Review protocol for review questions: For adults with a new episode of less severe depression or more severe depression, what are the relative benefits and harms of psychological, psychosocial, pharmacological and physical interventions alone or in combination?

Table 33. Review protocol

Topic	First-line treatment for adults with depression
Review questions	RQ. 2.1 For adults with a new episode of less severe depression, what are the relative benefits and harms of psychological, psychosocial, pharmacological and physical interventions alone or in combination? RQ. 2.2. For adults with a new episode of more severe depression, what are the relative benefits and harms of psychological, psychosocial, pharmacological and physical interventions alone or in combination?
Objectives	To identify the most effective first-line interventions for the treatment of a new episode of depression
Population	• Adults receiving first-line treatment for a new episode of depression, as defined by a diagnosis of depression according to DSM, ICD or similar criteria, or depressive symptoms as indicated by baseline depression scores on validated scales (and including those with subthreshold [just below threshold] depressive symptoms) If some, but not all, of a study's participants are eligible for the review, for instance, mixed anxiety and depression diagnoses, then we will include a study if at least 80% of its participants are eligible for this review. Baseline mean scores are used to classify study population severity according to less severe (RQ 2.1) or more severe (RQ 2.2) using the thresholds outlined below. These thresholds are derived using standardization of depression measurement crosswalk tables (Wahl 2014; Rush 2003; Carmody 2006; Uher 2008). An anchor point of 16 on the PHQ-9 was selected on the basis of alignment with the clinical judgement of the committee and eligibility criteria in published studies. If baseline mean scores are not available, severity will be classified according to the inclusion criteria of the study or the description given by the study authors (but only in cases where this is unambiguous, for example 'severe' or 'subthreshold' or 'mild'). Severity thresholds:

Scale	Threshold
HAMD (17-item, 21-item and 24-item)	16
MADRS (10-item)	22

Topic	First-line treatment for adults with depression
	PHQ-9: 16 BDI-I (21-item): 22 BDI-II (21-item): 30 CES-D (20-item): 36 QIDS (16-item): 12 HADS-D (7-item): 12
Exclude	• Trials of women with antenatal or postnatal depression • Trials of children and young people (mean age under 18 years) • Trials of people with learning disabilities • Trials of people with bipolar disorder • Trials of adults in contact with the criminal justice system (not solely as a result of being a witness or victim) • Trials where more than 20% of the population have psychotic symptoms • Trials where more than 20% of the population have a coexisting personality disorder • Trials where more than 20% of the population have chronic depression (chronic depression defined as depression for at least 2 years, or persistent subthreshold symptoms [dysthymia], or double depression [an acute episode of major depressive disorder superimposed on dysthymia]) • Trials of further-line treatment • Trials of people with Seasonal Affective Disorder (SAD) Trials that specifically recruit participants with a physical health condition in addition to depression (e.g. depression in people with diabetes)
Intervention	The following interventions will be included: **Psychological interventions**: • Behavioural therapies (including behavioural activation, behavioural therapy [Lewinsohn 1976], coping with depression group) • Cognitive and cognitive behavioural therapies (including CBT individual or group [defined as under or over 15 sessions], problem solving, rational emotive behaviour therapy [REBT] and third-wave cognitive therapies individual or group) • Counselling (including emotion-focused therapy [EFT], non-directive/supportive/ person-centred counselling and relational client-centred therapy) • Interpersonal psychotherapy • Psychodynamic psychotherapies (including individual or group-based short-term psychodynamic psychotherapy, long-term psychodynamic psychotherapy and psychodynamic counselling) • Psychoeducational interventions (including psychoeducational group programmes) • Self-help with or without support (including cognitive bibliotherapy with or without support, computerised CBT [CCBT] with or without support, computerised psychodynamic therapy with or without support)

Topic	First-line treatment for adults with depression
	• Art therapy • Music therapy • Eye movement desensitization and reprocessing (EMDR) (for depression, not PTSD) The following interventions are more appropriate for subgroups of adults with depression and as such will be considered only in pairwise comparisons (and not included in the NMA): • Couple interventions, including behavioural couples therapy (for people with problems in the relationship with their partner) **Pharmacological interventions**: To be included, pharmacological interventions needed to be licensed in the UK and in routine clinical use for the first-line treatment of depression. SSRIs • Citalopram • Escitalopram • Paroxetine • Sertraline • Fluoxetine • TCAs • Amitriptyline • Clomipramine • Lofepramine • Nortriptyline • Note: To improve connectivity, imipramine will be included in the network (because it has been used as a control in many trials) however it will not be considered as part of the decision problem SNRIs • Venlafaxine • Duloxetine Other antidepressant drugs: • Mirtazapine • Trazodone

Topic	First-line treatment for adults with depression
	Note that if necessary for connectivity in the network specific drugs that are excluded and 'any antidepressant' or 'any SSRI' or 'any TCA' nodes will be added where they have been compared against a psychological or physical intervention and/or combined with a psychological or physical intervention but they will not be considered as part of the decision problem. **Physical interventions**: • Acupuncture • Exercise (including yoga) • Light therapy (for depression, not SAD) **Psychosocial interventions**: • Peer support (including befriending, mentoring, and community navigators) • Mindfulness, meditation or relaxation (including mindfulness-based stress reduction [MBSR])
Comparison	• Other active intervention (must also meet inclusion criteria above) • Treatment as usual (TAU) • Waitlist • No treatment • Placebo If a study compares 'intervention + TAU vs TAU alone' it will be recoded as 'intervention vs no treatment'
Outcomes	**Critical outcomes:** **Efficacy** • Depression symptomatology (mean endpoint score or change in depression score from baseline) • Remission (usually defined as a cut off on a depression scale), this will be analysed for those randomised and for completers • Response (usually defined as at least 50% improvement from the baseline score on a depression scale), this will be analysed for those randomised and for completers The following depression scales will be included in the following hierarchy: • MADRS • HAMD • QIDS • PHQ • CGI (for dichotomous outcomes only)

Topic	First-line treatment for adults with depression

	• CES-D
	• BDI
	• HADS-D (depression subscale)
	• HADS (full scale)
	
	Only one continuous scale will be used per study
	
	• For studies reporting response and/or remission, the scale used in the study to define cut-offs for response and/or remission will be used
	• If more than one definition is used, a hierarchy of scales will be adopted (hierarchy listed above)
	
	For studies not reporting dichotomous data, a hierarchy of scales (see above) will be adopted for continuous outcomes
	
	Acceptability/tolerability
	• Discontinuation due to side effects (for pharmacological trials)
	• Discontinuation due to any reason (including side effects)
	
	Important, but not critical, outcomes:
	• Quality of life
	  • Quality of life (as assessed with a validated scale, including the 12-item/36-item Short-Form Survey [SF-12/SF-36], 26-item short version of the World Health Organization Quality of Life assessment [WHOQOL-BREF], EuroQoL [EQ5D], Quality of Life Depression Scale [QLDS], Quality of Life Enjoyment and Satisfaction Questionnaire [Q-LES-Q], Quality of Life Inventory [QoLI], and World Health Organization 5-item Well-Being Index [WHO-5])
	• Personal, social, and occupational functioning
	  • Global functioning (as assessed with a validated scale, including Global Assessment of Functioning [GAF], Global Assessment Scale [GAS], and Social and Occupational Functioning Assessment Scale [SOFAS])
	  • Functional impairment (as assessed with a validated scale, including Sheehan Disability Scale [SDS], Social Adjustment Scale [SAS], and Work and Social Adjustment Scale [WSAS])
	  • Sleeping difficulties (as assessed with a validated scale, including Insomnia Severity Index [ISI] and Pittsburgh Sleep Quality Index [PSQI])
	  • Employment (for instance, % unemployed)
	  • Interpersonal problems (as assessed with a validated scale, including Inventory of Interpersonal Problems [IIP])

Topic	First-line treatment for adults with depression
	• Outcomes will be assessed at endpoint and follow-up (data for all available follow-up periods of at least 1-month post-intervention will be extracted and will be grouped into categories for analysis, for instance, 1-3 months, 4-6 months, 7-9 months, 10-12 months, 13-18 months, 19-24 months, and >2 years).
Study design	• RCTs • Systematic reviews of RCTs
Include unpublished data?	Conference abstracts, dissertations and unpublished data will not be included unless the data can be extracted from elsewhere (for instance, from the previous guideline)
Restriction by date	All relevant studies from existing reviews from the 2009 guideline and from previous searches (pre-2016) will be carried forward. Studies published between 2016 and the date the searches are run will be sought.
Minimum sample size	N = 10 in each arm Studies with <50% completion data (drop out of >50%) will be excluded.
Study setting	Primary, secondary, tertiary and social care settings. Non-English-language papers will be excluded (unless data can be obtained from an existing review).
The review strategy	**Data Extraction (selection and coding)** Citations from each search will be downloaded into EndNote and duplicates removed. Titles and abstracts of identified studies will be screened by two reviewers for inclusion against criteria, until a good inter-rater reliability has been observed (percentage agreement =>90%). Initially 10% of references will be double-screened. If inter-rater agreement is good then the remaining references will be screened by one reviewer. All primary-level studies included after the first scan of citations will be acquired in full and re-evaluated for eligibility at the time they are being entered into a study database (standardised template created in Microsoft Excel). At least 10% of data extraction will be double-coded. Discrepancies or difficulties with coding will be resolved through discussion between reviewers or the opinion of a third reviewer will be sought. **Data Analysis** Pairwise comparisons (meta-analyses using random-effects models) will be conducted to combine results from similar studies. An intention to treat (ITT) approach will be taken where possible. Network meta-analysis (NMA) in a Bayesian framework will also be used to synthesise the data for all eligible interventions which are connected in a network of RCT comparisons. Interventions with similar effects (as determined by the committee) will be grouped into classes and class effects models will be fitted [Dias 2018]. The relative effects of the interventions within each class will be assumed to be distributed around a common class mean with a within-class variance, permitting the borrowing of strength across interventions within each class.

Topic	First-line treatment for adults with depression
	Classes which do not have enough evidence to estimate within-class variability of effects (i.e., a class with just 1 or 2 interventions) will share within-class variability with similar classes (as determined by the committee) where the variance can be estimated. For example, the individual cognitive and CBT class may borrow the within-class variance from the individual behavioural therapies class. If no such similar class is identified, we will assume zero variance in classes with only 1 or 2 interventions. In addition, the attention placebo, no treatment and TAU classes will share a within-class variance. If an 'any antidepressant' class is required to connect otherwise disconnected/excluded drugs to the network (as described under Intervention topic), its within-class variance will be equal to the maximum of the SSRI and TCA within-class variances. The random class effects assumption will be assessed by comparing the fit of fixed and random class effects models, where the former assumes the intervention effects within each class are the same (i.e., no within-class variability of effects). Continuous outcomes (SMDs) will be combined with dichotomous data to estimate intervention effects, using the methods described in the Appendix. The NMA will probably be restricted to critical outcomes at endpoint due to the likelihood of a lack of connectivity in a follow-up data network or in a network for important (but not critical) outcomes. The consistency of direct and indirect evidence will be assessed by fitting and comparing the fit of the NMA and unrelated mean effects (UME) models, the latter of which is equivalent to having separate, unrelated, meta-analyses for every pairwise contrast [Dias 2011]. Each data point's contribution to the posterior mean residual deviance for the NMA model will be plotted against that for the UME model, to visually assess if specific data points are contributing to inconsistency. If the UME suggests there is evidence of inconsistency, node-split models will be fitted to assist in identifying loops of evidence with inconsistency [Dias 2010]. Risk of bias will be assessed at the study level using the Cochrane risk of bias tool. This assessment includes: adequacy of randomisation (sufficient description of randomisation method, allocation concealment and any baseline difference between groups); blinding (of participants, intervention administrators and outcome assessors); attrition ('at risk of attrition bias' defined as a dropout of more than 20% and completer analysis used, or a difference of >20% between the groups); selective reporting bias (is the protocol registered, are all outcomes reported); other bias (for instance, conflict of interest in funding). Risk of bias will also be assessed at the outcome level using GRADE. For heterogeneity, outcomes will be downgraded once if I2>50%, twice if I2 >80%. For imprecision, outcomes will be downgraded using rules of thumb. If the 95% CI is imprecise i.e. crosses the line of no effect and the threshold for clinical benefit/harm, 0.8 or 1.25 (dichotomous) or -0.5 or 0.5 SMD (for continuous), the outcome will be downgraded. Outcomes will be downgraded one or two levels depending on how many lines it crosses. If the 95% CI is not imprecise, we will

Topic	First-line treatment for adults with depression
	consider whether the criterion for Optimal Information Size is met (for dichotomous outcomes, 300 events; for continuous outcomes, 400 participants), if not we will downgrade one level.
Heterogeneity (sensitivity analysis and subgroups)	Where possible, the influence of the following subgroups will be considered: • Primary care compared to secondary care • Inpatient compared to outpatient settings • Older adults (60 years and older) compared to younger adults (younger than 60 years) • BME populations • Men If the network structure allows, sensitivity analyses will be considered for depression symptoms (SMD, the primary outcome for the clinical analysis) and discontinuation for any reason and response in completers (the main outcomes for economic analysis), as follows: • Risk of bias as reflected by publication bias and study size using methods described in [Dias 2010]. We will assume possible bias in comparisons of active interventions vs inactive control and no bias between inactive control comparisons, as well as active intervention comparisons, except in comparisons where counselling is the control intervention (in which case bias against counselling will be assumed) • Validity of transitivity assumption will be explored by sensitivity analysis on SMD outcome that includes non-pharmacological trials only and examines any differences in magnitude of effects and ranking of non-pharmacological interventions compared to results from the mixed psychological, psychosocial, pharmacological and physical model Threshold analysis will be performed to assess the robustness of intervention recommendations due to bias [Phillippo 2018].
Notes	For interventions in the NMA it is assumed that any patient that meets all inclusion criteria is, in principle, equally likely to be randomised to any of the interventions in the synthesis comparator set. For defining routine usage of drugs, the national prescription cost data for England in 2017 - the most recent year for which relevant data existed - (Prescribing & Medicines Team, Health and Social Care Information Centre, 2017) was used. If a drug appeared in the top 15 it was included, with the exception of dosulepin which the BNF indicates should be initiated by a specialist. Cipriani 2018 network meta-analysis will be used as a source for studies and data. References for crosswalk tables:

Topic	First-line treatment for adults with depression
	Carmody, T. J., Rush, A. J., Bernstein, I., Warden, D., Brannan, S., Burnham, D., ... & Trivedi, M. H. (2006). The Montgomery Äsberg and the Hamilton ratings of depression: a comparison of measures. European Neuropsychopharmacology, 16(8), 601-611. Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., ... & Thase, M. E. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biological psychiatry, 54(5), 573-583. Uher, R., Farmer, A., Maier, W., Rietschel, M., Hauser, J., Marusic, A., ... & Henigsberg, N. (2008). Measuring depression: comparison and integration of three scales in the GENDEP study. Psychological medicine, 38(2), 289-300. Wahl, I., Löwe, B., Bjorner, J. B., Fischer, F., Langs, G., Voderholzer, U., ... & Rose, M. (2014). Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. Journal of clinical epidemiology, 67(1), 73-86. Assuming a normal distribution and using baseline mean and standard deviation data, we will explore the categorisation of less and more severe, including the percentage of studies 'definitely' within the correct category (≥70% of the study sample above cut-off) in order to aid the committee in interpreting the results. References for data analysis: Dias, S., Ades, A.E., Welton, N.J., Jansen, J.P., Sutton, A.J. (2018). Network meta-analysis for decision making. Hoboken, NJ: Wiley. Dias, S., Welton, N.J., Sutton, A.J., Caldwell, D.M., … & Ades, A.E. (2011). NICE DSU Technical Support Document 4: Inconsistency in networks of evidence based on randomised controlled trials. Dias, S., Welton, N.J., Caldwell, D.M., Ades A.E. (2010a). Checking consistency in mixed treatment comparison meta-analysis. Statistics in Medicine, 29(7-8), 932-44. References for heterogeneity: Dias, S., Welton, N.J., Marinho, V.C.C., Salanti, G., … & Ades A.E. (2010b). Estimation and adjustment of bias in randomised evidence by using mixed treatment comparison meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 173(3), 613-29.

Topic	First-line treatment for adults with depression
	Phillippo, D.M., Welton, N.J., Dias, S., Didelez, V., Ades A.E. (2018). Sensitivity of treatment recommendations to bias in network meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society), 181(3), 843-67.
Information sources – databases and dates	Database(s): Embase 1974 to Present, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present; Cochrane Library; WEB OF SCIENCE
Identify if an update	Update of CG90 (2009)
Author contacts	For details please see the guideline in development web site.
Highlight if amendment to previous protocol	For details please see section 4.5 of Developing NICE guidelines: the manual 2014
Search strategy – for one database	For details please see appendix B.
Data collection process – forms/duplicate	A standardised evidence table format will be used, and published as appendix D (clinical evidence tables) or H (economic evidence tables).
Data items – define all variables to be collected	For details please see evidence tables in appendix D (clinical evidence tables) or H (economic evidence tables).
Methods for assessing bias at outcome/study level	Standard study checklists were used to critically appraise individual studies. For details please see section 6.2 of Developing NICE guidelines: the manual 2014. The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/.
Criteria for quantitative synthesis	For details please see section 6.4 of Developing NICE guidelines: the manual 2014
Methods for quantitative analysis – combining studies and exploring (in)consistency	For details please see the methods chapter.
Meta-bias assessment – publication bias, selective reporting bias	For details please see section 6.2 of Developing NICE guidelines: the manual 2014.
Confidence in cumulative evidence	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual 2014
Rationale/context – what is known	For details please see the introduction to the evidence review.
Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the National Guideline Alliance (NGA) and chaired by Dr Navneet Kapur in line with section 3 of Developing NICE guidelines: the manual 2014. Staff from the NGA undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost effectiveness analysis where appropriate, and drafted the guideline in collaboration with the committee. For details please see the methods chapter.
Sources of funding/support	The NGA is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists.

Topic	First-line treatment for adults with depression
Name of sponsor	The NGA is funded by NICE and hosted by the Royal College of Obstetricians and Gynaecologists.
Roles of sponsor	NICE funds NGA to develop guidelines for those working in the NHS, public health and social care in England
PROSPERO registration number	CRD42019151328

BDI: Beck depression inventory; BME: black minority ethnic; BNF: British national formulary; (C)CBT: (computerised) cognitive behavioural therapy; CDSR: Cochrane Database of Systematic Reviews; CENTRAL: Cochrane Central Register of Controlled Trials; CES-D: Centre of epidemiology studies – depression; CGI: clinical global impressions; CI: confidence interval; DARE: Database of Abstracts of Reviews of Effects; DSM: Diagnostic and statistical manual; EFT: emotion-focused therapy; EMDR: eye movement desensitization and reprocessing; EQ-5D: European quality of life 5 dimensions; GAF: global assessment of functioning; GAS: global assessment scale; GRADE: Grading of Recommendations Assessment, Development and Evaluation; HADS-D: hospital anxiety and depression scale – depression; HAMD: Hamilton Depression Rating Scale; ICD: International classification of diseases; IIP: inventory of interpersonal problems; ISI: insomnia severity index; ITT: intention to treat; MADRS: Montgomery–Åsberg Depression Rating Scale; MBSR: Mindfulness-based stress reduction; MID: minimally important difference; NGA: National Guideline Alliance; NHS: National health service; NICE: National Institute for Health and Care Excellence; NMA: network meta-analysis; PHQ-9: patient health questionnaire-9; PSQI: Pittsburgh sleep quality index; PTSD: post-traumatic stress disorder; QIDS: quick inventory of depressive symptomatology; QLDS: quality of life depression scale; Q-LES-Q: quality of life enjoyment and satisfaction questionnaire QOLI: quality of life inventory RCT: randomised controlled trial; REBT: rational emotive behaviour therapy; RoB: risk of bias; SAD: seasonal affective disorder; SAS: Spielberger state/trait anxiety scale; SDS: Sheehan disability scale; SMD: standardised mean difference; SNRI: serotonin-noradrenaline reuptake inhibitor; SOFAS: social and occupational functioning assessment scale; SSRI: selective serotonin reuptake inhibitor; TAU: treatment as usual; TCA: tricyclic antidepressant; UME: unrelated mean effects; WHOQOL-BRIEF: World health organization quality of life assessment (brief); WHO-5: world health organization 5-item wellbeing index; WSAS: work and social adjustment scale