CrossMark

REVIEW

Complementary and Alternative Medicine Treatments for Generalized Anxiety Disorder: Systematic Review and Meta-analysis of Randomized Controlled Trials

Hrvoje Barić · Veljko Đorđević · Ivan Cerovečki · Vladimir Trkulja

ABSTRACT

Introduction: The objective was to evaluate efficacy/safety of complementary and alternative medicine (CAM) methods for generalized anxiety disorder (GAD) based on randomized controlled trials in adults.
Methods: *Data sources*. Six electronic databases ("generalized anxiety (disorder)" and "randomized trial") and reference lists of identified publications were searched to March 2017.

Enhanced content   To view enhanced content for this article go to https://doi.org/10.6084/m9.figshare.5896606.

Electronic supplementary material   The online version of this article (https://doi.org/10.1007/s12325-018-0680-6) contains supplementary material, which is available to authorized users.

H. Barić
Department of Neurosurgery, University Hospital Center Zagreb, Zagreb University School of Medicine, Zagreb, Croatia

V. Đorđević
Center for Palliative Medicine, Medical Ethics and Communication Skills (CEPAMET), Zagreb University School of Medicine, Zagreb, Croatia

I. Cerovečki
Croatian Institute of Public Health, Zagreb, Croatia

V. Trkulja (✉)
Department of Pharmacology, Zagreb University School of Medicine, Zagreb, Croatia
e-mail: vtrkulja@mef.hr

Study selection. Eligibility: full-text publications (English, German language); CAM versus conventional treatment, placebo/sham or no treatment; GAD diagnosed according to standard criteria; and a validated scale for disease severity. Of the 6693 screened records, 32 were included (18 on biologically-based therapies, exclusively herbal preparations; eight on manipulative and body-based therapies; and three on alternative medical systems and three on mind–body therapies). *Data extraction*. Cochrane Collaboration methodology was used for quality assessment and data extraction.
Results: Direct comparisons of Kava Kava (*Piper methysticum*) extracts to placebo (4 quality trials, $n = 233$) were highly heterogeneous. Network meta-regression reduced heterogeneity and suggested a modest Kava effect [end-of-treatment Hamilton Anxiety scale score difference adjusted for baseline scores and trial duration: -3.24 (95% CI -6.65, 0.17; $P = 0.059$), Kava Kava 4 arms, $n = 139$; placebo 5 arms, $n = 359$]. Lavender (*Lavandula angustifolia*) extract (1 quality trial, 10 weeks, $n = 523$) and a combination of extracts of *C. oxycantha*, *E. californica* and magnesium (1 quality trial, 12 weeks, $n = 264$) were superior to placebo and balneotherapy was superior to paroxetine (1 quality trial, 8 weeks, $n = 237$) indicating efficacy. All other trials were small and/or of modest/low quality and/or lacked assay sensitivity. Safety reporting was poor.

△ Adis

Conclusion: Evidence about efficacy/safety of most CAM methods in GAD is limited. Apparent efficacy of certain herbal preparations and body-based therapies requires further confirmation.

Keywords: Complementary and alternative medicine; Generalized anxiety disorder; Meta-analysis; Systematic review; Psychiatry

INTRODUCTION

It is estimated that anxiety disorders affect over a tenth of the population with increasing incidence [1, 2]. At the same time, they are under-recognized and under-treated and quality of care for the affected individuals is inadequate [3]. Anxiety is disabling for different dimensions of everyday life, reduces productivity and increases the risk of other diseases [4]. Among anxiety disorders, generalized anxiety disorder (GAD) has the highest prevalence: according to some reports, it affects 4–6% of the general population [5].

Recommended treatments for anxiety disorders include cognitive–behavioral therapy (CBT) and medications, primarily antidepressants and benzodiazepines. The latter have proven efficacy; however, they are associated with serious adverse effects and substantial limitations in application [6–8]. Non-pharmacological techniques are also efficient, some of them even superior to medication, yet, due to socioeconomic and other obstacles, a sizeable proportion of patients do not experience their benefits [9]. Combination of conventional treatments relieves symptoms in 50–65% of the patients, although many continue experiencing symptoms despite the treatment [10]. Consequently, there is a constant rise in interest for alternative treatment options. Some of them, in particular herbal remedies, are known as folk medicine and have been used for centuries [11, 12]. However, efficacy and safety of alternative methods have only been adequately addressed in clinical trials over the past 10–15 years [13]. According to the National Center for Complementary and Alternative Medicine of the National Institutes of Health

(NIH), complementary and alternative (CAM) treatment methods are classified into five groups: (1) natural remedies (food supplements, herbaceuticals, etc.); (2) mind and body medicine (meditation, acupuncture); (3) manipulative and body-based procedures (spinal manipulation, massage, etc.); (4) complete medical systems (traditional Chinese medicine, Ayurvedic medicine, etc.); and (5) other CAM methods (e.g., light therapy, etc.) [14, 15]. There are several major reasons to perform a comprehensive systematic review of CAM methods for the treatment of GAD in order to assess their efficacy and safety. Firstly, anxious individuals are prone to using CAM methods—it is estimated that half of them use a CAM treatment [16–19]. Secondly, it is estimated that half of GAD patients simultaneously use conventional and alternative treatments and there is a lack of studies which assess risks and benefits of such combined strategies [20]. Thirdly, the number of trials of CAM methods in GAD is rising, as is the number of the used CAM modalities [21].

The present systematic review aims to evaluate empirical evidence of clinical efficacy and safety of CAM methods in the treatment of GAD in adults, as assessed in randomized controlled trials (RCTs).

METHODS

Eligibility Criteria

Eligible for inclusion were RCTs comparing the efficacy and/or safety of any CAM treatment, alone or in a combination with another conventional or CAM treatments to a conventional treatment or a combination of treatments, placebo/sham treatments or no treatment in adults (≥ 18 years of age) with GAD diagnosed according to one of the defined criteria: Diagnostic and Statistical Manual of Mental Disorders (DSM), Chinese Classification of Mental Disorders (CCDM) or International Classification of Diseases (ICD). CAM interventions were defined as all treatments not listed as standard in the National Institute for Health and Care Excellence (NICE) guidelines [22]. Patients had to be free of psychiatric comorbidities such as

bipolar disorder, schizophrenia, major depressive disorder, posttraumatic stress disorder, organic brain syndrome or substance abuse, and condition severity had to be assessed using one of the established validated anxiety rating scales. Studies had to be published in full-text in the English or German languages.

Outcomes

When symptom alleviation was the trial objective, primary outcome was reduction of anxiety (vs. baseline) or alternatively severity of anxiety at the end of treatment, quantified using the scale defined as a "primary instrument" in the trial. We used the former outcome whenever reported so that standard deviation (SD) could be reliably extracted, but avoided rough approximations based on summary baseline and end-of-study data or SD imputations—in such cases, we preferred straightforward reported end-of-study anxiety scores. Secondary outcome was the proportion of patients responding to treatment (as per the definition in the trial). For "withdrawal trials" (patients with controlled symptoms switched to the test/control treatment to assess the ability to prevent relapse), the outcome was risk (hazard) of relapse. Incidence or incidence rate of adverse events (AEs) was considered in all trials.

Information Sources and Literature Search

We searched six electronic databases [Medline, Web of Science, EBSCO (Academic Search Complete, CINHAL and ERIC), Scopus—Health Sciences, Google Scholar and all Cochrane Library] up to March 2017 using the following key words: "generalized anxiety" OR "generalized anxiety disorder" AND "randomized trial". Such a broad and nonspecific strategy was used to ensure that relevant trials and all evaluated CAM interventions were identified. We also manually searched the reference lists of identified publications and previously published systematic reviews. Only published data were used in this review. In the case of multiple publications on the same trial, the one with the most complete data was used.

Study Selection and Data Extraction

Using the pre-specified strategy and eligibility criteria, literature search, study selection and data extraction were performed independently by two reviewers. Disagreements were resolved by a consensus. One reviewer entered data into a predefined spreadsheet and the second reviewer checked the entries for accuracy. The following was extracted from each trial: (1) data on participants (number, age, gender per group, diagnostic criteria); (2) trial data (design, duration); (3) intervention data (type, dosing and dosing schedule/mode of administration); and (4) predefined primary and secondary outcomes. For anxiety severity scores, data were extracted as mean $\pm$ SD for the number of patients reported; for the proportion of responders and the incidence of AEs, data were extracted as n/N using all patients who received the assigned treatment as a denominator; and for the risk of relapse, data were extracted as the effect measure. Outcome data extraction was carried out using the Cochrane Collaboration methodology [23].

Assessment of the Risk of Bias (Study Quality)

In a non-blinded manner, two investigators independently assessed study quality using the Cochrane Collaboration risk of bias tool to evaluate the quality of randomization, allocation concealment (sampling), blinding of participants and personnel (performance), blinding of outcome assessment (detection), completeness of outcome data (attrition) and reporting. Disagreements were resolved by a consensus.

Data Synthesis (Meta-analysis)

For direct pairwise comparative trials of a reasonable clinical homogeneity, we anticipated standard random-effects meta-analysis to generate pooled estimates of efficacy outcomes: weighted (or standardized) mean difference; and Mantel–Haenszel odds ratio and inverse variance method for (log) hazard ratios. However, we used the

Hartung–Knapp–Sidik–Jonkman correction for the standard error of the estimate [24]. Where three or more trials were available, we also determined prediction intervals as the best illustration of the heterogeneity of effects [25]. Only one treatment (Kava Kava) was evaluated in three or more trials of the same design and comparator (placebo), but the number of trials and patients was small; trials varied in duration and the reported primary efficacy outcome was end-of-treatment anxiety score (with different baseline scores across trials). Conventional pooled estimates of differences versus placebo were imprecise and highly heterogeneous. We attempted to improve the estimates by "borrowing" additional information from active-controlled Kava Kava trials and placebo-controlled trials of other biologically-based treatments, and by adjusting the estimates for baseline scores and trial duration using network meta-regression. For this purpose, we used the approach based on reconstructed patient-level data [26]. For binary outcomes, a study is reconstructed so that each contributing patient is represented by a record with a variable representing the study, a variable representing the treatment, and a variable depicting the outcome. For continuous outcomes (summarized as mean ± SD), a study is reconstructed so that for each arm a sample from a normal distribution with these parameters is drawn (n = number of subjects per arm) and a difference in mean (SD) of the drawn sample versus the reported parameters is adjusted for using linear transformation [26]. The method essentially provides the option of individual (notional) patient-level analysis. Patient-level covariates are not available, but may be substituted by average values by arm [26]. The method maintains randomization, allows that each patient contributes equally to the estimates, and allows for the inclusion of two- and multi-arm studies to generate direct, indirect, arm-level and combined estimates [26]. Originally, the method is a generalization of the fixed-effect meta-analysis. Considering clinical heterogeneity of the included studies, we considered it more appropriate to apply random-effects analysis (restricted maximum likelihood estimation) by fitting generalized linear mixed models to

binary [27] or general linear mixed models to continuous data [28] with study × treatment as a random effect. Where feasible, random-effects pooled estimates of treatment differences in incidence of AEs (Mantel–Haenszel relative risk) or incidence of AEs by treatment (Freeman–Tukey double-arcsine transformation) were generated. We used CMA v.3 (Biostat Englewood, NJ, USA) for standard meta-analysis and SAS for Windows 9.4 (SAS, Cary, NC, USA) for regression-based analysis (procglimmix, proc mixed).

Grading the Evidence

Paucity of RCTs of the same design and comparator prevented the intended formal assessment of "body of evidence" for individual CAM treatments using the GRADE system. However, we attempted to assess the level of (un)certainty about efficacy/lack of efficacy of a particular treatment by considering individual trials with respect to quality (risk of bias), precision and firmness/fragility (in particular for proportions) of the estimates (determined by sample sizes) and consistency of findings when more than one trial by treatment was available. For trials claiming efficacy based on "no statistically significant difference" versus a presumed active comparator, we considered evidence of assay sensitivity and the presence/absence of defined equivalence/non-inferiority margins.

Compliance with Ethics Guidelines

This article is based on previously conducted studies and does not involve any new studies on humans or animals performed by any of the authors.

RESULTS

Study Eligibility

Of the 7781 identified records (7734 electronic databases, 47 other sources) (Fig. 1), 6693 non-duplicates were screened, 54 were retrieved in full-text and 22 were excluded (Supplementary
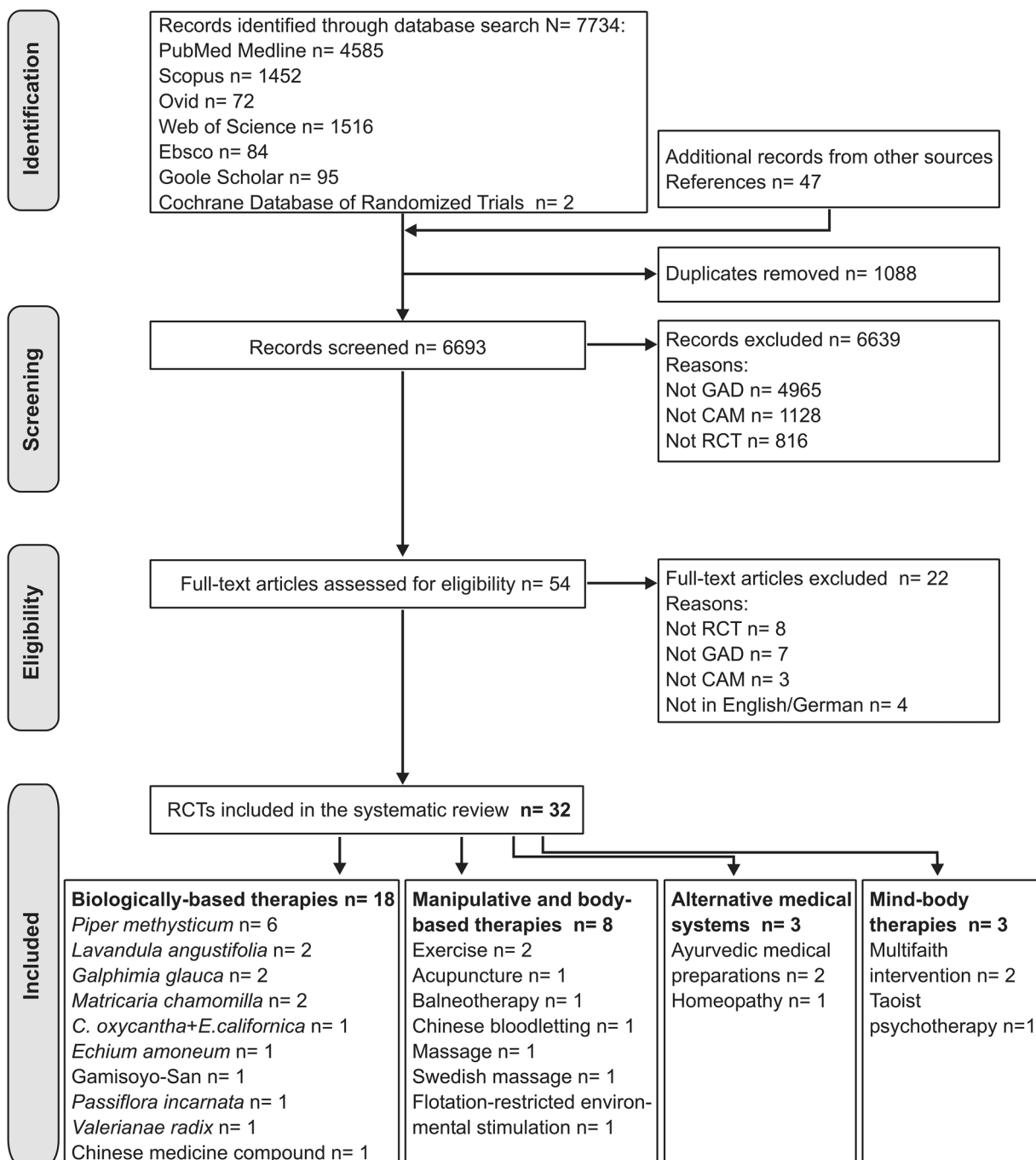
Fig. 1 PRISMA flow diagram. *GAD* general anxiety disorder, *CAM* complementary and alternative medicine, *RCT* randomized controlled trial, *C. oxycantha Crataegus oxycantha, E. californica Eschscholtzia californica*

eAppendix 1) resulting in 32 included RCTs [29–60] (Fig. 1). Most trials (*n* = 18) evaluated biologically-based therapies [29–46], i.e., oral herbal preparations with extracts of Kava Kava (*Piper methysticum*), Lavender (*Lavandula*

angustifolia), *Galphimia glauca* and Chamomile (*Matricaria chamomilla*) evaluated in more than one trial (Fig. 1). Manipulative and body-based therapies were evaluated in eight [47–54] and alternative medical systems [55–57] and

Table 1 Main characteristics of the included randomized controlled trials

Main study properties			Subjects		Treatments, number of randomized patients		Assessments	
Author, year (references)	Diagnostic criteria	Design	Women (%)	Age (years) mean ± SD; range	Test daily dose (unless stated otherwise)	Control	Primary	Secondary
Biologically-based therapies								
Volz [29]	DSM-III-R	24-week, DB, parallel, 2-arm	73	53.9 ± 16.3	Kava Kava extract 3 × 70 mg kava-lactones; n = 52	PBO n = 49	HAM-A	HAM-A subscores, CGI, SCL-90-R, Bf-S
Malsch [30]	DSM-III-R	5-week, DB, parallel, 2-arm	37.5	21–75	Kava Kava extract 1 × 35 to 3 × 70 mg kava-lactones; n = 20	PBO n = 20	HAM-A, Bf-S	EAAS, CGI
Connor [31]	DSM-IV	3-week, DB, parallel, 2-arm	82	51.7 ± 11.6; 31–75	Kava Kava extract 2 × 70 to 2 × 140 kava-lactones; n = 19	PBO n = 18	HAM-A, HADS	SARA
Sarris [32]	DSM-IV	6-week, DB, parallel, 2-arm	65	30.1 + 8.8	Kava Kava extract 2 × 60 to 2 × 120 mg kava-lactones; n = 27	PBO n = 31	HAM-A	BAI, MADRS
Wheatley [33]	DSM-IV	2-week, open, 2 × 2 cross-over, n = 24	37.5	41.4 ± 13.2; 23–66	Kava Kava extract 1 × 120 mg kava-lactones	Kava extract 3 × 45 mg	HAM-A	
Boerner [34]	ICD-10	8-week, DB, parallel, 3-arm	84.3	20–71	Kava Kava extract 1 × 120 mg kava-lactones; n = 43	BUSP 2 × 5 mg n = 43; OPIP 2 × 50 mg n = 43	HAM-A	BOEAS, SAS, CGI, Bf-S, SF-B, AL
Woelk [35]	DSM-IV	6-week, DB, parallel, 2-arm	76.6	21–65	*Lavandula angustifolia* extract 1 × 80 mg; n = 40	LORAZ 1 × 0.5 mg; n = 37	HAM-A	CGI, SAS, PSWQ-PW, Sf-36, Sleep Diary

Table 1 continued

Main study properties			Subjects		Treatments, number of randomized patients		Assessments	
Author, year (references)	Diagnostic criteria	Design	Women (%)	Age (years) mean ± SD; range	Test daily dose (unless stated otherwise)	Control	Primary	Secondary
Kasper [36]	DSM-IV-TR	10-week, DB, parallel, 4-arm	71.4	45.8 ± 12.0	Lavandula angustifolia extract 1 × 160 mg; n = 128 or 1 × 80 mg, n = 135	PAROX 1 × 20 mg n = 137 PBO n = 136	HAM-A	CAS, HAM-D, CGI, SDS, SF-36, PWC-20
Herrera [37]	DSM-IV	4-week, DB, parallel, 2-arm	76.9	37.8 ± 11.3	Galphimia glauca extract 2 × 1 348 µg galphimin B; n = 72	LORAZ 2 × 1 mg n = 80	HAM-A	CGI, PGE
Herrera [38]	DSM-IV	12-week, DB, parallel, 2-arm	85.3	40 ± 10.8	Galphimia glauca extract 2 × 1 to 2 × 2 175 µg galphimin B; n = 94	LORAZ 2 × 1 to 2 × 2 0.5 mg; n = 97	HAM-A	CGI, PGI, tolerability
Amsterdam [39]	DSM-IV	8-week, DB, parallel, 2-arm	49.7	45.7 ± 12.8	Chamomile extract 1 × 1 to 5 × 1 2.6 mg apigenin; n = 28	PBO n = 29	HAM-A	BAI, PGWB, CGI-S, tolerability
Mao [40]	DSM-IV	25-week, withdrawal DB parallel, 2-arm	69.9	47.3 ± 15.4	Chamomile extract 3 × 1 6 mg apigenin; n = 46	PBO n = 47	CGI-S, SCID-I	GAD-7, PGWB, HAM-A, BAI, TESS
Hanus [41]	DSM-III-R	12-week, DB, parallel, 2-arm	81	44.6; 18–82	Crataegus oxyacantha 75 mg + Eschscholtzia californica 20 mg extracts + Mg^{2+} 75 mg 2 × 2 capsules; n = 130	PBO = 134	HAM-A	Patient self-assessment VAS score, CGI
Sayyah [42]	DSM-IV-TR	8-week, add-on, DB, parallel, 2-arm	48.6	25.5 ± 3.3	Echium amoneum 3 × 750 mg extract + FLUOX 1 × 20 mg; n = 19	PBO + FLUOX 1 × 20 mg n = 18	HAM-A	

Table 1 continued

Main study properties			Subjects		Treatments, number of randomized patients		Assessments	
Author, year (references)	Diagnostic criteria	Design	Women (%)	Age (years) mean ± SD; range	Test daily dose (unless stated otherwise)	Control	Primary	Secondary
Park [43]	DSM-IV	8-week, DB, parallel, 3-arm	76.2	39.2 ± 11.4	Mixed extract[a] 10 herbs 3 × 7.7 g / Individual extraction; n = 49 / Simultaneous extraction; n = 49	PBO n = 49	HAM-A	K-STAI, PSQI, K-BDI, SCL-90-R, WHO-QOL-BREF
Akhondzadeh [44]	DSM-IV	4-week, DB, parallel, 2-arm	55.6	19–47	*Passiflora incarnata* extract 1 × 45 drops; n = 18	OXAZ 1 × 30 mg n = 18	HAM-A	
Andreatini [45]	DSM-III-R	4-week, DB, parallel, 3-arm	52.8	41.1 ± 9.3	Valeriana extract 1-3 × 50 mg; n = 12	DIAZ 1-3 × 2.5 mg n = 12 / PBO n = 12	HAM-A	STAI
Wang [46]	DSM-IV	24 +24-week withdrawal, open, parallel, add-on, 2-arm	43.9	37.3 ± 13.0	Cognitive therapy + 2 × 10 g crude powder mix of 14 herbs; n = 93	Cognitive therapy + PAROX 1 × 20 to 1 × 60 mg n = 109	HAM-A, SAS	
Manipuative and body-based therapies								
Eich [47]	ICD-10	4-week, DB, parallel, 2-arm	57.1	43.1 ± 13.5	Acupuncture × 10; n = 7	Sham acupuncture; n = 6	CGI	HAM-A, HAM-D, Bf-S, B-L
Merom [48]	DSM-IV	8-week, open, parallel, 2-arm	78.4	39.0 ± 11.9	Exercise-enhanced cognitive behavioral therapy; n = 11	Cognitive behavioral therapy + educational meetings; n = 15	DASS-21	Time walking for "recreation"

Table 1 continued

Main study properties			Subjects		Treatments, number of randomized patients		Assessments	
Author, year (references)	Diagnostic criteria	Design	Women (%)	Age (years) mean ± SD; range	Test daily dose (unless stated otherwise)	Control	Primary	Secondary
Dubois [49]	DSM-IV	8-week, open, parallel, 2-arm	76.4	51.7 ± 11.4	Balneotherapy daily 3 weeks; $n = 117$	PAROX 20-50 mg/day; $n = 120$	HAM-A	MADRS, CGI-S, CGI-I, BATE, STAI
Sherman [50]	DSM-IV	12-week, open, parallel, 3-arm	76.5	42.9 ± 11.4	Therapeutic massage × 10; $n = 22$	Thermotherapy × 10; $n = 23$ Relaxing room × 10; $n = 22$	HAM-A	STAI, HAM-D, QIDS-SR, POMS, Q-LES-Q
Herring [51]	DSM-IV	6-week, assessor B, parallel, 3-arm	100	23.5 ± 9.9	Twice/week exercise: aerobic $n = 10$ or resistance $n = 10$	No treatment (postponed treatment); $n = 10$	AIDS-IV	PDSQ, PSQW, BDI-II
Ma [52]	CCDM-3	4-week, assessor B, parallel, 2-arm	45.7	22-64	Chinese bloodletting thrice/week + PAROX 20 mg/day; $n = 35$	PAROX 20 mg/day; $n = 35$	SAS	SAS effective rate
Jonsson [53]	NA	24-week, open, parallel, 2-arm	72	43.0 ± 13.4	Flotation in salt-saturated water × 12 over 4 weeks; $n = 25$	No treatment (postponed treatment); $n = 25$	GAD-Q-IV, PSWQ	MADRS-S, PSQI, DERS, MAAS, EDN
Rappaport [54]	DSM-IV	12-week, assessor B, parallel, 2-arm	75	36.7 ± 16.8	Swedish massage daily; $n = 23$	Sham (light touch); $n = 24$	HAM-A	STAI, HAM-D, QIDS-SR, POMS, Q-LES-Q

Table 1 continued

Main study properties			Subjects		Treatments, number of randomized patients		Assessments	
Author, year (references)	Diagnostic criteria	Design	Women (%)	Age (years) mean ± SD; range	Test daily dose (unless stated otherwise)	Control	Primary	Secondary
Alternative medical systems								
Bonne [55]	DSM-IV	10-week, DB, parallel, 2-arm	59.1	46.1 ± 12.9	Homeopathic oral remedy 1 × day; n = 22	PBO; n = 22	HAM-A	HAM-D, BDI, STAI, BSI, PGWB, VAS
Tubaki [56]	DSM-IV TR	4-week, open, parallel, 3-arm	18.1	28.2 ± 5.7	Ayurvedic oral[b] 2 × 100 mg/day; n = 24; Oral + dripping medicated oil on forehead (1st week); n = 24	CLONAZ 0.25 + 0.5 mg/day; n = 24	HAM-A. CGI-I, BAI	BDI, ESS, WHO-QoL-BREF, CGI-S, CGI-I
Gupta [57]	DSM-IV TR	11-week, single blind[c] parallel, 2-arm	54.4	NA	Ayurvedic preparation oral 3 × 1 g/day; n = 57	Placebo; n = 57	HAM-A	
Mind–body therapies								
Zhang [58]	CCDM-2-R	24-week, open, parallel, 3-arm	44.1	34.8 ± 11.3	Chinese cognitive psychoth. (CTCP) 1-2 times/week; n = 46	DIAZ Eq. 10-20 mg/day; n = 48 DIAZ eq. + CTCP; n = 49	SCL-90	EPQ, CSQ, Type A personality scale

Table 1 continued

Main study properties			Subjects		Treatments, number of randomized patients		Assessments	
Author, year (references)	Diagnostic criteria	Design	Women (%)	Age (years) mean ± SD; range	Test daily dose (unless stated otherwise)	Control	Primary	Secondary
Koszycki [59]	DSM-IV	12-week, open, parallel, 2-arm	59.1	43.5 ± 14.4	Multifaith spiritually-based intervention 1/week; n = 11	Cognitive behavioral therapy 1/week; n = 11	HAM-A, PSWQ	CGI-S, IUS, BDI, SAS-SR,
Koszycki [60]	DSM-IV	12-week, open, parallel, 2-arm	65.2	42.4 ± 16.6	Multifaith spiritually-based intervention 1/week; n = 11	Supportive psychotherapy 1/week; n = 12	HAM-A, PSWQ	CGI-S, IUS, BDI, SAS-SR, DSES, AUIE

Assessor B assessor-blind, *BUSP* buspirone, *CCDM* Chinese Classification and Diagnostic Criteria for Mental Disorders (2nd edition, revised; 3rd edition), *CLONAZ* clonazepam, *DB* double-blind, *DIAZ* diazepam, *DSM (III, III-R, IV, IV-TR)* Diagnostic and statistical manual of mental disorders (3rd, 4th edition, revision, text revision), *FLUOX* fluoxetine, *ICD* International classification of diseases, *LORAZ* lorazepam, *OPIP* opipramol, *OXAZ* oxazepam, *PAROX* paroxetine, *PBO* Placebo

Psychiatric tools: *AIDS-IV* Anxiety Disorders Interview Schedule, *AL* quality of life questionnaire, *AUIE* age universal intrinsic–extrinsic scale, *BAI* Beck anxiety inventory, *BATE* bonis anxiety trait-state, *Bf-S* subjective well-being scale, *B-L* Beschwerden Liste, *BOEAS* boerner anxiety scale, *BSI* brief symptom inventory, *CAS* covi anxiety scale, *CGI* clinical global impression (*I* improvement, *S* severity), *CSQ* coping style questionnaire, *DASS-21* Depression Anxiety Stress Scale 21, *DERS* Dysfunctional Emotional Regulation Scale, *DSES* Daily Spiritual Experience Scale, *EAAS* Erlangen Anxiety and Aggression Scale, *EDN* Experienced Deviation from Normal scale, *EPQ* Eysenck Personality Questionnaire, *ESS* Epworth Sleep Scale, *GAD-7* Generalized Anxiety Disorder 7-item scale, *GAD-Q-IV* Dimensional scoring from the Generalized Anxiety Disorder Questionnaire, *HAM-A* Hamilton Anxiety Scale, *HAM-D* Hamilton Depression Scale, *IUS* Intolerance of Uncertainty Scale, *K-BDI* Korean version Beck Depression Inventory, *K-STAI* Korean State-trait Anxiety Inventory, *MAAS* Mindful Attention and Awareness Scale, *MADRS* Montgomery-Asberg Depression Rating Scale, *PDSQ* Psychiatric Diagnostic Screening Questionnaire, *PGE* Patient Global Evaluation scale, *PGI* Patient Global Impression scale, *PGWB* Psychological General Well-being index, *POMS* Profile of Mood States, *PSQI* Pittsburgh Sleep Quality Index, *PSWQ* Penn State Worry Questionnaire (PW past week), *PWC-20* Physician Withdrawal Checklist, *QIDS-SR* Quick Inventory of Depression Symptomatology-Self report, *Q-LES-Q* Quality of Life Enjoyment and Satisfaction Questionnaire, *SARA* Self-Assessment of Resilience and Anxiety, *SAS* Self-rating Anxiety Scale, *SCL-90-R* Symptom Checklist 90-revised, *SDS* Sheehan Disability Scale, *Sf-B* sleep questionnaire, *Sf-36* Health Survey Questionnaire, *STAI* State-trait Anxiety Inventory, *TESS* treatment-emergent Symptom Scale, *VAS* Visual Analogue Scale, *WHO-QOL-BREF* WHO quality of life scale abbreviated version

[a] Korean herbal preparation called Gamisoyo-San

[b] Ayurvedic oral preparation *Manasamitra Vataka*; medicated oil dripped on the forehead, called *Shirodhara*

[c] "Single blind"—unclear whether patients or investigators/outcome assessors. Ayurvedic medicine called *Sarasvata choorna*

mind–body therapies [58–60] in three trials each (Fig. 1).

All trials enrolled exclusively GAD patients or reported outcomes specifically for subsets of GAD patients, except for two [29, 30] (out of four) placebo-controlled RCTs of Kava Kava extract. One [29] was the first placebo-controlled RCT of Kava Kava specifically in anxiety spectrum disorders (agoraphobia, specific phobia, social phobia, GAD and adjustment disorder with anxiety). Use of Kava Kava in psychiatry attracted much attention in the 1990s, but this specific trial was the first to include patients in line with the DSM-III-R, and thus the first with clearly operationalized diagnoses as inclusion criteria. The prevalence of patients in the trial with individual specific conditions was not stated, but patients were mostly comorbid with phobia and GAD [29]. The other trial [30] was the second placebo-controlled trial of Kava Kava in anxiety which used the same criteria and included diagnoses as the previous one (14/40 patients suffered from social phobia, 12/40 suffered from GAD and 11/40 from simple phobia). Together with an additional two placebo-controlled [31, 32] and one active-controlled trial [34] which included exclusively GAD patients, these two represented the largest pool of RCTs of Kava Kava in psychiatric patients included in line with defined, standardized diagnostic criteria, and we reasoned that they could reasonably be considered a part of the "evidence base" for the evaluation of efficacy/safety of Kava Kava in GAD.

Study Characteristics and Quality (Risk of Bias)

All but one trial [33] were parallel-group trials of varying duration (up to 24 weeks), and in all but a few (Table 1) women prevailed. Biologically-based therapies were evaluated as mono-treatments except in two "add-on" trials [42, 46] (Table 1). All but two [33, 46] were double-blind (Table 1). One trial [40] was a withdrawal trial, while others included symptomatic patients (Table 1) with HAM-A as the primary assessment tool (Table 1). Only two out of twelve trials of other interventions were double-blind

[47, 55] (Table 1), and a further three were assessor blind [51, 52, 54] (Table 1). They evaluated a variety of interventions of which only a "multifaith intervention" occurred in more than one trial [59, 60] (Table 1).

The main quality issues were related to performance bias (open-label trials), lack of explicit statement of blinded outcome assessment (detection bias) particularly in open-label trials (Table 1) and attrition bias; four trials [37, 38, 46, 48] had a high risk of attrition bias and the level of risk was unclear in a further six [31, 39, 45, 49, 57, 58]. Detailed quality assessment is available in the supplementary material (Supplementary eFigure 1 and eAppendix 2).

Efficacy

Kava Kava (Piper methysticum) Extracts
Of the six included RCTs, one cross-over trial [33] (Table 1) reported no difference between two dosing schedules and was uninformative regarding efficacy since it lacked assay sensitivity. The remaining five parallel-group RCTs (3–24 weeks), all with an overall low risk of bias and with intention-to-treat (ITT) efficacy analysis, compared Kava Kava to placebo (four trials [29–32]) or versus active treatments (one trial [34]) using HAM-A as the main instrument for quantification of anxiety (Table 1).

Direct Comparison of Kava Kava versus Placebo
In each of the four trials [29–32], baseline HAM-A scores were comparable between treatments, but varied across trials (from 13 to 31.4 score points) (Fig. 2). Primary efficacy outcome in three trials was end-of-study HAM-A score, while one (Malsch 2001) [30] reported a greater median reduction with Kava Kava than with placebo (− 7.5 vs. 1, $P = 0.010$). Kava–placebo differences in the remaining three (a total of 96 patients on Kava Kava, 97 on placebo) were highly heterogeneous (from favoring Kava to favoring placebo), best illustrated by an extremely wide prediction interval (PI) around the pooled Kava–placebo difference (from − 31.7 to 29.6) and by high inconsistency ($I^2 = 80.2\%$) (Fig. 2), resulting in an imprecise (95% CI
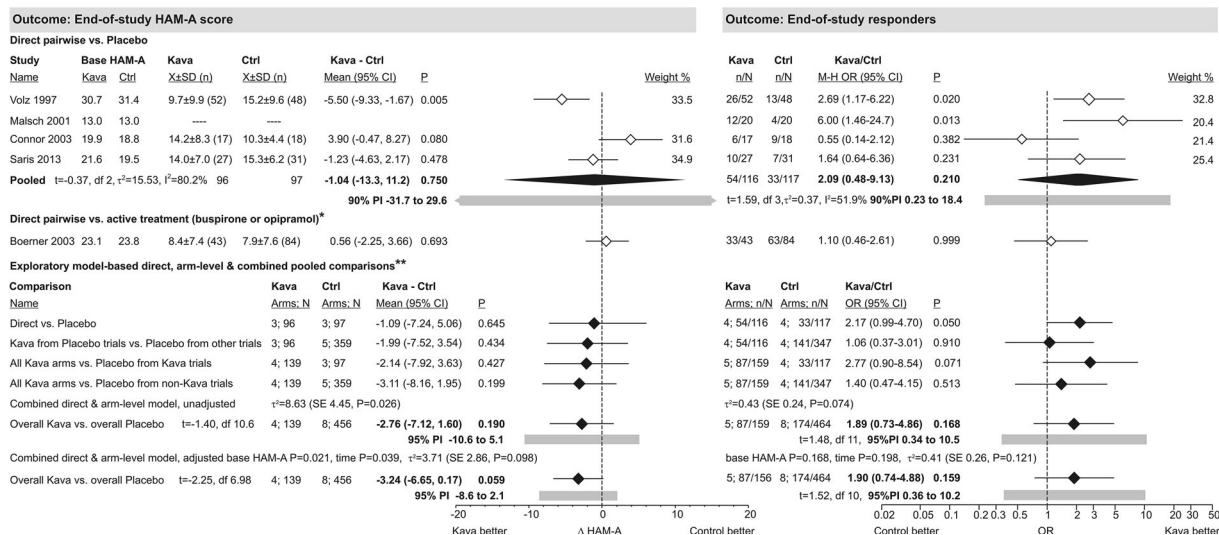
Fig. 2 Meta-analysis of RCTs comparing Kava Kava extracts (120 or 210 mg/day fixed or titrated to 240–280 mg/day) to placebo or active treatments. Meta-analysis of direct pairwise trials versus placebo is by a standard random-effects method (weighted mean difference or Mantel–Haenszel odds ratio) with Hartung–Knapp–Sidik–Jonkman correction. For the single trial versus active treatments, mean difference from an independent t test and a conditional maximum likelihood estimate of odds ratio (exact Fisher confidence interval) are shown. Exploratory direct, arm-level and combined meta-analysis and meta-regression was carried out by fitting linear mixed models with restricted maximum likelihood estimation. *Patients in the two active control groups (buspirone, opipramol) had virtually identical baseline HAM-A score (mean 23.6 and 23.9, respectively), end-of-study score (8.00 ± 7.56 and 7.74 ± 7.67, respectively) and proportion of responders ($\geq 50\%$ HAM-A reduction) (31/42 and 32/42, respectively), and were therefore pooled into one control group. **Characteristics of trials contributing additional placebo arm data are detailed in the Supplementary material eTable 1. *Open diamonds* individual trials, *solid diamonds* pooled estimates, *bars* indicate 95% confidence intervals and *gray boxes* prediction intervals (PI)

$- 13.3$ to 11.2) pooled estimate (Fig. 2). All four trials (116 patients on Kava Kava, 117 on placebo) reported on the proportion of responders ($\geq 50\%$ HAM-A reduction in three [30–32] and CGI-I "very much improved" in one [29] trial). Inconsistency of the results was lower ($I^2 = 51.9\%$) but heterogeneity was still considerable (PI around the pooled odds ratio, 0.23–18.4) (Fig. 2), with a pooled point-estimate in favor of Kava Kava (OR = 2.09), but largely imprecise (95% CI 0.48–9.13) (Fig. 2).

Direct Comparison of Kava Kava versus Buspirone or Opipramol

The trial [34] indicated no statistically significant difference between Kava Kava and active treatments in respect to end-of-study HAM-A score (primary efficacy outcome reported) or proportion of responders ($\geq 50\%$ HAM-A reduction) (Fig. 2); however, the estimated

differences had wide confidence intervals (imprecision) (Fig. 2). There was no explicit evidence of the assay sensitivity of the trial.

Network Meta-Analysis and Meta-Regression of Kava Kava versus Placebo

Analysis included the five Kava Kava trials (placebo and active-controlled) and a further five placebo-controlled trials of other biologically-based therapies (all with an overall low risk of bias) [36, 39, 41, 43, 45] (Table 1) (see Supplementary eTable 1 for data). It indicated (Fig. 2) that for end-study HAM-A score as well as for proportion of responders, Kava Kava–placebo differences estimated in direct comparisons largely overlapped with estimated Kava Kava–placebo differences from comparisons between: Kava arms from placebo-controlled Kava trials versus placebo arms from other (non-Kava) trials; Kava arms from all

Table 2 Summary outcomes of RCTs comparing Lavender (*Lavandula angustifolia*) extract to lorazepam over 6 weeks (Woelk 2010) [35] or to placebo and paroxetine over 10 weeks (Kasper 2014) [36]

Study/treatments	*n*	Base HAM-A	HAM-A ⇓	≥ 50% ⇓	Differences: HAM-A reduction (mean); response (RR) (95% CI; *P*)
Woelk [35] (6 weeks)					
Extract 80 mg/day	40	25.0 ± 4.0	− 11.3 ± 6.7	21 (52.5)	Mean = 0.3 (− 2.7, 3.3; *P* = 0.844); RR = 1.30 (0.80–2.1; 0.293)
Lorazepam 0.5 mg/day	37	25.0 ± 4.0	− 11.6 ± 6.6	15 (40.5)	Reference treatment
Kasper [36] (10 weeks)					
Extract 80 mg/day	135	25.8 ± 4.8	− 12.8 ± 8.7	70 (51.9)	Mean = − 3.3 (− 5.4, − 1.2; 0.002); RR = 1.37 (1.05–1.81; 0.020)
Extract 160 mg/day	121	26.0 ± 4.9	− 14.1 ± 9.3	73 (60.3)	Mean = − 4.6 (− 6.9, − 2.3; < 0.001); RR = 1.60 (1.24–2.08; < 0.001)
Paroxetine 20 mg/day	132	25.8 ± 4.9	− 11.3 ± 8.0	57 (43.2)	Mean = − 1.8 (− 3.9, 0.3; 0.085); RR = 1.14 (0.85–1.53; 0.368)
Placebo	135	25.1 ± 4.7	− 9.5 ± 9.0	51 (37.8)	Reference treatment

Study data are mean ± SD or count (%), all by intent-to-treat principle. Differences versus reference treatments (controls) were calculated for the purpose of this review

(placebo and active-controlled) trials versus placebo from Kava trials; and Kava arms from all trials versus placebo from non-Kava trials. It therefore appeared plausible to combine direct and arm-level Kava Kava versus placebo comparisons. Combined data (Fig. 2, unadjusted models) resulted in more precise estimates of Kava–placebo differences and narrower PIs, but still with considerable heterogeneity ($\tau^2 = 8.63$, $P = 0.026$ for end-study HAM-A and $\tau^2 = 0.43$, $P = 0.074$ for proportion of responders). Heterogeneity was reduced (particularly for HAM-A score) with adjustment for baseline HAM-A and trial duration, resulting in improved precision of the estimates and further narrowed PIs (Fig. 2, adjusted models). Data suggested a possible modest effect of Kava Kava (Fig. 2): lower end-of-study HAM-A scores with borderline statistical significance (mean difference = − 3.24, 95% CI − 6.65, 0.17; $P = 0.059$; PI − 8.6 to 2.1); and somewhat higher odds of response (OR = 1.90, 95% CI 0.74–4.88; $P = 0.159$; PI 0.36–10.2).

Evidence does not support the efficacy of Kava Kava extracts (120–280 mg/day kava-

lactones), but a modest effect cannot be unequivocally excluded: data are scarce and estimates are burdened with high uncertainty (imprecision, inconsistency, indirectness).

Lavender (*Lavandula angustifolia) Extracts*

One specific Lavender extract preparation (Silexan[®]) was evaluated in two trials [35, 36] (Table 1) with an overall low risk of bias and with ITT efficacy analysis. The first trial [35] reported no statistically significant difference between the extract (80 mg/day) and lorazepam (0.5 mg/day) regarding HAM-A reduction and proportion of responders (≥ 50% reduction in HAM-A score) at 6 weeks (Table 2). Although non-inferiority of the extract was claimed [35], the trial was small, with 95% CI around the difference in HAM-A reduction extending from 2.7 points to more than 3.3 points less reduction (Table 2), and the trial may have lacked assay sensitivity. Hence, it is uncertain whether the claimed non-inferiority should be considered as evidence of efficacy. In a larger 10-week trial [36], however, 80 mg/day and particularly 160 mg/day were clearly superior to placebo

Table 3 Summary outcomes of RCTs comparing *Galphimia glauca* extract to lorazepam over 4 (Herrera-Allerano 2007) [37] or 12 weeks (Herrera-Allerano 2012) [38]

Study/treatments	*n*	Base HAM-A	End HAM-A	Mean difference in end-of-study HAM-A score
Herrera-Allerano [37] (4 weeks)				
Extract 700 µg/day galphimin B	55	29.3 ± 4.7	9.0 ± 4.7	− 1.0 (95% CI − 3.6, 1.6; $P = 0.443$)
Lorazepam 2 mg/day	59	28.2 ± 8.7	10.0 ± 8.7	Reference treatment
Herrera-Allerano [38] (12 weeks)				
Extract 350–700 µg/daygalphimin B	52	29.1 ± 5.7	7.9 ± 5.7	− 1.5 (95% CI − 3.8, 0.8; $P = 0.096$)
Lorazepam 1–2 mg/day	57	28.2 ± 6.2	9.4 ± 6.2	Reference treatment

Study data are mean ± SD. Differences versus control were calculated for the purpose of this review

(Table 2). Numerically, both produced more effect than paroxetine 20 mg/day (Table 2).

Data suggest the efficacy of Lavender extract but require confirmation.

Galphimia glauca Extracts

Standardized *Galphimia glauca* extract was compared to lorazepam in two RCTs over 4 and 12 weeks [37, 38] (Table 1). Individual studies showed no statistically significant difference between the two treatments regarding end-of-study HAM-A scores (Table 3), thus suggesting efficacy of the extract through "comparability" to lorazepam. However, both trials suffered from a high risk of attrition bias (152 and 191 patients were randomized, efficacy reported for 114 and 109 completers, respectively) and there was no evidence of assay sensitivity. Pooled estimate (regression model with adjustment for baseline HAM-a) was highly imprecise (extract–lorazepam difference = − 4.0, 95% CI − 14.4, 6.4; $t = − 0.76$, $P = 0.450$). Therefore, it is highly uncertain whether the results should be interpreted as suggestive of efficacy of the tested preparation.

Chamomile (*Matricaria chamomilla*) Extracts

Chamomile extracts were evaluated in two placebo-controlled trials: one to reduce symptoms [39] and one withdrawal trial evaluating the risk of relapse in initial extract responders [40] (Table 1). Data showed greater HAM-A reduction and a trend towards higher proportion of responders at 8 weeks versus placebo [39] (Table 3), and a trend towards a lower risk of relapse over 25 weeks versus placebo [40] (Table 4). Both trials had an overall low risk of bias with ITT efficacy analysis. However, both trials were small with imprecise and fragile estimates. Together, they suggest efficacy of chamomile extract(s), but this would need to be proven in larger quality trials.

Extract of *Crataegus oxycantha* and *Eschscholtzia californica* Combined with Magnesium

This product was evaluated in a moderately-sized 12-week placebo-controlled trial [41] (Table 1) with an overall low risk of bias and with ITT efficacy analysis which showed significantly lower end-study HAM-A scores and a higher response rate with the tested product (Table 4). Data show efficacy, but require confirmation.

Other Biologically-Based Therapies

Echium amoneum extract was compared to placebo in an "add-on" (to fluoxetine) trial [42] (Table 1) with an overall low risk of bias and with ITT efficacy analysis showing lower HAM-A scores at 8 weeks versus placebo (Table 4).

Table 4 Main efficacy results of test (T) treatments evaluated in less than two controlled (Ctrl) trials of the same design/comparator

Study	Assessments	Treatments	Base score	Main reported end-treatment outcomes[a]
Biologically-based therapies				
Amsterdam [39]	HAM-A, score, response 8 weeks	T: Chamomile $n = 28$ Ctrl: Placebo $n = 29$	15.4 ± 4.2 14.3 ± 2.8	More reduction with T: $\Delta = -3.2$ (95% CI $-6.3, -0.45$; $P = 0.047$); more responders: 16/28 (57.1%) vs. 11/29 (37.9%); RR = 1.51 (95% CI 0.87–2.71; $P = 0.146$)
Mao [40]	CGI-S relapse over 25 weeks	T: Chamomile $n = 46$ Ctrl: Placebo $n = 47$	$\leq$ mild on CGI-S	Less relapse with T: 7/46 (15.2%) vs. 12/47 (25.5%); HR = 0.52 (95% CI 0.20–1.33; $P = 0.160$)
Hanus [41]	HAM-A, score, response, 12 weeks	T: *C. oxycantha* + *E. californica* +magnesium $n = 130$ Ctrl: Placebo $n = 134$	22.7 ± 2.9 22.4 ± 2.9	More reduction with T: $\Delta = -1.7$ (95% CI $-1.8, -1.6$; $P = 0.005$); more responders: 59/130 (45%) vs. 43/134 (32%); RR = 1.41 (95% CI 1.04–1.93; $P = 0.017$)
Sayyah [42]	HAM-A score, 8 weeks	T: *E. amoneum* + FLUOX $n = 19$ Ctrl: Placebo + FLUOX $n = 18$	37.2 ± 3.2 35.2 ± 2.8	Lower scores with T: 17.1 ± 3.2 vs. 23.1 ± 2.8; $\Delta = -5.0$ (95% CI $-8.0, -4.0$; $P = 0.018$)
Park [43]	HAM-A, score, response, 8 weeks	T: Gamisoyo-San[b] A $n = 49$ T: Gamisoyo-San[b] B $n = 49$ Ctrl: Placebo $n = 49$	29.1 ± 6.6 27.1 ± 7.6 27.9 ± 6.9	Similar scores with T A or B and Placebo: 19.6 ± 8.5 or 17.1 ± 6.6 vs. 19.3 ± 7.9; similar response : 17/49 (34.7%) or 13/49 (26.5%) vs. 13/49 (26.5%); For T A RR = 1.31 (95% CI 0.72–2.39)
Akhonzadeh [44]	HAM-A, score, 4 weeks	T: *Passiflora incarnata* $n = 18$ Ctrl: Oxazepam $n = 18$	19.6 ± 5.1 19.8 ± 5.1	Similar scores with T and Oxazepam: 5.7 ± 5.1 vs. 5.2 ± 5.1
Andreatini [45]	HAM-A score, 4 weeks	T: Valeriana extract $n = 12$ Ctrl: Diazepam $n = 12$ Ctrl: Placebo $n = 12$	22.8 ± 7.6 25.2 ± 4.5 25.1 ± 7.5	Similar scores with T, Diazepam and Placebo: 14.6 ± 9.8 vs. 14.2 ± 6.3 vs. 16.0 ± 6.1

Table 4 continued

Study	Assessments	Treatments	Base score	Main reported end-treatment outcomes[a]
Wang [46]	HAM-A, score, response, 24 weeks	T: CT + Chinese herbal $n = 93$ (?) Ctrl: CT + PAROX $n = 109$ (?)	27.9 ± 8.4 28.3 ± 8.6	Similar scores with T and Ctrl: 9.4 ± 6.5 vs. 10.1 ± 6.8; similar response: 76/93 (81.7%) vs. 84/109 (78.9%)
Manipulative and body-based therapies				
Eich [47]	CGI-S score, response, 4 weeks	T: Acupuncture $n = 7$ Ctrl: Sham acupuncture; $n = 6$	NA NA	More response with T: 6/7 vs. 2/6; RR = 2.57 (95% CI 0.99–9.06; $P = 0.053$)
Merom [48]	DASS-21 score, 8 weeks	T: CBT + exercise $n = 11$ C: CBT + education $n = 15$	19.0 ± 9.7 18.6 ± 10.7	No clear numerical data. Quote: "Results in patients with GAD remain questionable".
Dubois [49]	HAM-A score, 8 weeks	T: Balneotherapy $n = 117$ Ctrl: PAROX $n = 120$	24.4 ± 3.7 23.9 ± 3.4	More reduction with T: $- 12.0 \pm 4.8$ vs. $- 8.7 \pm 4.3$; $\Delta = - 3.3$ (95% CI $- 4.5, - 2.1$; $P < 0.001$)
Sherman [50]	HAM-A, score, response, 12 weeks	T: Th. massage $n = 23$ Ctrl: Thermotherapy $n = 22$ Ctrl: Relaxing room $n = 23$	24.8 ± 5.7 27.4 ± 7.0 26.2 ± 5.5	Somewhat less reduction with T vs. each Ctrl: $- 10$ vs. $- 13$ or vs. $- 11.1$ (reported as not significant); somewhat less response: 36.8% vs. 55% (RR = 0.55; 95% CI 0.34–1.26) or vs. 47.4% (RR = 0.76; 95% CI 0.38–1.50) (reported as not significant)
Herring [51]	AIDS-IV remiss., PSQW score, 6 weeks	T1: Resistance exercise $n = 10$ T2: Aerobic exercise $n = 10$ Ctrl: No treatment (wait) $n = 10$	63.8 ± 9.8 62.1 ± 6.4 64.3 ± 7.0	More remissions with T1 (6/10) but not with T2 (4/10) vs. Ctrl (3/10), reported significant for T1, but recalculated RR = 2.0 (95% CI 0.74–6.04); reported significantly greater score reduction (adjusted) for combined T1 and T2 vs. Ctrl $P = 0.039$
Ma [52]	SAS, score, response, 4 weeks	T: Ch. bloodlett. + PAROX $n = 35$ Ctrl: PAROX $n = 35$	62.8 ± 8.0 60.1 ± 8.3	Lower scores with T: 41.6 ± 9.6 vs. 46.9 ± 7.3; $\Delta = - 5.3$ (95% CI $- 9.4, - 1.2$; $P = 0.013$); more response: 29 (82.4%) vs. 18 (52.9%)/35; RR = 1.61 (95% CI 1.15–2.38; $P = 0.005$)

Table 4 continued

Study	Assessments	Treatments	Base score	Main reported end-treatment outcomes[a]
Jonsson [53]	GAD-Q-IV score, response 24 weeks	T: Flotation in water $n = 24$ Ctrl: No treatment (wait) $n = 22$	10.0 ± 2.2 9.9 ± 2.2	Lower scores with T: 7.1 ± 3.0 vs. 9.2 ± 3.4; $\Delta = -2.1$ (95% CI -4.0, -0.2; $P = 0.013$); more response: 9/24 (37.5%) vs. 3/22 (14.0%); RR = 2.75 (95% CI 0.94–8.65, $P = 0.066$)
Rappaport [54]	HAM-A, score, response, 12 weeks	T: Swedish massage $n = 21$ Ctrl: Sham (light touch) $n = 19$	20.1 ± 3.3 19.6 ± 4.9	More reduction with T: $\Delta = -3.3$ (95% CI -6.3, -0.2; $P = 0.035$); somewhat more response: 11/21 (52.4%) vs. 7/19 (36.8%); RR = 1.42 (95% CI 0.71–2.99; $P = 0.324$)
Alternative medical systems				
Bonne [55]	HAM-A, score, response, 10 weeks	T: Homeopathy $n = 22$ Ctrl: Placebo $n = 22$	31.4 ± 7.2 30.4 ± 7.6	Similar scores with T and Placebo: 21.7 ± 11.6 vs. 20.9 ± 9.2; same response: 8/22 (36%) vs. 8/22 (36%)
Tubaki [56]	HAM-A, score, response, 4 weeks	T1: Ayurvedic oral[c] $n = 22$ T2: Oral + topical[c] $n = 22$ Ctrl: Clonazepam $n = 21$	31.6 ± 3.2 32.6 ± 3.3 31.9 ± 4.3	Similar scores with T1, T2 and Clonazepam: 13.2 ± 4.9 vs. 12.4 ± 4.4 vs. 14.5 ± 7.1; similar response with T1 (14/22, 63.6%) and somewhat higher with T2 (20/22, 90.9%) vs. Clonazepam (16/21, 76.2%); for T2 RR = 1.19 (95% CI 0.90–1.68; $P = 0.191$)
Gupta [57]	HAM-A, score, response, 11 weeks	T: Ayurvedic oral[d] $n = 51$ Ctrl: Placebo $n = 51$	30.9 ± 7.0 31.3 ± 7.6	Similar reduction with T and Placebo: -15.8 ± 7.0 vs. -14.9 ± 6.7; similar response: 26/51 (51%) vs. 23/51 (45%)
Mind–body therapies				
Zhang [58]	SCL-90 score, 24 weeks	T1: Ch. cognitive therapy $n = 43$ T2: Cognitive + benzo. $n = 45$ Ctrl: Benzodiazepines $n = 43$	90.7 ± 52.5 107 ± 56.0 114 ± 66.0	Lower scores with T1 (49.3 ± 48.1) and T2 (47.2 ± 50.2) vs. Benzodiaz. (99.6 ± 67.7); for T1 $\Delta = -50.3$ (95% CI -75.5, -25.1; $P < 0.001$); for T2 $\Delta = -52.4$ (95% CI -77.7, -27.1; $P < 0.001$)

Table 4 continued

Study	Assessments	Treatments	Base score	Main reported end-treatment outcomes[a]
Koszycki [59]	HAM-A, score, response 12 weeks	T: Multifaith spiritual $n = 11$ Ctrl: CBT $n = 11$	23.6 ± 4.7 23.4 ± 5.8	Similar scores with T and Ctrl: 10.1 ± 8.9 vs. 8.9 ± 9.5; similar response: 8/11 (72.7%) vs. 7/11 (63.6%)
Koszycki [60]	HAM-A, score, response, 12 weeks	T: Multifaith spiritual $n = 11$ Ctrl: Support. psychoth. $n = 12$	20.1 ± 3.1 19.7 ± 3.0	Lower scores with T: 4.8 ± 3.1 vs. 11.0 ± 4.8; $\Delta = -6.2$ (95% CI -9.7, -2.7; $P < 0.001$); more response: 9/11 (82%) vs. 3/12 (25%); RR = 3.27 (95% CI 1.36–9.43; $P = 0.006$)

AIDS-IV Anxiety Disorders Interview Schedule, *CBT* cognitive behavioral therapy, *CGI-S* clinical global impression-severity, *Ch* Chinese, *CT* cognitive therapy, *DASS-21* depression anxiety stress scale 21, *GAD* generalized anxiety disorder, *GAD-Q-IV* Dimensional scoring from the Generalized Anxiety Disorder Questionnaire, *FLUOX* fluoxetine, *HAM-A* Hamilton Anxiety scale, *PAROX* paroxetine, *PSQW* Penn State Worry Questionnaire, *SAS* self-rating anxiety scale, *SCL-90* Symptom checklist 90-revised

[a] Outcomes were reported in different formats across trials: scale scores, score changes versus baseline (presented here as mean ± SD) or their differences (with confidence intervals), response or remission rates (based on cut-offs of specific scales). Where missing, and if possible based on reported data, we additionally calculated treatment differences (mean differences, Δ, or relative risks, RR, of response/remission)

[b] Korean herbal preparation consisting of a mix of 10 herbs—preparations A and B—different extraction

[c] Ayurvedic oral preparation *Manasamitra vataka* (T1); oral + medicated oil dripped on the forehead, called *Shirodhara* (T2)

[d] Ayurvedic medicine called *Sarasvata choorna*

However, the trial was small and the suggested efficacy would need to be proven in larger quality trials.

Gamisoyo-San, a Korean herbal mix (two different extraction methods) was compared to placebo in an 8-week trial [43] (Table 1) with an overall low risk of bias and with ITT efficacy analysis showing similar HAM-A scores and response rates (≥ 50% HAM-A reduction) versus placebo (Table 4). Data suggest no effect of the tested preparations, but small trials do not support finite conclusions.

Passiflora incarnata extract was compared to oxazepam [44] (Table 1) in a trial with an overall low risk of bias and with ITT efficacy analysis showing similar HAM-A scores for the two treatments at 4 weeks (Table 4). However, the trial was really small, with no criteria of equivalence/non-inferiority and with no clear

evidence of assay sensitivity. It is therefore highly uncertain whether the results should be interpreted as suggestive of efficacy of *Passiflora incarnata* extract.

Valeriana extract was compared to diazepam and placebo in a very small 4-week trial [45] (Table 1) with an overall low risk of bias and with ITT efficacy analysis showing similar end-study HAM-A scores for all three treatments (Table 4). The trial suggests no effect of *Valeriana* extract but also no effect of diazepam, and is uninformative regarding potential efficacy of *Valeriana* in GAD.

A *Chinese herbal compound* was compared to paroxetine in a medium-sized "add-on" (to cognitive therapy) trial [46] (Table 1) showing similar HAM-A scores and response rates for the two treatments at 24 weeks (Table 4). The trial was open-label with no evidence of blinded

outcome assessment (performance/detection bias), and there was uncertainty about the risk of attrition bias. There was no defined limit of equivalence/non-inferiority and no evidence of assay sensitivity. It is therefore highly uncertain whether the results should be considered indicative of efficacy of the tested herbal compound.

Manipulative and Body-Based Therapies

Acupuncture was compared to a sham procedure in a 4-week trial [47] (Table 1) with an overall low risk of bias and with ITT efficacy analysis showing a trend towards higher response rates with acupuncture (Table 3). However, the trial was extremely small, with imprecise and fragile estimates leaving a high level of uncertainty about the suggested efficacy of acupuncture.

Exercise was compared to education in an "add-on" 8-week trial (to cognitive-behavioral therapy) [48] (Table 1) which was burdened with extensive attrition and provided no numerical or other data on the efficacy outcomes in a subset of patients with GAD (Table 4).

Balneotherapy was compared to paroxetine in a medium-sized open-label trial [49] (Table 1) with no evidence of blinded outcome assessment (performance/detection bias), with ITT efficacy analysis showing a greater reduction in HAM-A score at 8 weeks with balneotherapy (Table 4). Data suggest efficacy of the specific balneotherapy procedure; however, non-inclusion of a sham procedure added to paroxetine and of paroxetine placebo added to the specific balneotherapy procedure might have introduced a form of a comparator bias ("relaxing therapy" vs. "standard pharmacological therapy"), as well as a kind of a selection bias: subjects more inclined to consenting to and responding to such an informal treatment might have been selected and as such might not have been "true representatives" of GAD patients in general. Therefore, the present observations would need to be clearly confirmed before a conclusion of efficacy of such a treatment is drawn.

Therapeutic massage was compared to thermotherapy and relaxing room treatment ("relaxing" parts of the therapeutic massage procedure, but without actual massage) in an open-label trial [50] (Table 1) with no evidence of blinded outcome assessment (performance/detection bias) with ITT efficacy analysis showing only slightly different HAM-A scores and response rates ($\geq$ 50% HAM-A reduction) for the three treatments at 12 weeks (Table 4). The trial was small with imprecise estimates, and a kind of comparator bias might have been introduced by a potential effect of the presumed control treatments: the trial does not rule out a possibility that massage could yield a difference versus, e.g., no treatment, or formal non-inferiority versus some established treatment, thus indicating efficacy. Therefore, the trial is inconclusive regarding (in)efficacy of therapeutic massage.

Resistance and *aerobic exercise* were compared to no treatment (postponed treatment) in a 6-week open-label trial [51] (Table 1), but with otherwise low risk of bias (blinded outcome assessment, ITT efficacy analysis, no attrition) showing a trend of more remissions (AIDS-IV scale) with resistance exercise versus no treatment and a greater reduction of worry (PSQW scale) for combined exercise groups versus no treatment (Table 4). However, the trial was small with imprecise and fragile estimates. Also, selection bias might have been introduced by the choice of no/postponed treatment as a control, since only patients prone to responding to such an informal treatment (and thus potentially not representative for GAD patients in general) might have been enrolled. Overall, the results should be considered as preliminary findings.

Chinese bloodletting added to paroxetine was compared to paroxetine in an open-label trial [52] (Table 1), but with otherwise low risk of bias (blinded outcome assessment, ITT efficacy analysis, no attrition) showing lower SAS scores and a higher proportion of responders at 4 weeks (Table 4). The trial was small with rather imprecise estimates, hence the suggested efficacy would need to be evaluated in larger quality trials, including a sham procedure for the specific bloodletting intervention.

Flotation in water was compared to no treatment (postponed treatment) in a 24-week trial [53] (Table 1) with ITT efficacy analysis that

showed lower anxiety scores and a trend towards higher response rate (GAD-Q-IV) versus no treatment (Table 4). However, it was an open-label trial with no evidence of blinded outcome assessment (performance/detection bias), small, with imprecise and fragile estimates. Also, selection bias might have been introduced by the choice of no/postponed treatment as a control: only patients inclined to responding to such an informal treatment might have been enrolled, not "typical" for GAD patients in general. Overall, the present results should be viewed as a preliminary finding.

Swedish massage was compared to a sham procedure [54] (Table 1) in a trial with an overall low risk of bias (blinded outcome assessment, ITT efficacy analysis, no attrition issue) and showed greater HAM-A reduction and a trend towards higher response rates at 12 weeks than the sham procedure (Table 4). The trial was small, with imprecise and fragile estimates and the suggested efficacy of Swedish massage should be confirmed in larger quality trials.

Alternative Medical Systems

Homeopathy was compared to placebo [55] (Table 1) in a small trial with an overall low risk of bias and with ITT efficacy analysis showing similar HAM-A scores and response rates at 10 weeks versus placebo (Table 4). Data suggest no effect of the tested preparation, but small trials do not support finite conclusions.

Two *Ayurvedic medications* were compared to clonazepam [56] (Table 1) showing similar HAM-A scores and proportion of responders at 4 weeks versus clonazepam, indicating efficacy (Table 4). However, it was a very small open-label trial with no evidence of blinded outcome assessment (performance/detection bias), no evidence of assay sensitivity, no definition of equivalence/non-inferiority limits, and with imprecise and fragile estimates, and is therefore inconclusive regarding the efficacy of the tested Ayurvedic medications.

Another oral *Ayurvedic* medication was compared to placebo in an 11-week trial [57] (Table 1) showing similar HAM-A score reduction and proportion of responders versus placebo (Table 4) suggesting no effect. However, it was a small trial with imprecise and fragile estimates, burdened with the risk of performance, detection and attrition bias. It is therefore inconclusive regarding the suggested inefficacy of the tested medication.

Mind–Body Therapies

Chinese cognitive therapy alone or combined with benzodiazepines was compared to benzodiazepines in a 24-week trial [58] (Table 1), and in both cases showed lower SCL-90 scores versus benzodiazepines alone (Table 4). It was a small open-label trial with no evidence of blinded outcome assessment (performance/detection bias) and with concerns about the risk of attrition bias. Therefore, the suggested efficacy of the evaluated procedure is uncertain and should be re-assessed in trials of better quality.

Multifaith spiritually-based intervention was compared to CBT in an open-label trial [59] (Table 1) with ITT efficacy analysis showing similar HAM-A scores and response rates at 12 weeks versus CBT (Table 4), thus implying efficacy. However, the trial was really small with imprecise and fragile estimates, with no evidence of blinded outcome assessment (detection bias) and of assay sensitivity, and with no defined limits of equivalence/non-inferiority. It is therefore inconclusive regarding the efficacy of the evaluated procedure. The same intervention was compared to supportive psychotherapy in an identically designed and sized trial [60] (Table 1) showing lower HAM-A scores and higher response rates (Table 4), indicating efficacy. However, considering the imprecision and fragility of the estimates and a high risk of detection bias, the trial should be considered as a preliminary finding.

Safety

Reporting on safety/tolerability greatly varied across trials (see Supplementary eTable 3 for details), with no reference to safety in 11/32 trials [47, 48, 51–53, 55–60] and with inconclusive and uninformative safety reporting in 2 additional trials [37, 38]. Most of the treatments

were evaluated in a single (most commonly small) trial with scarce and inconclusive safety/tolerability data, and a reasonably sound assessment was feasible for only a few.

Kava Kava

Of the four RCTs versus placebo [29–32], one reported that both treatments were "well tolerated" [31], while, based on the remaining three [29, 30, 32], the incidence of any AEs was consistently slightly lower with Kava Kava (total 12/99) than with placebo (total 20/101): pooled Mantel–Hanszel random-effects RR = 0.57 (95% CI 0.30–1.08; $P = 0.085$; $I^2 = 0\%$, 95% CI 0–73%). In one RCT [34] (total $n = 129$), the incidence of AEs was slightly higher with Kava Kava (32.5%) than with buspirone (23.8%) or opipramol (26.2%). Across all six Kava Kava trials (including the cross-over trial comparing two doses [33]), with a total of 166 patients exposed for 2–24 weeks, pooled random-effects incidence of any AE was 25.8% (95% CI 10.2–45.5; $I^2 = 86\%$, 95% CI 66–92%).

Lavender Extract

In one trial [35] (total $n = 77$), the incidence of any AE was similar with the extract (80 mg/day; 50%) and with lorazepam (0.5 mg/day; 48.6%). In another one [36] (total $n = 536$), it was similar for 80 mg/day extract (34.8%), 160 mg/day extract (25.0%) and placebo (31.6%), all numerically lower than with paroxetine 20 mg/day (40.9%).

Chamomile Extract

Based on two RCTs [39, 40], the incidence of any AEs was similar for the extract (total 36/74) and placebo (total 31/76): pooled RR = 0.85 (95% CI 0.62–1.18; $P = 0.338$).

Extract of *Crataegus oxycantha* and *Eschscholtzia californica* Combined with Magnesium

In one RCT [41], 130 treated patients experienced 22 AEs versus 15 AEs in 134 placebo-treated patients: rate ratio 1.51 (95% CI 0.75–3.13; $P = 0.214$), mostly due to poorer gastro-intestinal tolerability.

Chinese Herbal Preparation (Crude Mix of 14 Herbs)

In one RCT [46] (total $n = 202$), the incidence of AEs was lower with the preparation (16.1%) than with paroxetine (31.1%): RR = 0.52 (95% CI 0.30–0.87; $P = 0.013$).

Balneotherapy

In one RCT [49], 117 treated patients reported 70 AEs versus 162 AEs in 120 patients on paroxetine: rate ratio 0.44 (95% CI 0.33–0.59; $P < 0.001$).

DISCUSSION

Generalized anxiety disorder (GAD) is a chronic condition characterized by oscillations in symptoms and fluctuations between remissions and exacerbations [61]. Despite a rather wide range of recommended treatments (primarily pharmacological), GAD is difficult to treat with a substantial proportion of treatment-resistant patients and a rather high rate of relapses [61]. A recent meta-analysis [62] indicated that typically recommended [61, 63] first-line acute phase treatments (some of the selective serotonin or serotonin-noradrenaline re-uptake inhibitors; SSRI, SNRI) resulted in response rates of 68% (20 trials, 2311 patients) and remission rates of 40% (12 trials, 1502 patients) over the initial weeks of therapy (typically up to 12). For second-line (some benzodiazepines, buspirone, imipramine, pregabalin, bupropion) and third-line (some antipsychotics, citalopram, hydroxyzine) options envisaged for switching/augmentation strategies [61, 63], these rates were 10–15% lower [62]. A recent systematic review (eight trials) [64] demonstrated that these treatments also prevented relapses (24–76 weeks after response to the initial 8 to 26-week treatment); however, considerable numbers of treated patients did relapse: from 10–20% (SSRI,

SNRI, quetiapine, agomelatine, vortioxetine) to 42% (pregabalin) (vs. 31–65% on placebo). Furthermore, some pharmacological treatments may require a longer period of time to produce an effect, while some uncertainty exists about the long-term efficacy of the others, while the burden of side effects is significant [61]. CBT is effective in GAD, but the number of studies is small, the effect is lower than in other anxious disorders with questionable durability [65, 66], and combining pharmacological and psychological standard treatments did not result in a hypothesized additive effect [67]. In part, this suboptimal efficiency might be attributable to a high prevalence of comorbid mental disorders in GAD, a tendency to switch to other diagnoses, and inadequate (particularly long-term) treatment compliance [61].

CAM treatments are growing in popularity and are widely used by individuals suffering from mental illnesses, including anxiety disorders [68]. Nearly 38 percent of adults in the United States use CAM treatments to cope with mental disorders [69]. Considering, in addition, that anxiety is associated with lower treatment compliance [70], it seems plausible that CAM could be integrated into conventional treatment strategies with the aim of improving compliance and maximizing efficiency. These facts provide a rationale for the current review of evidence of efficacy and safety of CAM in GAD, and also for potential future research of combined CAM/conventional strategies for GAD.

The current research has several limitations at the review level. First, it is difficult to define CAM. Although many of the existing definitions seem straightforward, a lack of consistency of definitions is ubiquitous across the literature, and the reasons for defining therapies as CAM are not only scientific, but also political, social, and conceptual [71]. We opted for a pragmatic stepwise approach in which we firstly relied on a list of therapies defined as standard in the relevant literature, while all other therapies were considered alternative and then individually re-assessed according to published descriptive definitions [72]. Next, for efficacy assessment, we focused only on rating scales defined as primary in a particular study and did not consider other measures pertaining to, e.g., levels of depression, sleep quality and other domains which could provide a more complete picture of an "overall" effect of evaluated interventions. In this respect, we stayed with the rationale that the main purpose of treating GAD is a reduction of anxiety and that tests performed on a range of measurement scales are increasingly likely to yield spurious associations. Finally, we included only articles published in English and German, even though there is a substantial body of literature on traditional remedies and treatments originating mainly from Eastern Asia and published in local languages. However, although non-inclusion of studies due to language restrictions is generally an important limitation of systematic reviews, we considered that, in the present case, this should not be viewed as a major drawback. We assumed minor practical relevance/applicability of such treatments outside the specific traditional, cultural, philosophical and religious context from which they emerged. Consequently, we did not consider omission of these studies to have a relevant impact on the objectivity and comprehensiveness of the present review, particularly considering the highly non-specific, broad search strategy (which we consider to be a strength of the review) that resulted in identification of a variety of treatments fitting the predefined definition of CAM.

The main findings of the present review pertain primarily to the quality of the identified trials, in the sense of standard risk of bias assessment, quality of reporting and also methodological/design quality. Only 3 trials were rated as having a low risk of bias across all items, whereas 14 studies had a high risk of performance bias (blinding of participants and personnel), which were typically trials of interventions that were not biologically based. While it is understandable that blinding is a technical problem for such interventions, the lack of explicit blinded outcome assessment (and high or uncertain risk of detection bias) further increased the uncertainty about the validity of the reported outcomes. Attrition was a clear issue in at least two trials, while in several others it was unclear to what extent it could have biased the results. This is closely

connected to the quality of reporting. Low quality/incomplete reporting was noticeable, in particular related to safety. We consider this to be an important limitation, since adverse effects of some of the evaluated treatments have been previously documented [73]. In general, we noticed a higher reporting quality in studies on biologically-based therapies, in particular among recent publications. This improvement might be due to adherence to reporting guidelines. The use of CONSORT guidelines is associated with improved reporting quality, and a trend of improvement was shown for CAM treatments other than the ones addressed in this systematic review [74]. Most of the included trials were small. Small trials yield imprecise and fragile estimates particularly regarding proportions, where one or two responders more or less per arm may substantially change the overall conclusion; lack of "significance" in a small inequality trial could be simply due to a lack of power (only sporadically power and sample size considerations were reported), and "significance" could be simply by chance (small trials tend to report unrealistically large effects). Finally, 14 trials compared CAM to a (supposedly) active treatment. Showing superiority over a reference treatment in such trials does indicate efficacy; however, "lack of difference" is commonly (as in the majority of the present trials) perceived as an evidence of efficacy, which is conceptually erroneous. Ways of proving efficacy through showing non-inferiority (by a formal test) to a proven effective reference have been clearly defined [75], and since none of the present trials met them, we concluded there was a lack of assay sensitivity and qualified them as, in a sense, "inconclusive". In part, these methodological/quality issues are likely due to the fact that trials of CAM treatments are typically not sponsored by the industry: conducting a large-enough, quality trial is financially and operationally demanding. Indeed, the largest and the highest quality of the reviewed trials was actually industry-sponsored [36].

Considering the efficacy of individual treatments, meaningful conclusions could be drawn for only 5/22 evaluated. Despite some opinions [73], the present analysis does not support the efficacy of aqueous Kava Kava extracts, although a modest effect cannot be excluded. It could be objected that the network meta-analysis carried out to improve the estimates of Kava Kava versus placebo differences might have included additional placebo data, not only from other included trials of biologically-based CAM treatments, but also from trials of any "conventional pharmacological" treatment for GAD. We refrained from doing so, since CAM treatments, even if herbal preparations, are not regular, recognized medical treatments, and some GAD patients might be more inclined to "respond" to such treatments and, relatedly, more prone to consent to participate in trials for which they know they could receive such treatments. They could therefore differ from (placebo-treated) patients enrolled in trials of "conventional treatments" regarding this "benevolence towards CAM", and their inclusion in the network might have introduced a form of a selection bias that could not have been controlled for.

One specific *Lavender* extract (Silexan®) was clearly indicated as effective (superior to placebo and numerically more effective than paroxetine) in a large high-quality trial, and an additional similar (or longer) trial would confirm efficacy. The same is applicable for a combination of extracts of *Crataegus oxycantha* and *Eschscholtzia californica* and magnesium, and balneotherapy consisting of daily immersion (10 min) in a mineral water bubbling bath, with underwater massage by an experienced physiotherapist (10 min) followed by a pressurized shower (water massage) of the whole body. Finally, two small but quality placebo-controlled trials strongly indicate the efficacy of *Chamomile* extract—one in initial treatment and one in relapse prevention—but they only represent a reasonable justification for large(r) confirmatory trials.

CONCLUSION

In conclusion, at present, the body of evidence about the efficacy and safety of the overall CAM category of treatments in GAD is modest, both in size and quality. There are, however, two herbal

preparations and a specific balneotherapy regimen with demonstrated efficacy in single trials meeting the same standards implemented foir conventional recommended treatments, but they require confirmation before finite conclusions can be drawn, while, for *Chamomile* extract, it seems plausible to state that the two proof-of-the concept studies justify further research. Considering the circumstances (available standard treatments, required quality of evidence), it does not seem likely that any of the reviewed treatments would be investigated to the extent that would provide evidence to justify their *alternative* use (i.e., instead of the standard treatments), ; however, it appears feasible and justified to evaluate their *complementary* use (alongside standard treatments), in particular considering herbal preparations, as this could be evaluated through real-life pragmatic trials.

ACKNOWLEDGEMENTS

REFERENCES

1. Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of dsm-iii-r psychiatric disorders in the united states. Results from the national comorbidity survey. Arch Gen Psychiatry. 1994;51:8–19.

2. Nutt DJ, Kessler RC, Alonso J, et al. Consensus statement on the benefit to the community of ESEMeD (European study of the epidemiology of mental disorders) survey data on depression and anxiety. J Clin Psychiatry. 2007;68:42–8.

3. Stein M, Sherbourne C, Craske M, et al. Quality of care for primary care patients with anxiety disorders. Am J Psychiatry. 2004;161:2230–7.

4. Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. Arch Gen Psychiatry. 2005;62:617–27.

5. Alonso J, Angermeyer M, Bernert S, et al. Prevalence of mental disorders in Europe: results from the European Study of the epidemiology of mental disorders (ESEMeD) project. Acta Psychiatr Scand. 2004;109:21–7.

6. Hidalgo RB, Tupler LA, Davidson JR. An effect-size analysis of pharmacologic treatments for generalized anxiety disorder. J Psychopharmacol. 2007;21:864–72.

7. Schweitzer I, Maguire K, Ng C. Sexual side-effects of contemporary antidepressants: review. Aust N Z J Psychiatry. 2009;43:795–808.

8. Baldwin DS, Montgomery SA, Nil R, et al. Discontinuation symptoms in depression and anxiety disorders. Int J Neuropsychopharmacol. 2007;10:73–84.

9. Foa EB, Franklin ME, Moser J. Context in the clinic: How well do cognitive-behavioral therapies and medications work in combination. Biol Psychiatry. 2002;10:987–97.

10. Baldwin DS, Polkinghorn C. Evidence-based pharmacotherapy of generalized anxiety disorder. Int J Neuropsychoph. 2005;8:293–302.

11. Liu L, Liu C, Wang Y, et al. Herbal medicine for anxiety, depression and insomnia. Curr Neuropharmacol. 2015;13:481–93.

12. Kinrys G, Coleman E, Rothstein E. Natural remedies for anxiety disorders: potential use and clinical applications. Depress Anxiety. 2009;26:259–65.

13. Garcia-Garcia P, Lopez-Munoz F, Rubio G, et al. Phytotherapy and psychiatry: bibliometric study of the scientific literature from the last 20 years. Phytomedicine. 2008;15:566–76.

14. Williams J, Gierisch J, McDuffie J, et al. An overview of complementary and alternative medicine therapies for anxiety and depressive disorders: supplement to efficacy of complementary and alternative medicine therapies for posttraumatic stress disorder. Evidence-based synthesis program. Washington: Department of Veterans Affairs; 2011:1–23.

15. Sarris J, Moylan S, Camfield DA, et al. Complementary medicine, exercise, meditation, diet, and lifestyle modification for anxiety disorders: a review of current evidence. Evid Based Complement Alternat Med. 2012;2012:809653.

16. Jorm AF, Christensen H, Griffiths KM, et al. Effectiveness of complementary and self-help treatments for anxiety disorders. Med J Aust. 2004;181:S29–46.

17. Sarris J, Goncalves D, Robins-Wahlins T, et al. Complementary medicine use by middle-aged and older women: personality, mood and anxiety factors. J Health Psychol. 2010;16:314–21.

18. Eisenberg D, Davis R, Ettner S, et al. Trends in alternative medicine use in the United States, 1990–1997: results of a follow-up national survey. JAMA. 1998;280:1569–75.

19. Unutzer J, Klap R, Sturm R, et al. Mental disorders and the use of alternative medicine: results from a national survey. Am J Psychiatry. 2000;157:1851–7.

20. Kessler R, Soukup J, Davis R, et al. The use of complementary and alternative therapies to treat anxiety and depression in the United States. Am J Psychiatry. 2001;158:289–94.

21. Bazzan A, Zabrecky G, Monti D, et al. Current evidence regarding the management of mood and anxiety disorders using complementary and alternative medicine. Expert Rev Neurother. 2014;14:411–23.

22. Generalised anxiety disorder and panic disorder in adults: management | Guidance and guidelines | NICE. Niceorguk. 2017. Available at: https://www.nice.org.uk/guidance/cg113/chapter/1-Guidance#principles-of-care-for-people-with-generalised-anxiety-disorder-gad. Accessed April 25, 2017.

23. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. BMJ. 2011;343:d5928.

24. IntHout J, Ioannidis JPA, Borm GF. The Hartung–Knapp–Sidik–Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Method. 2014;14:25.

25. Riley R, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ. 2011;342:d549.

26. Kessels AGH, terRiet G, Puhan MA, et al. A simple regression model for network meta-analysis. OA Epidemol. 2013;1:7.

27. Brown H, Prescott R. Applied mixed models in medicine, 3rd ed, Wiley 2013:197-230.

28. Whitehead A. Meta-analysis of controlled clinical trials. New Jersey: Wiley; 2002. p. 131–6.

29. Volz H, Kieser M. Kava-kava extract WS 1490 versus placebo in anxiety disorders—a randomized placebo-controlled 25-week outpatient trial. Pharmacopsychiatry. 1997;30:1–5.

30. Malsch U, Kieser M. Efficacy of kava-kava in the treatment of non-psychotic anxiety, following pretreatment with benzodiazepines. Psychopharmacology. 2001;157:277–83.

31. Connor K, Davidson J. A placebo-controlled study of Kava kava in generalized anxiety disorder. Int Clin Psychopharmacol. 2002;17:185–8.

32. Sarris J, Stough C, Bousman C, et al. Kava in the treatment of generalized anxiety disorder. J Clin Psychopharmacol. 2013;33:643–8.

33. Wheatley D. Kava-kava (LI 150) in the treatment of generalized anxiety disorder. Primary Care Psychia. 2001;7:97–100.

34. Boerner R, Sommer H, Berger W, Kuhn U, Schmidt U, Mannel M. Kava-Kava extract LI 150 is as effective as opipramol and buspirone in generalised anxiety disorder—an 8-week randomized, double-blind multi-centre clinical trial in 129 out-patients. Phytomedicine. 2003;10:38–49.

35. Woelk H, Schläfke S. A multi-center, double-blind, randomized study of the Lavender oil preparation Silexan in comparison to lorazepam for generalized anxiety disorder. Phytomedicine. 2010;17:94–9.

36. Kasper S, Gastpar M, Müller W, et al. Lavender oil preparation Silexan is effective in generalized anxiety disorder—a randomized, double-blind comparison to placebo and paroxetine. Int J Neuropsychopharmacol. 2014;17:859–69.

37. Herrera-Arellano A, Jiménez-Ferrer E, Zamilpa A, Morales-Valdéz M, García-Valencia C, Tortoriello J. Efficacy and tolerability of a standardized herbal product from Galphimiaglauca on generalized anxiety disorder. A randomized, double-blind

clinical trial controlled with Lorazepam. Planta Med. 2007;73:713–7.

38. Herrera-Arellano A, Jiménez-Ferrer J, Zamilpa A, García-Alonso G, Herrera-Alvarez S, Tortoriello J. Therapeutic effectiveness of Galphimiaglauca vs. lorazepam in generalized anxiety disorder. A controlled 15-week clinical trial. Planta Med. 2012;78:1529–35.

39. Amsterdam J, Li Y, Soeller I, Rockwell K, Mao J, Shults J. A randomized, double-blind, placebo-controlled trial of oral *Matricaria recutita* (Chamomile) extract therapy for generalized anxiety disorder. J Clin Psychopharmacol. 2009;29:378–82.

40. Mao J, Xie S, Keefe J, Soeller I, Li Q, Amsterdam J. Long-term chamomile (*Matricaria chamomilla* L.) treatment for generalized anxiety disorder: a randomized clinical trial. Phytomedicine. 2016;23:1735–42.

41. Hanus M, Lafon J, Mathieu M. Double-blind, randomized, placebo-controlled study to evaluate the efficacy and safety of a fixed combination containing two plant extracts (*Crataegus oxyacantha* and *Eschscholtzia californica*) and magnesium in mild-to-moderate anxiety disorders. Curr Med Res Opin. 2004;20:63–71.

42. Sayyah M, Siahpoosh A, Khalili H, Malayeri A, Samaee H. A double-blind, placebo-controlled study of the aqueous extract of *Echium amoenum* for patients with general anxiety disorder. Iran J Pharm Res. 2012;11:697–701.

43. Park D, Kim S, Park Y, Kang W, Lee S, Jung I. The comparative clinical study of efficacy of Gamisoyo-San (Jiaweixiaoyaosan) on generalized anxiety disorder according to differently manufactured preparations: Multicenter, randomized, double blind, placebo controlled trial. J Ethnopharmacol. 2014;158:11–7.

44. Akhondzadeh S, Naghavi H, Vazirian M, Shayeganpour A, Rashidi H, Khani M. Passionflower in the treatment of generalized anxiety: a pilot double-blind randomized controlled trial with oxazepam. J Clin Pharm Ther. 2001;26:363–7.

45. Andreatini R, Sartori V, Seabra M, Leite J. Effect of valepotriates (valerian extract) in generalized anxiety disorder: a randomized placebo-controlled pilot study. Phytother Res. 2002;16:650–4.

46. Wang T, Ding J, Xu G, Zeng Y, Xiao S. Efficacy of Yiqiyangxin Chinese medicine compound combined with cognitive therapy in the treatment of generalized anxiety disorders. Asian Pac J Trop Med. 2012;5:818–22.

47. Eich H, Agelink M, Lehmann E, Lemmer W, Klieser E. AkupunkturbeileichtenbismittelschwerendepressivenEpisoden und Angststörungen–Ergeb-nisseeiner experimentellen Untersuchung. Fortschr Neurol Psychiatr. 2000;68:137–44.

48. Merom D, Phongsavan P, Wagner R, et al. Promoting walking as an adjunct intervention to group cognitive behavioral therapy for anxiety disorders—a pilot group randomized trial. J Anxiety Disord. 2008;22:959–68.

49. Dubois O, Salamon R, Germain C, et al. Balneotherapy versus paroxetine in the treatment of generalized anxiety disorder. Complement Ther Med. 2010;18:1–7.

50. Sherman K, Ludman E, Cook A, et al. Effectiveness of therapeutic massage for generalized anxiety disorder: a randomized controlled trial. Depress Anxiety. 2010;27:441–50.

51. Herring M, Jacob M, Suveg C, Dishman R, O'Connor P. Feasibility of exercise training for the short-term treatment of generalized anxiety disorder: a randomized controlled trial. Psychother Psychosom. 2012;81:21–8.

52. Ma H, Kui Y, Li Y, Huang B, Li S, Chen X. Blood-letting therapy combined with paroxetine hydrochloride for generalized anxiety disorder: A randomized controlled trial. 2013 IEEE International Conference on Bioinformatics and Biomedicine. 2013.

53. Jonsson K, Kjellgren A. Promising effects of treatment with flotation-REST (restricted environmental stimulation technique) as an intervention for generalized anxiety disorder (GAD): a randomized controlled pilot trial. BMC Complement Altern Med. 2016;16:108.

54. Rapaport M, Schettler P, Larson E et al. Acute Swedish massage monotherapy successfully remediates symptoms of generalized anxiety disorder. J Clin Psychiatry 2016:e883–e891.

55. Bonne O, Shemer Y, Gorali Y, Katz M, Shalev A. A randomized, double-blind, placebo-controlled study of classical homeopathy in generalized anxiety disorder. J Clin Psychiatry. 2003;64:282–7.

56. Tubaki B, Chandrashekar C, Sudhakar D, Prabha T, Lavekar G, Kutty B. Clinical efficacy of Manasamitra Vataka (an ayurveda medication) on generalized anxiety disorder with comorbid generalized social phobia: a randomized controlled study. J Altern Complement Med. 2012;18:612–21.

57. Gupta K, Mamidi P, Thakar A. Randomized placebo controlled study on Sarasvata choorna in generalised anxiety disorder. Int J Green Pharm. 2014;8:231–6.

58. Zhang Y, Young D, Lee S, et al. Chinese Taoist cognitive psychotherapy in the treatment of generalized anxiety disorder in contemporary China. Transcult Psychiatry. 2002;39:115–29.

59. Koszycki D, Raab K, Aldosary F, Bradwejn J. A multifaith spiritually based intervention for generalized anxiety disorder: a pilot randomized trial. J Clin Psychol. 2010;66:430–41.

60. Koszycki D, Bilodeau C, Raab-Mayo K, Bradwejn J. A multifaith spiritually based intervention versus supportive therapy for generalized anxiety disorder: a pilot randomized controlled Trial. J Clin Psychol. 2013;70:489–509.

61. Bandelow B, Zohar J, Hollander E, et al. World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the pharmacological treatment of anxiety, obsessive-compulsive and post-traumatic stress disorders. World J Biol Psychiatry. 2002;3:171–99.

62. Bereza BG, Machado M, Ravindran AV, Einarson TR. Evidence-based review of clinical outcomes of guideline-recommended pharmacotherapies for generalized anxiety disorder. Can J Psychiatry. 2012;57:470–8.

63. Baldwin D, Anderson I, Nutt D, et al. Evidence-based pharmacological treatment of anxiety disorders, post-traumatic stress disorder and obsessive-compulsive disorder: a revision of the 2005 guidelines from the British Association for Psychopharmacology. J Psychopharmacol. 2014;28:403–39.

64. Mochcovitch M, da Rocha Freire R, Garcia R, Nardi A. Can long-term pharmacotherapy prevent relapses in generalized anxiety disorder? A Systematic Review. Clin Drug Investig. 2017;37:737–43.

65. Bandelow B, Seidler-Brandler U, Becker A, Wedekind D, Rüther E. Meta-analysis of randomized controlled comparisons of psychopharmacological and psychological treatments for anxiety disorders. World J Biol Psychiatry. 2007;8:175–87.

66. Bandelow B, Reitt M, Röver C, Michaelis S, Görlich Y, Wedekind D. Efficacy of treatments for anxiety disorders. Int Clin Psychopharmacol. 2015;30:183–92.

67. Otto MW, Smits JA, Reese HE. Combined psychotherapy and pharmacotherapy for mood and anxiety disorders in adults: review and analysis. Clin Psychol Sci Pract. 2005;27:572–81.

68. Kessler RC, Davis RB, Foster DF, et al. Long-term trends in the use of complementary and alternative medical therapies in the United States. Ann Intern Med. 2001;135:262–8.

69. Barnes PM, Bloom B, Nahin RL. Complementary and alternative medicine use among adults and children: United States, 2007, National health statistics reports; no 12.Hyattsville, MD: National Center for Health Statistics; 2008.

70. Fontenelle L, Santana. A review of studies concerning treatment adherence of patients with anxiety disorders. Patient Prefer Adherence. 2011;5:427–39.

71. Jonas WB. Policy, the public, and priorities in alternative medicine research. Ann Am Acad Polit Soc Sci. 2002;583:29–43.

72. Complementary and alternative medicine in the United States. 1st ed. Washington, DC: The National Academies Press; 2005.

73. Sarris J, Stough C, Teschke R, et al. Kava for the treatment of generalized anxiety disorder RCT: analysis of adverse reactions, liver function, addiction, and sexual effects. Phytother Res. 2013;27:1723–8.

74. Plint AC, Moher D, Morrison A, Schulz K, Altman DG, Hill C, Gaboury I. Does the CONSORT checklist improve the quality of reports of randomized controlled trials? A systematic review. Med J Aust. 2006;185:263–7.

75. Nutt D, Allgulander C, Lecrubier Y, Peters T, Wittchen H. Establishing non-inferiority in treatment trials in psychiatry—guidelines from an Expert Consensus Meeting. J Psychopharmacol. 2008;22:409–16.