META-ANALYSIS

1 Empathy assessment in healthcare students is highly heterogeneous: A systematic review and meta-analysis (2012-2016)

Fragkos KC, Sotiropoulos I, Frangos CC
Empathy assessment in healthcare students is highly heterogeneous: A systematic review and meta-analysis (2012-2016)

Konstantinos C Fragkos, Ioannis Sotiropoulos, Christos C Frangos

ORCID number: Konstantinos C Fragkos (0000-0002-7677-7989); Christos C Frangos (0000-0002-1743-5962).

Author contributions: Fragkos KC conceptualization, methodology, acquisition of data, analysis and interpretation of data, drafting the article, final approval; Sotiropoulos I data analysis, interpretation of data, revising the article, final approval; Frangos CC data analysis, interpretation of data, critical revision, final approval.

Conflict-of-interest statement: The authors deny any conflict of interest.

PRISMA 2009 Checklist statement: The authors have read the PRISMA 2009 Checklist, and the manuscript was prepared and revised according to the PRISMA 2009 Checklist.

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

Manuscript source: Invited manuscript

Konstantinos C Fragkos, GI Services, University College London Hospitals NHS Foundation Trust, London NW1 2PG, United Kingdom

Konstantinos C Fragkos, Medical School, University College London, London WC1E 6BT, United Kingdom

Ioannis Sotiropoulos, Department of Accounting and Finance, University of Ioannina, Psathaki 481 00 Preveza, Greece

Christos C Frangos, Greek Research Institute for the Study of Quantitative, Social and Biomedical Problems, Athens 141 21, Greece

Corresponding author: Konstantinos C Fragkos, MBBS, MA, MSc, PhD, GI Services, University College London Hospitals NHS Foundation Trust, 250 Euston Road, London NW1 2PG, United Kingdom. constantinos.frangos.09@ucl.ac.uk
Telephone: +44-7960340489
Fax: +44-2034479217

Abstract

BACKGROUND
Clinical empathy leads to improved patient satisfaction and better clinical outcomes. Currently, there are multiple empathy scales with minimal or no efforts to produce an integrated definition of clinical empathy which can be assessed sufficiently by only a few scales. Moreover, there is an unclear overall reliability of these empathy scales, hence limiting comparative evaluation.

AIM
To examine which empathy scales have been used in healthcare students and to estimate their overall internal consistency.

METHODS
A systematic review was performed with inclusion criteria any empirical study with quantitative data examining empathy of healthcare students toward patients between 2012 and 2016. A random effects model was used to produce a pooled estimate of the Cronbach’s alphas. The Haskian-Whalen transformation was used for analyses based on the Rodriguez-Maeda method. Heterogeneity was quantified using the I² statistic and further investigated with subgroup analysis and meta-regression. Publication bias was assessed using funnel plots, Egger’s test, Begg’s test, and the trim and fill analysis.

RESULTS
Thirteen scales have been used to assess clinical empathy in healthcare students from forty nine studies with total sample size 49384 students. The most frequently used scale is the Jefferson Scale of Physician Empathy followed by Davis’ Interpersonal Reactivity Index. The overall reliability was 0.805 (95% CI 0.786-0.823), which is acceptable, but there was heterogeneity and publication bias. Some heterogeneity was explained by the different countries of the studies under investigation and student types but most heterogeneity remained unexplained.

CONCLUSION
The results indicate that scales have satisfactory internal consistency but there are a multitude of scales, definitions and empathy components. Future research should focus on standardizing scales and creating consensus statements regarding the definition of empathy and use of appropriate scales.

Key words: Empathy; Scale; Reliability; Systematic review; Meta-analysis; Heterogeneity

INTRODUCTION
At its core, empathy is a particular form of interpersonal understanding and stands for a basic sensitivity to the mindedness of other persons\(^1\). Empathy is commonly distinguished between cognitive empathy and affective empathy\(^2\). Cognitive empathy denotes the ability to ascribe mental states to others, such as beliefs, intentions, or emotions; and affective empathy essentially involves affect on the part of the empathizer\(^3\). Its archetypal form is empathy with a person in need (e.g., someone who is in pain, sad, or in an upsetting situation)\(^3\). An agreed definition of empathy remains an issue of controversy. A recent systematic review identified 109 suitable articles that could provide a definition of clinical empathy and they first identified three elements to its definition: thinking, feeling and acting. Each definition incorporates one or more of these elements\(^4\). Components of thinking that emerged from the systematic review were: cognition, perspective-taking, imagination/imaging, apprehension, understanding, seeing, perceiving, processing, comprehend, appreciation of, knowledge, recognize, identification, controlled, intellectually sense, role-taking, grasp, and identify with. Components of feeling were: Compassion, feeling, emotion, concern, joining with patient’s feelings, to enter into or join with feelings, socio-emotional, care, emotional participation, affective, vicarious emotional response, generation of similar feelings, sharing of emotions, sense, emotional contagion, sympathize, match/ experience someone’s emotional state, emotive, and specific feeling words: e.g., angry, enjoy, care, sad. And finally, components of acting were: Communication, conveying, behavioural, express, listen, interrupt, and eye contact\(^4\). However the authors didn’t arrive at a complete succinct definition.

Affective approaches to empathy are defined as having a matching reaction to the
emotions of another individual. More recently, integrated approaches comprising of both types of empathy have emerged and appear to be more pragmatic\(^5\). Empathy should not be confused with compassion which requires emotion and action on the part of respondents, based on love, vulnerability, and reciprocity, and is actualized by disempowering one’s self and empowering the subject\(^6\). Similar concepts are sympathy, which is a subjective response based on pity towards another, humanism, caring, altruism, respect and integrity\(^7\). Empathy also has a strong neurobiological basis, although this is not discussed in the present study\(^8\-\(^11\).

Assessing empathy is one of the most important aspects of investigating empathy with over 10 scales currently in use\(^12\-\(^14\). Differences between empathy measures depend on the conceptualization of empathy, factor structures and solutions and the target audience. For example, the Jefferson Scale of Physician Empathy traditionally is a 20-item scale and has three or four subscales depending on the sample under analysis: Perspective taking, compassionate care or standing in the patient’s shoes. Davis’ Interpersonal Reactivity Index consists of four subscales, Fantasy and Perspective Taking, which mirror the emotional aspect of empathy; and Empathetic Concern and Personal Distress, which reflect the cognitive dimension\(^15\).

In general, it is considered that clinical empathy leads to improved patient satisfaction, greater adherence to therapy, better clinical outcomes, greater quality of service delivery perception and lower malpractice liability\(^16\), while for clinicians and society it promotes the values of humanism\(^18\), reduces professional burnout and increases diagnostic accuracy, and increases public trust to the healthcare system\(^19\,20\). Empathy has been found to decline over the course of clinical education, with clinical contact, distress, mistreatment by superiors, vulnerability, lack of social support, high workload, unsuitable learning environments, unsatisfactory role models and short patient admissions contributing towards this\(^21\). A demographic that plays a role in empathy is gender with females having been reported to usually have higher scores in excess of about 1-10 units in the Jefferson Scale of Physician Empathy\(^21\-\(^23\).

Nevertheless, there are unresolved issues. First, there is a multitude of empathy scales with minimal or no efforts to produce an integrated definition of clinical empathy which can be assessed sufficiently by only a few scales. Next, there are 38 systematic reviews in the topic of empathy\(^4\,12\,13\,14\,16\,19\,20\,24\,28\), but only six of them are meta-analyses\(^4\,12\,13\,14\,16\,28\). The lack of quantitative synthesis does not allow for aggregation of research into meta-studies and umbrella reviews that are better in summarizing evidence and guiding policy and practice. Our hypotheses are that multiple empathy scales are currently in use and their results are heterogeneous. Hence, the present study’s objectives are to answer which empathy scales have been used in the last years (2012-2016) to assess and measure empathy in healthcare students, what their reliability is and which factors contribute to their heterogeneity.

**MATERIALS AND METHODS**

**Study eligibility criteria**

Inclusion criteria were any empirical study (full paper) with quantitative data examining empathy of healthcare students towards patients. There was a restriction to English language cross-sectional studies published between 2012 and 2016. Papers had to provide sufficient data to produce an effect measure for the meta-analysis. Empathy was only considered from healthcare students towards patients. Studies were excluded when they had other designs (case studies, pre-post experimental designs, empathy intervention studies), didn’t report quantitative data, were non-English, and investigated other types of empathy: empathy from professionals, non-healthcare students, teachers (teacher empathy), perceived empathy of patients from their healthcare professionals, empathy among adolescents and high school students, empathy towards HIV patients, cross-cultural empathy, multicultural empathy, and empathy towards particular medical conditions.

**Search strategy and terms**

PRISMA guidelines for systematic reviews and meta-analyses were followed\(^4\,29\). Electronic database searches were conducted in Google Scholar, PubMed/Medline, Scopus, CINAHL, EMBASE and ERIC. The keywords for searching were: empathy, caring, humanism, cognitive, emotional, healthcare, medical students, nursing students, allied healthcare professionals students, dental students, reliability, validity, psychometric properties, validity, definition, compassion, care, Cronbach’s alpha, reliability, and consistency. The date of search was 1 January 2017 and included articles from 1 January 2012 up until 31 December 2016. The bibliographies from all included manuscripts and hand searching of relevant healthcare education journals...
(Academic Medicine, Advances in Health Sciences Education, International Journal of Medical Education, Medical Education Online, Medical Teacher, Medical Education, Teaching and Learning in Medicine, Perspectives on Medical Education, BMC Medical Education) were further literature search strategies.

**Study selection, data extraction and quality assessment**

The abstracts of the resulting studies were evaluated in terms of inclusion criteria. The full text was consulted when reading the abstract was not adequate to warrant inclusion. We extracted the following data from the included studies: Country, sample size, age, gender distribution, field, study design, aim, main results, scale, Cronbach’s alpha and number of items. The quality of studies (risk of bias) was assessed with the criteria suggested by Reilly et al.[66].

**Statistical analysis**

Cronbach’s alpha (or alpha coefficient) is a measure of internal consistency of a test or scale and is expressed between 0 and 1. It essentially examines whether the scale under investigation measures what it claims to measure, by computing the inter-correlations of the items. When Cronbach’s alpha ranges from 0.70 to 0.95, a scale is considered to have acceptable internal consistency. Heterogeneous constructs, apparent lack of correlation between items, and too few scale items can lead to a low value of alpha[67,68].

Quantitative analysis was performed with Review Manager 5.3 and R 3.5.1. A random effects model was used to produce a pooled estimate of the Cronbach’s alphas. The Rodriguez-Maeda method is adopted to normalise the alpha coefficient by transforming it to the Hakstian-Whalen $T=(1-\alpha)^{1/3}$, which is a variable with standard normal distribution[69,70]. All analyses are performed using the transformed values and the results, such as the summary effect and its confidence interval, would then be converted back to alpha coefficients for presentation.

Heterogeneity between studies was quantified using the $I^2$ statistic ($P \leq 0.10$ for significance)[71], with values over 50 % indicating considerable heterogeneity[72-76]. Sources of heterogeneity were explored performing subgroup analysis and meta-regression. Publication bias was assessed using funnel plots, Egger’s and Begg’s tests, and the trim and fill analysis[77-79]. A funnel plot is a scatter plot of the effect estimates from individual studies against a measurement of the study’s sample size or precision[72]. Resemblance of a symmetrical inverted funnel supports that findings are due to sampling variation alone; thus, absence of bias[79]. In the present study, the funnel plot depicts the Hakstian-Whalen $T$ against the study’s sample size, since the standard error of the effect size is dependent of the effect size[70,80].

**RESULTS**

**Studies and quality assessment**

The flow chart of study selection is shown in Figure 1. The initial search revealed 2491 studies which after applying the inclusion criteria were reduced to 49 studies. In total, 14 scales of empathy were used. The total number of participants was 49384 students, mean age was 23.5 years and average male percentage was 44.6%. Sample sizes ranged from 44 to 5343. The distribution of student types are: dentistry (2 studies), healthcare (medicine, nursing, physiotherapy, occupational therapy, paramedics, midwifery, nutrition and dietetics) (1 study), medicine (31 studies), nursing (7 studies), osteopathic medical students (2 studies), paramedics (1 study), pharmacy and nursing (2 studies), pharmacy (1 study), physician assistant students (1 study), and speech and hearing sciences (1 study). Overall, medicine and dentistry students were 43028, nursing students were 3242 and allied professions students were 3114. Included studies’ information is shown in shown in Table 1. The overall quality of studies was satisfactory indicating low sampling bias and investigator bias with sufficient sample size in almost each study (Figure 2).

**Meta-analysis of Cronbach’s alpha**

The overall meta-analytic mean of Cronbach’s alpha is 0.805 (95%CI 0.786-0.823) (Figure 3A). The distribution of the alpha coefficients is shown in Figure 3B. There is heterogeneity ($I^2 = 98\%$) and publication bias is possibly present, as shown with partial asymmetry in the funnel plot (Figure 3C). However, Egger’s and Begg’s tests are not significant ($P = 0.076$ and $P = 0.648$, respectively). The average alpha coefficients per scale are shown in Table 2. Trim and fill funnel analysis, indicated that most likely 11 studies more are unpublished due to publication bias and this is significant ($P < 0.05$). The scale with lowest Cronbach’s alpha is the Balanced Emotional Empathy Scale (0.720) and the one with the highest Cronbach’s alpha is the...
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>n</th>
<th>mean Age</th>
<th>Male %</th>
<th>Field</th>
<th>Reliability</th>
<th>Items</th>
<th>Aims</th>
<th>Main results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced emotional empathy scale</td>
<td>Dehning et al[82]</td>
<td>237</td>
<td>21.4</td>
<td>87.3</td>
<td>Medical Students</td>
<td>0.72</td>
<td>30</td>
<td>To examine the differences in empathy between first year and final year medical students in Jimma University, Ethiopia</td>
<td>Male students had statistically significant lower empathy scores</td>
</tr>
<tr>
<td>Caring ability inventory</td>
<td>Ma et al[91]</td>
<td>598</td>
<td>20.9</td>
<td>6.4</td>
<td>Nursing Students</td>
<td>0.77</td>
<td>37</td>
<td>To investigate baccalaureate nursing students’ caring ability in the context of China and to explore the role of clinical practice learning in the development of students’ caring skills</td>
<td>Students in the clinical stage of training scored lower than students in the pre-clinical stage.</td>
</tr>
<tr>
<td>Caring behaviour inventory tool</td>
<td>Labrague et al[111]</td>
<td>586</td>
<td>22.3</td>
<td>10.1</td>
<td>Nursing Students</td>
<td>0.92</td>
<td>42</td>
<td>To identify the correlation between instructors’ and students’ caring behaviours and to explore the impact of instructors’ caring on students’ perceptions of their own caring behaviours</td>
<td>The highest self-reported subscale in the Caring Behaviour Inventory was assurance (mean = 4.796), and the lowest self-rated subscale was connectedness (mean = 4.541)</td>
</tr>
<tr>
<td></td>
<td>Loke et al[92]</td>
<td>657</td>
<td>20.3</td>
<td>13.2</td>
<td>Nursing Students</td>
<td>0.922</td>
<td>42</td>
<td>To evaluate the impact of Singapore’s pre-registration nursing programmes on students’ concept of caring</td>
<td>Results indicated a statistically significant reduction in the overall level of caring behaviour in first to final year students</td>
</tr>
<tr>
<td>Davis’ interpersonal reactivity index</td>
<td>Neumann et al[112]</td>
<td>44</td>
<td>22.8</td>
<td>54.5</td>
<td>Medical Students</td>
<td>0.721</td>
<td>28</td>
<td>To investigate the psychometric properties of two empathy scales</td>
<td>Reliability was satisfactory and comparable to international adaptations</td>
</tr>
</tbody>
</table>
To examine individual predictors of first semester medical students' attitudes toward the value of physician empathy in clinical encounters in univariate and multivariate analyses, discomfort with uncertainty, close-mindedness, dispositional empathy, elitism, medical authoritarianism, egalitarianism, self-concept and well-being predicted students' empathy.

Costa et al. (1)

Portugal, Brazil, UK, New Zealand, Ireland

To examine psychometric properties (reliability, factor structure) of two empathy scales and compare them. The Interpersonal Reactivity Index and Jefferson Scale of Physician Empathy are only weakly related, suggesting that they may measure different constructs (maximum correlation 0.313).

Emotional intelligence assessment scale - empathy

Senyuva et al. (1)

Turkey

To analyse the correlation of self-compassion and emotional intelligence of nursing students. There was a correlation between self-compassion and emotional intelligence (r = 0.400, P < 0.05) and that emotional intelligence has positive contributions to the features of nurses with developed self-compassion.

Jefferson scale of physician empathy
<p>| Study Description                                                                 | Country | Year | Gender | Ethnicity | Specialty | Mean Empathy Score | Study Population | Study Objective                                                                                       | Findings                                                                                                                                                                                                 |
|----------------------------------------------------------------------------------|---------|------|--------|-----------|-----------|--------------------|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Kimmelman et al. [114]                                                           | USA     | 415  | 26     | 54        | Osteopathic medical students | 0.83              | 20               | To determine differences according to year of schooling in mean levels of empathy among osteopathic medical students | There were no statistically significant differences by year of schooling in respondents' gender, ethnicity, or specialty orientation and no statistically significant differences by year of schooling in the mean empathy scores |
| Mandel and Schweinle [115]                                                       | USA     | 328  | 24     | 17.4      | Physician Assistant Students | 0.8               | 20               | To investigate empathy trends among physician assistant students through their education and included gender differences and specialty job interest | 62% had lower median empathy scores toward the end of their didactic training than at the time of matriculation ($P = 0.0001$). Female students were significantly more empathetic at the time of matriculation than men ($P = 0.0003$), while both genders appeared to lose empathy in a parallel fashion during didactic training ($P = 0.76$). There was no association between empathy scores and prospective job category interest |
| Neumann et al. [112] (2)                                                          | Germany | 44   | 22.8   | 54.5      | Medical Students            | 0.803             | 20               | To investigate the psychometric properties of two empathy scales | Reliability was satisfactory and comparable to international adaptations |</p>
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>N</th>
<th>Age (mean)</th>
<th>Specialty</th>
<th>Kaiser-Meyer-Olkin (KMO)</th>
<th>Sample Size</th>
<th>Objective</th>
<th>Methodological Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paro et al.</td>
<td>Brazil</td>
<td>299</td>
<td>61.7</td>
<td>Medical Students</td>
<td>0.84</td>
<td>20</td>
<td>To adapt the Jefferson Scale of Empathy to the Brazilian culture and to test its reliability and validity among Brazilian medical students</td>
<td>Principal component analysis confirmed the construct validity of the scale for three main factors: Compassionate Care (first factor), Ability to Stand in the Patient’s Shoes (second factor), and Perspective Taking (third factor). Gender differences with respect to empathy were not significant</td>
</tr>
<tr>
<td>Calabrese et al.</td>
<td>USA</td>
<td>373</td>
<td>26.1</td>
<td>Osteopathic medical students</td>
<td>0.84</td>
<td>20</td>
<td>To investigate correlations between empathy and interprofessional collaboration in osteopathic medical students and to examine differences in empathy and interprofessional collaboration scores by sex, class year, and specialty interest</td>
<td>Significant correlation was found between scores on the empathy and attitudes scales ($r = 0.42$, $P &lt; 0.01$). Women scored higher than men on the empathy scale (117.1 vs 111.9). No statistically significant difference on the scores of the 2 scales was observed among students who planned to pursue “people-oriented” specialties compared with those interested in “technology/procedure-oriented” specialties as well as in different years if education</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>N</td>
<td>Median</td>
<td>Gender</td>
<td>Age</td>
<td>Study Type</td>
<td>Empathy Scores</td>
<td>Longitudinal Model</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>-----</td>
<td>---------------------------------</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Costa et al.</td>
<td>Portugal</td>
<td>77</td>
<td>31.2</td>
<td></td>
<td></td>
<td>Medical Students</td>
<td>0.77</td>
<td>20</td>
</tr>
<tr>
<td>Gonçalves-Pereira et al.</td>
<td>Portugal</td>
<td>202</td>
<td>32.7</td>
<td></td>
<td></td>
<td>Medical Students</td>
<td>0.75</td>
<td>20</td>
</tr>
<tr>
<td>Hsiao et al.</td>
<td>Taiwan</td>
<td>613</td>
<td>23.3</td>
<td>Women</td>
<td>89.1</td>
<td>Nursing Students</td>
<td>0.93</td>
<td>20</td>
</tr>
<tr>
<td>Kiersma et al.</td>
<td>USA</td>
<td>216</td>
<td>20.5</td>
<td>Men</td>
<td>24.1</td>
<td>Pharmacy and nursing students</td>
<td>0.855</td>
<td>20</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Sample Size</td>
<td>Mean Age</td>
<td>Sample</td>
<td>Empathy Score</td>
<td>Item-total score</td>
<td>Comparison Notes</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------</td>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>---------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Preusche and Wagner-Menghin[1]</td>
<td>Austria</td>
<td>516</td>
<td>47.8</td>
<td>Medical Students</td>
<td>0.823</td>
<td>20</td>
<td>To adapt the Jefferson Scale of Physician Empathy into a German version, examine its psychometric properties, to compare the level of attitude towards empathy with other adaptations</td>
<td></td>
</tr>
<tr>
<td>Shariat and Habibi[11]</td>
<td>Iran</td>
<td>1187</td>
<td>22.6</td>
<td>36.1</td>
<td>Medical Students</td>
<td>0.79</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Wen et al[33]</td>
<td>China</td>
<td>753</td>
<td>36.8</td>
<td>Medical Students</td>
<td>0.83</td>
<td>20</td>
<td>To examine empathy among medical students in China</td>
<td></td>
</tr>
</tbody>
</table>

The scale had acceptable internal consistency and test re-test reliability with a three-structure solution emerging from factor analysis.

Female students had higher scores of empathy and empathy decreased with higher years of education.

The scale accounted for 48% of the variance. The mean empathy score was 109.60. The empathy score of medical students had significant differences between male and females ($P < 0.05$) and academic year ($P < 0.05$).
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>N</th>
<th>Mean Age</th>
<th>Empathy Score</th>
<th>Student Type</th>
<th>Empathy Scale Type</th>
<th>Methodology</th>
<th>Data Fit</th>
<th>Reliability Estimate</th>
<th>Sample Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Williams et al [8]</td>
<td>Australia</td>
<td>330</td>
<td>34.8</td>
<td></td>
<td>Paramedics</td>
<td>Jefferson Scale of Physician Empathy in paramedic students</td>
<td>To investigate psychometric properties of Jefferson Scale of Physician Empathy in paramedic students</td>
<td>Two-factor solution, “compassionate care” and “perspective taking”, accounted for 44.2% of the total variance. The 17-item two-factor model produced good model fit and good reliability estimates. Three of the original items did not fit the model.</td>
<td>0.75</td>
<td>20</td>
</tr>
<tr>
<td>Khademalhosseini et al [12]</td>
<td>Iran</td>
<td>260</td>
<td>20.9</td>
<td>46.2</td>
<td>Medical Students</td>
<td>Jefferson Empathy Scale</td>
<td>To measure empathy score among medical students</td>
<td>Empathy scores decreased with increase in the students' age (P = 0.001) and year of study (P = 0.030). Mean empathy score in basic science level (65.5) was higher than clinical level empathy (55.5). Female students had higher mean empathy score (65.53) than male students (59.02)</td>
<td>0.76</td>
<td>20</td>
</tr>
<tr>
<td>Leombruni et al [11]</td>
<td>Italy</td>
<td>257</td>
<td>20.6</td>
<td>44.4</td>
<td>Medical Students</td>
<td>Jefferson Empathy Scale</td>
<td>To examine psychometric and confirm factor structure of the Italian version of the Jefferson Empathy Scale in Italian medical students</td>
<td>The empathy scale showed an acceptable internal consistency (r = 0.76) and test-retest reliability (r = 0.72). Confirmatory factor analysis found that the 3-factor structure has acceptable data fit. Female medical students showed a higher mean empathy score than did males</td>
<td>0.76</td>
<td>20</td>
</tr>
</tbody>
</table>
To measure and examine empathy among a sample of undergraduate medical students of Bangladesh, mean empathy score was 110.41. There were significant associations between gender and empathy scores. The level of empathy in medical students gradually increases after clinical training in medical college. Non-significant difference were noted between empathy scores and specialty preferences.

van Ryn et al\[113\] (2) USA 4732 50.1 Medical Students 0.88 20 (see details above) (see details above)

Williams et al\[122\] Australia 1111 18.4 Healthcare students (medicine, nursing, physiotherapy, occupational therapy, paramedics, midwifery, nutrition and dietetics) 0.78 20 To examine self-reported empathy levels of students enrolled in different health disciplines from two large Australian universities. The mean female empathy score was significantly higher than the mean male score. Paramedic students had significantly lower empathy scores than all other participants except nursing students ($P < 0.0001$)
To explore the empathy profile of students across five years of medical training and to examine whether the Jefferson Scale for Physician Empathy correlated with a measure of cognitive empathy, the Reading the Mind in the Eyes Test and a measure of affective empathy, the Toronto Empathy Questionnaire, there was a significant correlation between the Jefferson Scale of Physician Empathy and the Toronto Empathy Questionnaire \((\rho = 0.48)\). There was a decline in medical student empathy scores over time. There was a weak little correlation between scores from the Reading the Mind in the Eyes Test and the Jefferson Scale of Physician Empathy. Female students demonstrated significantly higher scores on all three measures. The score distributions of the Jefferson Scale of Empathy tended to be moderately skewed and platykurtic. Women obtained a significantly higher mean score \((116.2 \pm 9.7)\) than men \((112.3 \pm 10.8)\) on the Jefferson Scale of Empathy \((P < 0.01)\). The tentative cut-off score to identify low scorers was ≤ 95 for men and ≤ 100 for women.
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>Group</th>
<th>Empathy Scale</th>
<th>及格分数</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeon and Cho [103]</td>
<td>South Korea</td>
<td>447</td>
<td>18.1</td>
<td></td>
<td>Pharmacy students</td>
<td>0.713</td>
<td>20</td>
<td>To validate an empathy sale and to investigate the empathy levels of pharmacy students in South Korea. The 3-factor model of the empathy scale was confirmed by confirmatory factor analysis and the convergent validity was also supported by its correlations with the interpersonal reactivity index subscales.</td>
</tr>
<tr>
<td>Montanari et al [104]</td>
<td>Italy</td>
<td>797</td>
<td>22.63</td>
<td>26</td>
<td>Nursing Students</td>
<td>0.78</td>
<td>20</td>
<td>To test the psychometric properties of the Jefferson Scale of Empathy-Health Professional Student’s version and to describe their empathic engagement. Fit for a three-factor solution for 14 items: compassionate care/emotional engagement, perspective-taking, and standing in the patient’s shoes. Confirmatory factor analysis on the second half of the sample showed good fit indexes for the 14-item solution and the 20 item solution of the scale, with the exception of one item.</td>
</tr>
<tr>
<td>Park et al [87]</td>
<td>South Korea</td>
<td>5343</td>
<td>26.4</td>
<td>61.5</td>
<td>Medical Students</td>
<td>0.83</td>
<td>20</td>
<td>To evaluate empathy in Korean medical students throughout the country and to make suggestions to improve empathy. Females and post-baccalaureate students had higher scores. Students from higher grade levels had lower scores than those from the lower grade levels.</td>
</tr>
<tr>
<td>Park et al [99]</td>
<td>South Korea</td>
<td>2692</td>
<td>24.7</td>
<td>62</td>
<td>Medical Students</td>
<td>0.715</td>
<td>20</td>
<td>To examine the relationship between stress, social support, and empathy among medical students. Empathy and social support were positively correlated, and empathy and stress negatively correlated. In the regression model, stress and social support predicted empathy.</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Country</td>
<td>N</td>
<td>Mean Age</td>
<td>Gender</td>
<td>Sample</td>
<td>Sample Size</td>
<td>Study Description</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>----</td>
<td>----------</td>
<td>--------</td>
<td>--------</td>
<td>-------------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td>Petek Ster and Selic</td>
<td>Slovenia</td>
<td>845</td>
<td>22.5</td>
<td>Medical Students</td>
<td>0.781</td>
<td>To re-validate the Jefferson Scale of Empathy (Student version) and its factor structure prior further research on empathy in medical students. Females achieved higher empathy scores. The three-factor structure of empathy was confirmed. A higher proportion of explained variation was observed with Perspective Taking and Standing in the Patient’s Shoes, and better internal consistency was noted in a reduced-item scale (16-18 items).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Williams et al</td>
<td>Malaysia</td>
<td>204</td>
<td>20</td>
<td>Medical Students</td>
<td>0.7</td>
<td>To examine empathy scores in undergraduate medical students. The mean empathy score for first year students was significantly higher than second year students ($P &lt; 0.05$). No significant difference relating to gender.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggarwal et al</td>
<td>India</td>
<td>978</td>
<td>21.6</td>
<td>Dentistry Students</td>
<td>0.677</td>
<td>To measure the self-reported empathy levels among dental undergraduate and postgraduate students and to review factors that could affect empathy. There were significant differences in empathy scores by gender and age ($P &lt; 0.01$).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costa et al</td>
<td>Portugal, Brazil, UK, New Zealand, Ireland</td>
<td>3069</td>
<td>38.5</td>
<td>Medical Students</td>
<td>0.69</td>
<td>To measure the self-reported empathy levels among dental undergraduate and postgraduate students and to review factors that could affect empathy. There were significant differences in empathy scores by gender and age ($P &lt; 0.01$).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferreira-Valente et al</td>
<td>Spain</td>
<td>1104</td>
<td>20.7</td>
<td>Medical Students</td>
<td>0.78</td>
<td>To examine the psychometric properties of a Spanish empathy scale. The Spanish scale had acceptable to good sensitivity, convergent validity and reliability. The confirmatory factor analysis supported the three-factor solution and the second order latent factor model.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>N</td>
<td>Gender</td>
<td>Year</td>
<td>Method</td>
<td>Results/Findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>---------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jordan and Foster[98]</td>
<td>USA</td>
<td>163</td>
<td></td>
<td>20</td>
<td>Examination of the interpersonal theory of clinical, personality, and social psychology to examine the construct of empathy and theorize about likely interpersonal correlates.</td>
<td>All factors of empathy were related to interpersonal warmth. Perspective taking and compassionate care were associated with submission. Walking in the patient’s shoes was correlated with social support and less loneliness.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mahoney et al[124]</td>
<td>Australia</td>
<td>281</td>
<td>26</td>
<td>42</td>
<td>To examine student and doctor empathy, and possible associations between empathy and the structure of clinical learning.</td>
<td>Empathy decreased during the course of each year, but no differences between years of clinical education.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Sng et al[93]                                                        | Singapore | 881  | 46.3   | 20     | To investigate psychometric properties of Jefferson Scale of Physician Empathy. | Empathy declined between preclinical and clinical years. Female and medical specialty interest respondents had higher scores but factor analysis suggested that the three factor model did not fit adequately.

Sng et al[93]                          Singapore | 881  | 46.3 | Medical Students | 0.83 | 20 | To investigate psychometric properties of Jefferson Scale of Physician Empathy. | Empathy declined between preclinical and clinical years. Female and medical specialty interest respondents had higher scores but factor analysis suggested that the three factor model did not fit adequately. |

Spasenoska et al[100]                                                  | Malaysia | 193  | 19.3   | 43     | To investigate psychometric properties of Jefferson Scale of Physician Empathy. | The scale was best interpreted as a two-factor solution of perspective taking and compassionate care. |

Stansfield et al[125]                                                  | USA     | 4797 | 23.7   | 50.5   | Analysis of factor structure of empathy and relations to other factors. | Components of empathy change over time during undergraduate medical education (towards the end of education, it is a four factor structure of feelings, importance, ease, and metacognitive effort). |

Kiersma-Chen empathy scale
Narcissism, aloofness, confidence, empathy (NACE) scale - empathy

Pitt et al.[85] Australia 133 27 14 Nursing Students 0.79 24 To describe the personal qualities of newly enrolled Bachelor of Nursing students; to determine if these change according to age, gender, and time. Females were significantly more conscientious, community orientated and involved while males had significantly higher narcissism and aloofness scores and lower empathy.

Patient-Practitioner orientation scale

Dockens et al.[128] USA 93 22.6 6.5 Speech and hearing sciences students 0.78 18 To determine preferences to patient-centeredness in pre-service speech and hearing students in the field of speech and hearing sciences. Across exposure levels, students exhibited high preference to patient centeredness with a mean empathy score of 4.13. A paired sample t-test revealed a significant difference ($P < 0.0001$) between the caring and sharing subscales of the empathy scale. No significant differences across levels of exposure for sharing subscale and caring subscale.

Professionalism assessment scale - empathy

Klemenc-Ketis and Vrecko[127] Slovenia 122 22.1 22.1 Medical Students 0.84 10 To develop and validate a scale for the assessment of professionalism in medical students based on students' perceptions of and attitudes towards professionalism in medicine. The scale was developed with 22 items. The Cronbach’s alpha of the scale was 0.88. Factor analysis revealed three factors: empathy and humanism, professional relationships and development and responsibility.

Pro-Social personality battery (Other-Oriented empathy)
Eley et al.[128] USA 145 24 40.7 Medical Students 0.77 18 
To examine personality trait profiles of rural longitudinal integrated clerkships students who intended and matched to family medicine showed the highest levels of Reward Dependence (warm sociability) and Other-Oriented Empathy compared to any other specialty.

Rural longitudinal integrated clerkships students who intended and matched to family medicine showed the highest levels of Reward Dependence (warm sociability) and Other-Oriented Empathy compared to any other specialty.

<table>
<thead>
<tr>
<th>Toronto composite empathy scale</th>
<th>Tsiantou et al.</th>
<th>Greece</th>
<th>460</th>
<th>20.7</th>
<th>34.8</th>
<th>Dentistry Students</th>
<th>0.75</th>
<th>52</th>
<th>To examine empathy among dentistry students in Greece and validate the Toronto Composite Empathy Scale. The scale has good discriminant and convergent validities. Test-retest reliabilities ranged from 0.478 to 0.779. Rotated factor analysis indicated that items loaded on two cognitive and three emotional factors. Females had significantly higher empathy scores.</th>
</tr>
</thead>
</table>

| Toronto Empathy Questionnaire | Youssuf et al. (2) | Trinidad and Tobago | 662 | 22.2 | 35 | Medical Students | 0.85 | 16 | (see details above) | (see details above) |

Caring Behaviour Inventory tool (0.921).

Analysis of heterogeneity, sensitivity analysis and cumulative meta-analysis
Subgroup analyses were performed with regards to scale (Figure 4A) and country (Figure 4B), and subject of study (Figure 4C). With respect to country, heterogeneity reduces to non-significant in Portugal, Australia, Italy, Malaysia, Iran, and Germany (Figure 4B). Regarding subject of study, heterogeneity is non-significant only in osteopathic medical students and pharmacy and nursing students (Figure 4C). Finally, a meta-regression of mean age and male percentage as predictors of reliability is not significant for both parameters (Table 3).

Sensitivity analysis did not identify any study which influenced the meta-analytic outcome significantly (Figure 5A). Cumulative meta-analysis shows that after the addition of the meta-analysis’ tenth study in 2013 by Preusche and Wagner-Menghin[81], the outcome overall has trivial variation, indicating stability of the results.

Narrative discussion of studies
Thirteen empathy scales were identified with most of them being used in one study apart from the Jefferson Scale of Physician Empathy, Davis’ Interpersonal Reactivity Index and Caring Behaviour Inventory Tool which were used in 34, 3, and 2 studies respectively. Many studies reported that males had lower empathy scores compared to females. This was observed with the Balanced Emotional Empathy Scale[82], the Toronto Composite Empathy Scale[83], the Toronto Empathy Questionnaire[84], the
Narcissism, Aloofness, Confidence, Empathy (NACE) scale–Empathy[85], and the Jefferson Scale of Physician Empathy[28,86-90]. Next, empathy decreased with progression through clinical education and empathy had decreased towards the clinical years with the Caring Ability Inventory[91], the Caring Behaviour Inventory Tool[92], the Toronto Empathy Questionnaire[84], and the Jefferson Scale of Physician Empathy[86,87,93,94].

Regarding correlations between concepts and scales, the Jefferson Scale of Physician Empathy was correlated to Davis’ Interpersonal Reactivity Index (0.313)[95], Kiersma-Chen Empathy Scale ($P < 0.05$)[96] and the Toronto Empathy Questionnaire (0.480)[84]. There was a correlation between self-compassion and emotional intelligence (0.400)[97], empathy and interpersonal warmth[98], and empathy and social support[99].

Main solutions of the factor analysis of Jefferson scale of Physician Empathy were a two[100] or three factor[101-104] solution comprising of mainly perspective taking and compassionate care followed standing in the patient’s shoes.

**DISCUSSION**

In the present study, a systematic review of empathy scales in healthcare students was performed. This is the first systematic review that also performed a meta-analysis of the internal consistency (Cronbach’s alpha) of the scales reviewed. Answering the two research questions, 13 scales have been used to assess clinical empathy in healthcare students form 49 studies and total sample size 49384 students. The most frequently used is the Jefferson Scale of Physician Empathy followed by Davis’ Interpersonal Reactivity Index[105]. And secondly, the overall reliability was 0.805 (95%CI 0.786-0.823), which is acceptable, but there was heterogeneity and publication bias. Some heterogeneity was explained by different countries and student types but most heterogeneity remained unexplained. The most reliable scale was the Caring Behaviour Inventory Tool with alpha 0.921; Balanced Emotional Empathy Scale had the lowest reliability of 0.720, while the Jefferson Scale of Physician Empathy and Davis’ Interpersonal Reactivity Index were near the average with values of 0.798 and 0.791, respectively.

The present study has certain limitations. First of all, there was heterogeneity in the meta-analysis. This was investigated with subgroup analyses and meta-regression,
and a random effects model which sought to reduce the impact of heterogeneity. Heterogeneity could have resulted from population differences and design biases possibly not taken into account. A further limitation was that the meta-analysis was performed at the level of the scale and not the level of subscales for similar scales. The next limitation is the presence of publication bias, as indicated by asymmetry in the

Figure 2 Risk of bias assessment. A: Risk of bias graph: review authors’ judgements about each risk of bias item presented as percentages across all included studies; B: Risk of bias summary: Review authors’ judgements about each risk of bias item for each included study.
Table 2  Meta-analytic means of Cronbach’s alpha per empathy scale

<table>
<thead>
<tr>
<th>Scale</th>
<th>N</th>
<th>k</th>
<th>Alpha</th>
<th>95%CI</th>
<th>l²</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced Emotional Empathy Scale</td>
<td>237</td>
<td>1</td>
<td>0.72</td>
<td>0.666</td>
<td>0.769</td>
<td>--</td>
</tr>
<tr>
<td>Toronto Composite Empathy Scale</td>
<td>460</td>
<td>1</td>
<td>0.75</td>
<td>0.716</td>
<td>0.781</td>
<td>--</td>
</tr>
<tr>
<td>Caring Ability Inventory</td>
<td>598</td>
<td>1</td>
<td>0.77</td>
<td>0.743</td>
<td>0.795</td>
<td>--</td>
</tr>
<tr>
<td>Pro-Social Personality Battery (Other-Oriented Empathy)</td>
<td>145</td>
<td>1</td>
<td>0.77</td>
<td>0.71</td>
<td>0.821</td>
<td>--</td>
</tr>
<tr>
<td>Patient-Practitioner Orientation Scale</td>
<td>93</td>
<td>1</td>
<td>0.78</td>
<td>0.707</td>
<td>0.839</td>
<td>--</td>
</tr>
<tr>
<td>Narcissism, Aloofness, Confidence, Empathy (NACE) scale - Empathy</td>
<td>133</td>
<td>1</td>
<td>0.79</td>
<td>0.734</td>
<td>0.838</td>
<td>--</td>
</tr>
<tr>
<td>Davis’ Interpersonal Reactivity Index</td>
<td>7845</td>
<td>3</td>
<td>0.791</td>
<td>0.742</td>
<td>0.835</td>
<td>96.40%</td>
</tr>
<tr>
<td>Jefferson Scale of Physician Empathy</td>
<td>37159</td>
<td>34</td>
<td>0.798</td>
<td>0.774</td>
<td>0.821</td>
<td>98.30%</td>
</tr>
<tr>
<td>Professionalism assessment scale - Empathy</td>
<td>122</td>
<td>1</td>
<td>0.84</td>
<td>0.794</td>
<td>0.879</td>
<td>--</td>
</tr>
<tr>
<td>Toronto Empathy Questionnaire</td>
<td>662</td>
<td>1</td>
<td>0.85</td>
<td>0.833</td>
<td>0.867</td>
<td>--</td>
</tr>
<tr>
<td>Kiersma-Chen Empathy Scale</td>
<td>216</td>
<td>1</td>
<td>0.86</td>
<td>0.831</td>
<td>0.886</td>
<td>--</td>
</tr>
<tr>
<td>Emotional Intelligence Assessment Scale - Empathy</td>
<td>471</td>
<td>1</td>
<td>0.87</td>
<td>0.85</td>
<td>0.887</td>
<td>--</td>
</tr>
<tr>
<td>Caring Behaviour Inventory Tool</td>
<td>1243</td>
<td>2</td>
<td>0.921</td>
<td>0.915</td>
<td>0.927</td>
<td>0%</td>
</tr>
<tr>
<td>Overall</td>
<td>49384</td>
<td>49</td>
<td>0.805</td>
<td>0.786</td>
<td>0.823</td>
<td>98%</td>
</tr>
</tbody>
</table>

N: Number of participants, k: Number of studies.

funnel plot. This could indicate that there are possibly studies performed that had low internal consistency in their scales and have not been published or accepted for publication. The next important limitation is that the mean values and standard deviations (overall, females, and males) for each scale were not recorded. This is important considering the gender effect in empathy and the variability that this could produce and hence explain heterogeneity. Other empathy moderators, not examined in the present study, were other psychological or personality constructs and differences in years. Quality assessment found only minimal impact from insufficient sample sizes. Nevertheless, the sample size in many studies is not sufficiently large which might have led to small study effects. This suggests caution when interpreting results because small studies are more likely to report better results.

The meta-analytic mean for Cronbach’s alpha was within the acceptable range (0.805 which is over 0.700) for scales as described in the literature. However, it would be preferable if it were closer to 0.900. The most frequent scale used was the Jefferson Scale for Physician Empathy which had an alpha of 0.798. The highest alpha was noted with the Caring Behaviour Inventory Tool which had 0.921. These findings corroborate the concerns mentioned in the Introduction, which suggested that empathy is assessed in many ways with no clear consensus at present with regards to a preferred optimum.

Based on the findings of this meta-analysis, it seems that Caring Behaviour Inventory Tool ought to be utilized more. However, it has been developed mainly for nurses and it hasn’t been adapted for other healthcare specialties. This needs to be addressed and further studies are required with this scale tested in other groups. Nonetheless, the Jefferson Scale of Physician Empathy is currently being used most frequently and is currently dominating the clinical empathy literature. Hence, it seems that this scale could be improved with addition of further items that will increase internal consistency, thus increasing the Cronbach’s alpha. The third scale currently being used most frequently is Davis’ Interpersonal Reactivity Index. Its meta-analytic Cronbach’s alpha has the same issues mentioned for the Jefferson Scale of Physician Empathy.

Another practical implication is the need for a consensus to adopt a certain group of scales with the best internal consistency and widespread use. The present Babel of scales is confusing and disruptive for sound research production. The Jefferson Scale of Physician Empathy seems to be close to an integrated definition of empathy incorporating both cognitive and affective traits. It might benefit from the addition of more items similar to Davis’ Interpersonal Reactivity Index, thus also increasing its alpha. Moreover, empathy is a massive field and its research extends to social neuroscience and neuropsychology. Hence, we believe that clinical empathy as a field might benefit from borrowing concepts from research studies in the field of social neuroscience and neurobiology. This will allow comparative analyses and possible improvements on definition, scales and development of theory. Finally, we believe that more meta-analyses and umbrella reviews are needed with various approaches and research questions in the field of clinical empathy. The data produced by the
literature is large and potentially unmanageable for individual researchers. At the same time, narrative reviews are bound by subjectivity, no matter how in-depth they are. The evidence-based approach will allow the most widely accepted definition and scales of clinical empathy to surface. Also, problematic areas will emerge and will stimulate further research.

In conclusion, this study presented the first meta-analysis of reliability for empathy scales in healthcare education. The results indicate that scales have satisfactory internal consistency but there is a confusion of scales, definitions and empathy components. There is evidently a need to standardize research in clinical empathy with meta-reviews\[110\]. Future research should focus on standardizing scales that are used throughout healthcare education and production of consensus statements on definition of empathy and use of appropriate empathy scales.
Table 3  Meta-regression results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient (95%CI)</th>
<th>Standard error</th>
<th>t</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean age</td>
<td>-0.00528 (-0.01813, 0.007571)</td>
<td>0.006301</td>
<td>-0.84</td>
<td>0.409</td>
</tr>
<tr>
<td>Male %</td>
<td>0.000527 (-0.00077, 0.00182)</td>
<td>0.000634</td>
<td>0.83</td>
<td>0.413</td>
</tr>
<tr>
<td>Constant</td>
<td>0.677161 (0.394914, 0.959408)</td>
<td>0.138389</td>
<td>4.89</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

**Figure 4** Subgroup analyses. A: Forest plot according to scale; B: Forest plot according to country; C: Forest plot according to field of study. IV: Inverse variance.
Empathy is a particular form of interpersonal understanding and stands for a basic sensitivity to the mindedness of other persons. In general, it is considered that clinical empathy leads to improved patient satisfaction, greater adherence to therapy, better clinical outcomes, greater quality of service delivery perception and lower malpractice liability, while for clinicians and society it promotes the values of humanism, reduces professional burnout and increases diagnostic accuracy, and increases public trust to the healthcare system. Assessing empathy is one of the most important aspects of investigating empathy with over 10 scales currently in use.

There is a multitude of empathy scales with minimal or no efforts to produce an integrated definition of clinical empathy which can be assessed sufficiently by only a few scales. Next, there are 38 systematic reviews in the topic of empathy but only 6 of them are meta-analyses. The lack of quantitative synthesis does not allow for aggregation of research into meta-studies and umbrella reviews that are better in summarizing evidence and guiding policy and practice.

Our hypotheses are that multiple empathy scales are currently in use and their results are heterogeneous. The present study’s objectives are to answer which empathy scales have been used in the years 2012 to 2016 to assess and measure empathy in healthcare students, what their reliability is and which factors contribute to their heterogeneity.

A systematic review was performed with inclusion criteria any empirical study with quantitative data examining empathy of healthcare students toward patients between 2012 and 2016. A random effects model was used to produce a pooled estimate of the Cronbach’s alphas. The Hakstian-Whalen transformation was used for analyses based on the Rodriguez-Maeda method. Heterogeneity was quantified using the $I^2$ statistic and further investigated with subgroup analysis and meta-regression. Publication bias was assessed using funnel plots, Egger’s test, Begg’s test, and the trim and fill analysis.
publication bias. Some heterogeneity was explained by the different countries of the studies under investigation and student types but most heterogeneity remained unexplained.

Research conclusions
This study is the first meta-analysis of reliability for empathy scales in healthcare education. The results indicate that scales have satisfactory internal consistency but there is a confusion of scales, definitions and empathy components. The meta-analysis for Cronbach's alpha was within the acceptable range for scales. The Jefferson Scale of Physician Empathy seems to be close to practice, there is need for consensus to adopt a certain group of scales with best internal consistency and widespread use. The Jefferson Scale of Physician Empathy seems to be close to an integrated definition of empathy incorporating both cognitive and affective traits. It might benefit from the addition of more items similar to Davis to interpersonal reactivity Index, thus also increasing its alpha. Finally, more meta-analyses and umbrella reviews are needed with various approaches and research questions in the field of clinical empathy.

Research perspectives
Future research should focus on standardizing scales that are used throughout healthcare education and production of consensus statements on definition of empathy and use of appropriate empathy scales.

REFERENCES


Fragkos KC et al. Empathy assessment in healthcare students


P- Reviewer: Qi XS, Zhang ZH S- Editor: Yan JP I- Editor: A E- Editor: Song H