National Institute for Health and Care Excellence

Consultation draft

Depression in adults: treatment and management

Appendix U2.1: Text from CG90 appendices that has been deleted

NICE Guideline <…>

Appendices

May 2018

Consultation draft

Developed by the National Guideline Alliance, hosted by the Royal College of Obstetricians and Gynaecologists

Disclaimer
Healthcare professionals are expected to take NICE clinical guidelines fully into account when exercising their clinical judgement. However, the guidance does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of each patient, in consultation with the patient and/or their guardian or carer.

National Institute for Health and Care Excellence [2017]. All rights reserved. Subject to Notice of rights.

Contents

Appendix U2.1: Superseded Appendices

Appendix 9: Clinical study data extraction form

CLINICAL STUDY DATA EXTRACTION FORM

Topic Area:		**Report reference ID:**			
Comparisons:				**Total N**	
Ref List checked		Rev Man		Study Database	
Data Checked		Reference Manager updated		Excluded (record reason in Notes below)	

Randomised?		**Blind?**	
Age:		Young/Elderly (mean age over 65) Mean Age % Women	
Setting:		In/Out/Mixed/Primary Care (80% patients)	
Analysis:		Completer/ITT (continuous data)	
Diagnosis		% Comorbid Axis I	
		% Comorbid Axis II	
Mean baseline			

Trial length:
Interventions (Dose): 1 2

Notes

Appendix 10: Quality checklists for clinical studies and reviews

Completed by:											
Report reference ID:											
1 TREATMENT GROUP:											
Leaving treatment early (any reason) (side effects)		**Leaving treatment early reporting)**		**Side Effects (total number**		**Remission [non-remission]**					
n	*N*	*n*	*N*	*N*	*N*	*n*	*N*				
Definition of remission **Definition of response**											
Post-treatment means											
	n	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*Mean*	*SD*
Other data	**Response [non-response]**											
	n	*N*		*n*	*N*		*n*	*Mean*	*SD*	*n*	*Mean*	*SD*

2 TREATMENT GROUP:							
Leaving treatment early (any reason) (side effects)		**Leaving treatment early reporting)**		**Side Effects (total number**		**Remission [non-remission]**	
n	*N*	*n*	*N*	*N*	*N*	*n*	*N*

Definition of remission **Definition of response**												
Post-treatment means												
	n	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*Mean*	*SD*
Other data												
	n	*N*		*n*	*N*		*n*	*Mean*	*SD*	*n*	*Mean*	*SD*

Comparisons entered:

3 TREATMENT GROUP:							
Leaving treatment early (any reason) (side effects)		**Leaving treatment early reporting)**		**Side Effects (total number**		**Remission [non-remission]**	
n	*N*	*n*	*N*	*N*	*N*	*n*	*N*
Definition of remission **Definition of response**							

Post-treatment means												
	n	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*Mean*	*SD*
Other data												
	n	*N*		*n*	*N*		*n*	*Mean*	*SD*	*n*	*Mean*	*SD*

4 TREATMENT GROUP:											
Leaving treatment early (any reason) (side effects)		**Leaving treatment early reporting)**		**Side Effects (total number**		**Remission [non-remission]**					
n	*N*	*n*	*N*	*N*	*N*	*n*	*N*				
Definition of remission **Definition of response**											
Post-treatment means											
	n	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*Mean*	*SD*	*n*	*Mean*	*SD*
Other data											
	n	*N*		*n*	*N*		*n*	*Mean*	*SD*	*n*	*Mean*	*SD*

Appendix 11: The classification of depression and depression rating scales/questionnaires

BACKGROUND

This appendix sets out an approach to the classification of depression that was used in the development of the guideline update (including the analysis of the evidence and the development of recommendations) and will be of value in routine clinical use.

Depression is a heterogeneous disorder in which a number of underlying presentations may share a common phenomenology but have different aetiologies. Despite considerable work on the aetiology of depression including neurobiological, genetic and psychological studies, no reliable classificatory system has emerged that links either to the underlying aetiology or has proven strongly predictive of response to treatment. A number of classification systems/subgroupings have been used, including reactive and endogenous depression, melancholia, atypical depression, depression with a seasonal pattern/seasonal affective disorder and dysthymia. These have been based on varying combinations of the nature, number, severity, pattern and duration of symptoms, and in some cases the assumed aetiology. Over time pragmatic definitions have emerged, enshrined in the current two major classification systems, DSM–IV-TR (APA, 2000c) and ICD–10 (WHO, 1992). These have defined a threshold of severity of clinical significance with further classification in terms of severity (for example, mild, moderate or severe as adopted in DSM–IV with regard to major depressive disorder), duration and course of the disorder (for example, recurrent, presence of residual symptoms) and subtype based on symptom profile (for example, melancholic, atypical). Other aspects of depression such as response to treatment (for example, treatment resistant, refractory) and aetiology (for example, preceding life events) do not feature specifically in the classifications and lack accepted definitions, although are used in clinical practice. The classification has some use in describing likely outcome and course (Khan et al., 1991; Barrett et al., 2001; Sullivan et al., 2003; Blom et al., 2007; Jackson et al., 2007; Conradi et al., 2008; Holma et al., 2008; Van et al., 2008) although social support, social impairment or personality factors also need to be taken into account. Lower severity and duration of a depressive episode predicts, to some extent, a greater likelihood of spontaneous or earlier and eventual improvement whereas greater severity, chronicity and number of previous episodes predict a higher chance of subsequent relapse.

The lack of a highly reliable or valid classificatory system has significant and practical clinical consequences, particularly in primary care where the full range of depression presents. A major concern is whether depression should be classified using dimensions or categories. Categories help distinguish cases from non-cases, while dimensions help distinguish severe disorder from mild (Cole et al., 2008). Clinicians are often required to make a categorical decisions – for example to treat with antidepressants or not, to refer for further interventions or not – and consequently there can be pressure to interpret data on a single dimension in a categorical way, for example, treat or not treat based solely on a symptom severity rating (for example, a PHQ-9 score alone). This conflicts with the recognised need to take multiple factors/dimensions into consideration within a consultation, including the patient's view on the cause of symptoms and acceptable treatment, and in the guideline update a major challenge has been to provide a useful categorisation that adequately captures the complexity.

CLASSIFICATION OF DEPRESSION AND NICE GUIDANCE

The approach adopted in the previous depression guideline (NICE, 2004a; NCCMH, 2004) was based on ICD–10 and rested on a dimensional approach based on a symptom count further elaborated by taking into account the presence of social role impairment and the duration of both symptoms and social impairment. The subsequent categorisation of depression into mild, moderate and severe has led to a number of concerns in practice. First this classification appears to have often been implemented with an emphasis on a symptom count alone with other important factors such as duration and social impairment ignored, although it should be noted that in general there is a relationship between the number of symptoms and severity of functional impairment (Faravelli et al., 1996). Second it implies that the different symptoms experienced are equivalent, although, in fact, symptom patterns may be important. Third, it does not take into account illness duration and course. This tendency may be exacerbated by the use of measures such as the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001) or Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983) under the Quality and Outcomes Framework (Department of Health, 2004).

A drawback inherent in using ICD–10 depression criteria is that most of the treatment research on which the guideline has to be based uses DSM–IV or previous, essentially similar, versions of DSM (DSM–III and DSM–III-R) criteria. As discussed below, the criteria are similar but not identical, and this has particular relevance for the 'threshold' of the diagnosis of a clinically significant depressive episode and therefore what are considered subthreshold depressive symptoms.

DIAGNOSIS OF A DEPRESSIVE/MAJOR DEPRESSIVE EPISODE

The criteria for diagnosing depressive episodes in ICD–10 and DSM–IV overlap considerably but have some differences of emphasis. In ICD–10 the patient must have two of the first three symptoms (depressed mood, loss of interest in everyday activities, reduction in energy) plus at least two of the remaining seven symptoms; while in DSM–IV the patient must have five or more out of nine symptoms with at least one from the first two (depressed mood and loss of interest). Both diagnostic systems require symptoms to have been present for at least 2 weeks to make a diagnosis (but can be shorter in ICD–10 if symptoms are unusually severe or of rapid onset). In both ICD–10 and DSM–IV the symptoms must result in impairment of functioning that increases with the episode severity. Table 143 compares the symptoms required in ICD–10 and DSM–IV.

DETERMINING SEVERITY OF A DEPRESSIVE/MAJOR DEPRESSIVE EPISODE

Both ICD–10 and DSM–IV classify clinically important depressive episodes as mild, moderate and severe based on the number, type and severity of symptoms present and degree of functional impairment. Table 144 shows the number of symptoms required by each diagnostic system, which are less specific than DSM–IV. The prescriptive symptom counting approach of ICD–10 tends to lend itself to using symptom counting alone to determine severity.

Table 143: Comparison of symptoms of depression in ICD–10 and DSM–IV

ICD–10	DSM–IV major/minor depressive disorder
Depressed mood*	Depressed mood by self-report or observation made by others*
Loss of interest*	Loss of interest or pleasure*
Reduction in energy*	Fatigue/loss of energy
Loss of confidence or self-esteem	

Unreasonable feelings of self-reproach or inappropriate guilt	Worthlessness/excessive or inappropriate guilt
Recurrent thoughts of death or suicide	Recurrent thoughts of death, suicidal thoughts or actual suicide attempts
Diminished ability to think/ concentrate or indecisiveness	Diminished ability to think/concentrate or indecisiveness
Change in psychomotor activity with agitation or retardation	Psychomotor agitation or retardation
Sleep disturbance	Insomnia/hypersomnia
Change in appetite with weight change	Significant appetite and/or weight loss

*Core symptoms

Table 144: Number of symptoms required in ICD–10 and DSM–IV for a diagnosis of depressive episode/major depression (but note they also need assessment of severity and functional impairment to ascertain diagnosis and severity)

	ICD–10 depressive episode	DSM–IV major depression
Mild	4	Minimal above the minimum (5)
Moderate	5–6	Between mild and severe
Severe	7+	Several symptoms in excess of 5

As ICD–10 requires only four symptoms for a diagnosis of a mild depressive episode, it can identify more people as having a depressive episode compared with a DSM–IV major depressive episode. One study in primary care in Europe identified two to three times more people as depressed using ICD–10 criteria compared with DSM–IV (11.3% versus 4.2%; Wittchen et al., 2001a). However another study in Australia (Andrews et al., 2008) found similar rates using the two criteria (6.8% versus 6.3%) but slightly different populations were identified (83% concordance), which appears to be related to the need for only one of two core symptoms for DSM–IV but two out of three for ICD–10. These studies emphasise that, although similar, the two systems are not identical and that this is particularly apparent at the threshold taken to indicate clinical importance.

DIAGNOSIS OF SUBTHRESHOLD DEPRESSIVE SYMPTOMS

Given how common milder forms of depression are, and the problems inherent in defining a 'threshold' of clinical importance because of the diagnostic system differences and the lack of any natural discontinuity identifying a critical threshold (Andrews et al., 2008), this guideline update has broadened its scope to include consideration of depression that is 'subthreshold', that is, does not meet the full criteria for a depressive/major depressive episode. A further reason is that subthreshold depression has been increasingly recognised as causing considerable morbidity and human and economic costs, is more common in those with a history of major depression and is a risk factor for future major depression (Rowe & Rapaport, 2006).

There is no accepted classification for this in the current diagnostic systems with the closest being minor depression, a research diagnosis in DSM–IV. At least two but less than five symptoms are required, of which one must be depressed mood or diminished interest. This includes ICD–10 depressive episode with four symptoms and, given the practical difficulty and inherent uncertainty in deciding thresholds for significant symptom severity and disability, there is no natural discontinuity between minor depression and mild major depression in routine clinical practice. There is however a danger of 'medicalising' distress by

adopting minor depression as a discrete diagnosis, which would inevitably broaden the concept of depression. For this guideline update the GDG therefore use the term 'subthreshold depressive symptoms' to avoid this problem while providing a way of describing this part of the depressive spectrum.

Both DSM–IV and ICD–10 do have the category of dysthymia, which consists of depressive symptoms which are subthreshold for major depression but which persist (by definition for more than 2 years). There appears to be no empirical evidence that dysthymia is distinct from subthreshold depressive symptoms apart from duration of symptoms.

ICD–10 has a category of mixed anxiety and depression, which is less clearly defined than minor depression, and is largely a diagnosis of exclusion in those with anxiety and depressive symptoms subthreshold for specific disorders. Not unexpectedly it appears to be a heterogeneous category with a lack of diagnostic stability over time (Wittchen et al., 2001b; Barkow et al., 2004). For this reason it has not been included in this guideline.

DURATION

The duration of a depressive episode can vary considerably among individuals. The average course of an untreated depressive episode is between 6 and 8 months with much of the improvement occurring in the first 3 months, and 80% recovered by 1 year (Coryell et al., 1994). There is evidence to suggest that patients who do not seek treatment for their depression may recover more quickly than those who seek but do not receive treatment (Posternak et al., 2006). There is also some evidence to suggest that people who do not seek help have a shorter mean duration of depressive episode (Posternak et al., 2006).

Traditionally the minimum duration of persistent symptoms for major depression is 2 weeks and for chronic depression (or dysthymia) 2 years. These conventional definitions have been adopted in the absence of good evidence as there is only a modest empirical base for the minimum duration (for example, Angst & Merikangas, 2001) and none that we could find for the 'cut-off' between acute and chronic depression. As with severity, duration is better thought of as a dimension with a decreased likelihood of remission with increasing chronicity over a given time frame (Van et al., 2008). The conventional criteria are therefore better viewed as guides rather than cut-offs. It is likely that that the minimum duration after which therapy provides more benefit than occurs by spontaneous improvement is somewhat longer than 2 weeks (possibly 2 to 3 months, Posternak et al., 2006), but this has never been tested empirically. By 2 years it does appear that outcome is poorer, supporting consideration of chronicity in describing the disorder; nevertheless the point at which acute becomes chronic is not clear, and indeed may not be a meaningful question. There is some evidence that outcome is poorer after about 1 year (for example, Khan et al., 1991). However there seems little to be gained by redefining duration for the guideline as long as it is recognised that the conventional definitions are merely signposts to include consideration of duration in relation to outcome and need for treatment.

COURSE OF DEPRESSION

An influential model of the course of major depression proposes that the onset of an episode of depression consists of a worsening of symptoms in a continuum going from depressive symptoms through to major depression. Phases of improvement with treatment consist of response (significant improvement) to remission (absence of depressive symptoms) which if stable for 4 to 6 months results in (symptomatic) recovery, meaning that the episode is over (Frank et al., 1991). It is important to distinguish this use of recovery from more recent concepts related to quality and meaning of life in spite of continued symptoms. After recovery a further episode of depression is viewed as a recurrence to distinguish it from a relapse of the same episode. There has been no consensus as to how long a period of remission should be in order to be able to declare recovery; different definitions result in different

definitions of episode length and time to full or subthreshold depressive recurrence (Furukawa et al., 2008). Therefore, in practice it can be difficult to distinguish between relapse and recurrence, particularly when people have mild residual symptoms. Follow-up studies of people with depression have shown that, overall, more time is spent with subthreshold depressive symptoms than major depression and there is a variable individual pattern ranging from persisting chronic major depression, through significant but not full improvement (partial remission), to full remission and recovery (Judd et al., 1998). DSM–IV defines full remission when there has been an absence of symptoms for at least 2 months. For partial remission, full criteria for a major depressive episode are no longer met, or there are no substantial symptoms but 2 months have not yet passed. DSM–IV specifies 'with full interepisode recovery' if full remission is attained between the two most recent depressive episodes and 'without full inter-episode recovery' if full remission is not attained. In DSM–IV, therefore, separate episodes are distinguished by at least 2 months of not meeting major depression criteria, which is in contrast to the more stringent ICD–10 requirements of 2 months without any significant symptoms. There is therefore some ambiguity as to whether full remission is required to define separate episodes.

Nevertheless the number of episodes and degree of symptom resolution have important implications for considering the course of an individual patient's depressive disorder. The risk of a further episode of major depression within a given time frame is greater with an increasing number of previous episodes (Solomon et al., 2000; Kessing & Andersen, 2005) and also if there has not been full remission/symptomatic recovery (Paykel et al., 1995; Kanai et al., 2003; Dombrovski et al., 2007). If someone presents with minor depressive symptoms it is therefore crucial to determine whether or not this directly follows an episode of major depression.

DEPRESSION SUBTYPES

Different symptom profiles have been described and are included in the classification systems. In DSM–IV, severe major depression can be without or with psychosis (psychotic depression) and there are specifiers that include melancholia, atypical features, catatonia, depression with a seasonal pattern (seasonal affective disorder) and post-partum onset. ICD–10 also provides specifiers for psychotic and somatic symptoms, the latter similar to DSM–IV melancholia. However, these subtypes do not form distinct categories (for example, Kendell, 1968; Angst et al., 2007) and they add a further complexity to the diagnosis of depression. The GDG judged that these specifiers were best considered where appropriate after the diagnosis of a depressive disorder is made and they are not discussed in detail here. Some specifiers, particularly psychosis and seasonal pattern depression, have potential treatment implications and are considered in the guideline update where evidence is available.

CLASSIFICATION OF DEPRESSION IN THE GUIDELINE UPDATE

The depression classification system adopted for the guideline update had to meet a number of criteria, notably the use of:

- a system that reflects the non-categorical, multidimensional nature of depression
- a system that makes best use of the available evidence on both efficacy and effectiveness
- a system that could be distilled for practical day-to-day use in healthcare settings without potentially harmful over-simplification or distortion
- terms that can be easily understood and are not open to misinterpretation by a wide range of healthcare staff and service users

- a system that would facilitate the generation of clinical recommendations.

These criteria led the GDG to adopt a classificatory system for depression based on DSM–IV criteria. When assessing an individual it is important to assess three dimensions to diagnose a depressive disorder – a) severity (symptomatology and social impairment), b) duration, and c) course – as linked, but separate, factors (see below). In addition there was recognition that a single dimension of severity was insufficient to fully capture its multidimensional nature.

As discussed above the following depressive symptoms require assessment to determine the presence of major depression. The symptoms need to be experienced to a sufficient degree of severity and persistence to be counted as definitely present. At least one core symptom is required; both core symptoms would be expected in moderate and severe major depression.

Core symptoms of depression

1) Depressed mood most of the day, nearly every day.

2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day.

Somatic symptoms

3) Significant weight loss when not dieting or weight gain (for example, a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day.

4) Insomnia or hypersomnia nearly every day.

5) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).

6) Fatigue or loss of energy nearly every day.

Other symptoms

7) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).

8) Diminished ability to think or concentrate, or indecisiveness, nearly every day.

9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

The symptoms are not due to the direct physiological effects of a substance (for example, a drug of misuse or a medication) or a general medical condition (for example, hypothyroidism) or better accounted for by bereavement.

There is evidence that doctors have difficulty in remembering the nine DSM–IV depressive symptoms (Rapp & Davis, 1989; Krupinski & Tiller, 2001), which has important implications for the application of these criteria. In addition there is need to be able to consistently diagnose depression in patients where physical symptoms may be due to medical illness. Zimmerman and colleagues (2006) and Andrews and colleagues (2008) have demonstrated that, compared with the diagnosis using the full DSM–IV criteria, there is a high agreement (94 to 97%) and good sensitivity (93%) and specificity (95 to 98%) when a reduced list (excluding the four somatic symptoms) is used with a requirement for three out of the remaining five symptoms.

It is therefore possible to use an abridged list, first asking about the two core symptoms of depression:

- persistent depressed mood
- markedly diminished interest or pleasure.

Then if either or both are present going on to ask about:

- feelings of worthlessness or guilt
- impaired concentration
- recurrent thoughts of death or suicide.

Three or more symptoms indicate a very high probability of major depression. This does not however replace the need to go on to assess somatic symptoms as an aid to determining severity and to help judge subsequent response to treatment. This limits the usefulness of the abridged list in practice and it may be most useful when there are confounding somatic symptoms due to physical illness.

Severity

While recognising that severity is not a unitary dimension, practically it is useful to make a judgement of severity consisting, at least, of number of symptoms, severity of individual symptoms and functional impairment. This leads to a classification of depression into the following severity groupings based on DSM–IV criteria, which should be viewed as exemplars not discrete categories. In the guideline update the term 'depression' refers to major depression:

- subthreshold depressive symptoms: fewer than five symptoms of depression
- mild depression: few, if any, symptoms in excess of the five required to make the diagnosis, and the symptoms result in only minor functional impairment
- moderate depression: symptoms or functional impairment are between 'mild' and 'severe'
- severe depression: most symptoms, and the symptoms markedly interfere with functioning; can occur with or without psychotic symptoms.

Symptom severity and degree of functional impairment correlate highly (for example, Zimmerman et al., 2008), but in individual cases this may not be the case and some mildly symptomatic individuals may have marked functional impairment while some people who are severely symptomatic may, at least for a time, maintain good function, employment and so on.

Duration

By convention the duration of persistent symptoms is required to be at least 2 weeks and once they have persisted for 2 years or more they are called chronic in the case of major depression or dysthymia in the case of subthreshold depressive symptoms. While the specific values may not be particularly helpful there are insufficient empirical data to change these:

1) Acute – meeting one of the severity criteria for a minimum of 2 weeks and not longer than 2 years.

2) Chronic – meeting one of the severity criteria for longer than 2 years.

Given that the cut-off of 2 years is arbitrary it is best in practice to consider the specific duration and degree of persistence of symptoms for an individual in the context of the severity and course of the disorder.

Course

This was not explicitly considered as a classificatory issue in the previous guideline but it has important treatment implications, particularly for the likelihood of relapse/recurrence:

1) Number of lifetime depressive episodes and the interval between recent episodes: the number varies from a single/first episode to increasingly frequent recurrences. At least 2 months of full or partial remission is required to distinguish episodes.

2) Stage of episode: this refers to where an individual is in the course of their depression. In an episode it is useful to determine if the depression is worsening, static or improving and whether subthreshold depressive symptoms may reflect partial remission from prior major depression.

Conventionally, classification has distinguished between a single episode and two or more episodes (recurrent depression) irrespective of how long there has been between episodes and how many recurrences have occurred. However, someone who has had two episodes separated by decades has a different clinical course from someone with three episodes in a few years, therefore, noting the number of episodes and their recent pattern is important. There is uncertainty about the duration and extent of the recovery that is required to distinguish between different episodes of depression and a fluctuating course of a single episode. In practice this is less important than recognising the risk of persistent symptoms and of major depressive relapse/recurrence.

CLASSIFICATION IN RELATION TO DEPRESSION RATING SCALES AND QUESTIONNAIRES

Depression rating scales and questionnaires give ranges that are proposed to describe different severities of depression. Some of these were described in Appendix 13 of the previous guideline. In reconsidering this for the update it quickly became apparent, not only that there is no consensus for the proposed ranges, but also that the ranges in different rating scales and questionnaires do not correspond with each other. In addition there is a variable degree of correlation between different scales, which indicates that they do not measure precisely the same aspects of depression. When these factors are added to the need to consider more than symptoms in determining severity, and more than severity in considering diagnosis, the GDG was concerned not to perpetuate a spurious precision in relating scores in depression rating scales and questionnaires to the diagnosis or severity of depression, which must in the end be a clinical judgement.

Nevertheless it is necessary to try and translate trial evidence (which may only provide rating scales or questionnaire scores) into a meaningful clinical context as well as relating this guideline update to the previous guideline which used the APA (2000a) cut-offs. The change to DSM–IV-based diagnosis and the inclusion of minor depression (subthreshold depressive symptoms) in the update means that the descriptors of ranges previously given are no longer tenable. Table 145 gives the descriptors and ranges used in this guideline update, with the important caveat that these must not be taken as clear cut-offs or a short-cut to classify people with depression.

IMPLICATIONS OF THE PROPOSED CLASSIFICATION

An important implication is that symptom counts alone (for example, using the PHQ-9) should not be used to determine the presence or absence of a depressive disorder although this is an important part of the assessment. The score on a rating scale or questionnaire can contribute to the assessment of depression and rating scales are also useful to monitor treatment progress.

Another very important point to emphasise is that making a diagnosis of depression does not automatically imply a specific treatment. Making and agreeing a diagnosis of depression is a starting point in considering the most appropriate way of helping that individual in their particular circumstances. The evidence base for treatments considered in this guideline are based primarily on RCTs in which standardised criteria have been used to determine entry into the trial. Patients seen clinically are rarely assessed using standardised criteria reinforcing the need to be circumspect about an over-rigid extrapolation from RCTs to clinical practice.

Table 146: Levels of depression in relation to the HRSD and BDI in the guideline update compared with those suggested by the APA (2000a)

17-item Hamilton Rating Scale for Depression (HRSD)					
Guideline	**Not depressed**	**Subthreshold**	**Mild**	**Moderate**	**Severe**
APA (2000a)*	Not depressed	Mild	Moderate	Severe	Very severe
Score	0–7	8–13	14–18	19–22	23+
Beck Depression Inventory (BDI)					
Guideline update	**Not depressed**	**Subthreshold**	**Mild to moderate**	**Moderate to severe**	
APA (2000a)*	Not depressed	Mild	Moderate	Severe	
Score	0–9	10–16	17–29	30+	

**Used in the previous guideline.

Diagnosis using severity, duration and course (see above) necessarily only provides a partial description of the individual experience of depression. People with depression vary in the pattern of symptoms they experience, their family history, personalities, pre-morbid difficulties (for example, sexual abuse), psychological mindedness and current relational and social problems – all of which may significantly affect outcomes. It is also common for people with depression to have a comorbid psychiatric diagnosis, such as anxiety, social phobia, panic and various personality disorders (Brown et al., 2001), and physical comorbidity, or for the depression to occur in the context of bipolar disorder (not considered in this guideline). Gender and socioeconomic factors account for large variations in the population rates of depression, and few studies of pharmacological, psychological and other treatments for depression control for or examine these variations. This emphasises that choice of treatment is a complex process and involves negotiation and discussion with patients. Given the current limited knowledge about which factors are associated with better antidepressant or psychotherapy response, most decisions will rely upon clinical judgement and patient preference until there is further research evidence. Trials of treatment in unclear cases may be warranted but the uncertainty needs to be discussed with the patient and benefits from treatment carefully monitored.

Appendix 12: Search strategies for the identification of health economic evidence

General search strategies

- MEDLINE, EMBASE, PsycINFO, CINAHL - Ovid interface

1. (depression or depressive disorder or depression, postpartum or depressive disorder, major or dysthymic disorder or mood disorders or seasonal affective disorder).sh,id.
2. (affective disorders or depression or depression, postpartum or depression, reactive or dysthymic disorder or seasonal affective disorder).sh,id.

3. (depression or agitated depression or atypical depression or depressive psychosis or dysphoria or dysthymia or endogenous depression or involutional depression or major depression or masked depression or melancholia or mood disorder or mourning syndrome or organic depression or postoperative depression or premenstrual dysphoric disorder or pseudodementia or puerperal depression or reactive depression or recurrent brief depression or seasonal affective disorder).sh,id. or "mixed anxiety and depression"/ or "mixed depression and dementia"/
4. (affective disorders or anaclitic depression or dysthymic disorder or endogenous depression or major depression or postpartum depression or reactive depression or recurrent depression or treatment resistant depression or atypical depression or pseudodementia or sadness or seasonal affective disorder).sh,id. or "depression (emotion)"/
5. (depress$ or dysphori$ or dysthym$ or melanchol$ or seasonal affective disor- der$).tw.
6. or/1–5

- NHS Economic Evaluation Database, Health Technology Assessment Database – Wiley interface

7. #1 MeSH descriptor Depression, this term only
8. #2 MeSH descriptor Depressive Disorder explode all trees
9. #3 MeSH descriptor Mood Disorders, this term only
10. #4 (depress* or dysphori* or dysthym* or seasonal affective disorder* or melan-chol*):ti or (depress* or dysphori* or dysthym* or seasonal affective disorder* or melanchol*):ab
11. #5 (#1 OR #2 OR #3 OR #4)

- OHE HEED - Wiley interface

12. AX = depress*
13. AX = dysthym*
14. AX = dysphori*
15. AX = seasonal AND affective AND disorder*
16. CS = 1 OR 2 OR 3 OR 4

Health economics and quality-of-life search filters

- MEDLINE, EMBASE, PsycINFO, CINAHL – Ovid interface[a]

17. (budget$ or cost$ or economic$ or expenditure$ or fee$1 or fees$ or financ$ or health resource$ or money or pharmacoeconomic$ or socioeconomic$).hw,id.
18. (health care rationing or health priorities or medical savings accounts or quality adjusted life years or quality of life or resource allocation or value of life).sh,id. or "deductibles and coinsurance"/ or "health services needs and demand"/
19. (budget$ or cost$ or econom$ or expenditure$ or financ$ or fiscal$ or funding or pharmacoeconomic$ or price or prices or pricing).tw.
20. (QALY$ or lifeyear$ or life year$ or ((qualit$3 or value) adj3 (life or survival))).tw.
21. ((burden adj3 (disease or illness)) or (resource adj3 (allocation$ or utilit$)) or (value adj5 money)).tw.
22. ec.fs.
23. (or/1–6)

a With respect to 2a, search request 6 was ANDed with or/1–4 from the general search strategy only.

Appendix 13: Quality checklist for economic studies

Author: Date:

Title:

	Study design	Yes	No	NA
1	The research question is stated	☐	☐	
2	The economic importance of the research question is stated	☐	☐	
3	The viewpoint(s) of the analysis are clearly stated and justified	☐	☐	
4	The rationale for choosing the alternative programmes or interventions compared is stated	☐	☐	
5	The alternatives being compared are clearly described	☐	☐	
6	The form of economic evaluation is stated	☐	☐	
7	The choice of form of economic evaluation used is justified in relation to the questions addressed	☐	☐	
	Data collection			
1	The source of effectiveness estimates used is stated	☐	☐	
2	Details of the design and results of effectiveness study are given (if based on a single study)	☐	☐	☐
3	Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies)	☐	☐	☐
4	The primary outcome measure(s) for the economic evaluation are clearly stated	☐	☐	
5	Methods to value health states and other benefits are stated	☐	☐	☐
6	Details of the subjects from whom valuations were obtained are given	☐	☐	☐
7	Indirect costs (if included) are reported separately	☐	☐	☐
8	The relevance of indirect costs to the study question is discussed	☐	☐	☐
9	Quantities of resources are reported separately from their unit costs	☐	☐	
10	Methods for the estimation of quantities and unit costs are described	☐	☐	
11	Currency and price data are recorded	☐	☐	
12	Details of currency, price adjustments for inflation or currency conversion are given	☐	☐	
13	Details of any model used are given	☐	☐	☐

14	The choice of model used and the key parameters on which it is based are justified	☐	☐	☐
	Analysis and interpretation of results			
1	The time horizon of costs and benefits is stated	☐	☐	
2	The discount rate(s) is stated	☐	☐	☐
3	The choice of rate(s) is justified	☐	☐	☐
4	An explanation is given if costs or benefits are not discounted	☐	☐	☐
5	Details of statistical tests and confidence intervals are given for stochastic data	☐	☐	☐
6	The approach to sensitivity analysis is given	☐	☐	☐
7	The choice of variables for sensitivity analysis is	☐	☐	☐
8	The ranges over which the variables are varied are	☐	☐	☐
9	Relevant alternatives are compared	☐	☐	
10	Incremental analysis is reported	☐	☐	☐
11	Major outcomes are presented in a disaggregated as well as aggregated form	☐	☐	
12	The answer to the study question is given	☐	☐	
13	Conclusions follow from the data reported	☐	☐	
14	Conclusions are accompanied by the appropriate caveats	☐	☐	

Validity score: Yes/No/NA:

Appendix 14: Data extraction form for economic studies

Reviewer: **Date of review:**

Authors:

Publication Date:

Title:

Country:

Language:

Economic study design:

- CEA
- CCA
- CUA
- CBA
- CA
- CMA

Modelling:

- No
- Yes

Source of data for effect size measure(s):

- Meta-analysis
- RCT
- Quasi experimental study
- Cohort study
- Mirror image (before-after) study
- Expert opinion

Comments

Primary outcome measure(s) (please list):

_____

Interventions compared (please describe):

Treatment:

Comparator:

Setting (please describe):

_____

_____

Patient population characteristics (please describe):

_____

_____

_____

Perspective of analysis:

- Societal
- Patient and family
- Healthcare system
- Healthcare provider
- Third party payer
- Other:

Time frame of analysis:

Cost data:

- Primary
- Secondary

If secondary please specify:

Costs included:

Direct medical	Direct non-medical	Lost productivity
direct treatment	social care	income forgone due to illness
inpatient	social benefits	income forgone due to death
outpatient	travel costs	income forgone by caregiver

- day care
- community healthcare
- medication
- caregiver out-of-pocket
- criminal justice
- training of staff

Or

- staff
- medication
- consumables
- overhead
- capital equipment
- real estate

Others:

Currency: **Year of costing:**

Was discounting used?

- Yes, for benefits and costs
- Yes, but only for costs
- No

Discount rate used for costs:

Discount rate used for benefits:

Appendix 15: Evidence tables for economic studies

Appendix 15 - See Appendix V2.2

Appendix 16: Clinical evidence profiles

Appendix 16a – See Appendix V2.3

Appendix 16b – See Appendix V2.4

Appendix 16c – See Appendix V2.5

Appendix 16d – See Appendix V2.6

Appendix 17: Clinical study characteristics tables

Appendix 17a – See Appendix V2.7

Appendix 17b – See Appendix V2.8

Appendix 17c – See Appendix V2.9

Appendix 17d – See Appendix V2.10

Appendix 19: Clinical evidence forest plots

Appendix 19a – See Appendix V2.11

Appendix 19b – See Appendix V2.12

Appendix 19c – See Appendix V2.13

Appendix 19d – See Appendix V2.14

Appendix 21: Methodology from 2004 guideline

3 Methods used to develop this Guideline

3.1 Overview

The development of this guideline drew upon methods outlined by NICE (NICE, 2001; Eccles & Mason, 2001). A team of experts, professionals and patients, known as the Guideline Development Group (GDG), with support from NCCMH staff, undertook the development of a patient-centred, evidence-based guideline. There are six basic steps in the process of developing a guideline:

● Define the scope, which sets the parameters of the guideline and provides a focus and steer for the development work

● Define clinical questions considered important for practitioners and patients

● Develop criteria for evidence searching and search for evidence

● Design validated protocols for systematic review and apply to evidence recovered by search

● Synthesise and (meta-) analyse data retrieved, guided by the clinical questions, and produce evidence statements

● Answer clinical questions with evidence-based recommendations for clinical practice

The clinical practice recommendations made by the GDG are, therefore, derived from the most up-to-date and robust evidence base for the clinical and cost effectiveness of the treatments and services used in the management of depression. In addition, to ensure a patient and carer focus, the concerns of patients and carers regarding clinical practice have been highlighted and addressed by good practice points and recommendations agreed by the whole GDG. The evidence-based recommendations and good practice points are the core of this guideline.

3.2 The Guideline Development Group

The GDG consisted of patients, and professionals and academic experts in psychiatry, clinical psychology and general practice. NCCMH staff undertook the clinical and health economics literature searches, reviewed and presented the evidence to the GDG, managed the process and contributed to the drafting of the guideline.

3.2.1 Guideline Development Group meetings

Twenty-six GDG meetings were held between November 2001 and October 2003. During each day-long GDG meeting clinical evidence was reviewed and assessed to develop statements and recommendations. At each meeting all GDG members declared any potential conflict of interests. Patient and carer concerns were routinely discussed as part of a standing agenda.

3.2.2 Topic groups

The GDG divided its workload along clinically relevant lines in order to deal with the large volume of evidence efficiently. GDG members formed three topic groups: the Service topic group covered questions relating to the presentation of services to users, including screening, exercise and guided self-help; the Pharmacology topic group covered pharmacological treatments for depression; and the Psychology topic group covered psychotherapies. Each topic group was chaired by a GDG member with expert knowledge of the topic area. Topic groups refined the clinical definitions of treatment interventions, reviewed and prepared the evidence with the NCCMH review team. Topic group leaders reported the status of their group's work as part of the GDG standing agenda. They also assisted in drafting the section of the guideline relevant to the work of each topic group.

3.2.3 Patients and carers

Individuals with direct experience of services gave an integral patient focus to the GDG and the guideline. The GDG included three patients. They contributed as full GDG members to writing the clinical questions, helping to ensure that the evidence addressed their views and preferences, highlighting sensitive issues and terminology associated with depression, and bringing service-user research to the attention of the GDG. In drafting the guideline, they contributed to the editing of the first draft of the guideline's introduction and identified good practice points from the patient and carer perspective; their suggestions were incorporated before distributing the draft to the GDG for further review.

3.2.4 Special advisers

Special advisers who had specific expertise in one or more aspects of treatment and management relevant to the guideline assisted the GDG, commenting on specific aspects of the developing guideline and making presentations to the GDG. Appendix 2 lists those who agreed to act as special advisers.

3.2.5 National and international experts

National and international experts in the area under review were identified through the literature search and through the experience of the GDG members. These experts were contacted to recommend unpublished or soon-to-be published studies in order to ensure up-to-date evidence was included in the evidence base for the guideline. Appendix 5 lists researchers who were contacted.

3.3 Clinical questions

Clinical questions were used to guide the identification and interrogation of the evidence base. The questions were developed using a modified nominal group technique. The process began by asking each member of the GDG to submit as many questions as possible. The questions were then collated and refined by the review team. At a subsequent meeting, the guideline chair facilitated a discussion to further refine the questions. At this point, the GDG members were asked to rate each question for importance. The results of this process were then discussed and consensus reached about which questions would be of primary importance and which would be secondary. The GDG aimed to address all primary

questions, while secondary questions would only be covered time permitting. Appendix 6 lists the clinical questions.

3.4 Systematic clinical literature review

The aim of the clinical literature review was to identify and synthesise systematically all relevant evidence in order to answer the clinical questions developed by the GDG. Thus, clinical practice recommendations are evidence-based as far as possible. Where an existing NICE Technology Appraisal addressed one of the clinical questions, the GDG was obliged to adopt the relevant existing recommendations. If evidence was not available, then informal consensus methods were used (see Section 3.4.4) and the need for future research was specified. A stepwise, hierarchical approach was taken to locating and presenting evidence to the GDG. The NCCMH developed the methodology for this process with advice from the National Guidelines Support and Research Unit (NICE) and after considering recommendations from a range of other sources. These included:

- Centre for Clinical Policy and Practice of the New South Wales Health Department (Australia)
- Clinical Evidence Online
- Cochrane Collaboration
- New Zealand Guideline Group
- NHS Centre for Reviews and Dissemination
- Oxford Centre for Evidence-Based Medicine
- Scottish Intercollegiate Guidelines Network (SIGN)
- United States Agency for Health Research and Quality
- Oxford Systematic Review Development Programme.

3.4.1 The review process

Since most of the clinical questions for this guideline concerned interventions, much of the evidence base was formed from high quality randomised controlled trials (RCTs). Although there are a number of difficulties with the use of RCTs in the evaluation of interventions in mental health, this research design remains the most important method for establishing treatment efficacy (see introductions to later chapters for fuller discussions of this issue).

The review process involved:

- Developing search filters
- Searching for existing systematic reviews
- Searching for new RCTs
- Selecting studies
- Synthesising the evidence.

3.4.1.2 Searching for existing systematic reviews

The NCCMH review team undertook searches for existing systematic reviews of RCTs published in English since 1995 (an arbitrary cut-off date to reduce the number of references

found and to ensure recency), which would answer the clinical questions posed by the GDG. The initial searches were undertaken in December 2001 and January 2002, with update searches being carried out every two months until May 2002. A search of PubMed (MEDLINE) was also undertaken weekly beginning in April 2003 until the end of the guideline development process. The following databases were searched: EMBASE, MEDLINE, PsycINFO, Cochrane Library, CINAHL, Web of Science.

Systematic reviews were assessed for quality and eligibility (Appendices 8 and 9) before being assessed by the GDG for relevance to a clinical question. Searches were undertaken for RCTs published too late to be included in chosen systematic reviews beginning two years before the publication date of the review in question. Where authors stated the date searches had been undertaken, the NCCMH review team undertook new searches from the beginning of that year. Each study included in an existing review was subjected to the same quality checks as those located through NCCMH searches, and the data were re-extracted according to NCCMH protocols (see below). Where existing reviews had been undertaken using Review Manager (any version) authors were approached for data sets, although any used were checked for accuracy. For clinical questions where no existing systematic review was identified, searches were undertaken for all relevant evidence.

3.4.1.3 Searching for RCTs

For Service and Pharmacology topic area clinical questions, searches for RCTs were undertaken for each clinical question individually. However, RCTs to answer the clinical questions posed by the Psychology topic group were searched for together. For all questions the following electronic databases were searched: EMBASE, MEDLINE, PsycINFO, Cochrane Library, CINAHL. For the pharmacological review of St John's Wort, AMED was also searched. In addition, hand searches were also made of the reference lists of all eligible RCTs, as well as of the list of evidence submitted by registered stakeholders (Appendix 3). Known experts in the field (see Appendix 5), based both on the references identified in earlier steps and on advice from GDG members, were approached for unpublished RCTs[b]. Studies were considered provided a full trial report was available. Studies published in languages other than English were used provided a native speaker was available.

If no RCTs were found to answer a clinical question the GDG adopted a consensus process (see Section 3.4.4). Future guidelines will be able to update and extend the usable evidence base starting from the evidence collected, synthesised and analysed for this guideline.

3.4.1.4 Study selection

All references located in searches of electronic databases were downloaded into Reference Manager (ISI ResearchSoft, 2002) and searched liberally to exclude irrelevant papers. The titles of excluded papers were double-checked by a second reviewer. All primary-level studies included after the first scan of citations were acquired in full and re-evaluated for eligibility. Appendix 8 lists the standard inclusion and exclusion criteria. Additional eligibility criteria were developed to assess trials of pharmacotherapy, and these are listed in Chapter 7. All eligible papers were critically appraised for methodological quality (see Appendix 10). The eligibility of each study was confirmed by at least one member of the appropriate topic group.

For some clinical questions, it was necessary to prioritise the evidence with respect to the UK context. To make this process explicit, the topic group members took into account the following factors when assessing the evidence:

- Participant factors (e.g. gender, age, ethnicity)

b Unpublished full trial reports were accepted where sufficient information was available to judge eligibility and quality.

● Provider factors (e.g. model fidelity, the conditions under which the intervention was performed, the availability of experienced staff to undertake the procedure)

● Cultural factors (e.g. differences in standard care, differences in the welfare system).

It was the responsibility of each topic group to decide which prioritisation factors were relevant to each clinical question in light of the UK context, and then decide how they should modify their recommendations.

3.4.2 Synthesising the evidence

3.4.2.1 Outcomes

The vast majority of data extracted were scores on the Hamilton Rating Scale for

Depression (HRSD), Montgomery-Asberg Depression Rating Scale (MADRS) and Beck Depression Inventory (BDI) at the end of treatment and, where available, at follow-up. Both continuous (e.g. mean endpoint scores) and dichotomised data (e.g. number of people achieving below the cut-off for remission) were extracted. The GDG felt it was important to extract a variety of measures since relying on only one can be misleading. For example, dichotomising scores into remission and non-remission creates an artificial boundary, with patients just over the cut-off score often being clinically indistinguishable from those just under the cut-off. The GDG would also have liked to have been able to use quality of life measures as outcomes, but these are rarely reported.

In addition, where possible, sub-analyses were performed for severity of depression. Because very few studies gave information about participants' baseline severity of depression in terms of number of symptoms using the ICD classification (see Chapter 2), the mean depression score at baseline (most commonly an HRSD score) was used as a proxy measure. Scores were categorised mild, moderate, severe or very severe according to American Psychiatric Association criteria (APA, 2000a). Where necessary different versions of the HRSD were standardised using the method for prorating suggested by Walsh *et al.* (2002). The GDG used these categories with caution, mindful of the problematic nature of this proxy measure, in particular the variation in the standard deviation around baseline mean scores. Details of the categories and further information about the depression rating scales are in Appendix 13. When drawing up recommendations the GDG related the APA categories to ICD categories. This method does not take account of the severity of individual symptoms but is nonetheless a rough approximation to clinical severity.

3.4.2.2 Data extraction

Where possible, outcome data from all eligible studies that met quality criteria were extracted using a data extraction form (Appendix 11) and input into Review Manager 4.2 (Cochrane Collaboration, 2003). Where trial reports contained incomplete data and it was possible to contact the original authors, additional information was sought. Where mean endpoint or change scores were extracted and trial reports did not provide standard deviations, standard conversion formulas were used (see Appendix 12).

All dichotomous outcomes were calculated on an intention-to-treat basis (i.e. a 'oncerandomised- always-analyse' basis). This assumes that those participants who ceased to engage in the study – from whatever group – had an unfavourable outcome. The effects of high attrition rates (defined as more than 50% of participants in a particular group leaving treatment early) were examined with sensitivity analyses, and studies were removed from efficacy outcomes if the possibility of bias was detected.

Consultation was used to overcome difficulties with coding. Data from studies included in existing systematic reviews were extracted independently by one reviewer directly into Review Manager and checked by a second reviewer. Where consensus could not be

reached, a third reviewer was consulted. Masked assessment (i.e. blind to the journal from which the article comes, the authors, the institution, and the magnitude of the effect) was not used since it is unclear that doing so reduces bias (Jadad *et al.*, 1996; Berlin, 1997).

Information describing each study was also extracted and input into Review Manager 4.2. This was used to generate evidence tables (see Appendix 17 on the CD).

3.4.2.3 Meta-analysis

Where possible, meta-analysis was used to synthesise data. If necessary, sub-analyses were used to answer clinical questions not addressed in the original studies or reviews.

The GDG was given a graphical presentation of the results using forest plots generated with Review Manager. Each forest plot displayed the effect size and 95% confidence interval (CI) for each study as well as the overall summary statistic with its 95% CI. The graphs were organised so that the display of data in the area to the left of the 'line of no effect' indicated a 'favourable' outcome for the treatment in question[c].

Dichotomous outcomes were presented as relative risks (RR) with the associated 95% CI (see Figure 1). A relative risk (or risk ratio) is the ratio of the treatment event rate to the control event rate. A RR of 1 indicates no difference between treatment and control. In Figure 1, the overall RR of 0.73 indicates that the event rate (i.e. non-remission rate) associated with intervention A is about half of that with the control intervention, or in other words, intervention A reduces non-remission rates by 27%. In addition, the 95% CI around the RR does not cross the 'line of no effect' indicating that this is a statistically significant effect. The CI shows with 95% certainty the range within which the true treatment effect should lie.

It had been planned to calculate the number needed to treat (NNT) (or number needed to harm (NNH)) for dichotomous outcomes with statistically significant effect sizes. However, when the baseline risk (i.e. control group event rate (CER)) or length of followup varies, NNT is a poor summary of the treatment effect, especially with low risk or where the CER is dissimilar across studies in a meta-analysis (Deeks, 2002). Since it was not possible to calculate the baseline risk for most outcomes NNT and NNH have not been calculated.

Continuous outcomes were analysed as weighted mean differences (WMD) or standardised mean differences (SMD) when different measures (or different versions of the same measure) were used in different studies to estimate the same underlying effect (see Figure 2).

Figure 1: Example of a forest plot displaying dichotomous data.

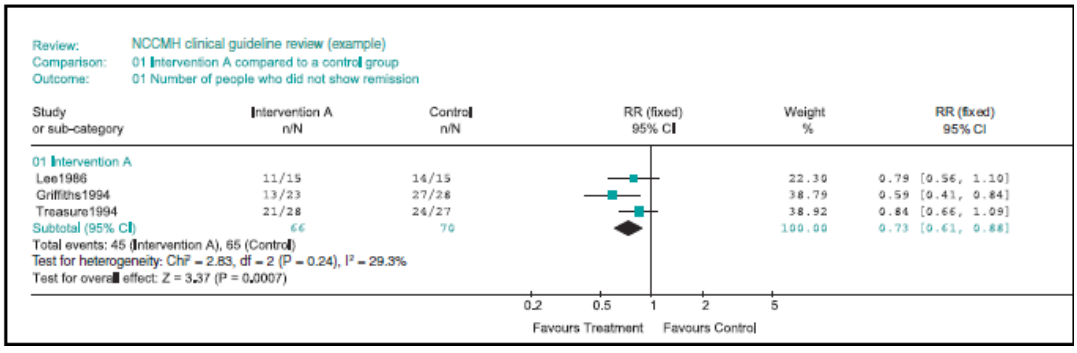
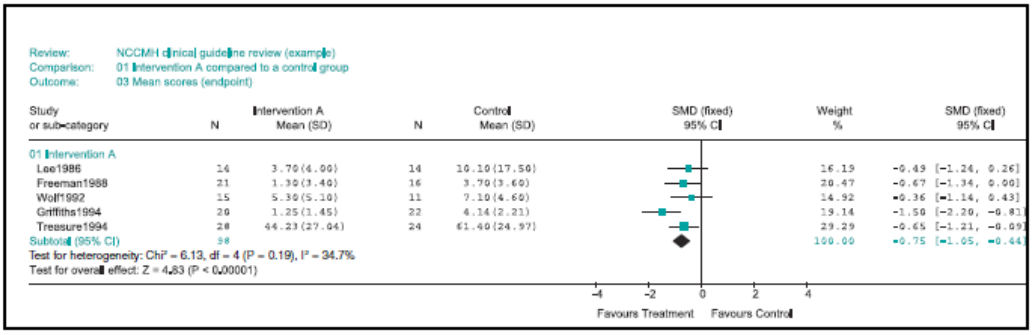


Figure 2: Example of a forest plot displaying continuous data.

c The exceptions to this are: the review of amitriptyline, for which the GDG were provided with a data set for an existing systematic review (Barbui & Hotopf, 2001), and the overview of TCA data.



To check for heterogeneity between studies, both the I2 test of heterogeneity and the chi-squared test of heterogeneity (p<0.10), as well as visual inspection of the forest plots, were used. The I2 statistic describes the proportion of total variation in study estimates that is due to heterogeneity (Higgins & Thompson, 2002). An I2 of less than 30% was taken to indicate mild heterogeneity and a fixed effects model was used to synthesise the results. This assumes that the underlying effect is the same (Egger *et al.*, 2001). An I2 of more than 50% was taken as notable heterogeneity. In this case, an attempt was made to explain the variation. If studies with heterogeneous results were found to be comparable, a random effects model was used to summarise the results (DerSimonian & Laird, 1986). In the random effects analysis, heterogeneity is accounted for both in the width of CIs and in the estimate of the treatment effect. With decreasing heterogeneity the random effects approach moves asymptotically towards a fixed effects model. An I2 of 30% to 50% was taken to indicate moderate heterogeneity. In this case, both the chi-squared test of heterogeneity and a visual inspection of the forest plot were used to decide between a fixed and random effects model.

To explore the possibility that the results entered into each meta-analysis suffered from publication bias, data from included studies were entered, where there were sufficient data, into a funnel plot. Asymmetry of the plot was taken to indicate possible publication bias and was investigated further.

3.4.3 Developing statements and graded recommendations

The summary statistics (effect sizes (ES)) and evidence tables formed the basis for developing clinical statements and recommendations.

3.4.3.1 Developing statements

For each outcome a clinical statement describing the evidence found was developed. To do this both the statistical and the clinical significance (i.e. the likely benefit to patients) of the summary statistic were taken into account.

Assessing statistically significant summary statistics

To assess clinical significance where a statistically significant summary was obtained (after controlling for heterogeneity) the GDG adopted the following ‘rules of thumb’,in addition to taking into account the trial population and nature of the outcome:

For dichotomous outcomes a RR of 0.80 or less was considered clinically significant (see Section 3.4.2.3).

For continuous outcomes for which an SMD was calculated (for example, when data from different versions of a scale are combined), an effect size of ~0.5 (a ‘medium’ effect size; Cohen, 1988) or higher was considered clinically significant. Where a WMD was calculated, a between group difference of at least three points (two points for treatment resistant depression) was considered clinically significant for both BDI and HRSD.

Once clinical significance had been established the strength of the evidence was

assessed by examining the 95% CIs surrounding the ES. For level I evidence, where the effect size was judged clinically important for the full range of plausible estimates, the result was characterised as 'strong evidence' (i.e. S1, Flowchart 1: Guideline Statement Decision Tree). For non-level I evidence or in situations where the CI also included clinically unimportant effects, the result was characterised as 'some evidence' (i.e. S2).

Where an ES was statistically significant, but *not* clinically significant and the CI excluded values judged clinically important, the result was characterised as 'unlikely to be clinically significant' (S3). Alternatively, if the CI included clinically important values, the result was characterised as 'insufficient to determine clinical significance' (S6).

Assessing non-statistically significant summary statistics

Where a non-statistically significant ES was obtained, the GDG reviewed the trial population, nature of the outcome, size of the effect and, in particular, the CI surrounding the result. If the CI was narrow and excluded a clinically significant ES, this was seen as indicating evidence of 'no clinically significant difference' (S4), but where the CI was wide this was seen as indicating 'insufficient evidence' to determine if there was a clinically significant difference or not (S5).

In order to facilitate consistency in generating and drafting the clinical statements the GDG utilised a statement decision tree (see Flowchart 1 overleaf). The flowchart was designed to assist with, but not replace, clinical judgement.

3.4.3.2 Developing graded recommendations

Once all evidence statements relating to a particular clinical question were finalised and agreed by the GDG, the associated recommendations were produced and graded. Recommendations were graded A to C based on the level of associated evidence, or noted as coming from a previous NICE guideline or health technology appraisal (see Table 1 overleaf).

Grading allowed the GDG to distinguish between the level of evidence and the strength of the associated recommendation. It is possible that a statement of evidence would cover only one part of an area in which a recommendation was to be made or would cover it in a way that would conflict with other evidence. In order to produce more comprehensive recommendations suitable for people in England and Wales, there were times when the GDG had to extrapolate from the available evidence based on their combined clinical experience. The resulting recommendations were then graded with a lower grade (e.g. a 'B' grade where data were based upon Level I evidence). This allowed the GDG to moderate recommendations based on factors other than the strength of evidence. Such considerations include the applicability of the evidence to the people in question, economic considerations, values of the development group and society, or the group's awareness of practical issues (Eccles *et al.*, 1998).

3.4.4 Method used to answer a clinical question in the absence of appropriately designed, high-quality research

In the absence of level I evidence (or a level that is appropriate to the question), or where the GDG were of the opinion (on the basis of previous searches or their knowledge of the literature) that there was unlikely to be such evidence, an informal consensus process was adopted. This process focused on those questions that the GDG considered a priority.

3.4.4.1 Informal consensus

The starting point for this process of informal consensus was that a member of the topic group identified, with help from the systematic reviewer, a narrative review that most directly addressed the clinical question. Where this was not possible, a brief review of the recent literature was initiated. This existing narrative review or new review was used as a basis for beginning an iterative process to identify lower levels of evidence relevant to the clinical question and to lead to written statements for the guideline. The process involved a number of steps:

1. A description of what is known about the issues concerning the clinical question was written by one of the topic group members.

2. Evidence from the existing review or new review was then presented in narrative form to the GDG and further comments were sought about the evidence and its perceived relevance to the clinical question.

3. Based on the feedback from the GDG, additional information was sought and added to the information collected. This may include studies that did not directly address the clinical question but were thought to contain relevant data.

4. If, during the course of preparing the report, a significant body of primary-level studies (of appropriate design to answer the question) was identified, a full systematic review was done.

5. At this time, subject possibly to further reviews of the evidence, a series of statements that directly addressed the clinical question was developed.

6. Following this, on occasions and as deemed appropriate by the development group, the report was then sent to appointed experts outside of the GDG for peer review and comment. The information from this process was then fed back to the GDG for further discussion of the statements.

7. Recommendations were then developed and could also be sent for further external peer review.

8. After this final stage of comment, the statements and recommendations were again reviewed and agreed upon by the GDG.

3.5 Evidence on safety and harm

In the UK the licensing and post-licensing safety monitoring of medicines is undertaken by the Medicines and Healthcare products Regulatory Agency (MHRA). During the development of this guideline the safety of some drugs used to treat depression (selective serotonin reuptake inhibitors (SSRIs), mirtazapine and venlafaxine) was formally reviewed by the MHRA on behalf of the Committee on Safety of Medicines (CSM). The CSM convened a working group to look at this issue (the SSRI Expert Working Group (EWG)). The EWG's findings were made available to the GDG, and used in addition to the efficacy and safety data reviewed during the guideline development process in drawing up recommendations. In particular, data on discontinuation/ withdrawal symptoms, cardiotoxicity, dose, and suicidality and self-harm, were used, together with information on changes to produce licences as a result of the EWG's report to the CSM (MHRA, 2004). The Marketing Authorisation Holder (the pharmaceutical company responsible for the drug in question) analysed data from clinical trials for each relevant drug, in accordance with a protocol specified by the EWG. These reviews formed the basis of the EWG's deliberations, and it should be noted that not all trial data were made available to the EWG (MHRA, 2004). The EWG used other data, including a number of analyses of the General Practice Research Database (for example, Jick *et al.*, 2004), along with spontaneous reporting of adverse drug reactions (via the MHRA's Yellow Card scheme).

Flowchart 1: Guideline Statement Decision Tree.

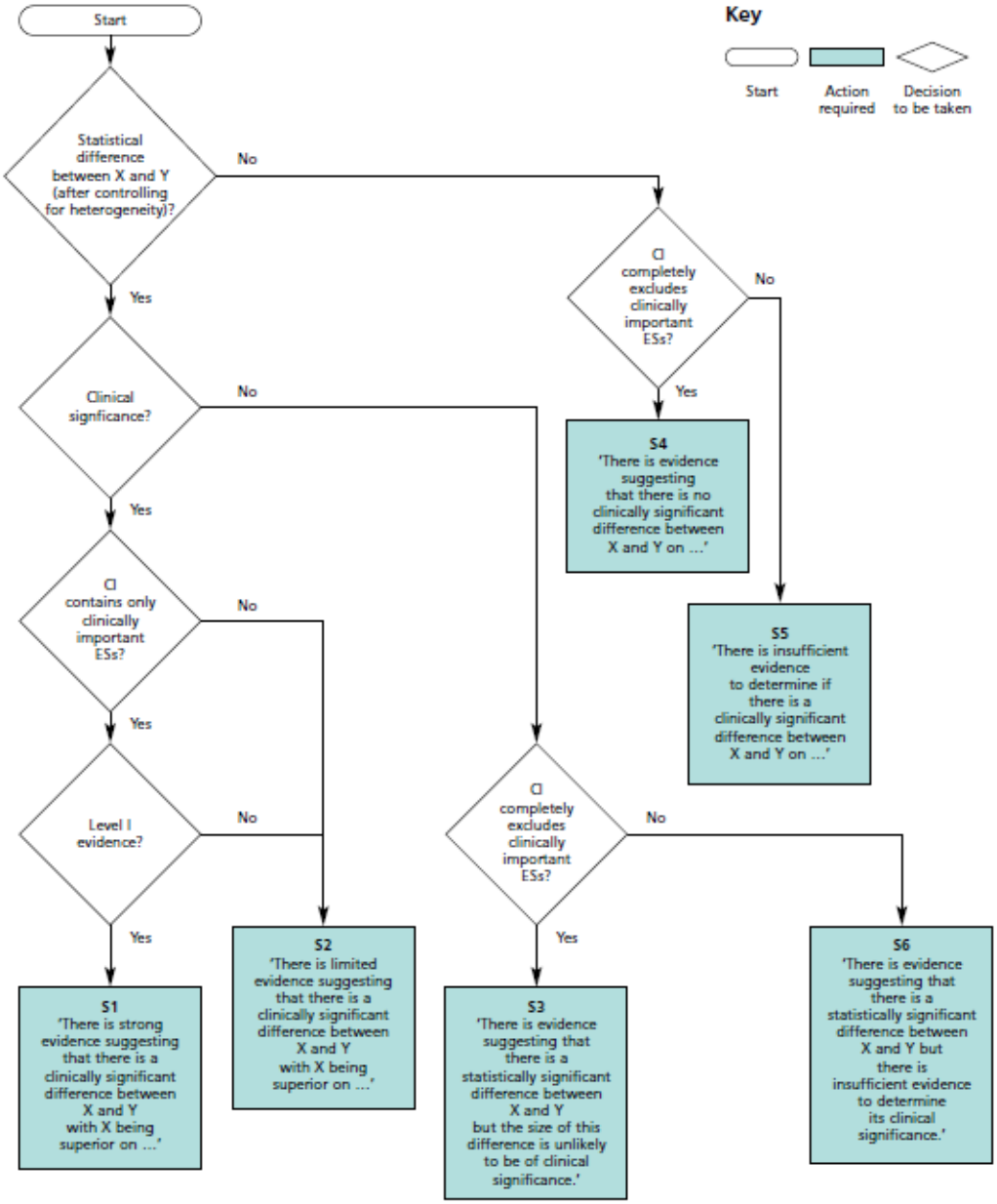


Table 1: Hierarchy of evidence and recommendations grading scheme.

Level	Type of evidence	Grade	Evidence
I	Evidence obtained from a single randomised controlled trial or a meta-analysis of randomised controlled trials	A	At least one randomised controlled trial as part of a body of literature of overall good quality and consistency addressing the specific recommendation (evidence level I) without extrapolation
IIa	Evidence obtained from at least one well-designed controlled study without randomisation	B	Well-conducted clinical studies but no randomised clinical trials on the topic of recommendation (evidence levels II or III); or extrapolated from level I evidence
IIb	Evidence obtained from at least one other well-designed quasi-experimental study		
III	Evidence obtained from well-designed non-experimental descriptive studies, such as comparative studies, correlation studies and case-control studies		
IV	Evidence obtained from expert committee reports or opinions and/or clinical experiences of respected authorities	C	Expert committee reports or opinions and/or clinical experiences of respected authorities (evidence level IV) or extrapolated from level I or II evidence. This grading indicates that directly applicable clinical studies of good quality are absent or not readily available
		GPP	Recommended good practice based on the clinical experience of the GDG
NICE	Evidence from NICE guideline or Technology Appraisal	NICE	Evidence from NICE guideline or Technology Appraisal

Adapted from Eccles, M. & Mason, J. (2001), How to develop cost-conscious guidelines. *Health Technology Assessment, 5*(16); Department of Health (1996), *Clinical Guidelines: Using clinical guidelines to improve patient care within the NHS*. Leeds: NHS Executive.

3.6 Health economics review strategies

The aim of the health economics review was to contribute to the guideline development process data on the economic burden of depression. Evidence of the cost-effectiveness of different treatment options for depression was collected and assessed in order to help the decision-making process. See Chapter 9, Health economics evidence, for the detailed review strategies.

3.7 Stakeholder contributions

Professionals, patients and companies have contributed to and commented on the guideline at key stages in its development. Stakeholders for this guideline include:

● Patient/carer stakeholders: the national patient and carer organisations that represent people whose care is described in this guideline

● Professional stakeholders: the national organisations that represent healthcare professionals who are providing services to patients

● Commercial stakeholders: the companies that manufacture medicines used in the treatment of depression

● Primary Care Trusts

● Department of Health and Welsh Assembly Government.

Stakeholders have been involved in the guideline's development at the following points:

● Commenting on the initial scope of the guideline and attended a briefing meeting held by NICE

● Contributing lists of evidence to the GDG

● Commenting on the first and second drafts of the guideline.

3.8 Validation of this guideline

This guideline has been validated through two consultation exercises. Drafts of the full and NICE versions of the guideline were submitted to the NICE Guidelines Review Panel and posted on the NICE website (www.nice.org.uk). Stakeholders and other reviewers nominated by the GDG were then informed that the documents were available.

The GDG reviewed comments from stakeholders, the NICE Guidelines Review Panel, a number of health authority and trust representatives and a wide range of national and international experts from the first round of consultation. The GDG then responded to all comments and prepared final consultation drafts of all three versions of the guideline – the full guideline, the NICE guideline, and the information for the public. These were made available on the NICE website, and stakeholders were informed. Following consultation, the drafts were amended and responses to any comments were made. The final drafts were then submitted to NICE to be signed off after review by the Guidelines Review Panel.

Appendices from 2009 short guideline

Appendix C: Assessing depression and its severity

As set out in the introduction to this guideline, the assessment of depression is based on the criteria in DSM-IV. Assessment should include the number and severity of symptoms, duration of the current episode, and course of illness.

Key symptoms:

- persistent sadness or low mood; and/or
- marked loss of interests or pleasure.

At least one of these, most days, most of the time for at least 2 weeks.

If any of above present, ask about associated symptoms:

- disturbed sleep (decreased or increased compared to usual)
- decreased or increased appetite and/or weight
- fatigue or loss of energy
- agitation or slowing of movements
- poor concentration or indecisiveness
- feelings of worthlessness or excessive or inappropriate guilt
- suicidal thoughts or acts.

Then ask about duration and associated disability, past and family history of mood disorders, and availability of social support

1. Factors that favour general advice and active monitoring:

- four or fewer of the above symptoms with little associated disability
- symptoms intermittent, or less than 2 weeks' duration
- recent onset with identified stressor
- no past or family history of depression
- social support available
- lack of suicidal thoughts.

2. Factors that favour more active treatment in primary care:

- five or more symptoms with associated disability
- persistent or long-standing symptoms
- personal or family history of depression
- low social support
- occasional suicidal thoughts.

3. Factors that favour referral to mental health professionals:

- inadequate or incomplete response to two or more interventions
- recurrent episode within 1 year of last one
- history suggestive of bipolar disorder
- the person with depression or relatives request referral
- more persistent suicidal thoughts
- self-neglect.

4. Factors that favour urgent referral to specialist mental health services

- actively suicidal ideas or plans
- psychotic symptoms
- severe agitation accompanying severe symptoms
- severe self-neglect.

Depression defifinitions (These are taken from DSM-IV. ICD-10 is similar but the threshold for mild depression is lower at 4 symptoms.)

Subthreshold depressive symptoms: Fewer than 5 symptoms of depression.

Mild depression: Few, if any, symptoms in excess of the 5 required to make the diagnosis, and symptoms result in only minor functional impairment.

Moderate depression: Symptoms or functional impairment are between 'mild' and 'severe'.

Severe depression: Most symptoms, and the symptoms markedly interfere with functioning. Can occur with or without psychotic symptoms.

Appendix D: Recommendations from NICE clinical guideline 23

The following recommendations have been taken from the previous NICE clinical guideline 23 'Management of depression in primary and secondary care'. Note that the evidence for these recommendations has not been updated. Any wording changes have been made for clarification only.

Recommendation number in updated NICE clinical guideline 90	Recommendation number in NICE clinical guideline 23
1.1.1.2	1.1.2.1/1.1.3.1/1.1.3.2
1.1.1.3	1.1.2.3
1.1.1.4	1.1.2.2
1.1.4.6	1.1.6.4/1.1.6.6
1.1.5.2	1.1.3.3
1.3.2.1	1.5.1.1
1.3.2.3	1.1.6.5
1.3.2.4	1.5.2.6/1.5.2.7
1.4.1.1	1.1.1.1
1.4.1.2	1.4.1.1
1.4.4.2	1.5.2.37/1.5.2.38
1.5.2.2	1.5.2.13/1.5.2.14
1.5.2.6	1.5.2.10
1.5.2.7	1.5.2.5
1.5.2.9	1.5.2.29

1.6.1.4	1.5.5.3/1.5.5.4
1.8.1.11	1.6.2.14
1.10.1.1	1.6.1.1
1.10.1.2	1.6.1.2
1.10.1.3	1.6.1.3
1.10.2.1	1.7.1.1
1.10.2.3	1.7.1.2
1.10.3.1	1.6.5.1