Assessing the effect of empathy-enhancing interventions in health education and training: a systematic review of randomised controlled trials

Rachel Winter 1, Eyad Issa, Nia Roberts, Robert I Norman, Jeremy Howick

ABSTRACT

Objective To estimate the effect of empathy interventions in health education and training from randomised controlled trials (RCTs).

Methods MEDLINE, PsycINFO, EMBASE, CINAHL and Cochrane databases were searched from inception to June 2019 for RCTs investigating the effect of empathy-enhancing interventions in medical and healthcare students and professionals. Studies measuring any aspect of ‘clinical empathy’ as a primary or secondary outcome were included. Two reviewers extracted data and assessed the risk of bias of eligible studies using the Cochrane Risk of Bias Tool. Random effects meta-analyses of the impact of empathy training on participants’ empathy levels were performed.

Results Twenty-six trials were included, with 22 providing adequate data for meta-analysis. An overall moderate effect on participant empathy postintervention (standardised mean difference 0.52, 95% CI 0.36 to 0.67) was found. Heterogeneity across trial results was substantial ($I^2=63\%$). Data on sustainability of effect was provided by 11 trials and found a moderate effect size for improved empathy up until 12 weeks (0.69, 95% CI 0.23 to 1.15), and a small but statistically significant effect size for sustainability at 12 weeks and beyond (standardised mean difference 0.34, 95% CI 0.11 to 0.57). In total, 15 studies were considered to be either unclear or high risk of bias. The quality of evidence of included studies was low.

Conclusion Findings suggest that empathy-enhancing interventions can be effective at cultivating and sustaining empathy with intervention specifics contributing to effect size. This review focuses on an important, growing area of medical education and provides guidance to those looking to develop effective interventions to enhance empathy in the healthcare setting. Further high-quality trials are needed that include patient-led outcome assessments and further evaluate the long-term sustainability of empathy training.

Protocol registration number PROSPERO (CRD42019126843).

INTRODUCTION

Rationale Clinical empathy has multiple benefits for patient care and practitioner health. Indeed, person-centred and empathic care are central to all professional healthcare education. Empathy in the clinical setting has been defined in various ways and can be considered as a multidimensional construct incorporating affective, cognitive, behavioural and moral components. A widely accepted definition of clinical empathy involves the ability to understand the patient’s situation, perspective and feelings, communicate that understanding to them and act on it in a helpful and therapeutic way. There is still, however, little consensus on the precise nature of clinical empathy, not least reflected in the variety of tools and scales available to measure it. No guidance exists on how to select measures for assessing clinical empathy, and choice of tools is likely to be led by the definition of empathy used or specific domain being measured.

A recent systematic review on empathy measurement tools for care professionals identifies certain measures as scoring highest for quality, but concedes even these had low scores in some of the criteria they used.

Although contested by some, there is evidence that empathy in medical and healthcare students declines during undergraduate education. Researchers agree that empathetic skills can be taught and cultivating empathy to protect against a possible decline would seem sensible. No standard empathy...
curriculum for healthcare training currently exists and empathy-based training does not appear routinely in healthcare education. Understanding what type of empathy training is most effective in healthcare at both cultivating and sustaining empathy would be a useful start in preparing one.

Four systematic reviews of empathy-promoting interventions have been conducted. Kelm et al47 conducted a qualitative synthesis of empathy-cultivating interventions for medical students or physicians. Their findings support the hypothesis that interventions can increase physician and medical student empathy. However, they also identified a lack of rigorous study design in most studies (such as lack of control groups). More recently, Kiosses et al50 published a systematic review of randomised controlled trials (RCTs) of empathy-promoting interventions for health professionals. However, only 2 out of 17 included reported change in empathy as a primary outcome, focusing instead on general communication skills. Hence, the review did not provide robust evidence of empathy-enhancing interventions. In 2019, Patel et al81 reviewed educational interventions aimed at enhancing both empathy and compassion. They included observational as well as randomised studies and looked only at physicians and physicians-in-training. They were not able to pool their results statistically and did not investigate whether potential benefits of empathy were sustained over time. With the most recent review, Fragkos and Crampton29 conclude that empathy interventions significantly increase empathy, but limit their study population to medical students only. In addition, they do not explore whether any improvement in empathy is sustained over time.

These problems listed above present barriers for medical educators looking to implement empathy training into their curricula. It is unclear how large the effect size of effective empathy training is; whether the effect is sustained over time or how best to train students and continuing learners from various health backgrounds. It is important to measure the effect of empathy training, both postintervention and sustainability of effect over time. Arthur et al83 found no effect of empathy training immediately after the training, but significant improvement 12 weeks after the end of the training. A delayed improvement in empathy could potentially be accounted for by participants only recognising the benefits of training once they are putting any lessons learnt into action.

In this systematic review and meta-analysis, we addressed these gaps, with an up-to-date synthesis of RCTs of interventions aimed at promoting empathy, delivered to both medical and healthcare students and professionals, with results that are generalisable to all healthcare contexts. In addition, we will consider both immediate and longer term impact of interventions on empathy.

Objectives
The overarching objective of this systematic review and meta-analysis is to combine data from all available RCTs of empathy-enhancing educational interventions in health education and training. This was achieved with two subsidiary objectives:

1. To assess the effectiveness of empathy-enhancing interventions aiming to enhance empathy in undergraduate and postgraduate health education and training.
2. To assess any lasting effect of empathy training.

We also had three secondary aims:
1. To identify the intervention type (eg, communication skills training) that is most effective at enhancing empathy.
2. To identify the duration of training that is most effective.
3. To identify the tools used to measure empathy levels in participants to consider differences in self-reported and observer-reported measures.

METHODS
Protocol and registration
In accordance with the Cochrane Handbook for systematic reviews of interventions,24 we published a protocol for this systematic review,25 registered with PROSPERO international prospective register of systematic reviews (registration number CRD42019126843). We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.26

Eligibility criteria
RCTs investigating the effect of empathy-enhancing interventions on medical and other healthcare students’ and professionals’ empathy levels as a primary or secondary outcome were eligible for inclusion. Trials measuring empathy via self-reported and/or observer-reported measures were included. See online supplemental eMethods for further details.

Information sources and search strategies
The following databases were searched from inception to 6 June 2019: MEDLINE, PsycINFO, EMBASE, CINAHL, and Cochrane. Search strategies are detailed within online supplemental eTable 1. Electronic searches were supplemented by hand-searching the references of retrieved papers.

Study selection
All studies retrieved through the search strategy were stored using EndNote with duplicates removed. Two authors (RW and EI) reviewed titles and abstracts to identify those meeting inclusion criteria. Full text manuscripts were retrieved for potentially relevant articles. If the decision to include or exclude was unclear, the study was discussed with a third author (JH) to reach a consensus. Seven papers were discussed with the third author. A PRISMA flow chart recorded the screening and selection process.
Data collection
One reviewer (RW) extracted, summarised and recorded data to assess quality and synthesise evidence from included studies. A second author (JH) independently extracted a random sample (10%) of studies to ensure agreement on the information extracted and summarised. See online supplementary eMethods for details on information extracted. If data was not reported, study authors were contacted.

Risk of bias in individual studies
Risk of bias was assessed using the Cochrane Collaboration’s Tool for assessing the risk of bias in clinical trials (see online supplemental eMethods for further details). Using the criteria provided by Higgins and Green, each item was scored as high, low or unclear risk of bias, and evidence from the study was used to justify each score given. Given that evidence increasingly suggests that sequence generation and allocation concealment are of particular importance in determining the overall risk of bias, a study was classed as being at high risk of bias if it scored as high or unclear risk on either of these domains.

Synthesis of results
We calculated the overall effect size of empathy interventions using the standardised mean difference (SMD) and 95% CIs based on the data provided in the studies: postintervention sample size, mean and SD for experimental versus control group (except where only mean difference and SD between preintervention and postintervention for the experimental and control groups were provided). We used a random effects model (REM) to allow for likely different (though related) intervention effects. If a study had more than one intervention arm, we used the results for the most comprehensive training intervention. If a study provided measures of empathy using different tools, the primary tool to measure empathy was used. If it was unclear which was the primary measure, we used the first reported measure of empathy.

Heterogeneity was anticipated between studies and assessed using Cochran’s Q Statistic (heterogeneity was declared if p value <0.10) and quantified using the I² statistic, with an I² value of 50% or more being considered to represent levels of heterogeneity.

Primary analysis included all studies providing the data needed to calculate the mean and SD (or SE) of the postintervention control and intervention groups. Where studies provided more than one point for outcome assessment, the data closest to the end-point of the intervention was used. Studies that provided no numerical data on empathy-related outcomes or data from which it was not possible to calculate mean values and SD were excluded from the meta-analysis.

Additional analyses
We performed a sensitivity analysis excluding studies that were considered to be at high risk of bias (scoring unclear or high risk of bias for either sequence generation or allocation concealment, with evidence suggesting these domains are of particular importance in establishing risk of bias). We conducted separate meta-analyses to look at: sustainability of the effects of the intervention; the intervention type that is most effective; the duration of intervention that is most effective; the outcome assessment tools (comparing objective and subjective outcome measures) and participant populations (effectiveness of interventions aimed at student populations compared with those aimed at professional populations). See online supplemental eMethods for further details.

Risk of bias across studies
Reporting bias was assessed qualitatively based on inspection of the characteristics of the studies included. A funnel plot was produced to investigate small study effects, which may indicate the presence of publication bias. The GRADE system was used to evaluate the overall quality of evidence for the primary outcome.

Patient and public involvement
This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient-relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

RESULTS
Study selection
The literature search resulted in 4904 citations with duplicates removed. Figure 1 provides an overview of the selection process (see online supplemental eResults for further details).

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. RCT, randomised controlled trial.
details). Seventy-two articles were retrieved for full-text review. Forty-six studies were excluded (see online supplemental eTable 2). Sixty-two trials were included (n=2900). Table 1 provides a summary of characteristics (online supplemental eTable 3 gives further details).

### Study characteristics

Study publication dates ranged from 1987 to 2019, with 15 out of 26 trials published in the last 5 years. Thirteen were carried out in the USA and Canada, seven in Europe, three in Iran, and one each in Australia, Ghana and China. Fourteen studies provided a definition of empathy.

#### Study design

Sample size ranged from 12 to 352 participants (median 90.5; IQR 49.25–154). Thirteen studies had 100 or more participants. Seven had fewer than 50 participants. Fifteen studies evaluated empathy interventions for student populations, including seven that looked at medical students, five with nursing students, two with student pharmacists, one with physiotherapy students, and one with a mixed nursing and midwifery student population. Ten trials used professional/qualified populations, with four of these focusing on physicians, one on nurses, and five with qualified care staff, including healthcare assistants. One study used a mixed student and professional population (nursing students and nurse practitioners). Five trials used multiple sites, and five were cluster RCTs. Ten studies defined both inclusion and exclusion criteria for the study. Thirteen defined inclusion criteria only and in three studies inclusion/exclusion criteria were either not given or not clear.

#### Study interventions

While the aims of eligible trials in this review were to enhance empathy through an educational intervention, a range of intervention types were employed. The most commonly used approach was a communication skill-based training intervention, with eight studies using this. Four studies used perspective-taking training, two had a psychotherapy-focused ‘Balint Group’ intervention on medical student empathy. The most frequently used mode of delivery was face-to-face, with 18 interventions using this. Six interventions were delivered online. One employed a self-directed mode of delivery and one a CD-ROM to deliver the intervention.

Studies ranged in duration of intervention (total time spent participating in the intervention) from 20 min to 42 hours. The mean duration was 10.2 hours (SD 8.8). Five studies did not explicitly state duration.

Training packages in six studies were considered to be short duration, lasting 3 hours or less; 10 were considered ‘medium duration’, lasting between 4 and 12 hours; and 5 were considered ‘long duration’, lasting more than 12 hours.

#### Outcome measures

Studies used either self-report or other-objective report measures to assess a change in participants’ empathy. Objective measures included those completed by patients or experts (eg, faculty staff or trained actors playing simulated patients). Most studies (18) used only self-report measures.

The Jefferson Scale of Empathy (JSE) was the most frequently used self-reported outcome measure, for example, Tusky et al. Other self-report tools used included the Balanced Emotional Empathy Scale (BEES), the Ekman Facial Decoding test, and the Toronto Empathy Questionnaire (TEQ).

The Consultation and Relational Empathy Scale (CARE) was the most frequently used objective measure of empathy, with three studies employing it. Other objective outcome measures included the Carkhuff Empathy Rating Scale. In addition, some studies developed their own measures of empathy, for example, Butow et al. They used a Likert scale with 10 items to assess perceived oncologist empathy.

#### Outcome assessment strategy

Timeframes for measuring outcomes varied. Fifteen studies did not specify a timeframe for postintervention measurements or were unclear. For example, Hastings et al reported measuring empathy 6 weeks post-randomisation but were not clear how long after the intervention had ended that this measurement was taken. For studies that were explicit, postintervention measures varied between 2 days and 6 months, with the
## Table 1: Summary of characteristics of included studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>No participants</th>
<th>Participant type</th>
<th>Intervention type</th>
<th>Intervention type</th>
<th>Duration of intervention (hours)</th>
<th>Outcome assessor</th>
<th>Outcome measure</th>
<th>Effect of intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alhassan et al.</td>
<td>2019</td>
<td>Ghana</td>
<td>210</td>
<td>Nursing and midwifery</td>
<td>Communication skills training</td>
<td>Self</td>
<td>12</td>
<td>JSE</td>
<td>No significant effect found</td>
<td></td>
</tr>
<tr>
<td>Arthur et al.</td>
<td>2017</td>
<td>UK</td>
<td>112</td>
<td>Healthcare assistants</td>
<td>Communication skills training</td>
<td>Self</td>
<td>12</td>
<td>JSE</td>
<td>No significant effect found</td>
<td></td>
</tr>
<tr>
<td>Arthur et al.</td>
<td>2017</td>
<td>USA</td>
<td>172</td>
<td>Healthcare professionals</td>
<td>Mixed</td>
<td>4</td>
<td>Self</td>
<td>VST</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Blair Irvine et al.</td>
<td>2012</td>
<td>USA</td>
<td>352</td>
<td>Medical students</td>
<td>Communication skills training</td>
<td>Self</td>
<td>10.5</td>
<td>JSE</td>
<td>No significant effect found</td>
<td></td>
</tr>
<tr>
<td>Buffel du Vauure et al.</td>
<td>2017</td>
<td>France</td>
<td>30</td>
<td>Physicians</td>
<td>Communication skills training</td>
<td>Self</td>
<td>15</td>
<td>JSE</td>
<td>No significant effect found</td>
<td></td>
</tr>
<tr>
<td>Buttow et al.</td>
<td>2007</td>
<td>Australia</td>
<td>25</td>
<td>Student pharmacists</td>
<td>Literature intervention</td>
<td>Self</td>
<td>2</td>
<td>CRP</td>
<td>No significant effect found</td>
<td></td>
</tr>
<tr>
<td>Collins et al.</td>
<td>1998</td>
<td>Canada</td>
<td>33</td>
<td>Nursing students</td>
<td>Communication skills training</td>
<td>NE</td>
<td>18</td>
<td>Self</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Daniels et al.</td>
<td>2016</td>
<td>USA</td>
<td>70</td>
<td>Medical students</td>
<td>Communication skills training</td>
<td>NE</td>
<td>10.5</td>
<td>Self</td>
<td>No significant effect found</td>
<td></td>
</tr>
<tr>
<td>Foster et al.</td>
<td>2016</td>
<td>Canada</td>
<td>33</td>
<td>Nursing students</td>
<td>Communication skills training</td>
<td>NE</td>
<td>8</td>
<td>Self</td>
<td>No significant effect found</td>
<td></td>
</tr>
<tr>
<td>Gholamzadeh et al.</td>
<td>2017</td>
<td>Iran</td>
<td>249</td>
<td>Nursing staff and healthcare assistants</td>
<td>Communication skills training</td>
<td>Self</td>
<td>4</td>
<td>JSE</td>
<td>No significant effect found</td>
<td></td>
</tr>
<tr>
<td>Gould et al.</td>
<td>2017</td>
<td>UK</td>
<td>236</td>
<td>Qualified care staff</td>
<td>Mixed</td>
<td>3</td>
<td>Self</td>
<td>JSE</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Hastings et al.</td>
<td>2018</td>
<td>UK</td>
<td>142</td>
<td>Qualified care staff</td>
<td>Mixed</td>
<td>12</td>
<td>Self</td>
<td>JSE</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Hattink et al.</td>
<td>2015</td>
<td>The Netherlands and UK</td>
<td>142</td>
<td>Student pharmacists</td>
<td>Communication skills training</td>
<td>NE</td>
<td>2.66</td>
<td>Observer</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Lor et al.</td>
<td>2014</td>
<td>USA</td>
<td>44</td>
<td>Nursing staff and students</td>
<td>Communication skills training</td>
<td>Self</td>
<td>2.66</td>
<td>Self</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Lort et al.</td>
<td>2014</td>
<td>USA</td>
<td>40</td>
<td>Medical students</td>
<td>Communication skills training</td>
<td>Self</td>
<td>1</td>
<td>Self</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Lobchuk et al.</td>
<td>2018</td>
<td>Canada</td>
<td>70</td>
<td>Medical students</td>
<td>Communication skills training</td>
<td>Self</td>
<td>1</td>
<td>Self</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>LoSasso et al.</td>
<td>2017</td>
<td>USA</td>
<td>70</td>
<td>Medical students</td>
<td>Communication skills training</td>
<td>Self</td>
<td>1</td>
<td>Self</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>County</td>
<td>No participants</td>
<td>Participant type</td>
<td>Intervention type</td>
<td>Duration of intervention (hours)</td>
<td>Outcome assessor</td>
<td>Outcome measure</td>
<td>Effect of intervention</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>
</tr>
<tr>
<td>Mueller et al 43</td>
<td>2001</td>
<td>USA</td>
<td>37</td>
<td>Physical therapy students</td>
<td>Mixed</td>
<td>11</td>
<td>Self</td>
<td>JSE</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Riess et al 44</td>
<td>2012</td>
<td>USA</td>
<td>99</td>
<td>Physicians</td>
<td>Empathy skills training</td>
<td>4</td>
<td>Observer</td>
<td>CARE, JSE</td>
<td>No significant effect found for CARE, JSE, BEES, Significant effect for EFDT</td>
<td></td>
</tr>
<tr>
<td>Shapiao et al 45</td>
<td>1998</td>
<td>USA</td>
<td>78</td>
<td>Medical students</td>
<td>Mindfulness training</td>
<td>17.5</td>
<td>Self</td>
<td>ECRS</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Sripada et al 46</td>
<td>2010</td>
<td>USA</td>
<td>12</td>
<td>Physicians</td>
<td>Psychotherapy intervention</td>
<td>NE</td>
<td>Observer</td>
<td>BLRI</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Sterkenburg and Vacaru 47</td>
<td>2018</td>
<td>The Netherlands</td>
<td>224</td>
<td>Qualified care staff</td>
<td>Serious game</td>
<td>0.33</td>
<td>Self</td>
<td>EQ</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Tulsky et al 48</td>
<td>2011</td>
<td>USA</td>
<td>48</td>
<td>Physicians</td>
<td>Communication skills training</td>
<td>NE</td>
<td>Observer</td>
<td>ES, EQ</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Vaghee et al 49</td>
<td>2018</td>
<td>Iran</td>
<td>127</td>
<td>Nursing students</td>
<td>Perspective-taking training</td>
<td>3</td>
<td>Self</td>
<td>JSE</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Wolf et al 50</td>
<td>1987</td>
<td>Canada</td>
<td>134</td>
<td>Medical students</td>
<td>Communication skills training</td>
<td>12</td>
<td>Self</td>
<td>HRI, MCI</td>
<td>Significant effect found</td>
<td></td>
</tr>
<tr>
<td>Wundrich et al 51</td>
<td>2017</td>
<td>Germany</td>
<td>158</td>
<td>Medical students</td>
<td>Empathy skills training</td>
<td>6</td>
<td>Self</td>
<td>JSE, OSCE</td>
<td>Mixed. No significant effect found for JSE. Significant effect found on OSCE scores</td>
<td></td>
</tr>
<tr>
<td>Yang et al 52</td>
<td>2018</td>
<td>China</td>
<td>177</td>
<td>Nursing students</td>
<td>Narrative medicine intervention</td>
<td>42</td>
<td>Self</td>
<td>JSE</td>
<td>Significant effect found</td>
<td></td>
</tr>
</tbody>
</table>

BEES, Balanced Emotional Empathy Scale; BLRI, Barrett-Lennard Relationship Inventory; CARE, Consultation and Relational Empathy Scale; CIC, Carkhuff Index of Communication; CRP, Coding Role Play; ECCS, Empathic Communication Coding System; ECRS, Empathy Construct Rating Scale; EFDT, Ekman Facial Decoding test; EO, Empathic Opportunities; EQ, Empathy Quotient; ES, Empathic Statements; HRI, Helping Relationship Inventory; IRI, Interpersonal Reactivity Inventory; JSE, Jefferson Scale of Empathy; MCI, Medical Communication Index; OSCE, Objective Structured Clinical Examination; PE, Perceived Empathy; SECBQ, Staff Empathy for People with Challenging Behaviour; VST, Video Situational Testing.
majority of measures taken within 2 weeks of the intervention.23 26–30 32 41 48 Eleven studies measured the effects at one or more follow-up points (in addition to the postintervention measurement).23 28 29 31 33 37 39 41 49 52 which varied between 4 weeks and 18 months.

**Risk of bias within studies**

In total, 11 studies23 28 31 36–39 43 45 47 48 were considered to be at low risk of bias overall (low risk of bias for sequence generation and allocation concealment).24 Thirteen were considered to be low risk for random sequence generation23 28 29 31 33 35 37 39 41 49 52 and 11 were low risk for allocation concealment.6–23 23–28 28–31 31–39 43 44 47 48 Blinding was not possible in the majority of studies due to the nature of the interventions (often described to participants as empathy-promoting) and the method of outcome assessment (eg, self-report questionnaires, making explicit what is being measured, such as the JSE). Full details of the risk of bias assessment are reported in the online supplemental eResultswith online supplemental eFigure 1 illustrating the overall findings.

**Results of individual studies**

The majority of studies (19/26) found that the tested intervention significantly improved empathy on at least one outcome measure.29 30 33–35 38–41 43–52 Seven studies did not find any significant increase in empathy.23 28 31 32 36 37 42 Of the studies that reported a significant improvement in empathy on at least one outcome measure, 11 were aimed at student populations (representing approximately 73% of student population studies)30 33–35 41 43 45 49–52 and 7 were aimed at professionals (representing 70% of professional population studies).29 38 39 44 46–48 50 52 Fifteen studies reported a significant improvement in empathy using a self-rated outcome measure (this represents 68% of studies (15/22) using a self-report outcome tool).29 30 33 35 38–41 43 45–47 49 50 52 Four studies reported an increase in empathy when using an objective measure (representing 50% (4/8) of studies using an objective outcome measure).24 44 48 51 Seventeen studies employed an educational intervention that had been specifically designed to foster empathy.25 32 34–37 39–44 46–48 50 52 Of these, 12 (70%) were successful.5 34 39–44 46–48 51 52 Four out of five studies that were classed as ‘long duration’ (lasting >12 hours) reported a significant improvement in empathy postintervention33 41 45 52; 50% of ‘medium duration’ studies (between 3 and 12 hours) reported a significant increase in empathy29 35 39 50 51 and 33% of ‘short duration’ studies (<3 hours) reported a significant improvement.47 49

**Synthesis of results**

Of the 26 studies included in this review, 4 were excluded from meta-analysis as they did not provide adequate data from which to calculate the SMD and SD.31 36 46 51 For the studies that were excluded from the primary analysis, Butow et al31 reported a positive but not statistically significant effect and Gould et al36 found no significant difference between control and intervention groups. Wündrich et al51 reported no significant influence of the intervention as measured by the JSE (student version) but did report a positive and statistically significant effect on the observer-assessed outcome. Sripada et al46 also reported a statistically significant positive effect. Of the 22 studies that had adequate data for pooling, all but one (Arthur et al23) showed a benefit of intervention. The primary analysis identified that the overall effect of empathy interventions in terms of improving participant empathy was statistically significant (SMD 0.52, 95% CI 0.36 to 0.67) (figure 2). The Q value indicated

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Weight</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
<th>Std. Mean Difference IV, Random, 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alhassen 2019</td>
<td>6.2%</td>
<td>0.19 [–0.11, 0.49]</td>
<td></td>
</tr>
<tr>
<td>Arthur 2017</td>
<td>3.8%</td>
<td>–0.02 [–0.60, 0.57]</td>
<td></td>
</tr>
<tr>
<td>Blair Irvine 2012</td>
<td>6.2%</td>
<td>0.33 [0.03, 0.63]</td>
<td></td>
</tr>
<tr>
<td>Buffel Du Vaure 2017</td>
<td>6.9%</td>
<td>0.08 [–0.15, 0.31]</td>
<td></td>
</tr>
<tr>
<td>Collins 2017</td>
<td>2.3%</td>
<td>0.24 [–0.62, 1.10]</td>
<td></td>
</tr>
<tr>
<td>Daniels 1998</td>
<td>3.9%</td>
<td>0.88 [0.31, 1.44]</td>
<td></td>
</tr>
<tr>
<td>Foster 2016</td>
<td>3.6%</td>
<td>0.88 [0.27, 1.49]</td>
<td></td>
</tr>
<tr>
<td>Gholamzadeh 2018</td>
<td>4.2%</td>
<td>0.93 [0.41, 1.46]</td>
<td></td>
</tr>
<tr>
<td>Hastings 2018</td>
<td>5.4%</td>
<td>0.35 [–0.03, 0.74]</td>
<td></td>
</tr>
<tr>
<td>Hattink 2015</td>
<td>2.4%</td>
<td>0.87 [0.01, 1.72]</td>
<td></td>
</tr>
<tr>
<td>Larit 2018</td>
<td>4.4%</td>
<td>1.44 [0.93, 1.94]</td>
<td></td>
</tr>
<tr>
<td>Lobchuck 2018</td>
<td>3.5%</td>
<td>0.60 [–0.02, 1.23]</td>
<td></td>
</tr>
<tr>
<td>Lor 2014</td>
<td>3.4%</td>
<td>0.81 [0.16, 1.46]</td>
<td></td>
</tr>
<tr>
<td>LoSasso 2017</td>
<td>4.6%</td>
<td>0.17 [–0.30, 0.64]</td>
<td></td>
</tr>
<tr>
<td>Mueller 2018</td>
<td>3.3%</td>
<td>0.48 [–0.19, 1.14]</td>
<td></td>
</tr>
<tr>
<td>Riess 2012</td>
<td>5.3%</td>
<td>0.31 [–0.09, 0.71]</td>
<td></td>
</tr>
<tr>
<td>Shapiro 1998</td>
<td>4.6%</td>
<td>0.47 [–0.01, 0.95]</td>
<td></td>
</tr>
<tr>
<td>Sterkenburg 2018</td>
<td>6.5%</td>
<td>0.32 [0.06, 0.59]</td>
<td></td>
</tr>
<tr>
<td>Tulszy 2011</td>
<td>3.9%</td>
<td>0.36 [–0.21, 0.94]</td>
<td></td>
</tr>
<tr>
<td>Vaghee 2018</td>
<td>4.5%</td>
<td>0.94 [0.46, 1.43]</td>
<td></td>
</tr>
<tr>
<td>Wolf 1987</td>
<td>5.6%</td>
<td>0.99 [0.63, 1.34]</td>
<td></td>
</tr>
<tr>
<td>Yang 2018</td>
<td>5.4%</td>
<td>0.41 [0.02, 0.80]</td>
<td></td>
</tr>
</tbody>
</table>

Total (95% CI) 100.0% 0.52 [0.36, 0.67]

Heterogeneity: Tau² = 0.08; Chi² = 56.26, df = 21 (P < 0.0001); I² = 63%

Test for overall effect: Z = 6.49 (P < 0.00001)

Figure 2: Meta-analysis of eligible studies providing adequate data to calculate standardised mean difference with 95% CI.
significant heterogeneity, with p equal to 0.0001 and $I^2$ equal to 63%. A summary of findings is presented in table 2.

### Additional analyses

#### Sensitivity analysis

For the sensitivity analysis of the least biased studies (table 2), 11 were judged to have low risk of bias for random allocation or allocation concealment and 9 of these provided sufficient data to be included in a meta-analysis (figure 3).

### Sustainability of improved empathy analysis

Eleven studies provided follow-up data assessing sustainability of changes to empathy, in addition to postintervention measurement. Eight were eligible for inclusion in a subgroup analysis (see online supplemental eResults 1 for further details), which found a moderate effect size for sustainability up to 12 weeks and a smaller, but still significant effect size for sustainability of impact of training at 12 weeks or later (figure 4 and table 2).

### Type of intervention analysis

A meta-analysis comparing subgroups of different types of intervention (see online supplemental eFigure 2 in the
online supplemental eResults for further details) found the greatest effect was with empathy training that was communication skill-based (table 2). The smallest effect reported was for interventions that were described as ‘mixed educational programmes’ and ones based in the arts and humanities (table 2). It is worth noting however that only two studies used arts and humanities interventions (compared with seven in the communication skill group) and this may well impact on the effect size.

Duration of intervention analysis
Interventions of medium and longer duration (online supplemental eFigure 3) were most effective. Interventions of short duration had the smallest effect size (table 2).

Participant population analysis
Studies using healthcare student participant populations appeared to have a larger effect size than those directed at professional/qualified populations (online supplemental eFigure 4). Studies included in a subanalysis of interventions for students showed a moderate effect size of training compared with a smaller but still significant effect size for training directed at professional/qualified populations (table 2).

Outcome assessor analysis
Studies using a self-assessment outcome scale showed a moderate and significant benefit to empathy for the intervention tested (online supplemental eFigure 5) compared with a small and statistically not significant effect size for observer-assessed outcome studies (table 2).

Risk of bias across studies
A funnel plot used in the primary meta-analysis (22 studies) did not reveal evidence of publication bias (figure 5). An evaluation of evidence using GRADE software found the quality of evidence was low (online supplemental eTable 4). This was due to a high or uncertain risk of bias based on random sequence generation and/or allocation concealment in a number of studies and a high degree of heterogeneity across studies.

DISCUSSION
Summary of evidence
Training healthcare practitioners and trainees improved their empathy by a modest amount. The effect of training seemed to diminish, but lasts to beyond 12 weeks.

Comparison with other evidence
Our review supports the evidence of previous similar reviews, finding benefits of empathy training.

Figure 4 Meta-analysis of studies that provided follow-up observation points to determine long-term effectiveness of intervention.
and that practitioner empathy training makes a difference to patients. Our study adds to this evidence by providing an estimate of empathy training from higher quality (randomised) trials and by showing that the effect lasts well beyond the intervention.

**Strengths and limitations**

This review, to the best of our knowledge, is the first systematic review and meta-analysis limited to RCTs of clinical empathy training for all healthcare students and professionals. This is an up-to-date review that excludes non-randomised studies, follows a published protocol and assesses both the immediate and longer term effects of empathy training. Our broad study population with both healthcare students and professionals means findings are generalisable to all areas of healthcare education and training.

We chose to include only the results of the primary measure of empathy reported by each study. Where it was unclear which was the primary measure, we used the measure that was reported first. We recognise that this might have been biased, as authors may have chosen to report the most positive outcomes first. However, we found that this was not necessarily the case. For example, the first measure of empathy reported by Buffel du Vaure et al (who did not specify which measure was primary) had a smaller effect than the second.

We recognise the heterogeneity of the studies in our review and anticipated this. This means that further research is required to identify the most effective empathy training methodology. Also, the strength of findings in this review may be limited by the reporting quality of some of the included studies. A sensitivity analysis of studies of highest quality found a slightly smaller but still significant effect size. Another limitation in reviewing the evidence in this field is the multiple tools used by investigators to measure clinical empathy. With the lack of a definitive definition of clinical empathy and a range of tools measuring different aspects of empathy, the impact of an intervention may vary depending on the measurement tool used. This is demonstrated by Riess et al who found a statistically significant improvement in empathy when measured using the CARE scale but no significant changes using the JSE. In contrast, Buffel du Vaure et al reported the opposite. Perhaps because of the larger sample size or other factors, our review found a benefit of training independently of how it was measured. A further limitation with this review is that we only identified four studies that followed participants up for at least 3 months. The trials identified however found a positive effect. Lastly, we did not measure the qualitative experiences of participants in this review.

**Implications for research and practice**

Interventions for cultivating student and trainee empathy should be further developed and implemented. Optimising implementation will require additional qualitative research on the experiences of empathy teachers and learners. Also, the longer term effects (>12 weeks) of empathy training have not been studied adequately and future research should address this. With competition for time and space in both undergraduate and postgraduate healthcare curriculums, future research in this area needs to be robust. Designers of future trials of empathy training in healthcare can use the results of this review as a guide to their intervention development.

**CONCLUSION**

Teaching students and other learners how to enhance empathy is moderately effective over a sustained period of time and is likely to benefit present and future patients. Future research should focus on empathy interventions with patient-led outcome assessment and on assessing effectiveness of training over more sustained periods of time. Medical educators and curriculum designers can use this research to think of ways to integrate empathy training into busy curricula.

**Contributors**

All authors meet the criteria for authorship as described in the ICMJE recommendations. RW and JH are responsible for substantial contributions to the conception and design of the work including the acquisition of data and analysis and interpretation of data. They have both made substantial contributions to revisions of the draft. RN has contributed to the conception, design and revising of the work for important intellectual content. EI and NR have made contributions to the acquisition, analysis and interpretation of the data. All authors have given final approval to the version to be published.

**Funding**

The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests**

All authors have completed the Unified Competing Interest form and declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

**Patient and public involvement**

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication**

Not required.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**Data availability statement**

Data are available in a public, open access repository.

**ORCID iD**

Rachel Winter http://orcid.org/0000-0002-8775-2461
REFERENCES


