination theory,[65] or the theory of planned behaviour.[36]

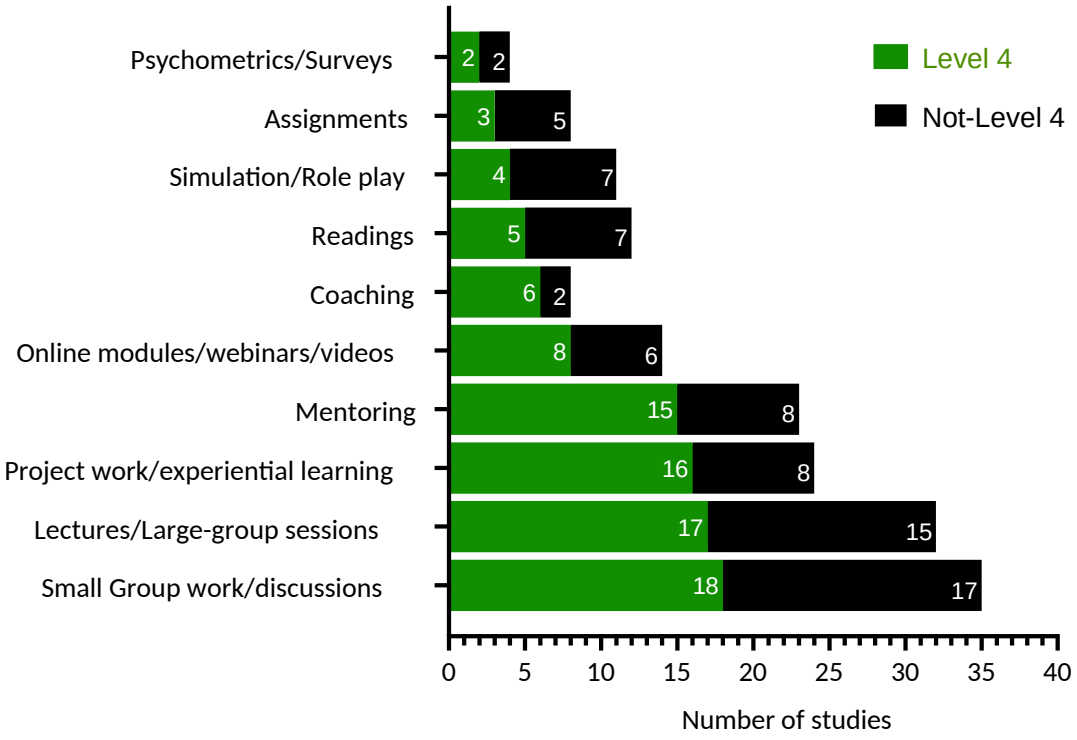
Educational content

Mentoring others and leadership behaviours were common content topics (Figure 2). Studies which covered content related to research (14/20) and organisational behaviour (5/8) were more common in the subgroup which reported organisational outcomes.

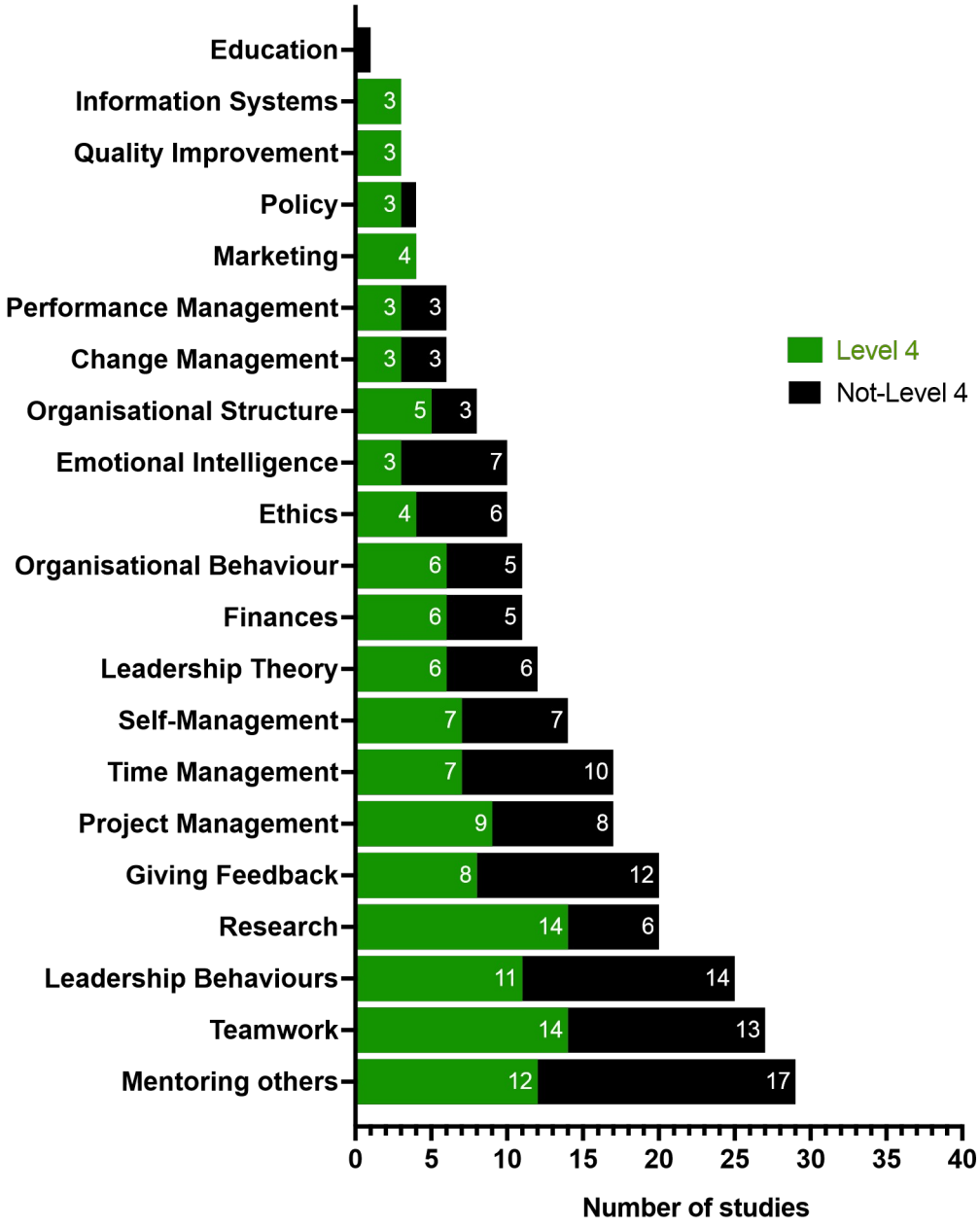
Studies which covered content related to giving feedback (12/20) were more common in the higher reliability subgroup. However, studies including research (6/20), emotional intelligence (3/10) and–surprisingly–leadership theory (2/12) were less common in higher reliability studies.

[Figure 2 to follow]

Educational Methods



Educational Content


Education
Information Systems
Quality Improvement
Policy
Marketing
Performance Management
Change Management
Organisational Structure
Emotional Intelligence
Ethics
Organisational Behaviour
Finances
Leadership Theory
Self-Management
Time Management
Project Management
Giving Feedback
Research
Leadership Behaviours
Teamwork
Mentoring others
3
3
3
4
3 3
3 3
5 3
3 7
4 6
6 5
6 5
6 6
7 7
7 10
9 8
8 12
14 6
11 14
14 13
12 17
Level 4
Not-Level 4
0 5 10 15 20 25 30 35 40
Number of studies

Study outcomes

After outcomes were summarised and organised into Kirkpatrick levels, we noted 163 (97%) of these outcomes were positive, three were neutral (2%) and two were negative (Table 2).

Meta-analysis of outcomes at Kirkpatrick level one (reaction) and two (learning)

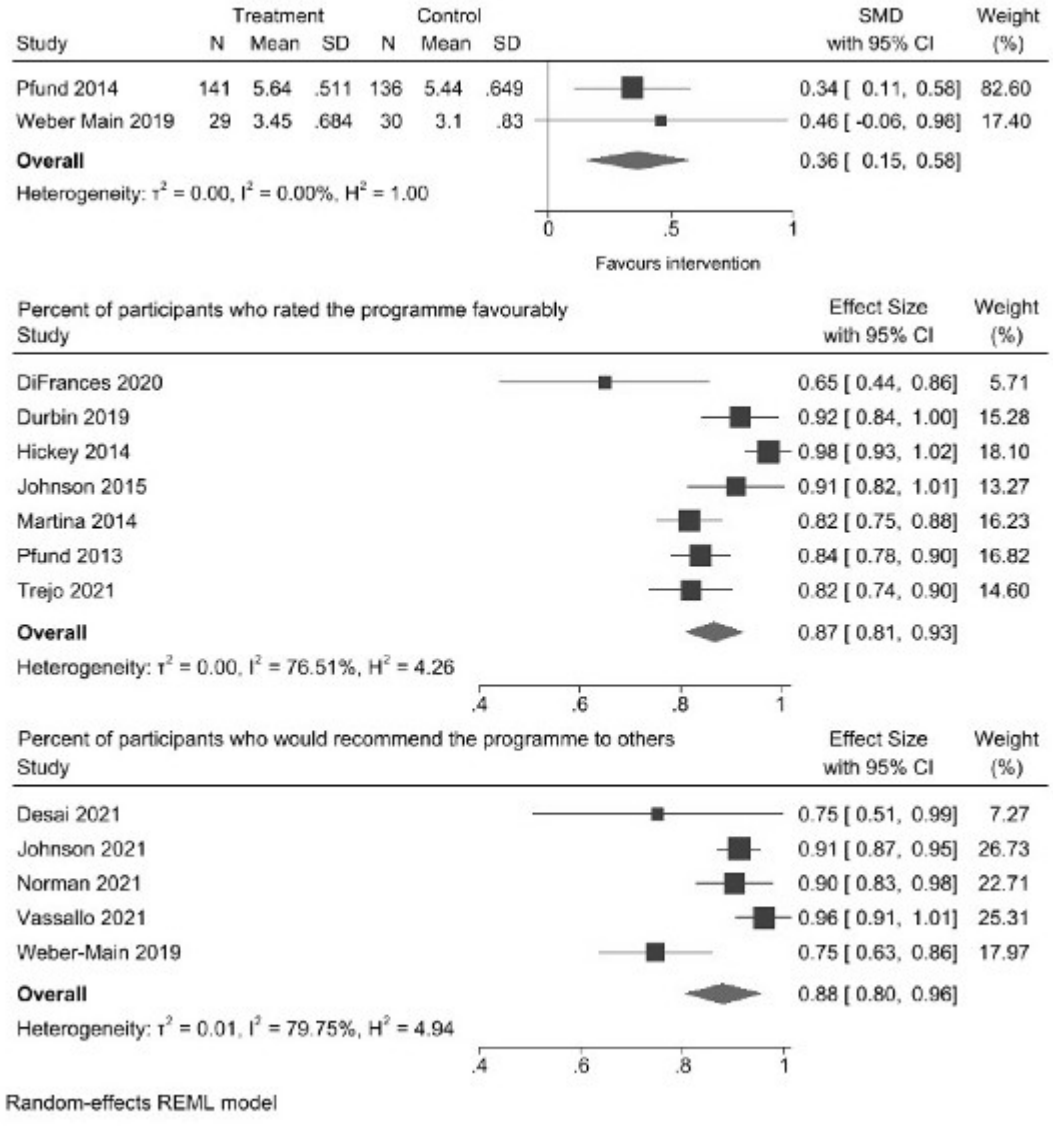
Two randomised controlled trials[25,26] (336 participants in aggregate) assessed the impact of mentoring training on participants' composite MCA scores. One trial compared the provision of 8 hours of mentoring training to a control group which received no training;[26] the other trial compared a 90-minute online mentoring module alongside five hours of workshops, to a mentoring tip-sheet.[25] Post-mentorship training, there was a significant difference in participants' composite skill scores, favouring the intervention (SMD 0.36, 95% CI 0.15 to 0.58, p=0.001) (Figure 3, top panel).

We pooled data on the percentage of participants who rated the leadership development programme favourably and those who would recommend the programme to others. Seven studies[27,42,47,49,70,72,73] included 506 participants in aggregate. Overall, 87% of respondents (95% CI: 81% to 93%) rated the leadership development programme favourably (e.g. rated the programme as excellent and/or effective and/or reported favourable appraisal). Heterogeneity was high (I^2 = 77%) (Figure 3, middle panel).

We pooled data on whether participants would recommend the programmes. Five

studies[25,33,48,61,63] included 362 participants in aggregate. Overall, 88% of respondents (95% CI: 80% to 96%) stated they would recommend the programmes to others. Heterogeneity was high (I^2 = 80%) (Figure 3, bottom panel).

[Figure 3 to follow]



Kirkpatrick level three (change in behaviour)

Several studies measured the proportion of participants who had applied for research leadership roles during/after the programme.[31,41–43] Göc and colleagues used network analysis to demonstrate increased inter-institutional collaboration.[57] Two studies evaluated changes in behaviour reported by others, either supervisors[38] or subordinates.[26]

Kirkpatrick level four (organisational results)

Half of the included studies reported organisational outcomes. Organisational outcomes were more common in studies which defined leadership (4/4), outlined a plan for transfer of learning (11/14) or delivered their intervention in a mixed location format (4/5). Few of these studies met criteria for higher reliability. Organisational outcomes were rarely reported in those programmes which had external-only faculty (3/11) or were held externally (1/4). Most of those papers with external faculty (8/11) or held externally (3/4) were rated as higher reliability.

Several studies collected data relating to objective academic outputs, including published peer-reviewed articles,[32,36,41,42,52] conference abstracts,[32,42,52] successful grant applications/funding,[32,42,43,46,52] and outcomes of programme projects.[41,46,52] Other studies focused on promotions won,[32,33,38] promotions of individuals who were mentored by programme participants,[33] and retention of participants as staff at their institution.[41,55] Durbin and colleagues reported expansion of their intervention at a national level.[49] Trejo and colleagues reported an impact on the workplace morale of staff from underrepresented backgrounds.[47] Finally, two studies reported improved team dynamics and productivity outside the programme, attributed to

the programme by participants.[43,46]

One study clearly presented their financial return on investment. Based on grant awards they calculated this to be a return of "20 to 1."[55] One study estimated the cost of their intervention at $4000 USD per participant,[52] and another reported their costs might not be feasible for some institutions.[53]

Qualitative analysis: meta-aggregation

Studies reporting qualitative findings generally provided limited primary data (such as quotations), or only higher-level summaries. Summaries were coded as qualitative results alongside participant quotes and are presented in Appendix 8.

Qualitative data were largely used in support of quantitative findings or simply to illustrate participants' responses to the programs studied. Additionally, some studies reported participants' rationale for enrolment, and facing/overcoming barriers (such as isolation or discrimination).

Many studies incorporated only superficial qualitative data, limiting our ability to meta-aggregate conclusions. Our meta-aggregation is therefore primarily descriptive, categorising data without further conclusions.

Discussion

We synthesised 48 studies of leadership development programmes in healthcare research. Interventions were mostly aimed at junior academics, in line with Kashiwagi and colleagues'

findings in their systematic review of mentoring programmes for clinician researchers.[75] Programmes varied in design, educational methods and content. Evaluations were principally questionnaires measuring self-assessed outcomes. Few studies included comparison groups which could have controlled for environmental effects. Reported study outcomes were overwhelmingly positive. Organisational outcomes were often related to academic publishing, grant funding, promotions and project outcomes.

Programmes of mixed or internal faculty more consistently reported organisational outcomes than those delivered by external faculty. We suggest external faculty for leadership development should be a supplement to internal faculty and not a replacement.

In line with findings from previous review of physician leadership development,[13] we found organisational outcomes were more aligned with educational methods than any specific educational content. Reviews of physician leadership development have also emphasised the importance of experiential learning.[4,6,8,11,13] We similarly found organisational outcomes were associated with the inclusion of coaching, project work or mentoring.

Studies which incorporated online learning were associated with organisational outcomes. A causal relationship between online learning and organisational outcomes seems unlikely given the emphasis on experiential learning in leadership development literature.[4,11] Nevertheless, given that the SARS-CoV-2 pandemic has driven a movement towards online learning, this association warrants further research.

Critical appraisal with MERQSI and JBI tools identified only half of the studies as being of

higher reliability. Studies tended to use mixed methods, with superficial qualitative components. There were methodological flaws and omissions of key reporting elements in many studies. As in previous reviews,[5,8,11,13] few evaluations conducted appropriate analysis beyond descriptive statistics, and even fewer quantitative studies had a control group. There is a need for more randomised control trials and high-quality qualitative research on leadership development programmes for healthcare researchers.

Many papers did not include all the relevant information needed to determine whether ideal practice in design, delivery and evaluation was followed.[76] We would encourage researchers to consult these critical appraisal tools and the TIDieR checklist[34] when designing and communicating their research, to facilitate more effective synthesis of knowledge. Where possible, study authors should publish their educational methods and content along with their evaluations, so future reviews can form inferences about optimal programme design. Ideally, this should include programme finances, to enable assessment of financial return on investment. Additionally, in line with best practices in healthcare research, authors should make data values relevant to key findings and figures available.[77]

Studies used a variety of assessment methods, with few validated instruments such as the MCA. This aligns with previous reviews, with marked heterogeneity in the quality and selection of methods.[4,5,8,13] The variety of assessment methods used highlights the lack of consensus regarding the evaluation of leadership development programmes. Whilst it can be appropriate to design bespoke approaches, the lack of consistent use of validated instruments limits our ability to learn from published research. There is a need for the development and publication of validated

instruments for evaluating healthcare leadership development programmes, alongside the increased use of existing validated instruments.

We noted fewer studies describing leadership development programmes for healthcare researchers (n=48) than in our previous review of programmes for physicians (n=117).[13] There was a similar tendency for studies to be based in the United States. For generalisability, it is important that studies in other countries are also published. We were pleased to see a larger proportion of papers including participants from lower and middle-income countries in this review.

Our meta-analysis showed that participants favourably rated the interventions evaluated and would recommend them to other colleagues (Figure 3). After categorisation of outcomes using Kirkpatrick's levels, we noted the connotations of 163/168 outcomes were positive. It is unlikely leadership development programmes are all this well-received and successful. Indeed, Straus and colleagues concluded there is at most a modest benefit from most healthcare leadership development programmes,[5] and Avolio and colleagues suggested that only 66% of programmes are successful.[78] It is likely therefore that the apparent success of the programmes evaluated represents a publication bias. For healthcare research leadership development to move forward, we must encourage the transparent publication of both positive and negative findings.

To the best of our knowledge, there is no critical appraisal tool specifically developed for mixed-methods research in medical education. To avoid incorrectly rating papers which included qualitative components we used a combination of the MERSQI and JBI. This approach still risks penalising papers which adopt a mixed-methods approach. Further, the MERSQI assigns three

points to the design of questionnaires, which were not always employed in evaluations.

As the first review of the field, we adopted an exploratory approach to our synthesis, to identify associations between reliability or educational methods/content and programme outcomes. We encourage future reviews to conduct statistical tests, including between core educational methods (coaching, mentoring and projects) and faculty source with organisational outcomes.

Conclusions

We identified the published literature as skewed towards the reporting of positive programme outcomes. Nevertheless, we noted opportunities to improve the design and evaluation of healthcare leadership development programmes. Educational methods appeared to be more important for organisational outcomes than specific educational content. Educational methods that were particularly associated with increased organisational outcomes included coaching, experiential learning/project work and mentoring. Programmes delivered by external faculty appeared less likely to be associated with organisational outcomes than those with internal or mixed faculty, though this needs to be investigated further. Finally, improving evaluation design will allow educators and evaluators to better understand factors which are reliably associated with organisational outcomes of leadership development.

Contributions:

HKS, CF, KK, AMS, KM and OL contributed to the conception and design of the review. HKS, CF, KK, SS, DF and OL screened studies for inclusion. HKS, CF, KK, SS, DF and OL abstracted, appraised and coded studies. HKS, KK, AMS and OL contributed to the data analysis. All authors contributed to the drafting and editing of the article. All authors reviewed and approved the final manuscript prior to submission, agreeing to be accountable for all aspects of the work.

Acknowledgements:

We would like to thank Nia Roberts, Oxford University Hospitals NHS Foundation Trust Librarian, for her invaluable assistance and advice in the search process. We would also like to thank Georgina White, University of Oxford Bodleian Healthcare Senior Library Assistant, for her assistance in sourcing full text publications. Finally, we would like to thank Eli Zuzovsky, University of Oxford Rhodes Scholar, for his assistance in text translation.

Funding/Support:

No specific funding was received for this research. Harry Kingsley-Smith was supported by University of Oxford's Clinical Academic Graduate School (OUCAGS), and a Horlock Trust Studentship (University of Oxford) during this work. Christian Farrier, Koot Kotze and Sarah Short were all supported by Rhodes Scholarships. Daniel Foran was also supported by OUCAGS.

Other disclosures:

None.

Ethical Approval:

Not Applicable

References

1. Cook I, Grange S, Eyre-Walker A. Research groups: How big should they be? *PeerJ*. 2015;3:e989. doi:10.7717/peerj.989

2. Goodall AH. Highly cited leaders and the performance of research universities. *Res Policy*. 2009;38(7):1079-1092. doi:10.1016/J.RESPOL.2009.04.002

3. Lichten CA, Marsden G, Pollitt A, Kiparoglou V, Channon KM, Sussex J. Does a biomedical research centre affect patient care in local hospitals? *Health Res Policy Syst*. 2017;15(1):1-25. doi:10.1186/S12961-016-0163-7/TABLES/3

4. Lucas R, Goldman EF, Scott AR, Dandar V. Leadership Development Programs at Academic Health Centers: Results of a National Survey. *Academic Medicine*. 2018;93(2):229-236. doi:10.1097/ACM.0000000000001813

5. Straus SE, Soobiah C, Levinson W. The impact of leadership training programs on physicians in academic medical centers: A systematic review. *Academic Medicine*. 2013;88(5):710-723. doi:10.1097/ACM.0b013e31828af493

6. West M, Armit K, Loewenthal L. *Leadership and Leadership Development in Healthcare: The Evidence Base.*; 2015.

7. Rosenman ED, Shandro JR, Ilgen JS, Harper AL, Fernandez R. Leadership training in health care action teams: A systematic review. *Academic Medicine*. 2014;89(9):1295-1306. doi:10.1097/ACM.0000000000000413

8. Frich JC, Brewster AL, Cherlin EJ, Bradley EH. Leadership Development Programs for Physicians: A Systematic Review. *J Gen Intern Med*. 2015;30(5):656-674. doi:10.1007/s11606-014-3141-1

9. Sadowski B, Cantrell S, Barelski A, O'Malley PG, Hartzell JD. Leadership Training in Graduate Medical Education: A Systematic Review. *J Grad Med Educ*. 2018;10(2):134-148. doi:10.4300/JGME-D-17-00194.1

10. Sultan N, Torti J, Haddara W, Inayat A, Inayat H, Lingard L. Leadership Development in Postgraduate Medical Education: A Systematic Review of the Literature. *Academic Medicine*. 2019;94(3):440-449. doi:10.1097/ACM.0000000000002503

11. Geerts JM, Goodall AH, Agius S. Evidence-based leadership development for physicians: A systematic literature review. *Soc Sci Med*. 2020;246:112709. doi:10.1016/j.socscimed.2019.112709

12. Kumar B, Swee ML, Suneja M. Leadership training programs in graduate medical education: A systematic review. *BMC Med Educ*. 2020;20(1):175. doi:10.1186/s12909-020-02089-2

13. Lyons O, George R, Galante J, et al. Evidence-based medical leadership development: a systematic review. *BMJ Leader*. Published online 2020. doi:10.1136/leader-2020-000360

14. Lyons O, Su'a B, Locke M, et al. A systematic review of leadership training for medical students. *N Z Med J*. 2018;131(1468):75-84. Accessed January 23, 2018. http://www.ncbi.nlm.nih.gov/pubmed/29346359

15. Onyura B, Crann S, Tannenbaum D, Whittaker MK, Murdoch S, Freeman R. Is postgraduate leadership education a match for the wicked problems of health systems leadership? A critical systematic review. *Perspect Med Educ*. 2019;8(3):133-142. doi:10.1007/s40037-019-0517-2

16. Webb A, Tsipis N, McClellan T, et al. A First Step Toward Understanding Best Practices in Leadership Training in Undergraduate Medical Education. *Academic Medicine*. 2014;89(11):1563-1570. doi:10.1097/ACM.0000000000000502

17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71

18. kingsley-smith H, Farrier C, Kotze koot, et al. Development of Research Leadership in Healthcare: A Systematic Review. Open Science Framework.

19. Clark J, Glasziou P, Del Mar C, Bannach-Brown A, Stehlik P, Scott AM. A full systematic review was completed in 2 weeks using automation tools: a case study. *J Clin Epidemiol*. 2020;121:81-90. doi:10.1016/j.jclinepi.2020.01.008

20. Kirkpatrick DL. Evaluation of training. In: Craig RL, Bittel LR, eds. *Training and Development Handbook*. McGraw-Hill Book Company; 1967.

21. Steinert Y, Mann K, Centeno A, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach*. 2006;28(6):497-526. doi:10.1080/01421590600902976

22. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *J Am Med Assoc*. 2007;298(9):1002-1009. doi:10.1001/jama.298.9.1002

23. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis. *Int J Evid Based Healthc*. 2015;13(3):179-187. doi:10.1097/XEB.0000000000000062

24. Aromataris E, Munn Z. *JBI Reviewer's Manual*. Joanna Briggs Institute; 2019. doi:10.46658/JBIRM-19-01

25. Weber-Main AM, Shanedling J, Kaizer AM, Connett JE, Lamere M, El-Fakahany EE. A randomized controlled pilot study of the University of Minnesota mentoring excellence training academy: A hybrid learning approach to research mentor training. *J Clin Transl Sci*. 2019;3:152-164. doi:10.1017/cts.2019.368

26. Pfund C, House S, Asquith P, et al. Training mentors of clinical and translational research scholars: a randomized controlled trial. *Acad Med*. 2014;89:774-782. doi:10.1097/acm.0000000000000218

27. Pfund C, House S, Spencer K, et al. A research mentor training curriculum for clinical and translational researchers. *Clin Transl Sci*. 2013;6(1):26-33. doi:10.1111/cts.12009

28. Gandhi M, Johnson M. Creating More Effective Mentors: Mentoring the Mentor. *AIDS Behav*. 2016;20 Suppl 2(Suppl 2):294-303. doi:10.1007/s10461-016-1364-3

29. Steen K, Vornhagen J, Weinberg ZY, et al. A structured professional development curriculum for postdoctoral fellows leads to recognized knowledge growth. *PLoS One*. 2021;16(11):e0260212-. doi:10.1371/journal.pone.0260212

30. Behar-Horenstein LS, Feng X, Prikhidko A, Su Y, Kuang H, Fillingim RB. Assessing Mentor Academy Program Effectiveness using Mixed Methods. *Mentor Tutoring*. 2019;27:109-125. doi:10.1080/13611267.2019.1586305

31. Gillespie GL, Gakumo CA, Von Ah D, Pesut DJ, Gonzalez-Guarda RM, Thomas TL. A summative evaluation of productivity and accomplishments of Robert Wood Johnson Foundation Nurse Faculty Scholars Program participants. *J Prof Nurs*. 2017;34:289-295. doi:10.1016/j.profnurs.2017.11.001

32. van Dongen L, Cardiff S, Kluijtmans M, et al. Developing leadership in postdoctoral nurses: A longitudinal mixed-methods study. *Nurs Outlook*. 2021;69:550-564. doi:10.1016/j.outlook.2021.01.014

33. Vassallo A, Walker K, Georgousakis M, Joshi R. Do mentoring programmes influence women's careers in the health and medical research sector? A mixed-methods evaluation of Australia's Franklin Women Mentoring Programme. *BMJ Open*. 2021;11(10):e052560-. doi:10.1136/bmjopen-2021-052560

34. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348. doi:10.1136/BMJ.G1687

35. Fleming M, House S, Hanson VS, et al. The Mentoring Competency Assessment: Validation of a New Instrument to Evaluate Skills of Research Mentors. *Acad Med*. 2013;88(7):1002. doi:10.1097/ACM.0B013E318295E298

36. Jaffe DM, Knapp JF, Jeffe DB. Final evaluation of the 2005 to 2007 National Pediatric Emergency Medicine Fellows' Conferences. *Pediatr Emerg Care*. 2009;25(5):295-300. doi:10.1097/PEC.0b013e3181a34159

37. Gianfredi V, Balzarini F, Gola M, et al. Leadership in Public Health: Opportunities for Young Generations Within Scientific Associations and the Experience of the "Academy of Young Leaders". *Front Public Health*. 2019;7:378. doi:10.3389/fpubh.2019.00378

38. Denicola C, Altshuler L, Denicola G, Zabar S. The Strategic Teamwork for Effective Practice Mentor Development Program (STEP-MDP): Expanding capacity for clinical and translational science by investing in research staff. *J Clin Transl Sci*. 2018;2:173-177. doi:10.1017/cts.2018.28

39. Deiner S. Expanding the Field of Surgical Researchers: The Jahnigen Career Development Award. *J Am Geriatr Soc*. 2017;65(10):e146-e150. doi:10.1111/jgs.14967

40. Hartvigsen J, Kawchuk G, Breen A, et al. Leadership and capacity building in chiropractic research: report from the first CARL cohort. *Chiropr Man Therap*. 2021;29:9. doi:10.1186/s12998-021-00363-8

41. Pillai G, Chibale K, Constable EC, et al. The Next Generation Scientist program: capacity-building for future scientific leaders in low- and middle-income countries. *BMC Med Educ*. 2018;18(1):233. doi:10.1186/s12909-018-1331-y

42. Hickey KT, Hodges EA, Thomas TL, et al. Initial evaluation of the Robert Wood Johnson Foundation Nurse Faculty Scholars program. *Nurs Outlook*. 2014;62(6):394-401. doi:10.1016/j.outlook.2014.06.004

43. Bragg M, Arshonsky J, Pageot Y, et al. Student-led research team-building program may help junior faculty increase productivity in competitive biomedical research environment. *BMC Med Educ*. 2021;21(1):3. doi:10.1186/s12909-020-02396-8

44. Abraham C, Kleinpell R, Godwin KM, Dolansky MA. The interprofessional Veterans Affairs Quality Scholars program pre- and postdoctoral nurse fellow outcomes. *Nurs Outlook*. 2021;69(2):202-211. doi:10.1016/j.outlook.2020.09.003

45. Hafsteinsdóttir TB, Schoonhoven L, Hamers J, Schuurmans MJ. The Leadership Mentoring in Nursing Research Program for Postdoctoral Nurses: A Development Paper. *Journal of Nursing Scholarship*. 2020;52(4):435-445. doi:10.1111/jnu.12565

46. Kingsley-Smith H, Short S, Kotze K, Lyons O. Next Generation Leaders Programme: A Multi-Methods Evaluation of a Leadership Development Programme for Biomedical Researchers. *Adv Med Educ Pract*. 2022;13:1547-1554. doi:10.2147/amep.S386961

47. Trejo J, Wingard D, Hazen V, et al. A system-wide health sciences faculty mentor training program is associated with improved effective mentoring and institutional climate. *J Clin Transl Sci*. 2021;6:e18-. doi:10.1017/cts.2021.883

48. Desai MM, Göç N, Chirwa T, et al. Strengthening the Mentorship and Leadership Capacity of HIV/AIDS and Tuberculosis Researchers in South Africa. *Am J Trop Med Hyg*. 2021;105(5):1317-1325. doi:10.4269/ajtmh.21-0072

49. Durbin DR, House S, Meagher EA, Rogers J. The role of mentors in addressing issues of work-life integration in an academic research environment. *J Clin Transl Sci*. 2019;3:302-307. doi:10.1017/cts.2019.408

50. Feldman MD, Steinauer JE, Khalili M, et al. A mentor development program for clinical translational science faculty leads to sustained, improved confidence in mentoring skills. *Clin Transl Sci*. 2012;5(4):362-367. doi:10.1111/j.1752-8062.2012.00419.x

51. McBride AB, Fagin CM, Franklin PD, Huba GJ, Quach L. Developing geriatric nursing leaders via an annual leadership conference. *Nurs Outlook*. 2006;54(4):226-230. doi:10.1016/j.outlook.2006.05.003

52. Galaviz KI, Narayan KM V, Manders OC, et al. The Public Health Leadership and Implementation Academy for Noncommunicable Diseases. *Prev Chronic Dis*. 2019;16:E49-. doi:10.5888/pcd16.180517

53. Stacciarini JMR, McDaniel AM. EMBRACE: Developing an inclusive leadership program with and for undergraduate nursing students. *Journal of Professional Nursing*. 2019;35(1):26-31. doi:https://dx.doi.org/10.1016/j.profnurs.2018.09.001

54. Butler 3rd J, Fryer CS, Ward E, et al. The Health Equity Leadership Institute (HELI): Developing workforce capacity for health disparities research. *J Clin Transl Sci*. 2017;1(3):153-159. doi:10.1017/cts.2017.6

55. Byington CL, Keenan H, Phillips JD, et al. A Matrix Mentoring Model That Effectively Supports Clinical and Translational Scientists and Increases Inclusion in Biomedical Research: Lessons From the University of Utah. *Acad Med*. 2016;91(4):497-502. doi:10.1097/acm.0000000000001021

56. Walsh KF, Fahme S, Reif LK, Mathad J, Konopasek L, Downs JA. Novel, Low-Cost Intervention to Promote Women's Advancement in Global Health Research. *Acad Med*. 2022;97(1):84-88. doi:10.1097/acm.0000000000004382

57. Göç N, Hassan S, Bani I, et al. Strengthening Public Health Scholarship in Sudan: The Role of Leadership and Mentorship Development. *Am J Trop Med Hyg*. 2022;107:1323-1330. doi:10.4269/ajtmh.22-0377

58. Williams NJ, Ravenell J, Duncan AF, Butler M, Jean-Louis G, Kalet A. Peer Mentor Development Program: Lessons Learned in Mentoring Racial/Ethnic Minority Faculty. *Ethn Dis*. 2020;30:321-330. doi:10.18865/ed.30.2.321

59. Gandhi M, Raj T, Fernandez R, et al. Mentoring the Mentors: Implementation and Evaluation of Four Fogarty-Sponsored Mentoring Training Workshops in Low-and Middle-Income Countries. *Am J Trop Med Hyg*. 2019;100:20-28. doi:10.4269/ajtmh.18-0559

60. McBride AB, Campbell JC, Woods NF, Manson SM. Building a mentoring network. *Nurs Outlook*. 2017;65:305-314. doi:10.1016/j.outlook.2016.12.001

61. Johnson MO, Fuchs JD, Sterling L, et al. A mentor training workshop focused on fostering diversity engenders lasting impact on mentoring techniques: Results of a long-term evaluation. *J Clin Transl Sci*. 2021;5:e116-. doi:10.1017/cts.2021.24

62. Aliyu MHD, Sani MU, Ingles DJMS, et al. Building Research Capacity in HIV and Noncommunicable Diseases in Africa: A Mentorship and Leadership Workshop Report. *Journal of Continuing Education in the Health Professions*. 2022;42(1):e106-e110. doi:10.1097/CEH.0000000000000380

63. Norman MK, Mayowski CA, Wendell SK, Forlenza MJ, Proulx CN, Rubio DM. Delivering What We PROMISED: Outcomes of a Coaching and Leadership Fellowship for Mentors of Underrepresented Mentees. *Int J Environ Res Public Health*. 2021;18:4793. doi:10.3390/ijerph18094793

64. Wides C, Mertz E, Lindstaedt B, Brown J. Building leadership among laboratory-based and clinical and translational researchers: the University of California, San Francisco experience. *Clin Transl Sci*. 2014;7(1):69-73. doi:10.1111/cts.12135

65. Nearing KA, Nuechterlein BM, Tan S, Zerzan JT, Libby AM, Austin GL. Training Mentor-Mentee Pairs to Build a Robust Culture for Mentorship and a Pipeline of Clinical and Translational Researchers: The Colorado Mentoring Training Program. *Acad Med*. 2020;95(5):730-736. doi:10.1097/acm.0000000000003152

66. Libby AM, Ingbar DH, Nearing KA, Moss M, Albino J. Developing senior leadership for clinical and translational science. *J Clin Transl Sci*. 2018;2(3):124-128. doi:10.1017/cts.2018.34

67. McMahon M, Brown A, Bornstein S, Tamblyn R. Developing Competencies for Health System Impact: Early Lessons Learned from the Health System Impact Fellows. *Healthc Policy*. 2019;15(Sp):61-72. doi:10.12927/hcpol.2019.25979

68. Adams J, Kawchuk G, Breen A, et al. Leadership and capacity building in international chiropractic research: introducing the chiropractic academy for research leadership (CARL). *Chiropr Man Therap*. 2018;26:5. doi:10.1186/s12998-018-0173-3

69. Mayowski CA, Norman MK, Schenker Y, Proulx CN, Kapoor WN. Developing a team science workshop for early-career investigators. *J Clin Transl Sci*. 2019;3:184-189. doi:10.1017/cts.2019.391

70. Di Frances CD, Childs E, Fetterman JL, et al. Implementing and Evaluating a Mentor Training to Improve Support for Early-Career Scholars in Tobacco Regulatory Science. *Nicotine Tob Res*. 2020;22(6):1041-1045. doi:10.1093/ntr/ntz083

71. Feldman MD, Huang L, Guglielmo BJ, et al. Training the next generation of research mentors: the University of California, San Francisco, Clinical & Translational Science Institute Mentor Development Program. *Clin Transl Sci*. 2009;2:216-221. doi:10.1111/j.1752-8062.2009.00120.x

72. Johnson M, Gandhi M. A mentor training program improves mentoring competency for researchers working with early-career investigators from underrepresented backgrounds. *Advances in Health Sciences Education*. 2015;20(3):683-689. doi:10.1007/s10459-014-9555-z

73. Martina CA, Mutrie A, Ward DS, Lewis V. A Sustainable Course in Research Mentoring. *Clin Transl Sci*. 2014;7:413-419. doi:10.1111/cts.12176

74. McBride AB, Campbell JC, Deming K. Does Having Been Mentored Affect Subsequent Mentoring. *J Prof Nurs*. 2018;35:156-161. doi:10.1016/j.profnurs.2018.11.003

75. Kashiwagi DT, Varkey P, Cook DA. Mentoring programs for physicians in academic medicine: A systematic review. *Academic Medicine*. 2013;88(7):1029-1037. doi:10.1097/ACM.0B013E318294F368

76. Thomas PA, Kern DE, Hughes MT, Chen BY. *Curriculum Development for Medical Education : A Six-Step Approach*.; 2016.

77. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA. Public Availability of Published Research Data in High-Impact Journals. *PLoS One*. 2011;6(9):e24357. doi:10.1371/JOURNAL.PONE.0024357

78. Avolio BJ, Reichard RJ, Hannah ST, Walumbwa FO, Chan A. A meta-analytic review of leadership impact research: Experimental and quasi-experimental studies. *Leadership Quarterly*. 2009;20(5):764-784. doi:10.1016/J.LEAQUA.2009.06.006

Table Titles and captions

Table 1: Kirkpatrick's Framework for evaluation of training programmes, adapted from Kirkpatrick[20] and Lyons et al.[13]

Table 2: Study outcomes, organised by Kirkpatrick level, and categorised as positive (green), neutral (amber) or negative (red). Where no symbol is present, a paper did not evaluate outcomes at that level. Half the included studies reported organisational outcomes (n=24). The majority also reported outcomes for one or more of: reactions (n=36, 75%); changes in attitude (n=32, 67%) and knowledge/skills (n=33, 69%); and self-reported changes in behaviour (n=28, 58%). A minority reported observed changes in behaviour (n=15, 31%).

Figure Legends

Figure 1. Flowchart showing authors' inclusion and exclusion data during the search of the literature around leadership development programmes for healthcare researchers. Adapted from the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement flow diagram.[17]

Figure 2. Stacked bar charts showing firstly the frequency of different educational methods used in included leadership development programmes and secondly, the frequency of different educational topics of the content delivered in included leadership development programmes (n=48). The frequency to which programmes that included an educational method or content topic and also included organisational outcomes (Kirkpatrick level four) is shown. "Level 4" (in green) highlights the number of studies which included an organisational outcome. "Not Level 4" (in black) highlights the number of studies which did not include organisational outcomes i.e. any combination of only Kirkpatrick levels one, two or three.

Figure 3. A meta-analysis of the outcomes of leadership development programmes for healthcare researchers. The figure shows three forest plots exploring: the impact of mentoring training on participants' composite mentoring competency scores from two studies aggregating 336 individuals (top panel), the percent of programme participants who rated the programme favourably from seven studies aggregating 506 individuals (middle panel), and the percent of programme participants who would recommend the programme to others from five studies aggregating 362 individuals (bottom panel).