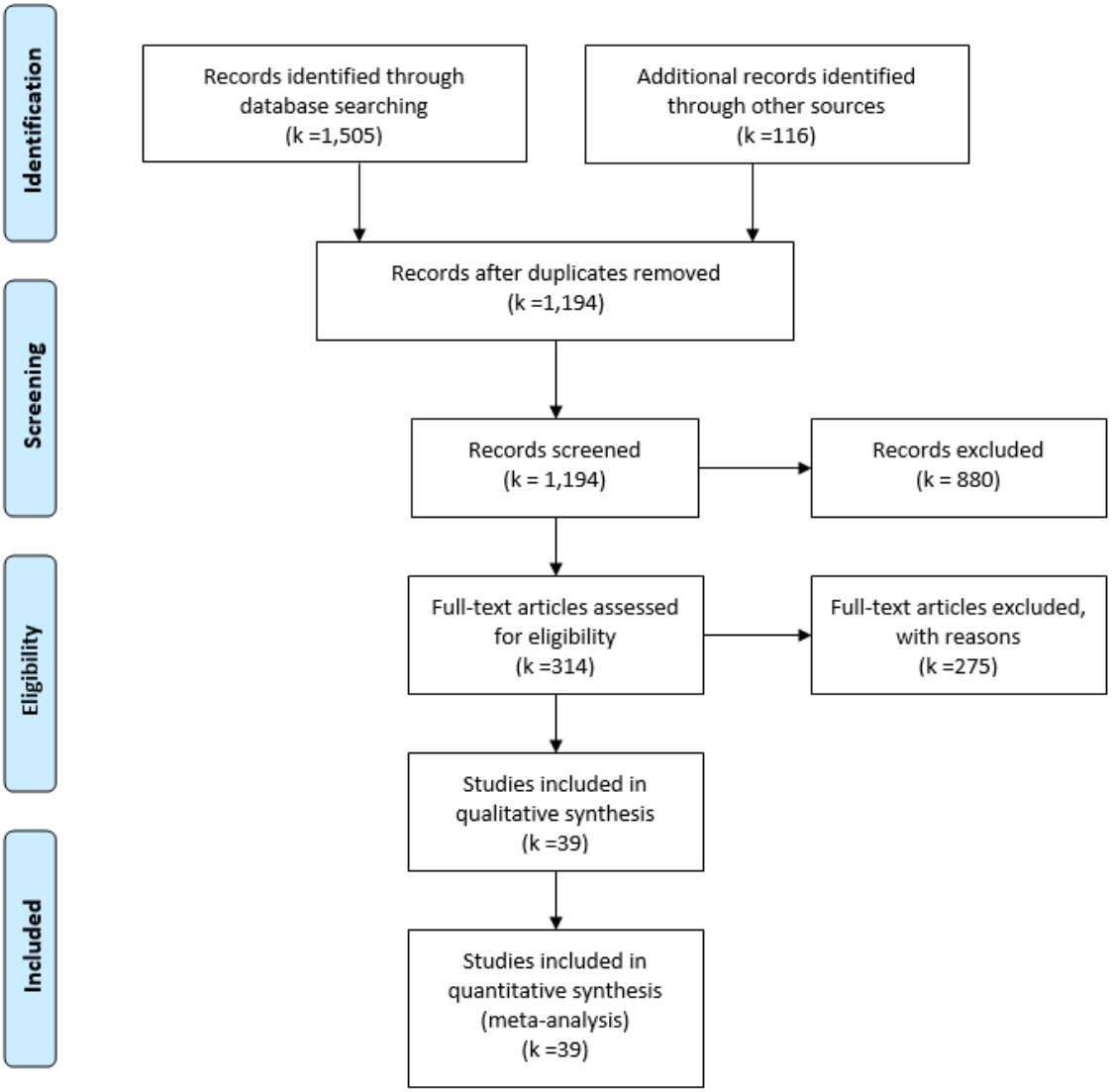
Appendix

Supplementary Table 1:
Electronic search strategy for the systematic review of suicide in medical students.

Tier	*Keyword*	*Hits*
1	"medic* school".ab,kw.ti	53,759
2	"medic* student".ab,kw,ti	15,540
3	"medic* graduat*".ab,kw.ti	6,000
4	"medic* doctor".ab,kw,ti	3,008
5	"medic* physician".ab,kw,ti	3,622
6	"student doctor".ab,kw,ti	70
7	"student physician".ab,kw,ti	85
8	"junior doctor".ab,kw,ti	1,142
9	"junior physician".ab.kw.ti	74
10	"train* doctor".ab.kw,ti	219
11	"train* physician".ab.kw.ti	1,048
12	"intern* doctor".ab,kw,ti	49
13	"intern* physician".ab,kw,ti	137
14	"residen* doctor".ab,kw,ti	152
15	"residen* physician".ab,kw,ti	1,301
16	"residen* year".ab,kw,ti	465
17	"foundat* doctor".ab,kw,ti	62
18	"foundat* physician".ab,kw,ti	9
19	"foundat* year".ab,kw,ti	796
20	"FY1".ab,kw,ti	343
21	"FY2".ab,kw,ti	176
22	"house officer".ab,kw,ti	1,364
23	"graduat* doctor".ab,kw,ti	46
24	"graduat* physician".ab,kw,ti	77
25	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24	83,814
26	exp suicide/	138,511
27	exp depression/	546,373
28	exp anxiety/	319,908
29	exp distress/	51,907
30	"self$harm".ab,kw,ti	141
31	"self$destruct".ab,kw,ti	68
32	"self$mutilat*".ab,kw,ti	76
33	"self$inflict*".ab,kw,ti	48
34	"self$injur*".ab,kw,ti	109
35	"self$poison*".ab,kw,ti	66
36	"auto$mutilat*".ab,kw,ti	314
37	"intent* injur*".ab,kw,ti	1,331
38	"intent* poison*".ab,kw,ti	507
39	"parasuicide*".ab,kw,ti	2,232
40	26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39	937,362
41	25 and 40	1,746
42	Limit 41 to humans	1,634
43	Limit 42 to English language	1,505

Supplementary Figure 1:
PRISMA study inclusion flow diagram.



Supplementary Table 2:
Methodological characteristics of the 39 studies included in this review by study design.

Author	Year	Trial Registration Number	Country	Methods	Participants	Intervention	Control	Duration of Intervention	Outcomes assessed
Randomised Controlled Trials (RCTs)									
Ball	**2002**	N/A	USA	Randomised controlled trial comparing a single session of a psycho-educational intervention with no treatment for the prevention of depression in medical students in their first year at one university.	A total of 29 young adult (mean age: 24.0 years; SD: 3.4 years) male and female (40.7% female) undergraduate medical students in their first year at one university.	A single session (approximately 1.5 hours) of a self-awareness and self-care intervention focusing on: providing written psychoeducation about depression, sleepiness, and alcohol use.	Information on content of the control condition not clearly reported.	Not clearly reported	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** Beck Depression Inventory (BDI-II). **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** alcohol use problems (AUDIT), quality of life (Medical Education Quality of Life Questionnaire), sleep problems (Epworth Sleepiness Scale), and physical health (Health Habits Survey).
Danilewitz	**2016**	N/A	Canada	Randomised controlled trial comparing an eight-week, adapted version of a manualised mindfulness-based stress reduction program with wait-list for the prevention of stress in medical students in their first and second years at one university.	A total of 30 young adult (mean age: not clearly reported) male and female (73.3% female) undergraduate medical students in their first year at one university.	Eight sessions (1.5 hours) of a manualised mindfulness-based stress reduction program consisting of: body scan, mindfulness, breathing exercises, stress management, yoga, loving kindness, meditation, and mindful communication. Students also received weekly homework meditation exercises. Sessions were co-facilitated by a trained medical student peer worker with a psychologist with training in mindfulness meditation.	No intervention (wait-list).	Not clearly reported, assume eight weeks	**Anxiety:** Depression, Anxiety, Stress Scale (DASS). **Burnout:** not assessed. **Depression:** DASS. **Stress:** DASS. **Suicidality:** not assessed. **Other outcomes:** altruistic behaviours (Adapted Altruism Scale), empathy (Jefferson Scale of Physician Empathy-Student Version), mindfulness skills (Five Facets of Mindfulness Questionnaire), satisfaction with the program (idiosyncratic five-point scale ranging from disagree [1] to agree [5]), and self-compassion (Self-Compassion Scale).
de Vibe	**2013**	NCT00892138	Norway	Parallel group randomised controlled trial comparing a seven-week course of a manualised mindfulness-based stress reduction program with no treatment for the prevention of burnout and stress in medical and psychology students undertaking their second or third years at two universities.	A total of 288 young adult (mean age: 23.0, SD: not reported) male and female (76.0% were female) medical and psychology students in their second and third years at two universities. No specific exclusion criteria were reported.	Consisted of seven weekly sessions (approximately 1.5 hours; session seven was a six hour, day-long session), of a manualised mindfulness-based stress reduction program consisting of: mindfulness, didactic teaching on mindfulness, stress management, and mindful communication, and group sessions to facilitate reflection. Sessions were facilitated by trained instructors (three male, three female). Participants also	Consisted of standard university courses with no intervention	7 Weeks	**Anxiety:** not assessed. **Burnout:** Maslach Burnout Inventory. **Depression:** not assessed. **Stress:** Perceived Medical School Stress Scale. **Suicidality:** not assessed. **Other outcomes:** distress (General Health Questionnaire, 12 item), well-being (Subjective Well Being Scale, 4 item version), mindfulness (Five Facet Mindfulness Questionnaire), and compliance (idiosyncratic sale based on self-reported home-based mindfulness practice)

						received weekly homework exercises to consolidate skills learned.			
Holtzworth-Munroe	**1985**	N/A	USA	Parallel group randomised controlled trial comparing a six sessions of a manualised group-based stress management with no treatment for the prevention of stress in first and second year medical students at one university.	A total of 40 young adult (mean age: not reported) male and female (proportion female not reported) undergraduate medical students in their first and second years at one university. No specific exclusion criteria were reported.	Consisted of six, hour long sessions of a manualised group-based stress management program, consisting of: cognitive restructuring, progressive muscle relaxation, and meditation. The session was facilitated by a doctoral student in clinical psychology	Consisted of standard university courses with no intervention.	6 Weeks	**Anxiety:** Spielberger Trait Anxiety Inventory. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** awareness of stress and coping ability (idiosyncratic scale).
Kiecolt-Glaser	**2011**	N/A	USA	Parallel group, double-blind, placebo controlled, randomised controlled trial comparing daily omega three fatty acid supplementation with placebo for the prevention of anxiety in first and second year medical students.	A total of 68 young adult (mean age: 23.6, SD: 1.9 years) male and female (44.1% were female) medical students in their first or second years at one university. Exclusion criteria were: high fish intake, fish oil or flaxseed supplements, smoking, alcohol or drug use, any chronic illness with an inflammatory or endocrine component, lipid-altering drugs, beta blockers, steroids, ACE-inhibitors, regular use of non-steroidal anti-inflammatories, and use of psychoactive drugs or mood altering medications.	Capsules containing 2.496 grams of an omega three fatty acid supplement. Each capsule contained 2085mg of eicosapentaenoic acid and 348mg of docosahexaenoic acid. Participants were instructed to take one capsule per day.	Capsules containing a mixture palm, olive, soy, canola, and coco butter oils. Participants were instructed to take one capsule per day.	12 Weeks	**Anxiety:** Beck Anxiety Inventory (BAI). **Burnout:** not assessed. **Depression:** Center for Epidemiological Studies Depression scale (CES-D). **Stress:** not assessed. **Suicidality:** not assessed.
McGrady	**2012**	N/A	USA	Parallel group randomised controlled trial comparing a four-month course of stress management and relaxation program with wait-list control for the prevention of anxiety and depression in medical students in their first year.	A total of 449 young adult (mean age: 23.4, SD: not provided) male and female (51.5% were female) medical students in their first year at one university. No specific exclusion criteria were reported.	Consisted of eight bi-monthly sessions (approximately 45 minutes) of a stress management program facilitated by an experiencing psychologist, counsellor, or physician. Program components focused on deep breathing, progressive relaxation, guided imagery, cognitive restructuring (termed 'survival thinking' in this trial), mindfulness, meditation, nutrition, coping, managing fatigue and anxiety, and balancing study and life.	No specific intervention was received.	4 Months	**Anxiety:** BAI. **Burnout:** not assessed. **Depression:** BDI-II. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** social adjustment (Social Readjustment Rating Scale-Revised [SRRS-R]), and an idiosyncratic scale developed to assess frequency of acute illness.
Moir	**2016**	N/A	New Zealand	Parallel group randomised controlled trial comparing a six-month course of weekly mindfulness sessions delivered by	A total of 275 young adult (mean age: 20.9, SD: not provided) male and female (53.0% were female) undergraduate medical	Consisted of weekly sessions of mindfulness based on the manualised Oxford University Peer Support Program with specific sessions on reflective	Consisted of treatment as usual, including usual health and mental health resources (e.g., the	6 Months	**Anxiety:** Generalized Anxiety Disorder Questionnaire (GAD-7). **Burnout:** not assessed. **Depression:** PHQ-9. **Stress:** not assessed.

				trained peer leaders with treatment as usual for the prevention of depression and anxiety in undergraduate medical students (year 2 and 3).	students in years 2 or 3 at one university. No specific exclusion criteria were reported.	listening, barriers and enables of help-seeking behaviour, enabling decision-making, values and judgements, identifying and labelling feelings, cultural competency, safety planning, developing limit-setting boundaries, assertiveness, crisis planning, suicide awareness, prevention, resources, and referral options. Two social events were also organised over the course of the program.	university counselling service and student medical clinic). Participants in the control group were asked not to undertake mindfulness or two attend the two social events, however, they were not prevented from doing so.		**Suicidality:** not assessed. **Other outcomes:** quality of life (Linear Analogue Self-Assessment), resilience (Wagnild Resilience Scale), academic self-concept (Perceived Competence Scale), academic motivation (Motivated Strategies for Learning Questionnaire)
Phang	**2015**	N/A	Malaysia	Parallel group randomised controlled trial comparing an adapted five-week mindfulness-based stress reduction program with wait-list for the prevention of stress in medical students in their first, second, or third years at one university.	A total of 75 young adult (mean age: 21.0; SD: 1.1 years) male and female (76.0% were female) undergraduate medical students in their first, second, or third years at one university. No specific exclusion criteria were reported.	Consisted of an adapted five-week program of a manualised mindfulness-based stress reduction program consisting of five weekly sessions (two hours) of: gratitude and cultivation of loving-kindness, stress reduction, cultivating the ability to pay attention to the present moment, and progressive muscle relaxation exercises. Sessions were facilitated by an experienced psychiatrist with more than 10 years' experience.	Participants randomised to the control group received the intervention materials in full on DVD at the conclusion of the six-month follow-up period.	5 Weeks	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** Perceived Stress Scale (PSS). **Suicidality:** not assessed. **Other outcomes:** compliance (idiosyncratic five-point Likert scale ranging from none at all [1] to almost every day [5]), general psychiatric symptomatology (General Health Questionnaire), mindfulness skills (Mindful Attention Awareness Scale), self-efficacy (General Self-Efficacy Scale).
Shapiro	**1998**	N/A	USA	Parallel group randomised controlled trial comparing an eight-week course of mindfulness with wait-list for the prevention of stress in medical and premedical students (years 1 and 2).	A total of 73 young adult (mean age: not provided, SD: not provided) male and female (56.2% were female) medical or premedical students in years 1 or 2 at one university. No specific exclusion criteria were reported.	Consisted of seven sessions (approximately 2.5 hours) of a mindfulness program based on the manualised Stress Reduction and Relaxation Program with specific sessions on meditation, attention on bodily sensations, hatha yoga, loving kindness, and forgiveness. Additional exercises in mindful listening, empathy, social support, as well as the completion of daily journals were also added.	No specific intervention was received.	8 Weeks	**Anxiety:** State-Trait Anxiety Inventory (STAI). Note, data were estimated from Figure 1a, p. 589. **Depression:** sub-scale 4 of the SCL-90-R. Note, data were estimated from Figure 1c, p. 589. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** empathy (Empathy Construct Rating Scale [ECRS]), psychological distress (Hopkins Symptom Checklist-90-Revised [SCL-90-R]), spirituality (Index of Core Spiritual Experiences-INSPIRIT), and compliance with meditation practice (from daily diary entries).
Velayudhan	**2010**	N/A	India	Parallel group randomised controlled trial comparing a course of counselling and relaxation (number of sessions, duration of sessions, and length of intervention period not reported) with no	A total of 120 young adult (mean age: not reported) male and female (50.0% were female) undergraduate medical students at one private medical college. No specific exclusion criteria were reported.	Consisted of a course of counselling and relaxation. The number of sessions, duration of sessions, and length of the intervention period were not reported.	No specific intervention was received.	Not clearly reported	**Anxiety:** BAI. **Burnout:** not assessed. **Depression:** BDI-II **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** not assessed.

				treatment for the prevention of stress in undergraduate medical students at one private medical college.					
Warnecke	**2011**	N/A	Australia	Parallel group randomised controlled trial comparing an eight-week CD-based, guided mindfulness program with wait-list for the prevention of stress in medical students undertaking their final two years at one university.	A total of 66 young adult (mean age: 23.9, SD: 3.2 years) male and female (64.6% were female) undergraduate medical students in their final two years at one university. No specific exclusion criteria were reported.	Consisted of an eight-week CD-based guided meditation program with each session lasting approximately 30 minutes. Participants were also required keep a daily diary to enable monitoring of adherence.	No specific treatment was received.	8 Weeks	**Anxiety:** DASS. **Burnout:** not assessed. **Depression:** DASS **Stress:** DASS and PSS. **Suicidality:** not assessed. **Other outcomes:** not assessed.
Whitehouse	**1996**	N/A	USA	Parallel group randomised controlled trial comparing a 19-week self-hypnosis and relaxation program with no treatment for the prevention of stress in first-year medical students.	A total of 35 young adult (mean age: not reported) male and female (60.0% were female) medical students in years 1 at one university. No specific exclusion criteria were reported.	consisted of 14 sessions (approximately 90 minutes) of self-hypnosis facilitated by trained senior psychiatrists with experience in the clinical use of hypnosis and relaxation techniques. Session one consisted of an assessment of each participant's hypnotic ability, whilst sessions two to 14 focused on practicing skills. Participants were also encouraged to engage in self-hypnosis for at least 15 minutes per day individually, and completed daily diaries.	No specific intervention was received. However, participants in this group were required to complete daily diaries enquiring about extent and quality of sleep, mood, use of medications, and any problems experienced.	19 Weeks	**Anxiety:** BSI. **Burnout:** not assessed. **Depression:** BSI. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** various immunological assays as well as somatization, obsessive-compulsive, interpersonal sensitivity, hostility, phobic anxiety, paranoia, psychoticism, general severity, positive symptoms, and psychological distress (all measured by the BSI).
Yusoff	**2015**	N/A	Malaysia	Parallel group randomised controlled trial comparing a four-hour session of a psychoeducational and problem-oriented stress management program (DEAL) with wait-list for the prevention of stress in undergraduate medical students (years 1 to 5).	A total of 171 young adult (mean age: not reported) male and female (57.9% were female) undergraduate medical students in years 1 to 5 at one university. No specific exclusion criteria were reported.	Consisted of a single four-hour workshop of a psychoeducational and problem-oriented stress management program (DEAL). Section one (one hour) focused on psychoeducation on stress, stressors, and coping mechanisms of relevance to medical students. Session two (one hour) focused on learning problem-solving techniques ad coping strategies to manage stress. Session three (one hour) focused on group-based exercises to practice these strategies, and session four (one hour) focused on sharing experiences and concluding the program.	No specific intervention was received.	4 Hours	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** BDI-II. **Stress:** total scores on the MSSQ-20, as well as scores on six domains: (1) Academic-Related Stressors; (2) Intrapersonal and Interpersonal Stressors; (3) Teaching and Learning-Related Stressors; (4) Social-Related Stressors; (5) Drive and Desire-Related Stressors, and; (6) Group Activities-Related Stressors. **Suicidality:** not assessed. **Other outcomes:** not assessed.

Non RCTs									
Chen	**2016**	N/A	USA	Non-randomised controlled trial comparing an 11-week course of mind-body skills training with no treatment for the prevention of stress in medical students in their first year.	A total of 20 young adult (mean age: not reported) male and female (proportion female not reported) medical students in their first year at one university. No specific exclusion criteria were reported.	Consisted of 11 sessions of a mind-body skills course. Core components included: meditation, guided imagery, and journal writing facilitated by two trained facilitators. The duration of these sessions was not reported.	No specific intervention was received.	No specific information reported	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** PHQ-9. **Stress:** PSS. **Suicidality:** not assessed. **Other outcomes:** empathy (Jefferson Scale of Physician Empathy-Student Version [JSPE-S])
Finkelstein	**2007**	N/A	USA	Non-randomised controlled trial comparing a 10-week course of mind-body skills training with no treatment for the prevention of anxiety and stress in medical students in their first year.	A total of 76 young adult (mean age: 25.0, SD: 2.3 years) male and female (51.3% were female) medical students in their first year at one university. No specific exclusion criteria were reported.	Consisted of 10 weekly sessions (approximately two hours) of a manualised mind-body skills course. Core components included: psychoeducation on stress response, meditation, imagery, advice on exercise and nutrition, and spirituality using both a small- and large-group format.	No specific intervention was received.	10 Weeks	**Anxiety:** subscale from the SCL-90. **Burnout:** not assessed. **Depression:** two-item Depression Index (DI-2). **Stress:** Perceived Stress of Medical School (PSMS). **Suicidality:** not assessed. **Other outcomes:** Profile of Mood States (POMS).
Kelly	**1982**	N/A	USA	Non-randomised controlled trial comparing a six session program of stress management training with wait-list for the prevention of stress in medical students in their first, second or third years.	A total of 48 young adult (mean age: not reported) male and female (33.0% were female) medical students in their first, second, or third years at one university. A minority of participants (20.0%) were residents and nurses. No specific exclusion criteria were reported.	six sessions (of between 60-90 minutes) of a stress management training program consisting of: didactic lectures on stress reduction techniques, group-based discussions, relaxation techniques, priority-setting, schedule-planning, focusing on engaging in leisure activities, exercise, cognitive modification skills training, and homework assignments to practice the stress reduction technique introduced.	A single, "almost identical" (p. 95) seminar (duration: not specified) of stress management training was received following the conclusion of the post-test period.	3 Weeks	**Anxiety:** Speilberger State-Trait Anxiety Inventory. Burnout: not assessed. **Depression:** Zung Self-Rating Depression Scale, using a cut off score of 60 or greater, indicating moderate to severe depression. **Stress:** Jenkins Activity Schedule, Type A subscale. **Suicidality:** not assessed. **Other outcomes:** compliance (daily activity log), knowledge about stress (Stress Knowledge Inventory, 26 item), situations that cause the most stress (idiosyncratic scale ranging from zero [not stressful] to 100 [extremely stressful]).
Kraemer	**2015**	N/A	USA	Non-randomised controlled trial comparing an 11-week course of mind-body skills training with no treatment for the prevention of stress in medical students.	A total of 22 young adult (mean age: 23.9 years; SD: not reported) male and female (68.2% were female) medical students in their first year at one university. No specific exclusion criteria were reported.	Consisted of 11 weekly sessions (90 minutes) of a mind-body skills course.	No specific intervention was received.	11 Weeks	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** not assessed. S**tress**: not assessed. **Suicidality:** not assessed. **Other outcomes:** distress (measure used not specified).
Michie	**1994**	N/A	USA	Non-randomised controlled trial comparing a three-week stress management program with wait-list for the prevention of stress in	A total of 302 young adult (mean age: not reported) male and female (proportion female not reported) medical students in their second year at one university. No specific	Three weekly sessions (approximately 2 hours) of a stress management program consisting of: psychoeducation on models of stress, stress management techniques, work	no specific intervention was received. Instead, "[a]ttendees acted as their own 'waiting list' control	3 Weeks	**Anxiety:** Speilberger State-Trait Anxiety Inventory. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** not assessed. **Suicidality:** not assessed.

				medical students in their first clinical year at one university.	exclusion criteria were reported.	management skills, assertiveness and communication skills, relaxation techniques, cognitive approaches, overcoming barriers, and longer-term planning.	group by asking those who had signed up for the course to complete the evaluation questionnaire at two time points whilst they waited for their course" (p. 529).		
Rosenzweig	**2003**	N/A	USA	Non-randomised controlled trial comparing a 10-week course of mind-body skills training with attention placebo for the prevention of anxiety and stress in medical students in their second year.	A total of 302 young adult (mean age: not reported) male and female (proportion female not reported) medical students in their second year at one university. No specific exclusion criteria were reported.	Consisted of 10 weekly sessions (approximately 90 minutes) of a manualised mind-body skills course. Core components included: body scanning techniques, meditation, guided imagery exercises, breathing exercises, and Hatha Yoga delivered in a group-based format.	Sessions of complementary and alternative medicine. Core components included: didactic exercises, demonstrations, group exercises, and presentations. Both the number of sessions and the duration of these sessions was not reported.	10 Weeks	**Anxiety:** subscale from the POMS. **Burnout:** not assessed. **Depression:** subscale from the POMS. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** anger, vigour, fatigue, and confusion subscales of the POMS as well as the overall POMS score.
Historically Controlled Studies									
Holm	**2010**	N/A	Norway	Historically controlled, retrospective study comparing 12 weekly group-based self-development program to prevent stress in undergraduate medical students at one university.	A total of 165 young adult (mean age: 23.6 years; SD: 3.4 years) male and female (59.4% were female) undergraduate medical students in their third year at one university. No specific exclusion criteria were reported.	12 weekly sessions (90 minute) of a group-based self-development program consisting of: identification of positive resources in the students' lives, building self-esteem, and personal insight. Sessions also focused on helping students to identify their typical patterns of relationships and how these may be restricting their ability to relate with others. Groups sessions were facilitated by qualified general practitioners and numbered between eight and 10 participants per group.	No specific intervention was received.	Not clearly reported, presume one year.	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** Perceived Medical School Stress (PMSS). Note, data were estimated from Fig.1, p.5. **Suicidality:** not assessed. **Other outcomes:** general mental health functioning (SCL-5).
Melo-Carillo	**2012**	N/A	Mexico	Historically controlled study comparing a single session (duration not reported) of a psychoeducation intervention to prevent depression in	A total of 1958 young adult (mean age: not provided, SD: not provided) male and female (proportion female not reported) undergraduate medical students at one	A single, annual session (duration not reported) of psychoeducation focused on providing information on common mental disorders among medical students, information on diagnosis, complications,	No specific intervention was received.	2 years	**Anxiety:** not assessed. **Burnout**: not assessed. **Depression:** BDI-II. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** not assessed.

				undergraduate medical students at one university.	university teaching hospital. No specific exclusion criteria were reported.	and treatment as well as the formation of a mental health support group, facilitated by six qualified psychiatrists (four male, two female), to provide confidential treatment to all students with mental health problems.			
Thompson	**2010**	N/A	USA	Historically controlled study comparing a package of interventions designed to prevent depression and suicidality in medical students at one university.	A total of 102 young adult (mean age: not provided, SD: not provided) male and female (proportion female was not reported) medical in years 1 to 3 at one university. No specific exclusion criteria were reported.	Consisted of one (30 minute) session of gate-keeper training for all faculty staff consisting of psychoeducation around student risk factors for depression and suicidality (e.g., sleep deprivation, isolation, and academic difficulties), symptoms of depression and suicidality, didactic exercises, and one-on-one coaching facilitated by psychiatric faculty staff. Students also received a well-being handbook which provided psychoeducation on common stressors, recognizing depression, advice on stress management, self-assessment exercises, along with a single, hour long, didactic discussion on managing stress during critical times of the year (e.g., examinations). All incoming first-year students could also participate in a well-being program with the following core components: discovering and nurturing wholeness, sharing grief and honouring loss, allowing awe in medicine, and the care of the soul.	No specific intervention was received	Not clearly specified, presume year-long.	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** cut-off score of greater than 21 on the CES-D. **Stress:** not assessed. **Suicidality:** cut-off score of one or greater on the PRIME-MD. **Other outcomes:** not assessed.
Uncontrolled Studies									
Bansal	**2013**	N/A	India	Uncontrolled longitudinal study comparing a one-month program of a daily yoga program for the prevention of stress in medical undergraduate students at one university undertaking a clinical placement in community medicine.	A total of 82 young adult (mean age: not reported; age range: 18-32 years) male and female (55.6% were female) undergraduate medical students at one university undertaking a clinical placement in community medicine.	Daily sessions (45 minutes) of a yoga program consisting of: a series of seven postures (asana), followed by breathing exercises (pranayama), and meditation. Sessions were facilitated by a trained yoga instructor.	No specific information	1 Month	**Anxiety:** GHQ-28, subscale. **Burnout:** not assessed. **Depression:** GHQ-28, subscale. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** general functioning (GHQ-28), satisfaction (qualitative comments), somatic symptoms (GHQ-28, subscale), and social dysfunction (GHQ-28, subscale).

Bond	**2013**	N/A	USA	Uncontrolled longitudinal study comparing an 11 session program of a manualised mind-body skills training intervention for the prevention of stress in medical students in their first and second years at one university.	A total of 27 young adult (mean age: not reported) male and female (57.6% were female) medical students in their first and second years at one university.	11 sessions (1.5 hours) of a manualised mind-body skills program consisting of: deep breathing, meditation, yoga, stress management, and a 30-minute didactic lecture on neuroscience and mind-body medicine. Participants also received peer-reviewed psychoeducation on mind-body medicine and undertook weekly homework exercises.	No specific information	11 Weeks	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** Cohen's Perceived Stress Scale. **Suicidality:** not assessed. **Other outcomes:** empathy (Jefferson Scale of Physician Empathy), self-regulation abilities (Self-Regulation Questionnaire), self-compassion (Self-Compassion Scale).
Bughi	**2006**	N/A	USA	Uncontrolled longitudinal study of a single psychoeducational lecture on stress management for the prevention of stress in medical students from two universities undertaking a one-month clinical rotation at one Diabetes/Endocrine Service at a tertiary referral service.	A total of 32 young adult (mean age: not reported) male and female (proportion female not reported) medical students in their third and fourth years at two universities who were undertaking a one-month clinical rotation at one Diabetes/Endocrine Service at a tertiary referral service.	A single psychoeducational lecture (duration not specified) consisting of: a review of the epidemiology of the stress response, information on the psychological and medical complications of stress, stress inoculation, deep diaphragmatic breathing (prolonged expiration or deep yoga breathing), self-control relaxation, and walking meditation.	No specific information	Not clearly reported, described as brief	**Anxiety:** General Well-Being Scale, subscale (lower scores indicative of higher anxiety). **Burnout:** not assessed. **Depression:** General Well-Being Scale, subscale (lower scores indicative of higher depression). **Stress:** General Well-Being Scale subscale (lower scores indicative of higher stress). For this study, a cut-off value of 60 or lower was interpreted as indicative of those with severe stress. **Suicidality:** not assessed. **Other outcomes:** not assessed.
Dutton	**2013**	N/A	USA	Uncontrolled, prospective, longitudinal study comparing an 11 session program of a manualised mind-body skills training intervention for the prevention of stress in medical students in their first year at one university.	A total of 59 young adult (mean age: not reported) male and female (57.6% were female) medical students in their first year at one university.	11 sessions (duration not reported) of a manualised mind-body skills program consisting of: self-awareness, relaxation, meditation, guided imagery, and biofeedback skills.	No specific information	11 Weeks	**Anxiety:** State-Trait Anxiety Index. **Burnout:** not assessed. **Depression:** subscale, Brief Symptom Index. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** distress (Brief Symptom Index), physical health (Brief Symptom Index), mindfulness skills (scale not clearly specified).
Gallagher	**2005**	N/A	USA	Uncontrolled longitudinal study comparing the establishment of a dedicated hotline for the prevention of anxiety in medical students in their third year at one university.	A total of 86 young adult (mean age: not reported) male and female (57.0% were female) medical students in their third year at one university.	A dedicated hotline, manned by a Masters-trained counsellor, designed to provide free, confidential advice to callers 24 hours a day, seven days a week. All students also received a laminated pocket card with further details and the hotline telephone number.	No specific information	1 Year	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** 10-point idiosyncratic scale ranging from one (least stressful year of life) to 10 (most stressful year of life). **Suicidality:** not assessed. **Other outcomes:** awareness of hotline (open-ended question), reassurance about existence of hotline (five-point idiosyncratic scale ranging from one [not at all reassuring] to five [very reassuring]), and importance of continuing the hotline (five-point idiosyncratic scale ranging from one [not at all important] to five [very important]).
Garneau	**2013**	N/A	USA	Uncontrolled longitudinal study comparing an adapted group-based	A total of 58 young adult (mean age: 26.0 years; SD: not reported) male and	twice weekly sessions of 2.5 hours each of a group-based mindfulness program based on	No specific information	4 Weeks	**Anxiety:** not assessed.

				mindfulness program for the prevention of stress in medical students in their fourth year at one university	female (74.0% were female) medical students in their fourth year at one university.	a manualised mindfulness-based program, consisting of: mindful communication, body-scan techniques, yoga, sitting meditation, imagery meditation, and breathing exercises. Participants also attended a single day six-hour long retreat day at the end of the course. Sessions were facilitated by two trained PhD-level psychologists and a palliative care physician. Participants also received a home practice manual and three CDs of materials to facilitate practice at home.			**Burnout:** Maslach Burnout Inventory, Human Services Survey (used Emotional Exhaustion subscale) **Depression:** BDI-II. **Stress:** Perceived Stress Scale-10. **Suicidality:** not assessed. **Other outcomes:** self-compassion (Self-Compassion Scale), mindfulness skills (Mindful Attention Awareness Scale), and wellness (Scales of Psychological Well-Being).
Greenson	**2015**	N/A	USA	Uncontrolled, prospective, longitudinal study comparing four-sessions of a mind-body program for the prevention of stress in medical students in their first, second, third, or fourth years at one university. Two post-graduate students studying at the same university were also included.	A total of 44 young adult (mean age not reported) male and female (65.9% were female) medical students in their first, second, third, fourth, or studying a combined Master/Doctorate degree at one university. No specific exclusion criteria were reported.	Consisted of an adapted, four week, program based on a manualised mind-body skills program. Each session (1.5 hours) commenced with facilitator-led meditation, mindfulness, relaxation breathing, guided imagery, drawing, body awareness, progressive muscle relaxation, and loving-kindness. Each session concluded with a didactic discussion on the science of mind-body medicine, a period of reflection, and information on stress management. Between sessions, students were encouraged to engage in further mind-body skills training (for six days per week, of up to 30 minutes each session; mean duration: 12 minutes) facilitated by a recorded guided meditation program. Students were also asked to maintain a weekly diary.	No specific information	4 Weeks	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression**: not assessed. **Stress:** PSS. **Suicidality:** not assessed. **Other outcomes:** mindfulness skills (Cognitive and Affective Mindfulness Scale-Revised).
Hassed	**2009**	N/A	Australia	Uncontrolled longitudinal study comparing an eight session mindfulness-based intervention for the prevention of stress in undergraduate medical	A total of 148 young adult (mean age: 18.8 years; SD: 1.1 years) male and female (57.4% were female) undergraduate medical students in their first year at one university.	Eight, one hour sessions of psychoeducation on the links between mental and physical health, mind-body practice, behaviour change strategies, and mindfulness therapies. Students also received six two-	No specific information	6 Weeks	**Anxiety:** subscale, SCL-90-R. **Burnout:** not assessed. **Depression:** subscale, SCL-90-R. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** hostility (subscale, SCL-90-R), physical health (subscale, WHOQOL-BREF),

				students in their first year at one university.		hour sessions of a manualised Stress Release Program facilitated by trained tutors. Program components consisted of: mindfulness meditation, and mindfulness-based cognitive tasks. Students also received a single hour-long session of a mindfulness program (ESSENCE Lifestyle) with specific components on: reflection, mindfulness, spirituality, coping, physical activity, nutrition, and social support. Weekly mindfulness homework exercises were also provided, and students were encouraged to record progress in a personal journal.			psychological health (WHOQOL-BREF). Also measured overall Global Severity Index.
Kötter	**2016**	N/A	Germany	Uncontrolled longitudinal study comparing a two-session progressive muscle relaxation training intervention for the prevention of stress in medical undergraduate students in their first and second years at one university.	A total of 122 young adult (mean age: 21.3 years; SD: 2.9 years) male and female (68.9% were female) undergraduate medical students in their first and second years at one university.	Two sessions (of 45 minutes duration each) of a progressive muscle relaxation intervention.	No specific information	Unclear "The first module...[was] followed by a refresher module several weeks later" (p. 3).	**Anxiety:** State-Trait Anxiety Index. **Burnout:** not assessed. **Depression:** subscale, Brief Symptom Index. **Stress:** Perceived Medical School Stress, German language translation. **Suicidality:** not assessed. **Other outcomes:** self-rated general health (single item: How would you describe your health in general, rated on a five-point scale from very good to very poor), mental health (Hospital Anxiety and Depression Scale, German language translation), professional commitment (Arbeitsbezogene Verhaltens-und Erlebensmuster Scale), resistance to stress (Arbeitsbezogene Verhaltens-und Erlebensmuster Scale), and emotional well-being at work (Arbeitsbezogene Verhaltens-und Erlebensmuster Scale).
Mercer	**2010**	N/A	USA	Uncontrolled longitudinal study comparing two sessions of a visual journaling intervention for the prevention of stress in medical students in their first year at one university.	A total of five young adult (mean age: not reported) medical students in their first year at one university and five (mean age: not reported) adult faculty members (proportion female not reported) at one university medical school.	Two sessions (duration not reported) of a visual journaling intervention. Sessions commenced with guided imagery visualisation focusing on breathing exercises and the identification of stress-producing emotions. In the first, participants drew images of their stressors and were provided with a series of exploration questions designed to help them understand the source(s) of their stress and what the imagery contained in	No specific information	2 Weeks	**Anxiety:** State-Trait Anxiety Index. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** negative affect (PANAS), positive affect (PANAS).

						their image may be conveying. In the second session, participants were guided to envision a less stressful image which was followed by another list of self-exploration questions.			
Simard	**2009**	N/A	Canada	Uncontrolled longitudinal study comparing a 16 week Hatha yoga program for the prevention of stress in medical students in their first year at one university.	A total of 16 young adult (mean age: 22.0 years; SD: 2.2 years) male and female (56.0% were female) medical students in their first year at one university.	16 weekly sessions (duration not reported) of a Hatha yoga program delivered by a certified Kripalu yoga teacher with over six years' experience.	No specific information	16 Weeks	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** CES-D. **Stress**: PSS. **Suicidality:** not assessed. **Other outcomes:** general health functioning (General Health Questionnaire, 12 item).
Wild	**2014**	N/A	Germany	Uncontrolled longitudinal study comparing a semester long program (number of weeks unclear) for the prevention of burnout in medical undergraduate students in their third and fourth years at one university.	A total of 42 young adult (mean age: 24.9 years; SD: 4.4 years) male and female (88.1% were female) undergraduate medical students in their third and fourth years at one university.	Weekly sessions (of two hours' duration each) of a autogenic training and progressive muscle relaxation program. Students also were encouraged to participate in twice daily independent practice sessions to review course components and elaborate on the autogenic training and progressive relaxation techniques taught.	No specific information	Unclear	**Anxiety:** STAI, German language version (STAI-G). **Burnout:** Burnout Symptom Scale, version two (BOS-II) (cognitive subscale). **Depression:** BDI-II. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** sense of coherence (Sense of Coherence Scale), and work related experiences (Arbetsbezogenes Verhaltens und Erlebensmuster).
Curriculum-Based Interventions									
Al-Faris	**2014**	N/A	Saudi Arabia	Historically controlled, retrospective study comparing a redesigned curriculum focusing on systemically addressing sources of stress with a traditional curriculum on depression in undergraduate medical students in their first and second years at one university.	A total of 707 young adult (age range: 18-21 years) male and female (34.5% female) undergraduate medical students in their first and second years at one university.	Consisted of a number of changes to the curriculum to systemically address sources of stress, including: restructuring traditional subjects into new integrated subjects organised by bodily system and adopted a problem-based learning strategy consisting of interactive lectures and self-directed learning delivered in a small group format. New subjects designed to provide students with early, hands-on contact with the health care system (e.g., ambulatory care training), and increases to the amount of time students spent on internships were also instituted.	Consisted of a traditional curriculum which focused on didactic lectures as the main teaching format. Students were taught traditional subjects (e.g., anatomy, physiology) with little integration of the various bodily systems, and undertook few electives. In addition, grading was largely based on summative assessment (i.e., grades ranged from A+ to Fail)	Not clearly reported	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** BDI-II. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** perceived quality of the educational environment (Dundee Ready Educational Environment Measure; DREEM).
Camp	**1994**	N/A	USA	Non-randomised controlled trial comparing a problem-based learning curriculum with a	A total of 232 young adult (mean age: not reported) male and female (proportion	consisted of a student-directed, problem-based learning curriculum which emphasises self-directed	Consisted of a lecture-based learning curriculum which emphases didactic	1 Year	**Anxiety:** not assessed. **Burnout:** not assessed.

				traditional, lecture-based learning curriculum on depression in medical students in their first year.	female not reported) medical students in their first year at one university. No specific exclusion criteria were reported.	learning using problem-solving strategies in a small group format facilitated by two faculty members. Students therefore have frequent one-on-one contact with faculty members throughout the year. Assessment also emphases understanding and took the format of an essay and/or performance-based examination every 10 weeks	teaching, typically in the format of a discipline-specific lectures, delivered to a large group of students. There is therefore limited one-on-one contact with faculty members. Assessment emphases the use of infrequent, multiple-choice examinations.		**Depression:** Zung Self-Rating Depression Scale, using a cut off score of 60 or greater, indicating moderate to severe depression. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** self-actualisation (California Psychological Inventory; CPI).
Moffat	**2004**	N/A	UK	Uncontrolled longitudinal study comparing a redesigned curriculum focusing on systemically addressing sources of stress in first year medical undergraduate students at one university.	A total of 275 young adult (mean age: 18.7; age range: 17.2-29.0 years) male and female (54.9% were female) undergraduate medical students at one university in their first year.	introduction of a problem-based curriculum which consisted of a number of changes to the curriculum to systemically address sources of stress, including: a reduction in the number of lectures in favour of group-based problem-oriented tutorial sessions with around half of the timetable dedicated to self-directed personal study.	No specific information	1 Year	**Anxiety:** not assessed. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** coping style (Brief COPE), exposure to stressors (idiosyncratic list of 59 potential stressors), and psychological morbidity (GHQ-12).
Slavin	**2014**	N/A	USA	Historically controlled study comparing a redesigned curriculum focusing on systemically addressing sources of stress with a traditional curriculum on anxiety and depression in medical students in their first and second years at one university.	Young adult (mean age: not reported) male and female (proportion female not reported) medical students in their first and second years at one university. No specific exclusion criteria were reported.	Consisted of a number of changes to the curriculum to systemically address sources of stress, including: changing the grading system to pass/fail grading system, elimination of norm-referenced grading systems, a reduction in contact hours, the introduction of fewer, longer-term electives (of one half-day session every two weeks for a total of 12 days per academic year), the establishment of five learning communities (service and advocacy, research, global health, wellness, and medical education) composed of students and faculty members, the introduction of a compulsory resiliency and mindfulness program (of six hours over one semester) focusing on mindfulness cultivation, energy management, stress reduction, cognitive restructuring,	Consisted of: an honours/near honours/pass/fail grading system, use of norm-reference grading systems for all subjects, a number of short-term electives (of one half-day session per week over a seven-week period) without any tie to any learning community, minimal student-faculty contact, and no resiliency training, mindfulness training, or formalised social events.	Not clearly reported	**Anxiety:** Speilberger State-Trait Anxiety Inventory. **Burnout:** not assessed. **Depression:** CES-D. **Stress:** PSS. **Suicidality:** not assessed. **Other outcomes:** cohesion (Perceived Cohesion Scale), satisfaction with the wellness program (Association of American Medical Colleges' Graduation Questionnaire).

						adopting optimistic explanatory styles, and fostering character strengths, and the introduction of social events.			
Zuardi	**2008**	N/A	Brazil	Non-randomised controlled trial comparing a redesigned curriculum focusing on systemically addressing sources of anxiety with a traditional curriculum on anxiety in undergraduate medical students in their first, second, and third years at two campuses of one university.	A total of 637 young adult (mean age: not reported) male and female (proportion female not reported) undergraduate medical students in their first and second years at one university.	Consisted of a number of changes to the curriculum to systemically address sources of anxiety, including: restructuring traditional subjects into new integrated subjects organised by bodily system, the introduction of new subjects designed to provide students with early, hands-on contact with the health care system, and increases to the amount of time students spent on internships	Consisted of a traditional curriculum which focused on traditional subjects (e.g., anatomy, physiology), limited contact with the health care system outside of internships, and shorter internships.	Not clearly reported	**Anxiety:** Speilberger State-Trait Anxiety Inventory, Portuguese translation. Note, data were estimated from Fig.1, p.138. **Burnout:** not assessed. **Depression:** not assessed. **Stress:** not assessed. **Suicidality:** not assessed. **Other outcomes:** not assessed.

Supplementary Table 3:
Assessment of study quality for randomised controlled trials (RCTs).

Author	Year	Adequate Sequence Generation	Allocation Concealment	Participant Blinding	Clinical Personnel Blinding	Outcome Assessor Blinding	Incomplete Outcome Data	Selective Outcome Reporting
Ball	2002	**Quote:** "a total of 29 medical students… were randomized to receive the self-awareness intervention" (p. 913). **Judgement:** no specifics on the method(s) used. **Rating:** unclear risk.	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Judgement:** available case data analyses were undertaken without imputation of missing data. No data on the number that completed follow-up assessments were reported. **Rating:** unclear risk.	**Quote:** "The Health Habits Survey assessed the prevalence [sic] of different types of sleep habits…The AUDIT [assessed] alcohol consumption and problems…the Beck Depression Inventory-II…the Medical Education Quality of Life Questionnaire… and the Epworth Sleepiness Scale [were also assessed]" (p. 912) **Judgement:** data on each of these outcomes reported on pp. 914 to 915. **Rating**: Low risk
Danilewitz	2016	**Quote:** "The study used a randomized waitlist (WL) control design" (p. e32). **Judgement:** no further specific information on the method(s) used to generate the randomisation sequence. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Quote:** "Efficacy analysis was performed on the intent-to-treat sample. For students with missing post-study assessments, we carried forward their baseline scores as an approach to imputing missing data (i.e., last observation carried forward - LOCF)" (p. e34). **Judgement:** missing data were imputed using the last observation carried forward method, which is known to introduce bias. Additionally, outcome	**Quote:** "Efficacy outcomes included the following self-report scales: The Depression Anxiety and Stress Scale...the Jefferson Scale of Physician Empathy-Student Version...The Five Facets of Mindfulness Questionnaire...the Self-Compassion Scale...the Adapted Altruism Scale...[and a] five point Likert scale, ranging from strongly disagree (1) to strongly agree (5)...[which] was used to assess student satisfaction..." (p. e33).

							data was missing for 40.0% of the intervention group (the proportion of data missing for the control group was not reported) (40.0% for the study overall). **Rating**: High risk	**Judgement:** data on each of these outcomes were reported for post-intervention in Tables 2. **Rating**: Low risk
de Vibe	2013	**Quote:** "After participants completed the T1 questionnaire, a computer program (Java-based random number generator) was used to randomly assign students either to the intervention group or to the control group" (p. 2). **Judgement**: use of a computerised, random number generator is likely to have minimised bias in the generation of the randomisation sequence. **Rating**: Low risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Quote:** "An email message sent two weeks prior to the Intervention informed the study participants of their group allocation" (pp. 2-3). **Judgement:** successful participant blinding was not attempted, likely due to the nature of the intervention. **Rating**: High risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "The head technician at the Norwegian Knowledge Centre for the Health Sciences assigned each participant an identity (ID) number which was then assigned to their online questionnaires to ensure that the data remained anonymous. Only the head technician had access to the data that showed the link between the student identities and the ID numbers, and he was not involved in the study in any other way" (p. 3). **Judgement:** outcome assessor blinding likely to have been successfully achieved. **Rating**: Low risk	**Quote:** "Completer and dropout comparisons were...examined...Data were missing from the responses of five students in the intervention group and seven in the control group respectively. The last observation carried forward method of imputation was chosen...Intention-to-treat analyses and per protocol analyses yielded very similar results and we have therefore presented only the former" (p. 4). **Judgement:** although available missing data were imputed using the last observation carried forward method, which is known to introduce bias, outcome data was missing for only 3.5% of the intervention group and 4.9% of the control group (4.2% for the study overall). **Rating**: Low risk	**Quote:** "...outcome measures were chosen that would capture the possible intervention effects on different aspects of psychological health, including mental distress, study stress, student burnout, subjective well-being, and mindfulness...Mental distress was measured using the 12-item General Health Questionnaire....Student burnout was measured using a version of the 15-item Maslach Burnout Inventory...Study stress was measured using the 13-item Perceived Medical School Stress scale...Subjective Well-Being (SWB) was measured using a 4-item version of the SWB scale....Mindfulness was measured using the Five Facet Mindfulness Questionnaire..." (pp. 3-4). **Judgement**: data on each of these outcomes were reported for the longest

								follow-up period in Tables 1 and 2. **Rating**: Low risk
Holtzworth-Munroe	1985	**Quote:** "...student volunteers were randomly assigned by the authors to a treatment group (those in the program) or to a control group..." (p. 418). **Judgement:** no further specific information on the method(s) used to generate the randomisation sequence. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Quote:** "Post-test data were available for 18 treatment group subjects and 16 control subjects...[f]ollow-up data were available for only 15 treatment group subjects and nine control subjects" (p. 418). **Judgement**: available case data analyses were undertaken without any imputation of missing data. Outcome data were missing for 25.0% of the intervention group by the final follow-up assessment and 55.0% of the control group by the final follow-up assessment (40.0% for the study overall). **Rating**: High risk	**Quote:** "The subjects completed a variety of instruments on three occasions: before random assignment (pre-test), during the week following the end of the treatment program (posttest), and after a 10-week follow-up period. These instruments consisted of the Spielberger Trait Anxiety Inventory...Additionally, at the posttest and follow-up periods, the subjects completed five rating scale evaluations of their awareness of stress and their ability to handle stress. Treatment group subjects also evaluated the usefulness of the program" (p. 417). **Judgement**: data on these outcomes reported on p. 418. **Rating**: Low risk

Kiecolt-Glaser	2011	**Quote:** "...students were randomized to n-3 or placebo using a computer generated permuted block randomization sequence, with six students per block" (p. 1727). **Judgement**: use of a computerised, random number generator is likely to have minimised bias in the generation of the randomisation sequence. **Rating**: Low risk	**Quote:** "The data manager who prepared and maintained the randomization sequence had no involvement in other aspects of the research, including data collection and biological laboratory analyses, and she was the only person who had the randomization list" (p. 1727). **Judgement:** the use of an independent researcher to allocate participants into the intervention and placebo arms ensures allocation concealment is likely to have ensured. **Rating**: Low risk	**Quote:** "OmegaBrite (Waltham, MA) supplied both the n-3 and the matching placebo; all pills were coated with a fuchsia coloring. OmegaBrite added a mild fish flavor to the placebo to help disguise any differences between the n-3...pills and the placebo, and we told participants about the fish flavoring to promote blindness." (p. 1726). **Judgement:** use of identical colouring and flavouring would have ensured participant blindness ensured. This is corroborated by "[t]he James' blinding index for participants at the end of the study was 0.55 (95% CI: 0.43-0.66, n=67). For primary experimenters the James' blinding index was 0.80 (95% CI: 0.71-0.90, n=67). Blinding is considered adequate if the index is greater than 0.5" (p. 1729). **Rating:** Low risk	**Judgement:** no specific information reported. **Rating:** Unclear risk	**Judgement:** no specific information reported. **Rating:** Unclear risk	**Judgement:** available case data analyses were undertaken without any imputation of missing data. However, outcome data were missing for 0.0% of the intervention group by the final follow-up assessment and 0.0% of the control group by the final follow-up assessment (0.0% for the study overall). **Rating**: Low risk	**Quote:** "...students completed the Women's Health Initiative Food Frequency Questionnaire...the Pittsburgh Sleep Quality Index...The Seven-Day Physical Activity Recall...The modified version of the Health Review...The Center for Epidemiological Studies Depression Scale...The Beck Anxiety Inventory...Lipids...[and] changes in serum and stimulated production of IL-6 and TNF-[alpha]" (p. 1727). **Judgement:** data on each of these outcomes were reported in Tables 3, 5, and 6 and in Figure 3. **Rating:** Low risk
McGrady	2012	**Quote:** "Students who elected to participate were randomly assigned to either the experimental group...or	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Quote:** "Missing data was dealt with by using case wise deletion..." (p. 255).	**Quote:** "The following measures were used: the Beck Depression Inventory (BDI-II), Beck Anxiety Inventory (BAI), Social

		to the wait list control group" (p. 254). **Judgement:** no further specific information on the method(s) used to generate the randomisation sequence. **Rating**: Unclear risk					**Judgement:** the use of case wise deletion to account for missing data has been found to introduce bias if data are not missing completely at random (KW to locate a reference on that at a later point). In the present review, data was missing for between 69.0% (for BAI scores) to 72.4% (for acute illness scores) of the randomised participants, meaning that data were unlikely to have been missing completely at random. **Rating**: High risk	Readjustment Rating Scale-Revised (SRRS-R), and a brief questionnaire indicating the frequency of acute illness…The experimental group completed the measures before (pre- or assessment time 1-August), and after (post- or assessment time 2-December) participation in the program and in May for follow up (assessment time 3-May)" (p. 255). **Judgement:** data on the SRRS-R not reported at any time point. **Rating**: High risk
Moir	2016	**Quote:** "Computer-generated individual randomization was undertaken, stratified by ear of medical training..." (p. 297). **Judgement**: use of a computer-generated sequence describes an adequately random component in the sequence generation process. **Rating**: Low risk	**Quote:** "...randomization was undertaken...by a researcher not involved in the recruitment or assessment procedures. Opaque sealed envelopes and Group A or B terms were used to ensure allocation concealment prior to enrolment and baseline assessment" (p. 297). **Judgement:** use of sequential, opaque, sealed envelopes describes an adequate method of ensuring allocation concealment. **Rating**: Low risk	**Quote:** "...'blinding was not possible" (p. 299). **Judgement**: the nature of the intervention would have precluded successful blinding of participants or intervention personnel (i.e., the peer-leaders delivering the mindfulness intervention). **Rating:** High risk	**Quote:** "...'blinding was not possible" (p. 299). **Judgement**: the nature of the intervention would have precluded successful blinding of participants or intervention personnel (i.e., the peer-leaders delivering the mindfulness intervention). **Rating:** High risk	**Quote:** "...'blinding was not possible" (p. 299). **Judgment:** the nature of the intervention would have made successful blinding of outcome assessors difficult, and it would appear they were not blind to treatment allocation. **Rating**: High risk	**Judgement**: although available case data analyses were undertaken without any imputation of missing data, outcome data was missing for only 16.5% of the intervention group and 14.8% of the control group (15.6% for the study overall). **Rating**: Low risk	**Quote:** "...the primary outcome measures chosen to assess mental health in the study were depression scores, assessed using the Primary Health Questionnaire (PHQ-9), and anxiety score assessed using the Generalized Anxiety Disorder Questionnaire (GAD-7) …[t]he secondary outcome measures included quality of life (Linear Analogue Self-Assessment [LASA], resilience (25-item questionnaire), academic self-concept (Perceived Competence Scale) and academic motivation (the Motivated Strategies for Learning

								Questionnaire) …Outcome measures were assessed at baseline and 6-month follow-up" (pp. 296-297). **Judgement:** data on each of these outcomes were reported for each time point in Table 3 **Rating**: Low risk
Phang	2015	**Quote:** "[The s]tratified random sampling method was used to assign participants to experimental groups. The participants were stratified according to the years of medical studies...In each year of studies, participants were randomly allocated to experimental groups with the help of an online computer program..." (p. 1122). **Judgement:** use of a computer-generated sequence describes an adequately random component in the sequence generation process. **Rating:** Low risk	**Quote:** "In order to avoid bias in experimental groups allocation, the random numbers generation and matching of numbers with students' name were done by a research staff [sic] who was not involved in the study" (p. 1122). **Judgement:** although no specific information on the method(s) used to achieve allocation concealment, the authors would suggest that treatment allocation was successfully concealed. **Rating:** Low risk	**Quote:** "This is a non-blinded randomized controlled study. The trainer of the intervention was not blinded to the experimental groups. In order to minimize experimental bias, assessments of outcomes were conducted using self-rated questionnaires sealed in envelopes distributed by class representatives; instead of the trainer/investigator. Participants in the control group were not explicitly informed that they were in an experimental control group. They were informed that they would receive the intervention in the form of a DVD scheduled at a different time; 6 months later which is after the follow-up period of the study" (p. 1122).	**Quote:** "This is a non-blinded randomized controlled study. The trainer of the intervention was not blinded to the experimental groups. In order to minimize experimental bias, assessments of outcomes were conducted using self-rated questionnaires sealed in envelopes distributed by class representatives; instead of the trainer/investigator. Participants in the control group were not explicitly informed that they were in an experimental control group. They were informed that they would receive the intervention in the form of a DVD scheduled at a different time; 6 months later which is after the follow-up period of the study" (p. 1122).	**Quote:** "In order to minimize experimental bias, assessments of outcomes were conducted using self-rated questionnaires sealed in envelopes distributed by class representatives; instead of the trainer/investigator. Participants in the control group were not explicitly informed that they were in an experimental control group. They were informed that they would receive the intervention in the form of a DVD scheduled at a different time; 6 months later which is after the follow-up period of the study" (p. 1122). **Judgement**: if participants in the wait-list control group were informed that they would be receiving the intervention in six	**Quote:** "We conducted intention-to-treat (ITT) analyses on the data...Missing values were…replaced with 'full information maximum likelihood (FIML) method" (p. 1123). **Judgement:** outcome data were only missing for 4.0% for the study overall (the proportion of data missing for the intervention and control groups individually was not clearly reported). **Rating:** Low risk	**Quote:** "For the main analyses, we evaluated the effect of experimental group...on each of the dependent outcome measures of the MAAS, PSS, GHQ, and GSE scores" (p. 1123). **Judgement:** data on these outcomes at baseline, one week and six-month follow-up assessments in Tables 3 and 4. **Rating:** Low risk

				Judgement: if participants in the wait-list control group were informed that they would be receiving the intervention in six months' time, then it is unlikely they would not have known they were participating in an experimental study. **Rating:** High risk	**Judgement**: if participants in the wait-list control group were informed that they would be receiving the intervention in six months' time, then it is unlikely they would not have known they were participating in an experimental study. **Rating:** High risk	months' time, then it is unlikely they would not have known they were participating in an experimental study. **Rating:** High Risk		
Shapiro	1998	**Quote:** "The design was a matched randomized experiment in which participants were assigned to a 7-week mindfulness-based intervention or a wait-list control group" (p. 585). **Judgement:** no further specific information on the method(s) used to generate the randomisation sequence. **Rating:** Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Quote:** "To avoid experimenter effects, assessment measures were administered and collected by an undergraduate research assistant not involved in the design of the research or intervention. Further, all participants were assigned a confidential identification number to which the primary investigator did not have access" (p. 586). **Judgement**: assume complete blinding of the outcome assessor ensured. **Rating**: Low risk	**Quote**: "One student did not complete the intervention due to severe medical problems...Four of the participants in the control group did not complete the post-measures" (p. 588). **Judgement**: although available case data analyses were undertaken without any imputation of missing data, outcome data was missing for only 2.6% of the intervention group and 10.2% of the control group (6.4% for the study overall). **Rating**: Low risk	**Quote:** "Participants completed the following measures to assess the six principle quantitative dependent variables: empathy...adapted version (half of the original version of 84 items) of the Empathy Construct Rating Scale (ECRS)...psychological distress [from t]he Hopkins Symptom Checklist-90 Revised...Depression [from s]ubscale 4 of the SCL-90...State and Trait Anxiety [from t]he State-Trait Anxiety Inventory...Spirituality [from t]he Index of Core Spiritual Experiences INSPIRIT...there were two ancillary measures included. A daily journal was used to measure compliance with meditation practice...Also, evaluation packets were filled out by

								participants upon completion of class to assess the course and to gain written qualitative reports on the impact of the course" (pp. 587-588). **Judgement:** data on psychological distress (SCL-90-R) missing from Figure 1. **Rating**: High risk
Velayudhan	2010	**Quote:** "The students were randomly selected from the population..." (p. 43). **Judgement:** no further specific information on the method(s) used to generate the randomisation sequence. **Rating:** Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information provided. However, outcome data was missing for 0.0% of the intervention group and 0.0% of the control group by the post-intervention assessment (0.0% for the study overall). **Rating:** Low risk	**Quote:** "Tools...[consisted of] Beck Anxiety Inventory...Beck Depression Inventory..." (p. 43). Judgement: Data on these outcomes at the post-intervention assessment are provided in Table 2 and 4. **Rating:** Low risk
Warnecke	2011	**Quote:** "Eligible participants were randomised centrally, using block randomisation with blocks of two, to the intervention arm or the usual care control arm**"** (p. 383). **Judgement:** no further specific information on the method(s) used to generate the randomisation sequence. **Rating:** Unclear risk	**Quote:** "All packs contained a CD cover so that trial packs in the two arms of the study looked identical. The purpose of this was to conceal allocation" (p 383). **Judgement:** the use of identical, presumably sealed opaque packs, with CD covers inside both those sent to participants randomised to the intervention arm as well as to those randomised to the control arm, is likely to have ensured	**Quote:** "Randomisation was not blinded to the individual participant because of the nature of the intervention" (p. 383). **Judgement:** due to the nature of the intervention, successful participant blinding was not able to be achieved. **Rating**: High risk	**Quote:** "Randomisation was not blinded to the individual participant because of the nature of the intervention" (p. 383). **Judgement:** due to the nature of the intervention, successful participant blinding was not able to be achieved. **Rating**: High risk	**Quote:** "Both the research assistant who scored and entered data and the statistician who analysed the results were blinded to group allocation" (p. 383). **Judgement:** outcome assessor blinding was able to be successfully achieved. **Rating:** Low risk	**Quote:** "Results were analysed on an intention-to-treat basis" (p. 383). **Judgement:** no specific information provided on the method(s) used to account for missing data. Additionally, outcome data was missing for 37.5% of the intervention group and 5.9% of the control group by the final follow-up assessment (21.5% for the study overall). **Rating**: High risk	**Quote:** "Outcome tools used were the Perceived Stress Scale (PSS) and the Depression, Anxiety and Stress Scale (DASS)...Data from these two questionnaires were collected at baseline...and at the end of the 8-week trial period" (pp.383-384). **Judgement:** Data on these two scales were reported in Table 1 (for baseline) and Table 2 (for post-intervention). **Rating**: Low risk

			allocation concealment was successfully achieved. **Rating:** Low risk					
Whitehouse	1996	**Quote:** "Twenty-one subjects were randomly selected for training in the use of self-hypnosis as a coping skill and were encouraged to practice regularly and to maintain daily diary records relation to mood, sleep, physical symptoms, and frequency of relaxation practice. An additional 14 subjects received no explicit training in stress-reduction strategies" (p. 249). **Judgement**: no further specific information on the method(s) used to generate the randomisation sequence. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Quote:** "All assays were performed on freshly drawn blood by laboratory personnel who were blind to subjects' identities and their assignment to experimental condition" (p. 251). **Judgement:** it is unclear whether outcome assessors investigating psychological measures were also blind to treatment assignment **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk
Yusoff	2015	**Quote:** "We performed stratified randomization method to allocate the consenting students into intervention and control groups by draw lots" (p. 85). **Judgement:** use of drawling lots describes an adequately random component in the sequence generation process.	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating**: Unclear risk	**Quote:** "To ensure the researchers were blinded during analysis, data were collected and entered into a data sheet by a research assistant and study subjects were assigned with a unique code throughout the study" (p. 86). **Judgement**: Blinding of outcome assessors ensured, and it is	**Judgement:** although available case data analyses were undertaken without any imputation of missing data, outcome data was missing for only 17.0% of the intervention group and 3.6% of the control group (10.5% for the study overall). **Rating**: Low risk	**Quote:** "We measured three main outcomes which were depression symptoms, coping strategies, and perceived stressors...BDI...MSSQ-20...Brief COPE...We collected data at five different intervals: the baseline measurement...the post-intervention measurements were performed at 1 week...8

		Rating: Low risk				unlikely blinding could have been broken **Rating**: Low risk		weeks...16 weeks...and 32 weeks (p. 84). **Judgement:** data on each of these outcomes were reported for each time point in Tables 3, 4, and 5. **Rating**: Low risk

Supplementary Table 4:
Assessment of study quality for non-randomised controlled trials (non-RCTs).

Author	Year	Bias due to Confounding	Bias in Selection of Participants	Bias in Classification of Intervention	Deviation(s) from Intended Intervention	Bias due to Missing Data	Bias in Measurement of Outcome(s)	Selective Outcome Reporting	Any other threats to internal/external validity
Camp	1994	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "…medical students enrolled in either an LBL or PBL curriculum at the same medical school" (p.25) **Judgement:** No specific inclusion/exclusion criteria **Rating:** Unclear risk	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "A total of 250 students completed the follow-up SDS test in November (90.9% compliance)" (p.26) Judgement: Less than 10% missing data **Rating**: Low risk	**Judgement:** The only outcome reported was assessed via self-report (SDS) **Rating:** High risk	**Quote:** "The SDS is a 10-item self-report instrument of the subjects' current symptoms of depression...A measure of students' self-actualization was [also] obtained from the self-actualization scale of the California Psychological Inventory (CPI)" (p. 25). **Judgement:** Data on both these outcomes at baseline and at post-intervention is reported in Table 1. **Rating**: Low risk	**Judgment**: Less than 10% missing data **Rating**: Low risk
Chen	2016	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Quote:** "Thirteen… medical students… self-selected into MBM" (p. 1). **Judgement:** included participants may be	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "At the beginning of their fall semester…122 first-year medical students completed the surveys;	**Judgement:** participants self-selected into the MBM program and all outcomes were	**Quote:** "[Participants] received the Jefferson Scale of Physician Empathy – Students (JSPE-S), the Perceived Stress Scale (PSS), and the Personal	**Judgement**: Missing data for 77.9% **Rating**: High risk

			more motivated and/or have greater insight than average student. However, the authors write: "We concluded that enrolment in the MBM course was…independent of gender, ethnicity, race, marital status, and religion" (p. 2). **Rating:** unclear risk.			however, only 27 completed the post-test surveys" (p. 2). **Judgement:** data missing for 77.9%. **Rating:** high risk.	assessed via self-report. **Rating:** high risk.	Health Questionnaire" (p. 2). **Judgement:** data on these outcomes reported in figures 1 to 3. **Rating:** low bias.	
Finkelstein	2007	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Quote: "**Year 2 (pre-clinical) medical students and graduate nursing students were eligible to enrol. Students were given the opportunity to participate on a first-come, first-served basis. Although some nursing students participated in the course, they were excluded from the study group because of insufficient numbers. The class filled within 3 days of being offered; 38 medical students formally requested to enrol. Due to space limitations only the first 30 students were enrolled" (p. 259) **Judgement**: There were no statistically significant	**Judgement:** no specific information reported. **Rating:** unclear risk.	**Quote:"** Intervention group attracted a higher percentage of 'possibly depressed' students as identified by the 2-item Depression index than there were in the comparison group" (p.261) **Judgement**: The intervention and comparison groups also differed significantly with respect to anxiety levels at the beginning **Rating**: High risk	**Quote:** "Of the 30 medical students who initially enrolled, 2 did not attend the first class and 2 dropped the class after 1 session; leaving 26 study subjects (87% participation rate). At the beginning and end of the course (times 1 and 2) we achieved 100% response rates from the 26 study participants. Three months following the course completion (time 3), 88.5% of study participants completed all 4 study instruments. The remaining 154 students in the medical school class were eligible to participate in the comparison group. By 10 days after the	**Judgement**: Participants volunteered to take part in the study, and all outcomes were assessed via self-report **Rating**: High risk	**Quote:** "Study and comparison groups were surveyed using 4 validated instruments at time 1 (beginning of the quarter), time 2 (the end of the quarter) and time 3 (3 months later). The instruments used were the SCL-90 Anxiety Subscale, the POMS, the 2-item Depression Index and the PSMS" (p. 260). **Judgement:** it is unclear whether the data on depression as reported on p.261 relate either to the post-intervention or longest follow-up assessment, however, data for one of these time points is missing. **Rating**: High risk	**Quote**: "…Neither the intervention nor the comparison group was randomised… It is also unknown whether the intervention group's decreased anxiety levels would be sustainable throughout the rest of medical school and into clinical training" (p.263) At T1, there was 13% missing data, at T2 there was 0%. At T3 there was 11.5% missing data. **Judgement**: High risk for non-randomisation, and bias concerning outcome measure **Rating**: High risk/ Low risk concerning missing data

			differences in age or gender between groups. Although the proportion of women enrolled in the mind) body elective (75%) was slightly higher than that in the comparison group (63%), the difference was not significant **Rating**: unclear risk			beginning of the course (time 1), 46 students had responded (participation rate 30%). Subsequent comparison groups were drawn from this original group, resulting in response rates of 54% at time 2 and 86.9% at time 3" (p. 260) **Judgement**: Adequate response rate **Rating**: Low risk			
Kelly	1982	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "Self-referred medical students…[t]hose who telephoned to indicate interest and later attended the first session were subjects" (p. 92). **Judgement:** included participants may be more motivated and/or have greater insight than average student. However, the authors do report "preliminary analyses were conducted to evaluate systematic differences between the control group scores and the scores of the stress management group before they received	**Judgement:** no specific information reported. **Rating:** unclear risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "Only those in the stress management group for whom both pretraining and posttraining data were obtained and who turning in at least 50 percent of the daily log forms were included in the results analysis (n=21)" (p. 94). **Judgement:** reads as available case analysis with 38.2% drop-out rate. However, the authors argue "[a]ttrition did not appear to be related differentially to any background variable" (p. 94). **Rating:** unclear risk.	**Judgement:** participants self-selected into the program and all outcomes were assessed via self-report. **Rating:** high risk.	**Quote:** "The measures administered were: Stress Knowledge Inventory…Jenkins Activity Schedule…Spielberger State-Trait Anxiety Inventory…Stressful situations rating…self-monitoring logs" (p. 93). **Judgment:** data on these outcomes reported in Tables 1 and 2 and in Figure 1. **Rating:** low risk.	**Judgement**: No specific information about missing data. However, as data were based only on those 21 participants with pretraining and posttraining data, there was a 38.2% drop-out rate. **Rating**: high risk

			training. Multivariate analysis of variance indicated no significant differences between the two groups" (p.95). **Rating:** unclear risk.						
Kraemer	2015	**Judgement:** no specific information reported. **Rating:** unclear risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "…students in the intervention group (n=11; 63.6% female; Mage=24.27 years) and no intervention control group (n=11; 72.7 female; Mage=23.64) …" (p.83) **Judgement:** no specific information on classification **Rating**: Unclear risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Judgement**: Participants volunteered to take part in the study, and all outcomes were assessed via self-report **Rating**: High risk	**Judgement:** Students in the intervention group and non-intervention control group completed self-report measures pre and post the 11-weeks. Intervention group students answered open-ended questions post-intervention for quotes on their group experiences. All outcomes based on self-report measures **Rating:** High risk	**Judgement**: No specific information about missing data or dropout rate **Rating**: Unclear risk
Michie	1994	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote: "**The course is offered to medical students at a London medical school, as part of its Occupational Health Service. The evaluation was carried out in the first year of this service, when it was offered only to first year clinical students; this year being chosen as it has been found to be the most stressful year of medical training. All students	**Quote:** "There were two self-selected groups, course attendees and non-attendees (p.529) **Judgment:** Participants self-selected themselves into groups **Rating:** High risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "Of the 19 attendees, 17 completed questionnaires at the beginning of the year. Of the 92 non-attendees, 40 completed questionnaires at the beginning of the year and 27 also completed the end of the year questionnaire" (p.530) **Judgement:** Less than 20% missing data **Rating:** Low risk	**Judgment:** All outcomes were assessed using self-reported questionnaires **Rating:** High risk	**Quote: "**Students completed a questionnaire including six items measuring anxiety (the shortened version of the Speilberger State-Trait Anxiety Inventory; Marteau & Bekker 1991) and seven questions, rated on a four point scale, validated on health care staff counselled at the Occupational Health Unit (Michie 1992). These questions covered anxiety, depression, satisfaction with themselves, their work and their life outside work, and perceived functioning at work and	**Judgment:** Missing data less than 20% **Rating**: Low risk

			were also offered access to an individual, confidential counselling service" (p.529) **Judgment**: No specific exclusion criteria **Rating**: Low risk					outside work. There was also an open-ended question about what students are finding particularly stressful at present. This questionnaire asked about reasons for not using the service, and about its accessibility, helpfulness, content and organization" (p.530) **Judgment:** same measures were used to assess outcomes in both groups, but groups were self-selected **Rating:** Moderate risk	
Rosenzweig	2003	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote: "**All 2nd-year students at Jefferson Medical College during the years 1996 to 2000 were eligible to participate in a MBSR program, offered as one choice among approximately 10 elective seminars" (p.3) Judgement: No specific exclusions criteria Rating: Low risk	**Quote:** "group assignment was nonrandomized. This gave rise to a significant difference between intervention and control groups at baseline" (p.5) **Judgement**: Students with greater overall mood disturbance enrolled in the MBSR program **Risk**: High risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote**:" Three hundred two 2nd-year medical students participated in the study between fall 1996 and fall 2000. One hundred forty students received MBSR training, and 162 students served as parallel cohort controls. The average number of students who participated in MBSR in a given year (40) represented approximately 18% of the entire 2nd-year medical student class" (p.3) **Judgement**: No specific information about dropout **Rating**: Unclear risk	**Judgement**: Participants volunteered to take part in the study, and all outcomes were assessed via self-report **Rating**: High risk	**Quote:** "The Profile of Mood States (POMS) was administered to all participants at the beginning and end...[t]his instrument is a factor analytically derived inventory that measures six identifiable mood or affective states: tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment. In additional to these six subscale scores, a total mood disturbance (TMD) score may be obtained from the POMS..." (p. 3). **Judgement**: data on each of these subscales, as well as the total score, are reported in Table 1. **Rating**: Low risk	**Judgment**: No specific information on dropout rate and missing data **Rating**: Unclear risk

Zuardi	2008	**Judgement:** no specific information reported. **Rating:** unclear risk	**Quote:** "The scale was administered inside the classrooms…students attending regular activities were not invited to voluntarily participate in the study. Those who were not willing to participate should simply return the scales unmarked…" (p. 137) **Judgement**: No significant differences between the two samples and between courses. **Rating**: Low risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Judgement:** no specific information reported. **Rating:** unclear risk	**Judgement**: Participants volunteered to take part in the study, and all outcomes were assessed via self-report **Rating**: High risk	**Quote:** "Anxiety levels had been assessed through Spielberg's State- Trait Anxiety Inventory9 - trait form, which has been validated into Portuguese by Biaggio et al.10 (STAI-T)" (p.137) **Judgement**: Data for the outcomes up to 4 years post intervention was estimated from the Figure 1. as authors only present data up to two-years post-intervention (i.e., up to where the changes are still significant which is suspicious **Rating**: Unclear risk	**Quote**: "This study has the shortcoming of using a single scale for anxiety evaluation instead of more specific instruments for the assessment of stress, burnout and depressive and anxious symptoms in undergraduate students. Moreover, interference from external factors cannot be completely disregarded" (p.138) No specific information about dropout and missing data **Judgment**: Other factors may have influenced results **Rating**: High risk

Supplementary Table 5:
Assessment of study quality for historically controlled studies.

Author	Year	Representativeness of Exposed Cohort	Selection of the non-exposed Cohort	Ascertainment of Exposure	Demonstration the Outcome of Interest Was Not Present at Start of the Study	Comparability between Intervention and Historical Cohort	Assessment of Outcome	Duration of Follow-up sufficient for Outcomes to Occur
Al-Faris	2014	**Quote**: "The study participants included two cohorts of students: a) all 1st and 2nd year students enrolled in the College of Medicine, KSU during the year 2007/ 2008 studying in the traditional curriculum b) all 1st and 2nd year students enrolled in the College of Medicine, KSU during the year 2010/ 2011 studying in the SBC. No sampling method was used as the whole population was invited to participate" (p.3)	**Judgement**: Non-exposed cohort was drawn from the same community as the exposed cohort (undergraduate medical students receiving traditional curriculum) **Rating**: One star	**Judgement**: Written self-report inventories were administered. These included The DREEM inventory and the BDI-II Inventory **Rating**: No star	**Quote**: "In order to address the first and third objectives, the mean total score of the DREEM and the mean score of the five domains of the inventory were calculated and compared for curricula cohorts, year of study and gender. In order to address the second and third objective, comparison of the mean BDI score was made between the two curricula cohorts, across the two academic years and the two genders" (p.3) **Judgment**: assessment of depression and	**Judgment**: The controls were exposed to the traditional curriculum which focused on didactic lectures as main teaching format **Rating**: One star	**Judgement**: self-administered inventories rather than structured interviews for exploration and clinical diagnosis **Rating**: One star	**Judgement**: Duration of follow-up was up to two years. Sufficient amount of time **Rating**: Yes

		Judgement: Cohort representative of population in question **Rating**: No star			educational environment present at baseline **Rating**: Unclear			
Holm	2010	**Quote**: "Two subsequent classes of third-year medical students at the University of Bergen, Norway participated in this quasi-experimental study. Medical students in Bergen follow a "traditional" curriculum, with two years of preclinical studies, followed by four years of clinical training. Problem-based learning is not a part of the curriculum. The intervention took place in the initial months of the students' clinical training. The intervention group (n = 129) enrolled at the university in 2001, and a second group (n = 152) enrolled in 2002, acted as the control group. Both groups were assessed twice **Judgment**: Cohort representative of population in question **Rating**: No star	**Quote:** "The intervention groups and the control group were from two different student classes and were assessed during two different calendar years. It is possible that the pressures on the intervention and control groups have been different, so also the motivation to participate in the study" (p. 7) **Judgement**: Bias possible **Rating**: One star	**Judgment**: Distress was measured using the Perceived Medical School Stress (PMSS) and Symptom Check List-5 (SCL-5) assessments. Written self-report measures **Rating**: No star	**Judgement**: Stress and symptoms measured at T1, meaning outcomes were present at baseline Rating: Unclear	**Judgement**: Lack of a randomized, controlled trial design. It is possible that because the control and intervention group came from two different classes and assessed at different time periods, that the pressures and motivations between groups may have varied **Rating**: One star	**Judgment**: Self-reported measures rather than structured interviews for exploration and clinical diagnosis **Rating**: One star	**Judgment**: A three-month follow-up showed that the intervention had a positive effect on perceived medical school stress among the students, and further analyses showed this was due to participation in self-development groups **Rating**: yes
Melo-Carillo	2012	**Judgment**: Cohort representative of population in question (medical students), but not comparable to	**Judgement**: General population, drawn from a different population **Rating**: One star	**Judgment**: Self-report questionnaires administered: BDI **Rating**: No star	**Judgement**: Depressive symptoms present at the beginning of study **Rating**: Unclear	**Judgment**: Exposed individuals not matched in historical cohort **Rating**: No star	**Judgment**: Self-report measures of depression assessed **Rating**: One star	**Judgment**: Not clearly specified, but analysed as post-intervention **Rating**: Unclear

		general public. Selected group of users **Rating**: No star						
Slavin	2014	**Judgment**: Cohort representative of population in question (medical students), but not comparable to general public. Selected group of users, did not conduct experiment with random assignment **Rating**: No star	**Judgment:** No comparison group to the intervention **Rating**: No star	**Judgment:** Measures included: CES-D, STAI, PSS and PCS. All written self-report **Rating:** No star	**Judgment:** outcomes of interest present at the start, which is why they wanted to reform the curriculum **Rating:** Unclear	**Judgment:** No comparison group, so unable to determine **Rating:** Unclear	**Judgement**: Self-report outcomes with no reference to medical or health records to confirm the outcome **Rating**: One star	**Quote: "**At the end of years one and two, a clear trend emerged in the post change classes compared with the pre change classes—the post change classes exhibited lower rates of moderate to severe depression symptoms. Anxiety symptoms followed a similar pattern—a substantial decrease in mean anxiety scores in the post change classes—as did stress levels—progressive decreases in the mean stress levels of the post change classes" (p.576) **Judgment:** sufficient time passed **Rating:** Yes
Thompson	2010	**Judgment**: Cohort representative of population in question (medical students), but not comparable to general public. Selected group of users **Rating**: No star	**Judgment**: Both cohorts were medical students from the same university, just at measured at different time points (2003-2004). Drawn from the same community as the exposed cohort **Rating**: One star	**Quote**: "Both years, we administered the anonymous surveys during colloquia, a mandatory monthly meeting for third year medical students. We asked the students to sit apart, consistent with usual test-taking arrangements. Cover letters described the project, solicited participation, and provided students with a means to shield their	**Judgement**: Outcome of interest (depression, suicidality) present at start **Rating**: Unclear	**Judgement**: No differences between intervention and historical cohort expect time point **Rating**: One star	**Judgement**: Self-report outcomes with no reference to medical or health records to confirm the outcome **Rating**: One star	**Quote**: "After one year of exposure to the newly implemented approach (i.e., a discussion at the beginning of the school year, heightened faculty awareness, and receipt of the well-being handbook), the rates of depressive symptoms and suicidal ideation in the next third-year class (class of 2004) were markedly lower. Only one in four students (n 14; 24.1%) described symptoms of

				answers for added privacy (the cover sheet also contained contact information should the student want to seek immediate psychiatric counselling). After completing the forms, participants placed them in a box and received a snack in appreciation for returning the survey. The colloquia facilitator sent the sealed envelopes containing the forms to the primary investigator" (p.1636) **Judgment**: Written self-report **Rating**: No star				mild or probable depression, representing a significant decrease from the previous year (2 12.84, df 2, P .01). Only one student (3%) reported suicidal ideation, representing a 10-fold decrease (2 13.05, df 1, P .001)" (p.1637) **Judgment**: Adequate amount of time passed **Rating**: Yes

Supplementary Table 6:
Assessment of study quality for uncontrolled longitudinal studies.

Author	Year	Pre-intervention: Bias due to confounding	Pre-intervention: Bias in selection of participants into the study	During intervention: Bias in the measurement of interventions	Post-intervention: Bias due to departure from intended intervention	Post-intervention: Bias due to missing data	Post-intervention: Bias in the measurement of outcome(s)	Post-intervention: Bias in Selective Reporting	Any other threats to internal/external validity
Bansal	2013	**Judgment**: No specific information. **Rating**: Unclear	**Judgement:** Participants included 90 students of MBBS 3rd semester who came for their 1st posting in the department of community medicine in different batches. Each batch comprised of around 10 students **Rating:** Low RoB	**Judgment:** GHQ-28 was used to assess the impact on general and mental well-being and was applied at baseline, and at the end of the study on all the participated students. **Rating:** Low RoB	**Judgment:** No specific information **Rating:** Unclear	**Judgment:** Out of the 90 students posted in the community medicine, eight students were not regular. They were absent either at the time of the pre or post-test, the final analysis included 82 students **Rating**: Low RoB	**Judgment:** All outcomes reported, all self-report measures **Rating:** High RoB	**Judgment**: No specific information **Rating**: Unclear	None
Bond	2013	**Quote**: "Student in the study were fairly high in empathy at baseline, suggesting a possible ceiling effect for this measure" (p.7) **Judgment**: **Rating**:	**Judgment**: Course enrollment was limited to 27 medical students in their first or second year of medical school at BUSM. No specific exclusion criteria **Rating**: Low RoB	**Judgment**: The pre-and post-course online survey included four scales: 1) Jefferson Scale of Physician Empathy; 2) Cohen's Perceived Stress Scale; 3) Self-Regulation Questionnaire; and 4) Self-Compassion Scale **Rating**: Low RoB	**Judgment**: The 27 students attended a median of 11 classes (range 9-11). The differences in intervention dosage may have affected the outcomes **Rating**: Moderate RoB	**Judgment**: All enrolled students completed the pre-course survey; 24 filled out the post course survey. No dropout rate **Rating**: Low RoB	**Judgment**: All outcomes reported, all self-report measures. Also, no control group-impossible to ascertain whether observed changes were due to the course or to other factors **Rating**: High RoB	**Judgment**: No specific information **Rating**: Unclear	None
Bughi	2006	**Judgment:** No specific information whether an individual receives one or the other	**Quote:** "The study sample included third and fourth year medical students and represents an accumulation of four	**Quote: "**The students were given a self-report questionnaire, the General Well Being Scale (GWBS), as a	**Judgment:** All students were tested during a similar period of time, no other information	**Judgment**: No specific information **Rating**: Unclear	**Quote:** "The students were given a self-report questionnaire, the General Well Being Scale (GWBS), as a pre-test in their	**Judgment:** No specific information, fully reported results **Rating:** Low RoB	None

		intervention of interest. **Rating:** Unclear	to six medical students per one-month rotation in our service over the last three years" (p.2) No specific exclusion criteria. **Judgment:** No specific information about selection **Rating**: Low RoB	pre-test in their psycho-educational lecture on stress. The prevalence of stress (research aim # 1) was measured in the entire group of students (N=104). The variation of stress based on the time of testing (beginning vs. end of rotation), academic year (third vs. fourth year), and gender was measured in the same group" (p.3). **Judgment:** **Rating:**	reported on deviances in intervention. **Rating:** Low RoB		psycho-educational lecture on stress" (p.4) **Judgment:** One outcome measured using self-report **Rating:** High RoB		
Dutton	2013	**Judgment**: No specific information **Rating**: Unclear	**Judgment**: 60 students out of a class of 192. No specific information on selection **Rating**: Low RoB	**Judgment**: Design consisted of pre- and post- test assessment of distress using the BSI, STAI **Rating**: Low RoB	**Judgment**: No specific information **Rating**: Unclear	**Judgment:** One survey excluded due to incompletion **Rating**: Low RoB	**Judgment**: No specific information **Rating**: Unclear	**Judgment**: No specific information **Rating**: Unclear	None
Gallagher	2005	**Judgement**: No pre intervention reports were administered in relation to the outcome measures assessed post intervention **Rating**: Low RoB	**Judgment**: Hotline was targeted for 3rd year students, but allowed other year students to call **Rating**: Low RoB	**Quote**: "The service was evaluated by a 10-point idiosyncratic scale ranging from one (least stressful year of life) to 10 (most stressful year of life), and awareness of hotline (open-ended question), reassurance about	**Judgment**: No specific information regarding this bias **Rating**: Unclear	**Quote**: "83% of the 104-year end surveys were returned" (p. **Judgment**: 17% missing data **Rating**: Low RoB	**Judgement**: No specific information **Rating**: Unclear	**Judgement**: No specific information **Rating**: Unclear	None

				existence of hotline (five-point idiosyncratic scale ranging from one [not at all reassuring] to five [very reassuring]), and importance of continuing the hotline (five-point idiosyncratic scale ranging from one [not at all important] to five [very important])" (p. **Judgment**: No information on bias **Rating**: Low RoB					
Garneau	2013	**Quote: "**the students were interviewed and matched to a residency program around the same time as the course evaluations were completed" (p.475) **Judgment:** This may of impacted ratings on preliminary outcome measures **Rating:** Moderate RoB	**Judgment:** 58 4th-year medical students between 2009-2012. No specific inclusion/exclusion criteria mentioned **Rating**: Low RoB	**Quote:** "Fifty-eight 4th-year medical students completed on-line questionnaires pertaining to depression, burnout, stress, wellbeing, self-compassion and mindfulness one week before and one week after the course" (p.470). **Judgment:** all outcomes measured **Rating:** Low RoB	**Judgment**: No specific information regarding this bias **Rating**: Unclear	**Judgment**: No specific information regarding this bias **Rating**: Unclear	**Judgment:** MBI-HSS, PSS-10, SPWB, SCS, MAAS, BDI-II, and follow-up questionnaire were administered to all participants. All self-reported measures **Rating:** High RoB	**Judgment**: No specific information **Rating**: Unclear	None
Greeson	2015	**Judgment**: No specific information **Rating**: Unclear	**Judgment**: the study enrolled a relatively small, self-selected sample of 44 medical	**Judgment**: Used quantitative (CAMS-R, PSS) and qualitative measures to assess	**Quote**: "it is possible that nonspecific factors such as social support, positive	**Quote**: "besides having students voluntarily report on adherence to weekly home practice	**Quote**: "although validated survey measures of mindfulness and perceived stress	**Judgement**: No specific information **Rating**: Unclear	**Judgement**: study lacked long-term follow-up, so the durability of positive changes in perceived

			students, no specific exclusion criteria **Rating**: Low RoB	intervention. Pre- and post-workshop surveys were distributed at the beginning and the end of the first and final group sessions, respectively **Rating**: Low RoB	expectancy, and/or attention from an empathic instructor may have accounted, in part, for some of the beneficial effects observed" (p.191) **Judgment**: These factors may have influenced outcomes and intervention **Rating**: Moderate RoB	exercises, the use of mind–body skills was not formally tracked; daily written or electronic logs or smartphone records could be used to track actual use in future studies" (p.191) **Judgment**: This may have resulted in missing data that wasn't recorded **Rating**: Moderate RoB	were used, self-report assessment is prone to social desirability bias and can fluctuate in reliability and validity" (p.191) **Judgment**: Self-reported outcome measures **Rating**: High RoB		stress, mindfulness, self-care behavior, and use of stress management skills remains to be determined. 80% response rate **Rating**: Low RoB
Hassed	2009	**Judgment**: Motivation, need and insight vary between participants **Rating**: Low RoB	**Judgment**: Participants were of an eligible cohort of 270 first year undergraduate medical students **Rating**: Low RoB	**Judgment**: Participants were measured twice (mid-semester one and 6 weeks later) using the SCL-90-R with 3 subscales and the (WHOQOL)-BREF **Rating**: Low RoB	**Judgment**: All participants received same intervention, but with different tutors **Rating**: Low RoB	**Quote**: "A total of 148 of an eligible 270 students returned data at T1 and T2 giving a response rate of 55% The lower response rate at T2 was attributed to the close proximity of exams as well as ''evaluation-fatigue'' considering the number of questionnaires students complete during routine curriculum evaluation. Also no follow-up conducted" (p. 394). **Judgment**: Low response rate **Rating**: Moderate RoB	**Quote**: "Seasonal effects would also count against a trend towards better mental health in the middle of the year, being winter in Australia. There may also be variation in outcomes across groups which were not measured reflecting individual group dynamics or differences in tutor performance. It is also unknown whether the findings would be similar in an older cohort of post-graduate students" (p.396). **Judgement**: These factors may have affected outcomes **Rating**: High RoB	**Judgment**: No specific information **Rating**: Unclear	**Judgment**: 45% missing data. The present data did not permit identification of the particular components of the HEP that might have been most useful or the ways in which these components might act synergistically

Kötter	2016	**Quote**: "Logistic regression revealed that being female, higher levels of *anxiety* and *emotional distancing*, as well as a lower level of *career ambition* were statistically significant predictors of participation in the intervention. Age did not prove a statistically significant predictor" (p.6) **Judgement**: **Rating**: Moderate RoB	**Quote**: "We invited all medical freshmen from the 2011 cohort at the University of Luẽbeck. The t1 survey was taken in June 2012, at the end of the freshman year. The intervention took place at the beginning of the summer semester 2013. The follow-up survey (t2) was taken in June 2013, at the end of the sophomore year. Both surveys were web-based. There were no exclusion criteria" (p.2) **Judgment**: No specific information regarding bias **Rating**: Low RoB	**Judgment**: Used the PMSS-D, self-rated health and the HADS as measures for the study. All included participants completed the same measures **Rating**: Low RoB	**Quote:** "Of the 122 students surveyed in this study, 75 (62%) took part in an introductory module on PMR. For a full participation as defined above, participation was required in both the introductory *and* the refresher module, which was the case for about one third of all respondents (n = 45; 37%)" (p. 4). **Judgment**: 38% dropout rate during intervention **Rating**: High RoB	**Quote**: "After exclusion of incomplete data-sets, 122 t2-cases could be matched to t1-cases (93% of the 131 t2-respondents and 66% of all students matriculated at t2)" (p.4) **Judgment**: High initial participation rate on one hand, and high dropout rate during the intervention **Rating**: Moderate RoB	**Quote**: "To reduce potential drop-out rates, participants received a book voucher to the value of 5 Euro per attended PMR-session and per completed questionnaire (t1 and t2)." (p.3) **Judgment**: All outcomes reported, though high rates of missing data and all are self-report measures **Rating**: High RoB	**Judgement**: All outcomes reported regardless of missing data **Rating**: Low RoB	**Judgment**: 38% dropout rate during intervention (T2)
Mercer	2010	**Quote**: "it was impossible to control for outside influences in the students' schedules that may have affected their stress levels" (p.147) **Judgment**: **Rating**: Moderate RoB	**Quote**: "recruitment was initially limited to only second-year medical students. The reason for this limitation was that studies showed that the second year of medical study was the most stressful. Through this limited recruitment I was only able to get one student, and it was necessary to amend my protocol to open	**Judgement**: many participants had personal experiences during the two weeks between the sessions that affected their stress and mood levels. The measures used to measure effects of journaling intervention were STAI-Y and the PANAS-X	**Judgment**: No specific information **Rating**: Unclear	**Judgment**: No specific information **Rating**: Unclear	**Judgement**: Not possible to show statistical significance due to low number of participants **Rating**: High RoB	**Quote:** "Due to low number of participants in the study, it is not possible to show statistical significance. Nor is it possible to generalize from the participants in this study to other EVMS staff and students or to members of the health professionals as a whole" (p.144) **Judgment**: unable to report statistical significance **Rating**: High RoB	None

			recruitment to all students, as well as staff. The amendment process limited the time available to complete the group interventions" (p.147) **Judgment**: no specific exclusion criteria **Rating**: Low RoB	**Rating**: Moderate RoB					
Moffat	2004	**Quote: "**Glasgow curriculum is that the act of simply participating in something new may be an engaging experience, creating positive attitudes by psychological mechanisms that are unrelated to the theory, structure or content of the curriculum… A further factor may be that our students might have actively selected to study in Glasgow because of its newly designed course" (P.488) **Judgment:** participation engagement and motivation may	**Quote: "**All Glasgow medical students who entered the course in 1997 (n ¼ 275) were simultaneously asked to complete a questionnaire survey, with one reminder, midway through term 1 of first year. It was repeated midway through term 3, 5 weeks prior to the end of the year" (p.483) **Judgment:** No specific exclusion criteria **Rating:** Low RoB	**Quote:** "The questionnaire comprised the 12-item General Health Questionnaire (GHQ-12)16 to measure psychological morbidity, a 59-item list of potential stressors, grouped into 14 themes, and the Brief COPE17 to determine coping styles" (p.483). **Judgment:** All participants completed same measures **Rating:** Low RoB	**Judgment**: No specific information **Rating**: Unclear	**Judgment**: No specific information **Rating**: Unclear	**Judgment:** All outcome measures reported, (GHQ-12 and COPE) **Rating:** Low RoB	**Judgment**: No specific information **Rating**: Unclear	**Quote**: "Despite the high response rates, a further limitation may be that of non-response bias. It would have been advantageous to interview a sample of non-respondents to assess their experience and psychological status" **Judgment:** No control group **Rating:** Moderate RoB

		have been influenced **Rating:** Moderate RoB							
Simard	2009	**Quote**: "Fifteen students (94%) were women and eight (50%) had Previous experience with yoga, but none had regularly practiced yoga. Although 54% reported being involved in more than 3.5 h of physical activity per week at the program's onset, the corresponding energy expenditures were modest with a metabolic equivalent (MET)(Ainsworth et al. 2000) of only 158.3 calories per day (SD 76.2; range 47–296) "(p.951) **Judgment**: Previous experience with yoga may influence outcomes **Rating**: Moderate RoB	**Judgment**: Sample included 16 medical students of a total of 204 first-year McGill University medical students in 2007, no specific exclusion criteria **Rating**: Low RoB	**Judgment**: All participants answered the GHQ, PSS, CED-S and student satisfaction scale **Rating**: Low RoB	**Quote**: "Students attended, on average, 18 yoga sessions (SD 4.58) (mean attendance rate: 65%)" (p.951) **Judgment**: The varying amount of intervention dose may have influenced outcomes **Rating**: Moderate RoB	**Judgment**: 14 out of 16 participants completed baseline, mid-term and end of program evaluations. High response rate **Rating**: Low RoB	**Judgment**: Same methods used to assess outcomes. All self-report measures **Rating**: High RoB	**Judgment**: No specific information **Rating**: Unclear	None.
Wild	2014	**Judgment:** No specific information **Rating:** Unclear	**Quote: "**A total of 39 medical students (classical curriculum, clinical section, 5th	**Judgment:** Psychometric data were collected from 11 students	**Quote:** "To generate a control group for the students	**Quote:** "Two participants in the first survey (summer term 2012) did not	**Judgment:** all outcome measures administered to both intervention and	**Judgement**: All outcomes reported regardless of missing data **Rating:** Low RoB	None

			to 8th semester) and three psychology students (6th semester) participated in Relacs during the 2012 summer term and 2012/2013 winter term. The course took place in small groups (max. 12 students) during one semester and with one session (2h) per week and group" (p.3) **Judgment:** No specific exclusion criteria, and no randomization **Rating:** Moderate RoB	at one point in time and from 31 students at two points. Students interviewed using BOS-II and STAI-G, AVEM-44, BDI-II and SOC-L9 **Rating:** Low RoB	participating in the Relacs course, other 8th-semester medical students were surveyed. To establish comparability of data acquisition times with the Relacs group, these assessments were held at the beginning (first week) of the semester and at the end of the semester (one week before final exams). The assessments were conducted with the same psychological questionnaires as listed above" (p.4) **Judgment:** Controlled for deviances in intervention **Rating:** Low RoB	fully complete their questionnaires, so that not all parameters were analyzed for them" (p.3) **Judgment:** Low drop-out rate **Rating:** Low RoB	control group for comparability. All outcome measures reported **Rating:** Low RoB		

Supplementary List 1:
Full reference information for the 39 studies included in this review.

Randomised Controlled Trials (RCTs):

Ball S, Bax A. Self-care in medical education: Effectiveness of health-habits interventions for first-year medical students. *Acad Med.* 2002;**77**:911-7.

Danielwitz M, Bradwejn J, Koszycki D. A pilot feasibility study of a peer-led mindfulness program for medical students. *Can Med Educ J*. 2016;**7**:e31-7.

de Vibe M, Solhaug I, Tyssen R, *et al*. Mindfulness training for stress management: A randomised controlled study of medical and psychology students. *BMC Med Educ.* 2013;**13**:107.

Holtzworth-Munroe A, Munroe M, Smith R. Effects of a stress-management training program on first- and second-year medical students. *Acad Med*. 1985;**60**:417-9.

Kiecolt-Glaser J, Belury M, Andridge R, *et al*. Omega-3 supplementation lowers inflammation and anxiety in medical students: A randomized controlled trial. *Brain Behav Immun.* 2011;**25**:1725-34.

McGrady A, Brennan J, Lynch D, *et al*. A wellness program for first year medical students. *Appl Psychophysiol Biofeedback.* 2012;**27**:253-60.

Moir F, Henning M, Hassed C, *et al*. A peer-support and mindfulness program to improve the mental health of medical students. *Teach Learn Med*. 2016;**28**:293-302.

Phang C, Mukhtar F, Ibrahim N, *et al*. Effects of a brief mindfulness-based intervention program for stress management among medical students: The Mindful-Gym randomized controlled trial. *Advances in Health Sciences Education* 2015;**20**:1115-34.

Shapiro S, Schwartz G, Bonner G. Effects of mindfulness-based stress reduction on medical and premedical students. *J Behav Med*. 1998;**21**:581-99.

Velayudhan A, Gayatridevi S, Bhattacharjee R. Efficacy of behavioral intervention in reducing anxiety and depression among medical students. *Ind Psychiatry*. 2010;**19**:41-6.

Warnecke E, Quinn S, Ogden K, *et al*. A randomised controlled trial of the effects of mindfulness practice on medical student stress levels. *Med Ed*. 2011;**45**:381-8.

Whitehouse W, Dinges D, Orne E, *et al*. Psychosocial and immune effects of self-hypnosis training for stress management throughout the first semester of medical school. *Psychosom Med*. 1996;**58**:249-63.

Yusoff M, Esa A. A DEAL-based intervention for the reduction of depression, denial, self-blame and academic stress: A randomized controlled trial. *Journal of the Taibah University Medical Sciences.* 2015;**10**:82-92.

Non-Randomised Controlled Trials (non-RCTs):

Camp D, Hollingsworth M, Zaccaro D, *et al*. Does a problem-based learning curriculum affect depression in medical students? *Acad Med.* 1994;**21**:e31996.

Chen A, Kumar A, Haramati A. The effect of a Mind-Body-Medicine course on medical student empathy: A pilot study. *Med Ed Online*. 2016;**21**:e31196.

Finkelstein C, Brownstein A, Scott C, et al. Anxiety and stress reduction in medical education: An intervention. *Med Educat.* 2007;**41**:258-64.

Kelly J, Bradlyn A, Dubbert P, *et al*. Stress management training in medical school. *J Med Educ.* 1982;**57**:91-9.

Kraemer K, Luberto C, Wasson R, *et al*. Does mind-body skills training help medical students to more effectively tolerate distressing emotions? *Integrative Medical Research*. 2015;**4**:83-4.

Mitchie S, Sandhu S. Stress management for clinical medical students. *Med Educ.* 1994;**28**:528-33.

Rosenzweig S, Reibel D, Greeson J, *et al*. Mindfulness-based stress reduction lowers psychological distress in medical students. *Teach Learn Med.* 2003;**15**:88-92.
Zuardi A, De Guerra Prota F, CM D-B. Reduction of the anxiety of medical students after curricular reform [Redução da ansiedade de estudantes de medicina após reforma curricular]. *Revista Brasileira de Psiqiatria*. 2008;**30**:136-8.

Historically Controlled Studies:

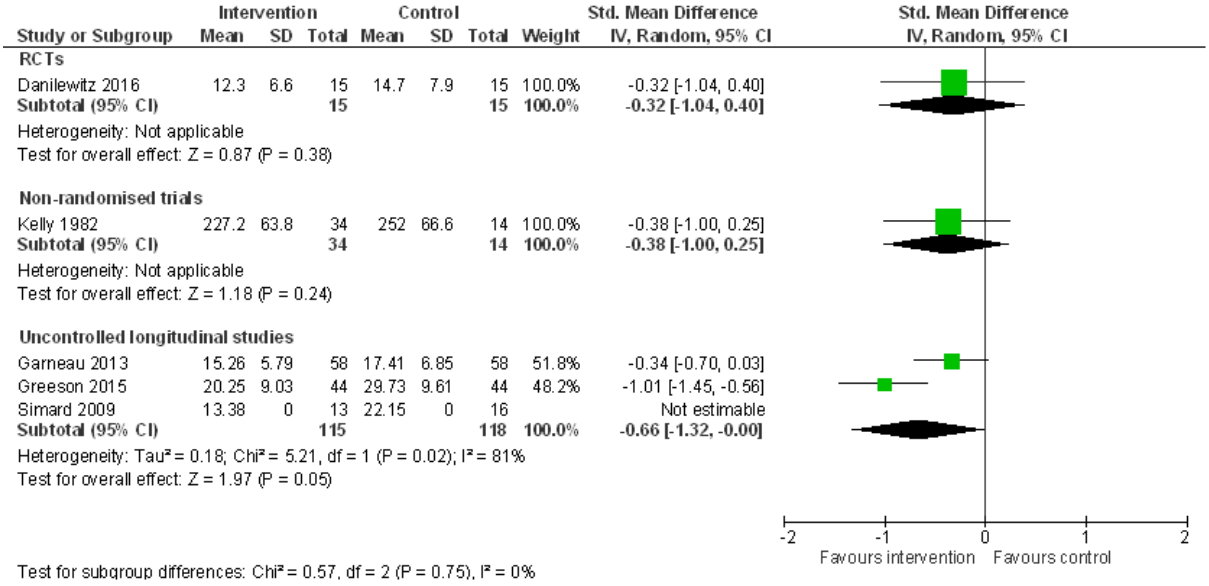
Al-Faris E, Naeem N, Irfan F, *et al*. Student centered curricular elements are associated with a healthier educational environment and lower depressive symptoms in medical students. *BMC Medical Education.* 2014;**14**:192.
Holm M, Tyssen R, Stordal K, *et al*. Self-development groups reduce medical school stress: A controlled intervention study. *BMC Med Educ*. 2010;**10**:23.
Melo-Carrillo A, Van Oudenhove L, Lopez-Avila A. Depressive symptoms among Mexican medical students: High prevalence and the effect of a group psychoeducation intervention. *J Affect Disord*.2102;**136**:1098-103.
Slavin S, Schindler DL, Chibnall JT. Medical student mental health 3.0: Improving student wellness through curricular changes. *Acad Med.* 2014;**89**:573-7.
Thompson D, Geobert D, Takeshita J. A program for reducing depressive symptoms and suicidal ideation in medical students. *Acad Med*. 2010;**85**:1635-9.

Uncontrolled Longitudinal Studies:

Bansal R, Gupta M, Agarwal B, *et al*. Impact of short term yoga intervention on mental well being of medical students posted in community medicine: A pilot study. *Indian J Comm Med*. 2013;**38**:105-8.
Bond AM, HF, Lemaster C, Shaw S, *et al*. Embodied health: The effects of a mind-body course for medical students. *Med Ed Online*. 2013;**18**:20699.
Bughi S, Sumcad J, Bughi S. Effect of brief behavioral intervention program in managing stress in medical students from two Southern California universities. *Med Edu Online*. 2006;**11**:4593.
Dutton M, Arun P, Talley J, *et al*. Mind-body skills training for improving emotional well-being in medical students. *Explore*. 2013;**9**:328.
Gallagher T, Munro J, Kahl L. Development and implementation of a clerkship counseling hotline. *Teach Learn Med.* 2005;**17**:80-4.
Garneau K, Hutchinson T, Zhao Q, *et al*. Cultivating person-centered medicine in future physicians. *European Journal for Person Centered Healthcare*. 2013;**1**:468-77.
Greeson J, Toohey M, Pearce M. An adapted, four-week mind-body skills group for medical students: Reducing stress, increasing mindfulness, and enhancing self-care. *Explore*. 2015;**11**:186-92.
Hassed C, de Lisle S, Sullivan G, *et al.* Enhancing the health of medical students: Outcomes of an integrated mindfulness and lifestyle program. *Adv in Health Sci Educ*. 2009;**14**:387-9.
Kötter T, Pohontsch N, Voltmer E. Stressors and starting points for health-promoting interventions in medical school from the students' perspective: A qualitative study. *Perspect Med Educ.* 2015;**4**:128-35.
Mercer A, Warson E, Zhao J. Visual journaling: An intervention to influence stress, anxiety and affect levels in medical students. *The Arts in Psychotherapy*. 2010;**37**:143-8.
Moffat K, McConnachie A, Ross S, *et al*. First year medical student stress and coping in a problem-based learning medical curriculum. *Med Ed.* 2004;**38**:482-91.
Simard A-A. Impact of a short yoga intervention on medical students' health: A pilot study. *Med Teach.* 2009;**31**:10.

Wild K, Scholz M, Ropohl A, *et al*. Strategies against burnout and anxiety in medical education - implementation and evaluation of a new course on relaxation techniques (Relacs) for medical students. *PLoS One*. 2014;**9**:e114967.

Supplementary Figure 2:
Random effects standardised mean difference (SMD), and accompanying 95% confidence interval (CI), on self-reported stress scores at the post-intervention assessment.



Supplementary Figure 3:
Random effects standardised mean difference (SMD), and accompanying 95% confidence interval (CI), on self-reported stress scores at the final follow-up assessment.