Cochrane Library

Cochrane Database of Systematic Reviews

Interventions to improve return to work in depressed people (Review)

Nieuwenhuijsen K, Verbeek JH, Neumeyer-Gromen A, Verhoeven AC, Bültmann U, Faber B

Nieuwenhuijsen K, Verbeek JH, Neumeyer-Gromen A, Verhoeven AC, Bültmann U, Faber B.
Interventions to improve return to work in depressed people.
Cochrane Database of Systematic Reviews 2020, Issue 10. Art. No.: CD006237.
DOI: 10.1002/14651858.CD006237.pub4.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Interventions to improve return to work in depressed people

Karen Nieuwenhuijsen[1], Jos H Verbeek[2], Angela Neumeyer-Gromen[3], Arco C Verhoeven[4], Ute Bültmann[5], Babs Faber[6]

[1]Department of Public and Occupational Health, Coronel Institute of Occupational Health, Amsterdam Public Health research institute, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands. [2]Cochrane Work, Department of Public and Occupational Health, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands. [3]Deutsche Krankenhausgesellschaft (The German Hospital Federation), Berlin, Germany. [4]SGBO, Radboud University Medical Centre, Nijmegen, Netherlands. [5]Department of Health Sciences, Community and Occupational Medicine, University Medical Center Groningen, University of Groningen, Groningen, Netherlands. [6]ENRGY in Business Group, Amsterdam, Netherlands

Contact address: Karen Nieuwenhuijsen, k.nieuwenhuijsen@amsterdamumc.nl.

Editorial group: Cochrane Work Group.
Publication status and date: Edited (no change to conclusions), published in Issue 11, 2020.

Citation: Nieuwenhuijsen K, Verbeek JH, Neumeyer-Gromen A, Verhoeven AC, Bültmann U, Faber B. Interventions to improve return to work in depressed people. *Cochrane Database of Systematic Reviews* 2020, Issue 10. Art. No.: CD006237. DOI: 10.1002/14651858.CD006237.pub4.

ABSTRACT

Background

Work disability such as sickness absence is common in people with depression.

Objectives

To evaluate the effectiveness of interventions aimed at reducing work disability in employees with depressive disorders.

Search methods

We searched CENTRAL (The Cochrane Library), MEDLINE, Embase, CINAHL, and PsycINFO until April 4th 2020.

Selection criteria

We included randomised controlled trials (RCTs) and cluster-RCTs of work-directed and clinical interventions for depressed people that included sickness absence days or being off work as an outcome. We also analysed the effects on depression and work functioning.

Data collection and analysis

Two review authors independently extracted the data and rated the certainty of the evidence using GRADE. We used standardised mean differences (SMDs) or risk ratios (RR) with 95% confidence intervals (CI) to pool study results in studies we judged to be sufficiently similar.

Main results

In this update, we added 23 new studies. In total, we included 45 studies with 88 study arms, involving 12,109 participants with either a major depressive disorder or a high level of depressive symptoms.

Risk of bias

The most common types of bias risk were detection bias (27 studies) and attrition bias (22 studies), both for the outcome of sickness absence.

Work-directed interventions

Work-directed interventions combined with clinical interventions

A combination of a work-directed intervention and a clinical intervention probably reduces sickness absence days within the first year of follow-up (SMD -0.25, 95% CI -0.38 to -0.12; 9 studies; moderate-certainty evidence). This translates back to 0.5 fewer (95% CI -0.7 to -0.2) sick leave days in the past two weeks or 25 fewer days during one year (95% CI -37.5 to -11.8). The intervention does not lead to fewer persons being off work beyond one year follow-up (RR 0.96, 95% CI 0.85 to 1.09; 2 studies, high-certainty evidence). The intervention may reduce depressive symptoms (SMD -0.25, 95% CI -0.49 to -0.01; 8 studies, low-certainty evidence) and probably has a small effect on work functioning (SMD -0.19, 95% CI -0.42 to 0.06; 5 studies, moderate-certainty evidence) within the first year of follow-up.

Stand alone work-directed interventions

A specific work-directed intervention alone may increase the number of sickness absence days compared with work-directed care as usual (SMD 0.39, 95% CI 0.04 to 0.74; 2 studies, low-certainty evidence) but probably does not lead to more people being off work within the first year of follow-up (RR 0.93, 95% CI 0.77 to 1.11; 1 study, moderate-certainty evidence) or beyond (RR 1.00, 95% CI 0.82 to 1.22; 2 studies, moderate-certainty evidence). There is probably no effect on depressive symptoms (SMD -0.10, 95% -0.30 CI to 0.10; 4 studies, moderate-certainty evidence) within the first year of follow-up and there may be no effect on depressive symptoms beyond that time (SMD 0.18, 95% CI -0.13 to 0.49; 1 study, low-certainty evidence). The intervention may also not lead to better work functioning (SMD -0.32, 95% CI -0.90 to 0.26; 1 study, low-certainty evidence) within the first year of follow-up.

Psychological interventions

A psychological intervention, either face-to-face, or an E-mental health intervention, with or without professional guidance, may reduce the number of sickness absence days, compared with care as usual (SMD -0.15, 95% CI -0.28 to -0.03; 9 studies, low-certainty evidence). It may also reduce depressive symptoms (SMD -0.30, 95% CI -0.45 to -0.15, 8 studies, low-certainty evidence). We are uncertain whether these psychological interventions improve work ability (SMD -0.15 95% CI -0.46 to 0.57; 1 study; very low-certainty evidence).

Psychological intervention combined with antidepressant medication

Two studies compared the effect of a psychological intervention combined with antidepressants to antidepressants alone. One study combined psychodynamic therapy with tricyclic antidepressant (TCA) medication and another combined telephone-administered cognitive behavioural therapy (CBT) with a selective serotonin reuptake inhibitor (SSRI). We are uncertain if this intervention reduces the number of sickness absence days (SMD -0.38, 95% CI -0.99 to 0.24; 2 studies, very low-certainty evidence) but found that there may be no effect on depressive symptoms (SMD -0.19, 95% CI -0.50 to 0.12; 2 studies, low-certainty evidence).

Antidepressant medication only

Three studies compared the effectiveness of SSRI to selective norepinephrine reuptake inhibitor (SNRI) medication on reducing sickness absence and yielded highly inconsistent results.

Improved care

Overall, interventions to improve care did not lead to fewer sickness absence days, compared to care as usual (SMD -0.05, 95% CI -0.16 to 0.06; 7 studies, moderate-certainty evidence). However, in studies with a low risk of bias, the intervention probably leads to fewer sickness absence days in the first year of follow-up (SMD -0.20, 95% CI -0.35 to -0.05; 2 studies; moderate-certainty evidence). Improved care probably leads to fewer depressive symptoms (SMD -0.21, 95% CI -0.35 to -0.07; 7 studies, moderate-certainty evidence) but may possibly lead to a decrease in work-functioning (SMD 0.5, 95% CI 0.34 to 0.66; 1 study; moderate-certainty evidence).

Exercise

Supervised strength exercise may reduce sickness absence, compared to relaxation (SMD -1.11; 95% CI -1.68 to -0.54; one study, low-certainty evidence). However, aerobic exercise probably is not more effective than relaxation or stretching (SMD -0.06; 95% CI -0.36 to 0.24; 2 studies, moderate-certainty evidence). Both studies found no differences between the two conditions in depressive symptoms.

Authors' conclusions

A combination of a work-directed intervention and a clinical intervention probably reduces the number of sickness absence days, but at the end of one year or longer follow-up, this does not lead to more people in the intervention group being at work. The intervention may also reduce depressive symptoms and probably increases work functioning more than care as usual. Specific work-directed interventions may not be more effective than usual work-directed care alone. Psychological interventions may reduce the number of sickness absence days, compared with care as usual. Interventions to improve clinical care probably lead to lower sickness absence and lower levels of depression, compared with care as usual. There was no evidence of a difference in effect on sickness absence of one antidepressant medication compared to another. Further research is needed to assess which combination of work-directed and clinical interventions works best.

PLAIN LANGUAGE SUMMARY

What are the best ways to help people with depression go back to work?

What is depression?

Depression is a common mental health problem that can cause a persistent feeling of sadness and loss of interest in people, activities, and things that were once enjoyable. A person with depression may feel tearful, irritable, or tired most of the time, and may have problems with sleep, concentration, and memory.

Depression may affect people's ability to work. People with depression may be absent from work (off sick), or feel less able to cope with working.

Going back to work

Reducing depressive symptoms may help people with depression to go back to work. Treatments include medications and psychological (talking) therapies, or a combination of both. Changes at the workplace could also help, such as:

changing a person's tasks or working hours;

supporting them in a gradual return to work; or

helping them to cope better with certain work situations.

Why we did this Cochrane Review

Work can improve a person's physical and mental well-being; it helps build confidence and self-esteem, allows people to socialise, and provides money. We wanted to find out if workplace changes and clinical programmes could help people with depression to return to work.

What did we do?

We searched for studies that looked at whether workplace changes and clinical programmes affected the amount of sick leave taken by people with depression. Clinical programmes included: medicines (anti-depressants); psychological therapies; improved healthcare by doctors; and other programmes such as exercise and diet.

Search date: we included evidence published up to 4 April 2020.

What we found

We found 45 studies in 12,109 people with depression. The studies took place in Europe (34 studies), the USA (8), Australia (2) and Canada (1).

The effects of 'care as usual' were compared with those of workplace changes and clinical programmes to find out:

how many days people with depression were on sick leave

how many people with depression were off work;

people's symptoms of depression; and

how well people with depression could cope with their work.

What are the results of our review?

Our main findings within the first year of follow-up, for workplace changes or treatments compared with usual care, are listed below.

Workplace changes combined with a clinical programme:

probably reduce the number of days on sick leave (on average, by 25 days for each person over one year; 9 studies; 1292 participants);

do not reduce the number of people off work (2 studies; 1025 participants);

may reduce symptoms of depression (8 studies; 1091 participants); and

may improve ability to cope with work (5 studies; 926 participants).

Workplace changes alone:

may increase the number of days on sick leave (2 studies, 130 participants);

probably do not lead to more people off work (1 study; 226 participants);

probably do not affect symptoms of depression (4 studies; 390 participants); and

may not improve ability to cope with work (1 study; 48 participants).

Improved healthcare alone:

probably reduces the number of days on sick leave, by 20 days (in two, well-conducted studies in 692 participants, although not in all 7 studies, in 1912 participants);

probably reduces symptoms of depression (7 studies; 1808 participants); and

may reduce ability to cope with work (1 study; 604 participants).

Psychological therapies alone:

may reduce the number of days off work, by 15 days (9 studies; 1649 participants); and

may reduce symptoms of depression (8 studies; 1255 participants).

We are uncertain if psychological therapies alone affect people's ability to cope with work (1 study; 58 participants).

How reliable are these results?

Our confidence in these results is mostly moderate to low. Some findings are based on small numbers of studies, in small numbers of participants. We also found limitations in the ways some studies were designed, conducted and reported.

Key messages

Combining workplace changes with a clinical programme probably helps people with depression to return to work more quickly and to take fewer days off sick. We need more evidence to assess which combination of workplace changes and clinical programmes works best.

Improved healthcare probably also helps people with depression to take fewer days off sick.

Summary of findings 1. Work-directed plus clinical intervention compared to care as usual in depressed people, medium-term follow-up

Work-directed plus clinical intervention compared to care as usual (medium-term) in depressed people

Patients: Depressed persons
Setting: Various: workplaces, outpatient and occupational healthcare
Intervention: Work-directed plus clinical
Control: Care as usual (medium-term)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with care as usual	Risk with work-directed intervention plus clinical intervention				
Sickness absence days	-	SMD 0.25 SD lower (0.38 lower to 0.12 lower)	-	1292 (9 RCTs)	⊕⊕⊕⊖ MODERATE [1]	The SMD translates back to -0.5 days per 2 weeks (CI -0.7 to -0.2) or -24.7 days in 12 months (-37.5 to -11.8).
On sick leave	417 per 1.000	401 per 1.000 (355 tot 455)	RR 0.96 (0.85 to 1.09)	1025 (2 RCTs)	⊕⊕⊕⊕ HIGH	
Depressive symptoms-	-	SMD 0.25 SD lower (0.49 lower to 0.01 lower)	-	1091 (8 RCTs)	⊕⊕⊖⊖ LOW [2 3]	
Work functioning	-	SMD 0.19 SD lower (0.43 lower to 0.06 higher)	-	926 (5 RCTs)	⊕⊕⊖⊖ LOW [1 4 5]	

The risk in the intervention group (and the 95% CI) is based on the risk in the control group and the **relative effect** of the intervention (and the 95% CI).

CI: Confidence interval; **RCT:** Randomised controlled trial; **RR:** Risk ratio; **SMD:** Standardised mean difference.

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]A majority of the studies in the meta-analysis (in terms of weights) showed high or unclear risk on the randomisation items (sequence and concealment), blinded outcome assessment or attrition. We therefore rated down one level due to a high risk of bias.
[2]Depression is self-reported and participants were not blinded. We rated down one level due to a high risk of bias.
[3]Study effects varied with some clearly indicating beneficial results and some not. We rated down one level due to imprecision.
[4]Rated down one level due to inconsistency (I^2 61%).
[5]Pooled effect size includes small harmful effec. Rated down one level due to wide CI (imprecision)

Summary of findings 2. Work-directed intervention compared to care as usual in depressed people, medium-term follow-up

Work-directed intervention compared to care as usual in depressed people

Patient or population: Depressed persons
Setting: Workplace and occupational healthcare
Intervention: Work-directed
Comparison: Care as usual (medium-term)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with care as usual	Risk with work-directed intervention				
Sickness absence days, medium-term follow-up	-	SMD 0.39 higher (0.04 higher to 0.74 higher)	-	130 (2 RCTs)	⊕⊕⊖⊖ LOW [1 2]	The SMD translates back to + 0.7 days in two weeks (95% CI 0.1 to 1.3) or + 38 days in 12 months (95% CI 3.9 to 73).
Off work, medium-term follow-up	708 per 1.000	658 per 1.000 (545 to 786)	RR 0.93 (0.77 to 1.11)	226 (1 RCT)	⊕⊕⊕⊖ MODERATE [3]	
Depressive symptoms, medium-term follow-up	-	SMD 0.1 lower (0.3 lower to 0.1 higher)	-	390 (4 RCTs)	⊕⊕⊕⊖ MODERATE [4]	
Work functioning, medium-term follow-up	-	SMD 0.32 lower (0.9 lower to 0.26 higher)	-	48 (1 RCT)	⊕⊕⊖⊖ LOW [3 5]	

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RCT:** Randomised controlled trial; **RR:** Risk ratio; **SMD:** Standardised mean difference.

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]One study with unclear risk and one with serious risk of bias. Rated down one level due to high risk of bias.
[2]Two studies with 130 participants. CI includes harms and benefits. Rated down one level due to imprecision.
[3]Based on one study with small number of participants, rated down one level due to to imprecision.
[4]Includes studies with high risk of bias. Rated down one level due to high risk of bias.
[5]One study with unclear risk of bias. Rated down with one level due to high risk of bias.

Summary of findings 3. Psychological intervention compared to care as usual in depressed people, medium-term follow-up

Psychological intervention compared to care as usual in depressed people

Patient or population: Depressed persons
Setting: Various: workplaces, primary care, insurance institute and academic hospital
Intervention: Psychological intervention
Comparison: Care as usual

Outcomes	Anticipated absolute effects* (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with psychological intervention			
Sickness absence days, medium-term follow-up	SMD 0.15 lower (0.28 lower to 0.03 lower)	1649 (9 RCTs)	⊕⊕⊖⊖ LOW [1 2]	The SMD translates back to -0.3 days per 2 weeks (95% CI -0.5 to -0.1) or -14.7 days in 12 months (95% CI -27.6 to -3.0).
Depressive symptoms, medium-term follow-up	SMD 0.3 lower (0.45 lower to 0.15 lower)	1255 (8 RCTs)	⊕⊕⊖⊖ LOW [2 3]	
Work ability, medium-term follow-up	SMD 0.05 higher (0.46 lower to 0.57 higher)	58 (1 RCT)	⊕⊖⊖⊖ VERY LOW [4 5]	

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RCT:** Randomised controlled trial; **SMD:** Standardised mean difference.

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]In most studies, the outcome was self-reported, leading to risk of bias in outcome assessment. There was also large attrition. Rated down one level due to high risk of bias.
[2]Funnel plot shows missing small studies with no effect or harmful effect. Rated down one level due to risk of publication bias.
[3]Outcomes self-reported in unblinded studies. Rated down one level due to high risk of bias
[4]CI includes appreciable harms and benefits. Sole study. Rated down two levels due to imprecision.
[5]One study with unclear risk of bias. Rated down one level due to high risk of bias.

Summary of findings 4. Improved care compared to care as usual in depressed people, medium-term follow-up

Improved care compared to care as usual in depressed persons

Patient or population: Depressed persons
Setting: Primary Care and community mental health
Intervention: Improved Care
Comparison: Care as usual

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with care as usual	Risk with improved care				
Sickness absence days, medium-term follow-up	-	SMD 0.06 lower (0.15 lower to 0.04 higher)[1]	-	1912 (7 RCTs)	⊕⊕⊕⊖ MODERATE [2]	The SMD translates back to -0.1 days per 2 weeks (95% CI -0.3 to 0.1) or -5.9 days in 12 months (95% CI -14.8 to 3.9). The SMD of the sensitivity analysis[1] translates back to -0.4 days per 2 weeks (95% CI -0.6 to -0.1) or -19.7 days in 12 months (95% CI -34.5 to -4.9).
Off work, medium-term follow up	496 per 1.000	516 per 1.000 (402 to 655)	RR 0.97 (0.77 to 1.22)	362 (1 RCT)	⊕⊕⊖⊖ LOW [3,4]	

Depressive symptoms, medium-term follow-up	-	SMD 0.21 SD lower (0.35 lower to 0.07 lower)	-	1808 (7 RCTs)	⊕⊕⊕⊖ MODERATE [2]
Work functioning, medium-term follow-up	-	SMD 0.5 higher (0.34 higher to 0.66 higher)	-	604 (1 RCT)	⊕⊕⊕⊖ MODERATE [5]

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the __relative effect__ of the intervention (and its 95% CI).

CI: Confidence interval; **RCT:** Randomised controlled trial; **RR:** Risk ratio; **SMD:** Standardised mean difference.

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

1 A sensitivity analysis revealed that two RCTs with a lower risk of bias found a SMD of 0.20 lower (0.35 lower to 0.05 lower); moderate-certainty evidence).
2 Majority of studies at high risk; downgraded with one level due to high risk of bias.
3 One study at high risk of bias, downgraded with one level due to high risk of bias.
4 One study with less than 400 participants, downgraded with one level due to imprecision
5 Study with unblinded outcome assessment, rated down one level due to high risk of bias.

BACKGROUND

Description of the condition

Depression is a major public health problem, with 298 million cases of major depressive disorders at any time point in 2010 (Ferrari 2013). The worldwide point prevalence of depressive disorder was 4.4% in both 2005 and 2010 (Ferrari 2013). Symptoms of depressive disorder include the presence of one or two core symptoms of low mood and loss of interest, coupled with other symptoms such as feelings of inadequacy and hopelessness, sleep disturbance, weight change, fatigue, impaired concentration, agitation or slowing down of movement and thought, and suicidal ideation (APA 2013). Depressive disorders can be classified along a continuum by the levels of symptom severity, number of mental or physical symptoms, and duration. Corresponding diagnostic categories range from persistent depression (dysthymia) and subclinical states (minor depressive disorder) to major depressive disorder (APA 1994; APA 2013).

Besides the serious consequences in terms of individual suffering, depression has a large impact on social functioning and the ability of patients to work (Evans-Lacko 2016; Hirschfeld 2000; Lerner 2008). In a population of US workers, the 12-month prevalence of major depressive disorder was found to be 6% and was associated with 27.2 lost workdays per ill worker per year (Kessler 2006). The economic burden of depressed individuals in the US was US dollars (USD) 210.5 billion in 2010, of which 50% were attributable to workplace costs (Greenberg 2015). The high prevalence of depressive disorders, combined with the impact on work disability, has extensive societal consequences. In 1990, major depressive disorders were the 15th leading contributor to the global burden of disease in terms of Disability Adjusted Life Years (DALYs), which is the sum of years of productive life lost due to premature mortality and the years of productive life lost due to disability. Data from the Global Burden of Disease study showed that depressive disorders were ranked the 11th leading contributor (Murray 2012).

While working is important from a societal point of view, work is also an important aspect of the quality of life of individuals (Bowling 1995). Work provides income, structure, and social interactions. One salient consequence of depression is absenteeism, but depression can also affect the at-work productivity for workers (Lerner 2008). Depressed workers experience specific limitations in their ability to function at work. These limitations include performing mental and interpersonal tasks (Adler 2006; Burton 2004). The quality of work performance can also be affected, as was shown in studies focusing on errors and safety issues (Haslam 2005; Suzuki 2004). Depressed workers may need to make an extra effort to be productive during their work (Dewa 2000), which may lead to spill-over effects of fatigue after work.

Description of the intervention

Work disability of depressed workers can be targeted by interventions. First of all, work-directed interventions aim to ameliorate the consequences of the depressive disorder on the ability to work. These types of interventions either target the work itself, by modifying the job task, or (temporarily) reduce the working hours. Work-directed interventions can also support the worker in dealing with the consequences of their depression at the workplace.

Second, clinical interventions aimed at reducing depression symptoms may improve work ability (Hees 2013b). Current clinical practice guidelines for the treatment of major depressive disorder recommend pharmacotherapy, psychotherapy, or a combination of both (APA 2010; NICE 2010). Pharmacologic treatment for major depressive disorder includes antidepressant medication such as tricyclic antidepressants (TCAs), selective serotonin reuptake inhibitors (SSRIs), monoamine oxidase inhibitors (MAO inhibitors), and selective norepinephrine reuptake inhibitors (SNRIs). With regard to psychotherapy, cognitive behavioural therapy (CBT) and interpersonal therapy in particular are considered effective treatment options (NICE 2010). Exercise has been increasingly used as an alternative to pharmacological or psychotherapeutic interventions (Cooney 2013).

How the intervention might work

Work-directed interventions are deemed to reduce work disability by creating a work environment better suited for a depressed worker, such as modifying work tasks or working hours. Moreover, the worker can be supported in dealing with the depression at work by a gradual return to work program or by enhancing skills to cope with work situations (Lagerveld 2012). Clinical interventions may reduce work disability by reducing depressive symptoms, thereby eliminating the obstacles to working.

Why it is important to do this review

Considering the impact of depressive disorders on the occupational health of many affected workers, it is vital to know what types of interventions are effective in improving occupational health. In the first version of this review, in 2008, we concluded that there was an urgent need to evaluate interventions that address work issues in future research. Since then, several such studies have been published underpinning the need for an update of the review.

OBJECTIVES

The goal of this review was to evaluate the effectiveness of interventions aimed at reducing sickness absence in employees with depressive disorders.

We considered the effectiveness of two types of interventions:

1. work-directed interventions, i.e. addressing the work or the work-worker interface as part of the clinical treatment or as a stand-alone intervention; and
2. clinical interventions, i.e. treatment of depressive disorder without a focus on work.

METHODS

Criteria for considering studies for this review

Types of studies

We included all randomised controlled trials (RCTs), including cluster-RCTs, in this review. We did not use any language restrictions.

Types of participants

Patient characteristics and setting

The population was limited to adult (i.e. more than 17 years old) workers (employees or self-employed). We included participants

from occupational health settings, primary care, or outpatient care settings. We based the selection of the studies on the primary outcome only. We still included studies if less than 50% of the participants were not employed.

Diagnosis

We defined depressive disorder as a main diagnosis fulfilling the criteria of the Diagnostic and Statistical Manual (DSM-IV) (APA 1994; APA 2013), the Research Diagnostic Criteria (RDC) (Spitzer 1979), or the International Classification of Disease (ICD-10) (WHO 1992) for one of the following disorders: dysthymic disorder, minor depressive disorder, or major depressive disorder. We also included studies that defined depressive disorder as a level of depressive symptoms assessed by validated self-report instruments published in peer-reviewed journals. An example is the Beck Depression Inventory (BDI) (Beck 1987); or clinician-rated instruments such as the Hamilton Depression Rating Scale (HDRS) (Hamilton 1967) or the Montgomery–Åsberg Depression Rating Scale (MADRS) (Montgomery 1979).

Exclusion criteria

We excluded studies involving workers with a primary diagnosis of a common mental disorder other than a depressive disorder. We did not exclude workers with a co-morbidity from other common mental disorders (such as anxiety disorders), but we did exclude workers with bipolar disorders or depressive disorders with psychotic features.

Types of interventions

We included all interventions aimed at reducing work disability. Naming and classifying interventions that aim to improve return to work is difficult. Health-care interventions aiming to enhance return to work are mainly based on two mechanisms. One is improving conditions related to work, such as helping workers with depressive symptoms to overcome barriers that prevent them from working such as reducing work hours, changing tasks, light duty, graded work exposure addressing causes of depression at work such as a conflict, or supporting the worker in coping with the consequences of their depression in the workplace. We called these types of interventions 'work-directed interventions' and we did not use any subcategories of these interventions. The other mechanism is through improvement of depressive symptoms as is usual in treatment situations, assuming that the symptoms are the main barrier for not being at work. We called these interventions 'clinical interventions.' For clinical interventions we made distinctions among the following treatment modalities: psychological or psychiatric treatment, antidepressants, a combination of these two, and other interventions such as improved care, exercise and diet.

We compared work-directed interventions, clinical interventions, and a combination of both types against any other intervention, no intervention or care as usual.

Types of outcome measures

In this review, we operationalised reduction in work disability as a reduction in sickness absence and as enhancement in work functioning.

Primary outcomes

The main outcome measure in this review was sickness absence, either measured as sickness absence days during the follow-up period or employment status after a period of time, categorised as being 'off work' or 'at work.' Sickness absence data could be extracted from the employee attendance records, the files of a compensation board, or it could be self-reported.

Secondary outcomes

When available, we included the following secondary outcomes from the included studies.

1. Depression (either dichotomously or continuously measured).
2. Work functioning (Nieuwenhuijsen 2010). Examples of work functioning measures are the Endicott Work Productivity Scale (EWPS) (Endicott 1997), the Sheehan Disability Scale (SDS) (Sheehan 1996) and the Work Role Functioning Questionnaire (WRFQ) (Abma 2012). We only included instruments that separately measured work functioning (instead of work and other activities combined). The outcome 'work ability' (Ilmarinen 2005) was also considered as a work functioning outcome.

We did not include other outcomes such as employee satisfaction, general social functioning (not work-specific), or quality of life scales.

We considered the effects measured with all the above instruments on the following time-scales:

- short-term, up to two months;
- medium-term, over two months to one year; and
- long-term, over one year.

Search methods for identification of studies

Electronic searches

We conducted the original search strategy for the first version of this review in 2006, using no limits on publication date (Appendix 1). We updated the search for the 2014 update and used this search strategy again for the 2020 update (Appendix 2). For this update, we searched the following electronic databases: CENTRAL (The Cochrane Library), MEDLINE, PsycINFO, Embase, and CINAHL up to the 4 April 2020. We used three types of terms: depression-related words combined with work-related words and database-specific methodological filter terms. We adapted the search terms for PsycINFO, Embase, and CINAHL from the MEDLINE search to fit the specific requirements of those databases. For CENTRAL, we replaced the methodological filter by a filter to identify trials.

We based the selected work-related search terms on previous studies. Work* and occupation* are sensitive single terms used to locate occupational health studies, as advocated by Verbeek (Verbeek 2005). Furthermore, we selected database-specific terms relevant to our objective from a study testing which work-related search terms are best suited for literature searching on chronic disease (rheumatoid arthritis, diabetes mellitus, hearing problems, and depression) and work (Haafkens 2006).

Searching other resources

We checked the reference lists of all articles that we retrieved as full papers and of all retrieved systematic and narrative reviews in order to identify further potentially eligible studies.

Data collection and analysis

Selection of studies

Pairs of review authors decided if a study did not fulfil the criteria for selection, and we excluded the study at that point. We excluded studies in this phase only if the study did not include participants with depressive disorders or it was not a controlled intervention study. When it was not clear whether sickness absence was measured, we retrieved the full article before deciding upon exclusion. We then examined the full text publications of the remaining studies in order to decide which studies fulfilled all inclusion criteria. We documented the reasons for exclusion at that stage. The two review authors discussed any disagreement about the inclusion of studies until they reached consensus. If they could not resolve their difference of opinion, they consulted a third review author (JV). All articles published in languages other than English were translated or assessed for inclusion by a native speaker.

Data extraction and management

We constructed a data extraction form that enabled the review authors to extract the data from the included studies. For each study, one review author filled out the forms; this form was checked by a second review author (AN, AV, BF, CF, HH, KN, and UB participated in data extraction). Review authors solved differences of opinion by discussion. When only a proportion of the study population was workers, we extracted the data for that subgroup from the article. When these data were not reported, we asked the original study authors to provide the data for this subgroup. We used the same procedure for studies where only a proportion of the study population was depressed.

Assessment of risk of bias in included studies

Pairs of review authors independently assessed the risk of bias of the included studies. We used the following items to assess risk of bias in the included studies: random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), and selective reporting (reporting bias). We evaluated risk associated with incomplete outcome data or with blinding of outcome assessments separately for sickness absence, depressive symptoms and where applicable also for work functioning. As the latter outcome was not often used, we did not report the risk of bias for work functioning separately in the risk of bias tables. We assessed the risk of bias in RCTs and cluster-RCTs by using Cochrane's 'Risk of bias' tool (Higgins 2011).

With regard to the risk of attrition bias, we calculated the percentage lost to follow up taking the number randomised as the starting point and the number analyzed at the latest follow-up measurement as the endpoint. We assigned a high risk of attrition bias to studies with a percentage of participants lost to follow up of more than 20%, and a low risk for studies with less than 10% lost to follow. The risk of attrition bias for studies with 10% to 20% lost to follow up depended on whether the analyses of results accounted for attrition sufficiently.

We rated each potential source of bias as 'high risk' of bias, 'low risk' of bias, or 'unclear risk' of bias in the 'Risk of bias' table. Next, we constructed a 'Risk of bias' summary figure together with an overview 'Risk of bias' graph as described in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). Where information on risk of bias related to unpublished data or correspondence with a researcher, we noted this in the 'Risk of bias' table.

Measures of treatment effect

We plotted the results of each trial as means and standard deviations (SD) for continuous outcomes. For each timescale (short-term, medium-term, and long-term), we selected the last available observation within this period for the meta-analysis. For the primary outcome measure, that is sickness absence days, we transformed the number of days or hours worked during the follow up into days of sickness absence. To do so, we extracted the hours or days worked from the mean number of hours a full-time employee would work in that specific country. When transforming the data from days worked to days not worked, the SDs did not need to be transformed. When transforming the data from hours to days, we divided both the means and SDs by eight. Studies used different time spans during which they measured the number of sickness absence days. Therefore, for sickness absence days we used the standardized mean difference (SMD) with a 95% confidence interval (CI) between the intervention and control groups as the summary effect measure. In order to aid interpretability of these SMDs, we also present the translation of the SMD in terms of the two most commonly used outcome measures; sickness absence days over a two-week period and over a one-year period. To this end, we multiplied the SMD by the median of the SDs of the intervention groups using these outcomes.

For the secondary outcome measures, we also used SMDs because it is likely that these outcomes were measured with different instruments. We chose to treat ordinal variables using a scale of more than five categories as continuous variables (it should be noted that this choice was based on arbitrary criteria). We dichotomised scales with fewer than five categories. For dichotomous data, we calculated the risk ratios (RRs) and 95% CIs.

For depression data, where studies presented both dichotomous and continuous data, we preferred the continuous outcome measures since the majority of the studies presented these.

Unit of analysis issues

For studies that employed a cluster-randomised design and did not consider the design effect in the analyses, we planned to calculate the design effect by following the methods presented in Donner 2002 based on a fairly large assumed intra-cluster correlation of 0.10. However, the cluster-RCTs included in the review reported negligible intra-cluster correlations. Therefore, we did not adjust the measures of effect presented by the authors.

Dealing with missing data

If the SDs (continuous data) or numbers of outcomes for each group (dichotomous data) were not presented in the publication, we contacted the authors with a request to provide these data. Whenever authors were unable or unwilling to provide this information, we calculated SDs from P values and CIs following the instructions of the *Handbook* (Higgins 2011).

We sought additional information regarding study details, statistical data, or both, from the authors of 20 studies. We received information from 15 authors. Ten of the authors provided statistical data that had not been published in their articles, which enabled us to include nine of these studies in the meta-analyses. In the case of two studies the correspondence led to the exclusion of the study because essential information on the primary outcome measure could not be provided (Simon 2000; Stant 2009). Whenever essential information concerning the risk of bias could not be obtained within four weeks of contacting the authors, we listed the corresponding details as 'unclear.'

Assessment of heterogeneity

For clinical heterogeneity, we had the following considerations for similarity or heterogeneity between studies.

- We considered interventions to have a similar mechanism and effect in all types of participants.
- We considered the effects and mechanisms for all work-directed interventions as similar.
- The three subcategories of clinical interventions, anti-depressants, psychological interventions or exercise were considered as having different effects and mechanisms.
- All various sickness absence outcomes and all various depression outcomes were considered similar.
- Follow-up times of up to two month (short-term), from over two months to one year (medium-term) and over one year (long-term) were considered different.

We assessed statistical heterogeneity in the meta-analyses with the I^2 statistic. If we observed considerable heterogeneity (I^2 > 75%), we refrained from statistical pooling of the studies within that comparison. Substantial inconsistency (I^2 statistic) also led to downgrading of the certainty of the evidence (see Data synthesis for details).

Assessment of reporting biases

We produced funnel plots for visual inspection of possible publication bias.

Data synthesis

For each predefined comparison, we analyzed data for each outcome measure separately. Whenever interventions belonged to the same category in the comparison but two review authors (KN and JV, or KN and BF) judged them dissimilar, we defined subcategories for these types of intervention. We conducted meta-analysis if two review authors (KN and BF) judged a group of trials sufficiently homogeneous in terms of participants, interventions, and outcomes to provide a meaningful summary. In such cases, we calculated pooled SMDs for the predefined outcome measures using the Review Manager 5 software (RevMan 2014) with a random-effects model. We chose a random-effects model as we expected statistical heterogeneity to occur as a result of the clinical and methodological heterogeneity in research on sickness absence. For three-armed trials contributing evidence to two different comparisons, we divided the number of participants of the arm used in both comparisons by two.

Subgroup analysis and investigation of heterogeneity

We considered that there could be difference in the way psychological treatment was administered and we compared studies with a personal therapist or in-person therapy with web-based or telephone-based studies without personal guidance of a therapist. In addition, we planned to analyze if studies with mostly women (> 80%) had different effects from studies with mostly men (> 80%). However, we did not include a sufficient number of studies with such an uneven gender-distribution to allow for this analysis.

Sensitivity analysis

We planned to conduct sensitivity analyses by excluding:

1. studies with a high risk of bias (defined as at least three of the 'Risk of bias' items were judged to present a low risk of bias: random sequence generation; allocation concealment; blinding of participants/personnel; blinding of outcome assessment;
2. studies with skewed data; and
3. studies in which workers were a small subgroup of the study population.

However, the small numbers of studies in each comparison only allowed for the sensitivity analyses of risk of bias and then only in the comparisons with the highest number of studies.

Summary of findings and assessment of the certainty of the evidence

We used the GRADE approach to assess the certainty of a body of evidence regarding the primary outcome category of the comparisons addressed in the review. At the start of the GRADE assessment process, we assumed high certainty for all studies and we downgraded the certainty of the evidence for each comparison by one to three levels depending on the seriousness of the violations in each domain.

To assess the risk of bias for a comparison, we considered the 'Risk of bias' tables for each study in that comparison. We saw items related to selection bias, detection bias, and attrition bias as prerequisites for high certainty. We only considered studies with low risks on these items to have a low risk of bias. For each comparison we considered the risk of bias serious (-1) if a majority of the evidence in the studies included in the meta-analysis (in terms of weights) were of low quality. We applied a -2 downgrade in cases where the majority of the studies did not have adequate random sequence generation and allocation concealment. For consistency, we considered an I^2 value of 50% to 75% to indicate substantial inconsistency, which lead us to downgrade (-1). If the I^2 value exceeded 75%, we refrained from pooling the results and we analyzed the results for each study separately. Indirectness of the evidence was not an issue in our review as all comparisons in the included studies directly addressed the comparison. For imprecision of results, we judged serious imprecision leading to downgrading (-1) if a comparison either included a number of fewer than 400 participants or a wide CI around the effect estimate. For a non-significant effect, we considered a CI to be wide if it included an SMD of both 0 and a moderate effect size (SMD > 0.5 or < -0.5). For I a significant effect, we considered a CI to be wide if it included both a small and large effect size (SMD small = -0.2 or 0.2; SMD large = 0.8 or -0.8). If in addition to a wide CI, the comparison included one study only, we downgraded with two levels (-2).

The resulting interpretation of the certainty of the level of evidence per comparison was as follows.

- **High certainty:** We are very confident that the true effect lies close to that of the estimate of the effect
- **Moderate certainty:** We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
- **Low certainty:** Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
- **Very low certainty:** We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

We created 'Summary of Findings' (SOF) tables with GRADEpro software (GRADEpro 2008) for the main comparisons.

RESULTS

Description of studies

Results of the search

Figure 1 displays a PRISMA study flow chart of the inclusion process up to 2014. Figure 2 displays the flow chart of the 2020 update. The electronic searches between 2014 and 2020 resulted in 5951 new hits. We assessed the titles and abstracts for eligibility. This resulted in the full text assessment of 104 publications. We excluded 82 studies after further scrutiny (see Characteristics of excluded studies). This resulted in the inclusion of 22 new studies additional to the 23 studies already in the review. In addition, we identified five ongoing studies in the first and none in the 2020 update (see Characteristics of ongoing studies).

Figure 1. PRISMA Study flow diagram of the study selection process until 2014.

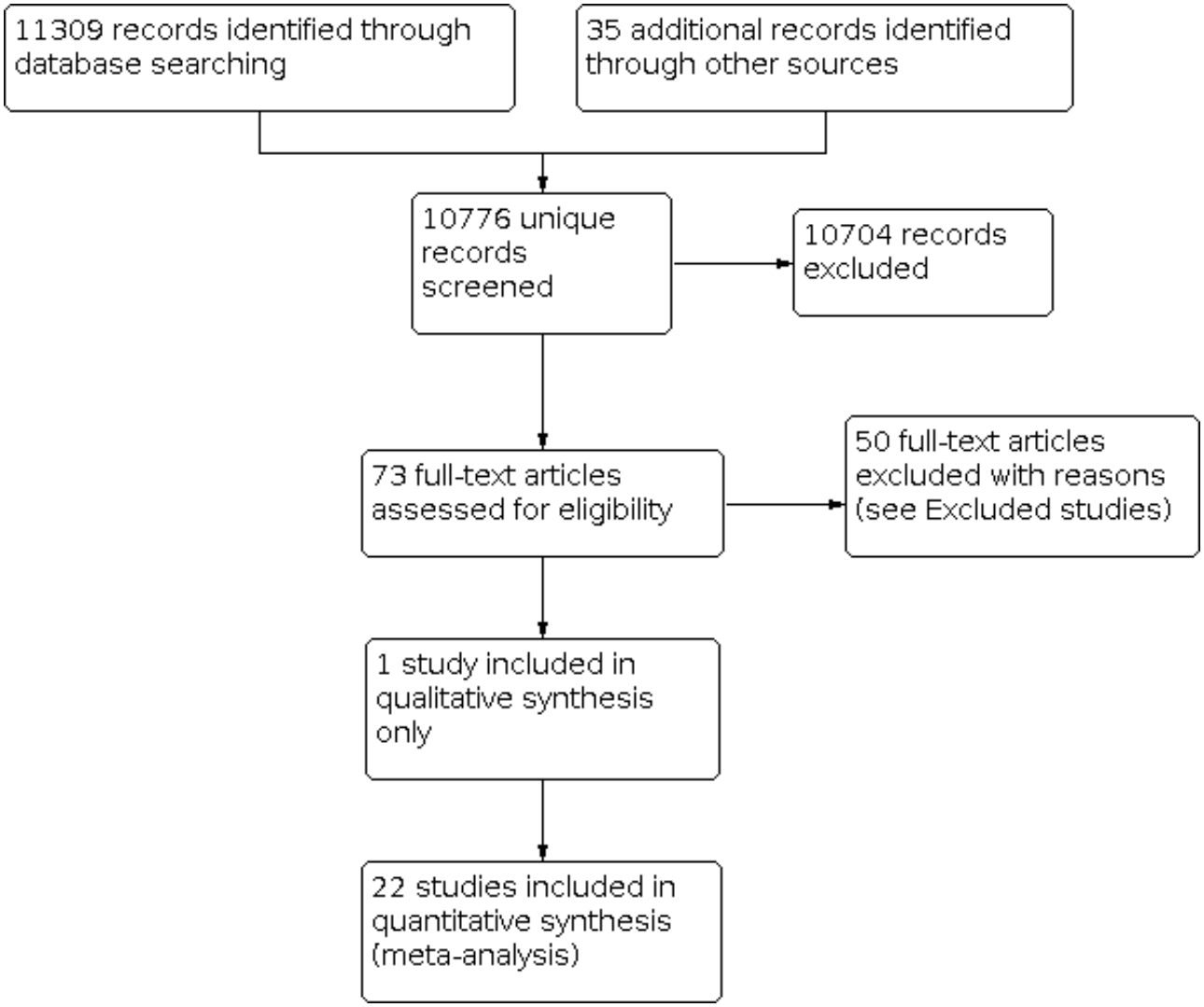
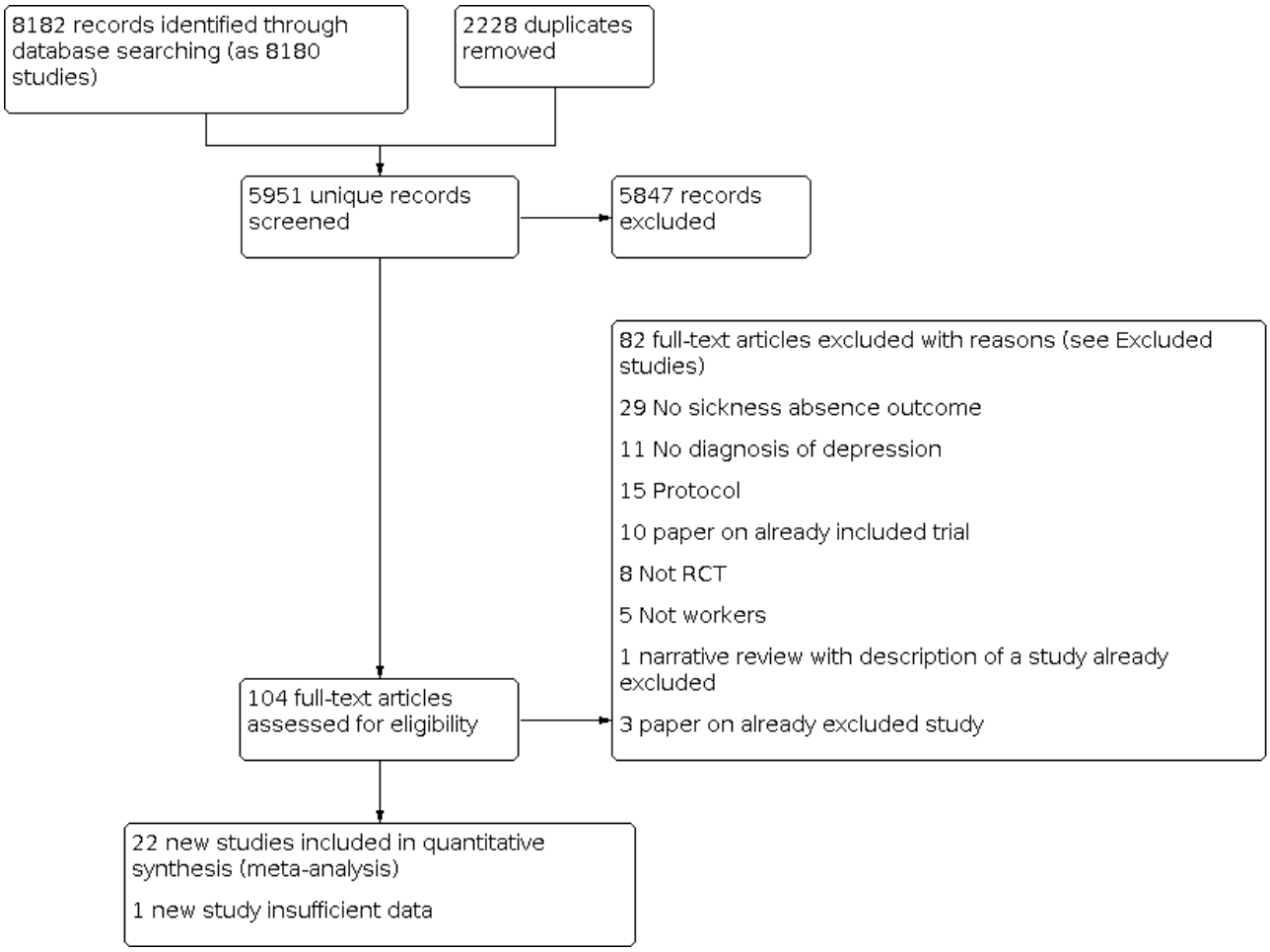


Figure 2. PRISMA Study flow diagram of the study selection process 2014-2020.



Included studies

We included 45 studies in the review (see Characteristics of included studies). Four of these studies included three study arms (Kaldo 2018; Kendrick 2005; Knekt 2013; Krogh 2009) and one had four study arms (Finnes 2017). In our analyses we combined two interventions groups of the Kendrick 2005 and two of the Knekt 2013 study. Therefore, in the end we analyzed a total of 93 study arms in this review.

Designs

Of the included studies, 40 were RCTs and five were cluster-RCTs (Noordik 2013; Rost 2004; Schoenbaum 2001; Björkelund 2018; Volker 2015). Intra-class correlations for four of these studies were reported to be negligible and therefore we did not adjust the data. However the Björkelund 2018 study did not report the ICC and we therefore adjusted undertook an unplanned sensitivity analysis (see effects of interventions).

Sample sizes

The total number of participants in the included studies was 13,669. The number of participants included in the analysis was lower (12,109) as we reported on the subgroup of 'employed and depressed participants only' in cases where studies included other subgroups as well. The number of participants in the smallest study (sub)group was lower than 50 in 20 study arms, between 50 and 99 in 22 study arms, between 100 and 199 in 23 study arms, and 200 or more in 18 study arms.

Time period, setting and participants

Three studies were published before 2000, 15 between 2000 and 2010, and 27 after 2010. Eight studies were conducted in the US, 34 were conducted in Europe, one in Canada and two in Australia. Participants were recruited in primary care settings (13 studies), outpatient settings (15 studies), workplace settings (five studies), occupational health care (five studies), through health insurance companies (two), a managed care setting (one study), an unemployment centre (one study), a community mental health centre (one study), a hospital (one study), and through an academic institution (one study). In 32 studies, all participants had a depressive disorder. In 13 studies (Bee 2010; Finnes 2017; Hellstrom 2017; Kendrick 2005; Knapstad 2020; Knekt 2013; McCrone 2004; Meuldijk 2015; Noordik 2013; Reme 2015; Reme 2019; Volker 2015; Wormgoor 2020) depressed patients constituted a subgroup of the study participants.

Interventions

Work-directed interventions, or work-directed interventions combined with a clinical intervention

We identified 17 work-directed interventions in 14 studies. Thirteen interventions, reported in 11 studies, were a combination of a work-directed and a clinical intervention (Finnes 2017; Geraedts 2014; Hees 2013; Kaldo 2018; Lerner 2012; Lerner 2015; Lerner 2020; Reme 2015; Schene 2006; Vlasveld 2013; Volker 2015). Four interventions were work-directed only (Finnes 2017; Hellstrom 2017; Noordik 2013; Reme 2019).

All four work-directed interventions included multiple meetings with intervention providers, three specified meetings in the Finnes 2017 study and multiple meetings in the other three. In Noordik 2013, the number depended on the time it took to return to work and the Hellstrom 2017 and Reme 2019 studies provided unlimited support, depending on the individual need of the participants.

The work-directed interventions in Finnes 2017, Hellstrom 2017 and Noordik 2013 all included contact of the intervention provider with the supervisor of the worker. In the Finnes 2017 study, however, this was most structured as it included a stepwise method with separate worker and supervisor interviews and one convergence meeting with both.

The Noordik 2013 study compared an exposure-based return to work intervention (RTW-E) conducted by occupational physicians (OPs), gradually exposing the participants to more demanding work situations, with regular support by the OP. The RTW-E program provided workers with several homework assignments aimed at preparing, executing, and evaluating an exposure-based RTW plan. In the Finnes 2017 study, the work-directed Intervention aimed to facilitate dialogue between the participant and the workplace through a series of steps involving the participant and the nearest supervisor, resulting in a return-to-work plan. Providers were either clinical or behavioural psychologists, or psychiatric nurses. In the Hellstrom 2017 study, the intervention followed the Individual Placement and Support (IPS) model. Workers could be out of a job for a longer time (up to three years) and return to work in those workers included return to a new job. The intervention included career counselling and contact with employers to help participants obtain jobs and keep them. Providers were mentors (nurses, social workers or occupational therapists) and career counsellors. In the Reme 2019 study, the intervention also followed the IPS model and included personalised benefits counselling, rapid job search (starting within one month), systematic job development, and time unlimited and individualized support. Providers were employment specialists.

Three of the work-directed interventions compared that intervention with other work-directed interventions; in the Hellstrom 2017 study, this included services as offered by the job centres in Denmark, for instance, courses, company internship programmes, wage subsidy jobs, skill development and guidance, mentor support or gradual return to employment. The work-directed 'care as usual' by OPs in the Noordik 2013 study was based on a national guideline. Care as usual included both work modification and support. In the Reme 2019 study, the work-directed 'care as usual' involved a referral to either work with assistance by a personal facilitator, and included finding suitable work, negotiating wage and employment conditions, modified duties, and follow-up at the work place or to a traineeship in a sheltered business. The work-directed intervention in the Finnes 2017 study was compared to care as usual from a medical doctor.

Of the 13 interventions that combined a work-directed intervention with a clinical intervention, the main mode of intervention delivery was face-to-face meetings in five studies with seven interventions (Finnes 2017; Hees 2013; Reme 2015; Schene 2006; Vlasveld 2013); online in three studies (Geraedts 2014; Kaldo 2018; Volker 2015); and by telephone in three studies (Lerner 2012; Lerner 2015; Lerner 2020). The number of meetings in interventions that included face-to-face or telephone meetings ranged from four (Lerner 2012); six to 12 (Vlasveld 2013); eight (Lerner 2015; Lerner 2020); 37 (Hees 2013) to 44 meetings in the Schene 2006 study.

The combined work-directed and clinical interventions most often based the clinical interventions on cognitive behavioural therapy (CBT) principles (Kaldo 2018; Lerner 2012; Lerner 2015; Lerner 2020; Reme 2015), on the principles of problem-solving therapy (PST) (Vlasveld 2013) or a combination of CBT and PST (Geraedts 2014; Volker 2015). The clinical part of the Hees 2013 and Schene 2006 studies included psychiatric clinical management according to American Psychiatric Association guidance, which also included antidepressants. The Vlasveld 2013 study also included antidepressant treatment with PST. Finnes 2017 was the only study that used acceptance and commitment therapy (ACT) as the basis for their clinical intervention. The work-directed interventions ranged from an elaborated and highly structured program provided by occupational therapists in the Hees 2013 and Schene 2006 studies, through facilitating dialogue between the participant and the workplace in three meetings in the Finnes 2017 study, and work-directed care modelled after Individual Placement and Support (IPS) in the Reme 2015 study.

Providers of the clinical part of the interventions were psychiatric residents in the Hees 2013 and Schene 2006 studies, clinical psychologists in the Finnes 2017 and Reme 2015 studies, counsellors in the Lerner 2012 and Lerner 2015 studies, occupational physicians in the Vlasveld 2013 study, The web-based clinical interventions included support from a psychology student in the Geraedts 2014 study, support from an occupational physician in the Volker 2015 study, and support from clinical psychologists or supervised psychology students in the Kaldo 2018 study.

The work-directed part of the intervention was delivered by the same providers in the Geraedts 2014, Kaldo 2018, Lerner 2012, Lerner 2015, Vlasveld 2013, and Volker 2015 studies. The work-directed part of the intervention was delivered by another provider in the Finnes 2017 study (clinical psychologist, behavioural therapist, or psychiatric nurse), the Hees 2013 and Schene 2006 studies (occupational therapist), the Lerner 2020 study (doctoral-level psychologist) and the Reme 2015 study (employment specialist).

Two studies combining a work-directed intervention with a clinical intervention employed multiple comparisons. In the Finnes 2017 study, comparisons were 1) the work-directed component only, 2) the clinical component only, and 3) care as usual provided by access to a medical doctor, psychologist, social worker, physical therapist, or nurse). Kaldo 2018 used both exercise (aerobic) and care as usual (primary care standard treatment for depression determined by the patient's general practitioner) as comparisons.

The Hees 2013 and Schene 2006 studies compared the work-directed intervention combined with a clinical intervention with the clinical intervention alone, whereas the Reme 2015, Vlasveld 2013, and Volker 2015 studies used the work-directed intervention alone as the comparison.

Three studies compared the work-directed intervention combined with a clinical intervention to care as usual which could include various providers, but these were not specified (Lerner 2012; Lerner 2015; Geraedts 2014) or included a team of various providers (Lerner 2020; psychologists, nurses and social workers).

Clinical interventions

We included 31 studies reporting the effects of clinical interventions for depressed workers.

Psychological interventions

Twelve studies assessed the effect of a psychological intervention (Bee 2010; Beiwinkel 2017; Birney 2016; Eriksson 2017; Finnes 2017; Hollinghurst 2010; Kendrick 2005; Knekt 2013; Mackenzie 2014; McCrone 2004; Phillips 2014; Wormgoor 2020). Four of these studies looked at an intervention that was delivered face-to-face (Finnes 2017; Kendrick 2005; Knekt 2013; Wormgoor 2020). One intervention was delivered by telephone only (Bee 2010); one offered telephone guidance alongside an online intervention (Eriksson 2017); and three studied an online intervention and provided guidance through text messages with a provider (Beiwinkel 2017; Hollinghurst 2010; Mackenzie 2014). A further three were online programmes delivered without guidance (Birney 2016; McCrone 2004; Phillips 2014). The intensity of these interventions varied from five or six sessions (Finnes 2017; Mackenzie 2014; Phillips 2014), eight (Kendrick 2005; McCrone 2004), ten (Hollinghurst 2010) and 12 sessions (Bee 2010; Beiwinkel 2017) to 20 sessions or more (Knekt 2013; Wormgoor 2020). In two studies, the number of sessions was not specified (Birney 2016; Eriksson 2017).

Eight of the 12 interventions were based on the principles of CBT. One was based on ACT (Finnes 2017); one was based on PST (Kendrick 2005); one was based oThank n psychodynamic therapy (Knekt 2013), and one focused on normalisation and coping (Wormgoor 2020).

Intervention providers were clinical psychologists in the ACT arm of the Finnes 2017 study. Beiwinkel 2017 and Eriksson 2017 used both psychologists and psychotherapists to provide guidance alongside the online CBT, while the psychotherapy intervention in Knekt 2013 study, and the coping-focussed therapy in Wormgoor 2020 were delivered by psychotherapists alone. The telephone CBT was provided by mental health workers in Bee 2010 and mental health specialists provided the email guidance alongside the online CBT in the Mackenzie 2014 study. In Kendrick 2005, the intervention was delivered by community mental health nurses.

Six studies (Bee 2010; Eriksson 2017; Finnes 2017; Hollinghurst 2010; Kendrick 2005; McCrone 2004) compared their intervention with care as usual in general practice.

Three studies compared their online interventions to directing workers with text based information on depression (Beiwinkel 2017; Birney 2016; Phillips 2014). The Knekt 2013 study compared psychodynamic psychotherapy with PST and Wormgoor 2020 compared their coping focused therapy to brief psychotherapy. The online CBT in the Mackenzie 2014 study was compared with a waiting list condition.

Psychological interventions plus antidepressant medication

Two studies included interventions with a combination of psychological interventions and antidepressant medication. One study (Burnand 2002) compared the effect of psychodynamic therapy combined with TCA medication with TCA medication alone. The intervention included face-to-face individual sessions by a nurse combined with clomipramine for a duration of 10 weeks. The frequency of the psychotherapy sessions was not fixed. This was compared with a group receiving the same medication and who received supportive care (an individual session with empathic listening, guidance, and support). One study (Sarfati 2016) compared a combination of SSRI medication and a telephone-administered CBT programme with medication and adherence enhancing phone calls. The CBT programme included eight 30-minutes sessions provided by PhD- or Master's degree-level experienced therapists. In the control condition, a research coordinator provided a 10-minute structured telephone call weekly for eight weeks, with enquiry about progress and reminders to take medication properly

Antidepressant medication

Six studies examined the effectiveness of antidepressant medication, of which one compared the antidepressant medication with a placebo condition (Agosti 1991) and the other five with another antidepressant medication (Fantino 2007; Fernandez 2005; Miller 1998; Romeo 2004; Wade 2008). Three studies compared a SSRI with SNRI medication (Fernandez 2005; Romeo 2004; Wade 2008), one study compared a SSRI with TCA (Miller 1998), one study compared two different SSRIs (Fantino 2007), and one study compared TCA or MAO inhibitors with placebo (Agosti 1991).

Improved care

Eight studies (Björkelund 2018; Knapstad 2020; Meuldijk 2015; Rost 2004; Schoenbaum 2001; Simon 1998; Wang 2007; Wikberg 2017) looked at the effects of improving care management for depressed workers rather than evaluating one specific clinical intervention.

Five studies (Björkelund 2018; Rost 2004; Schoenbaum 2001; Simon 1998; Wikberg 2017) compared enhanced primary care with primary care as usual. In these types of interventions general practitioners were enrolled in a quality improvement program and were expected to provide enhanced care including antidepressant medication and psychological interventions, according to primary care guidelines. In the Björkelund 2018 study, this included a nurse acting as a care manager to assist the general practitioner in providing care. The care manager would have one face-to-face meeting and five to seven follow-up meetings by telephone. In the Wikberg 2017 study, general practitioners were taught how to use the MADRS-S to monitor changes in depressive symptoms. Workers made four visits to the general practitioner before which the worker completed the MADRS-S.

One study (Wang 2007) compared a structured telephone outreach and care management program with usual managed care. Workers were enrolled after screening offered to various work organisations that took part in a managed behavioural health care program. The telephone outreach systematically assessed needs for treatment, facilitated entry into in-person treatment (both

psychotherapy and antidepressant medication), monitored and supported treatment adherence, and (for those declining in-person treatment) provided a structured psychotherapy intervention by telephone. Intervention participants declining in-person treatment and experiencing significant depressive symptoms after two months were offered a structured eight-session cognitive behavioural psychotherapy program.

In the Meuldijk 2015 study, concise and protocolised psychotherapy and pharmacotherapy care were provided within seven weeks and compared with psychotherapy and pharmacotherapy that was provided without limitations to the number of sessions. In the Knapstad 2020 study, Prompt Mental HealthCare (PMHC) was provided as part of primary care. This meant that clients could directly contact PMHC and have access to mental health treatment (within 48 hrs).

Exercise

Three exercise interventions were included; strength training (Krogh 2009), aerobic training (Krogh 2009; Krogh 2012) or a program including either light, moderate or vigorous exercise (Kaldo 2018).

The exercise interventions were compared with relaxation (Krogh 2009; Krogh 2012) or standard treatment for depression determined by the patient's general practitioner (Kaldo 2018).

Art

In the Blomdahl 2018 study, a protocolised art therapy was compared with care as usual, which could include acupuncture, cognitive–behavioural therapy, electroconvulsive therapy, interpersonal therapy, occupational therapy, pharmacological therapy, physiotherapy, psychodynamic therapy,and supportive therapy.

Diet

The Chatterton 2018 study assessed the effect of a dietary intervention comprising of personalised dietary advice and nutritional counselling support, including motivational interviewing, goal setting and mindful eating, from a clinical dietician in order to support optimal adherence to the recommended diet. The dietary intervention was compared with a social support control group in which trained personnel befriended participants by discussing neutral topics of interest to the participant, such as sport, news or music.

Outcomes

Studies were only selected if they reported on sickness absence. Of the 45 included studies, seven studies (Agosti 1991; Bee 2010; Krogh 2012; Miller 1998; Schene 2006; Wang 2007; Lerner 2020) reported days or hours worked instead of days of sickness absence. These measures were transformed into sickness absence days as described in the 'Methods' section (see Measures of treatment effect).

We were able to collect data on depression for all but five of the included studies (Agosti 1991; Kaldo 2018; Mackenzie 2014; Meuldijk 2015; Volker 2015). Of all studies reporting on depression, one study (Schoenbaum 2001) presented only dichotomous depression data while all others presented continuous data.

Nine studies (Agosti 1991; Burnand 2002; Hees 2013; Lerner 2012; Lerner 2020; Miller 1998; Rost 2004; Wade 2008; Wang 2007) reported on work functioning using a (sub)scale that separately measured work instead of work and other activities combined. The SD's around the mean scores for work functioning could not be retrieved in the Rost 2004 study, therefore this outcome was not included in the meta-analysis.

Two studies reported on work ability (Finnes 2017; Kaldo 2018) another study reported on both work functioning and work ability (Sarfati 2016).

Ten studies (Geraedts 2014; Hellstrom 2017; Kaldo 2018; Knapstad 2020; Krogh 2012; Reme 2015; Reme 2019; Schoenbaum 2001; Wade 2008; Wormgoor 2020) reported on 'not working' or 'working' or sickness absence status at the end of follow up. We recalculated all outcomes so that they represent the proportion of workers off work at follow-up. The Schoenbaum 2001 study presented only % of those employed at baseline, and these baseline numbers for the total group. The actual numbers of participants at work was calculated bases on the assumption of an equal distribution of baseline employment between study arms. The Blomdahl 2018 reported on a % of workers with sickness absence during follow-up.

Follow up

(a) Short-term

Four of the included studies had the last outcome measurement within one month Agosti 1991; Birney 2016; Fantino 2007; Fernandez 2005).

(b) Medium-term

In 32 studies, the last follow-up measurement was between one month and a year after inclusion. Five studies had the last follow-up measurement later than one year but provided data on earlier time points as well (Hees 2013; Hellstrom 2017; Knekt 2013; Rost 2004; Schene 2006). We included these outcomes in the medium-term analysis. We used the last available observation within the first year for this purpose.

(c) Long-term

In nine studies, the last follow-up measurement was later than one year after inclusion, of which three reported on a follow-up period of 18 months (Hees 2013; Reme 2015; Reme 2019), four on 24 months (Hellstrom 2017; Rost 2004; Schoenbaum 2001; Wormgoor 2020), one on 42 months (Schene 2006), and one on five years (Knekt 2013). However, only depression data and not sickness absence days were reported at two years in the Schoenbaum 2001 study. We therefore refrained from using the depression data at this time point, leaving six studies with long-term outcome data.

Excluded studies

We excluded a total of 82 studies from the review. Reasons for excluding studies were:

- sickness absence not measured as an outcome (Aasvik 2017; Ahola 2012; Amore 2001; Barbui 2009; Bejerholm 2017; Boyer 1998; Brandes 2011; Carlin 2010; Castillo-Pérez 2010; Dalgaard 2014; Danielsson 2019; Dean 2017; Dunlop 2011; Endicott 2014; Erkkilä 2011; Finley 2003; Fournier 2015; Han 2015; Hirani 2010; Hobart 2019; Hollon 2016; Johansson 2019; Kennedy 2016; Kennedy 2019; Knekt 2016; Kojima 2010; Kroenke 2001; Kuhs

1996; Lam 2012; Löbner 2018; Martinez 2011; Meyer 2009; Mundt 2001; Oakes 2012; Salminen 2008; Saloheimo 2016; Sandahl 2011; Shawyer 2016; Simon 2000; Sir 2005; Soares 2019; Stant 2009; Zwerenz 2017);

- participants had a mild depressive disorder or were not diagnosed with a depressive disorder (Aasdahl 2017; Aasdahl 2018; Aelfers 2013; Arends 2014; Bakker 2007; Becker 1998; Bejerholm 2015; Blonk 2007; Brouwers 2007; Dalgaard 2017; Dalgaard 2017a; Ebert 2014; Furukawa 2012; Hackett 1987; Jansson 2015; Lagerveld 2012; Lexis 2011; Mino 2006; Morgan 2011; Reavley 2018; Salomonsson 2017; Warmerdam 2007; Zeeuw 2010; Zwerenz 2017a);
- not a RCT design (Bech 2000; Eklund 2012; Evans 2016; Knekt 2011; Schmitt 2008; Wisenthal 2018; Zambori 2002);
- no worker population (Alexopoulos 2011; Folke 2012; Forman 2012; Gunnarson 2018; Twamley 2019);
- study took place in an inpatient care setting (Dick 1985; Hordern 1964);
- not able to define a subgroup of depressed patients (Beurden 2013; Gournay 1995);
- a double publication (deVries 2015; Maljanen 2016; Schoenbaum 2002; Wells 2000; Winter 2015);
- publication was a study protocol (Dean 2014; Ebert 2014a; Eisendrath 2014; Kooistra 2014; Zwerenz 2015);
- study was prematurely terminated because of massive reorganizations and reimbursement changes in mental health care in the Netherlands during the study period (Heer 2013).

Studies awaiting assessment

There are four ongoing studies that have not reported yet: Deady 2018; Imamura 2018; Kouvonen 2019; Poulsen 2017

Risk of bias in included studies

In Figure 3 and Figure 4 an overview of the risk of bias per study is presented. For details see the 'Risk of bias' tables that form part of the Characteristics of included studies.

Figure 3. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

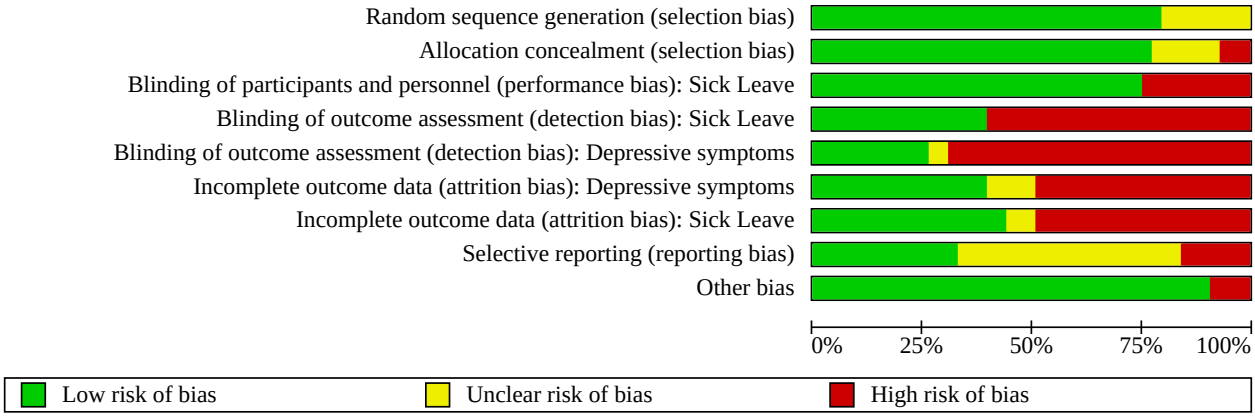


Figure 4. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

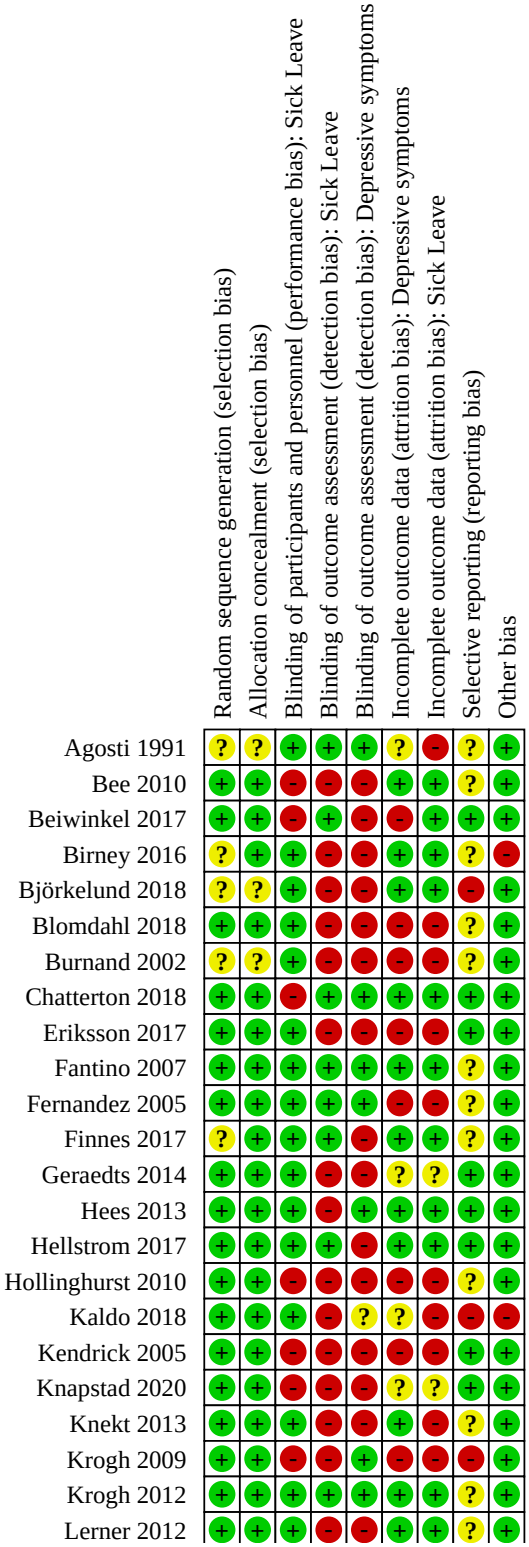
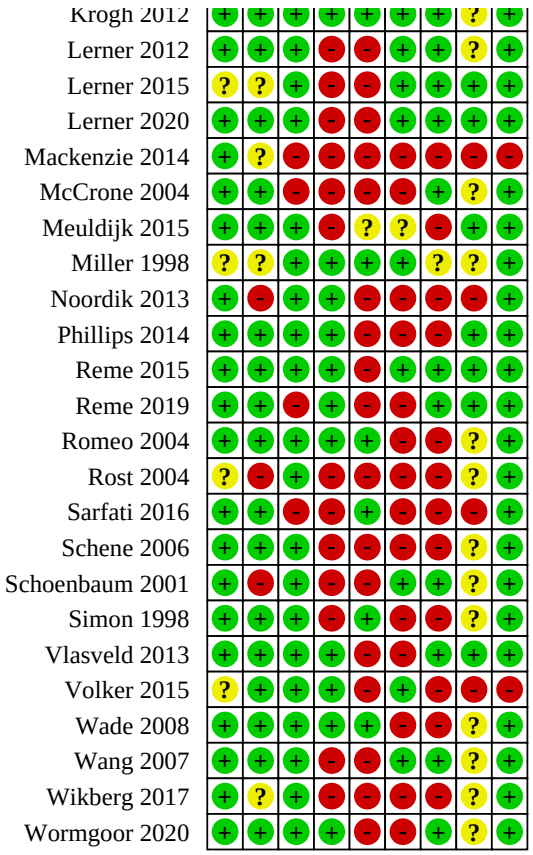


Figure 4. (Continued)



Allocation

The method for generating random numbers posed a low risk of bias in 36 studies and was unclear in nine.

In three cluster-RCTs (Noordik 2013; Rost 2004; Schoenbaum 2001) allocation concealment was not adequate, which was probably indicative of the non-feasibility of allocation concealment in this type of design. In seven further studies (Agosti 1991; Björkelund 2018; Burnand 2002; Lerner 2015; Mackenzie 2014; Miller 1998; Wikberg 2017) information on allocation concealment could not be retrieved, leading to a judgment of unclear risk of bias. In 35 studies, the allocation concealment was adequate and therefore posed a low risk of bias.

Blinding

Risk of performance bias was low in studies using a double-blind design (blinding of participant and care provider). This design was feasible in studies comparing the occupational health effects of antidepressant medications. This type of study has a low risk of performance bias (Agosti 1991; Fantino 2007; Fernandez 2005; Miller 1998; Romeo 2004; Wade 2008). In other clinical interventions, such as psychological or exercise interventions and in work-directed interventions, blinding of the participant or care provider was not feasible. However, we considered the risk of performance bias high only in those studies where the control intervention could be considered less desirable by participants or care provider (Bee 2010; Beiwinkel 2017;Chatterton 2018;Hollinghurst 2010; Kendrick 2005; Knapstad 2020; Krogh 2009; Mackenzie 2014; McCrone 2004; Reme 2019; Sarfati 2016).

Our primary outcome measure (sickness absence days) could be measured either by self-report or retrieval from attendance records and national registries. In the case of self-report, the outcome could be biased by unblinded participants' knowledge of the intervention. In 27 studies we considered the risk of detection bias to be high. With regard to the secondary outcome depression, 31 studies had a high risk of bias, and for two studies the risk was unclear. Our secondary outcome work functioning was measured in 11studies only. For reasons of clarity of the risk of bias table, the findings for this outcome were reported in Table 1.

Incomplete outcome data

We found nine of the 45 studies to have a low risk of attrition bias for both sickness absence and depressive symptoms, and 15 studies with a high risk of attrition bias for both outcomes, the other 21 studies showed different levels of risk of bias for sickness absence and depressive symptoms or had an unclear risk of attrition bias for either or both outcomes. Studies with attrition between 10% and 20% could still be classified as having low risk of attrition bias if adequate analyses were conducted to take selective attrition into account. Examples of such analyses are multiple imputation methods or sensitivity analyses. Our secondary outcome work functioning was measured in only 11 studies in a way that the findings could be included in the meta-analyses. To maintain clarity

in the 'Risk of bias' table, we reported the findings for this outcome in Table 1.

Selective reporting

For 23 studies, no design paper or trial registration could be identified in order to assess the risk of selective reporting. In 15 studies we considered the risk to be low (Beiwinkel 2017; Chatterton 2018; Eriksson 2017; Geraedts 2014; Hees 2013; Hellstrom 2017; Kendrick 2005; Knapstad 2020; Lerner 2015; Lerner 2020; Meuldijk 2015; Phillips 2014; Reme 2015; Reme 2019; Vlasveld 2013). In seven studies, the risk of reporting bias was deemed high, in the Björkelund 2018 study the trial was retrospectively registered. In Mackenzie 2014 and Kaldo 2018, the trial protocol was retrospectively registered and work participation was not mentioned as an outcome. In Krogh 2009, no report was made regarding the third treatment group (relaxation) in the study protocol. In the protocol of the Sarfati 2016 study, an assessment at 6 months is announced, but this is not reported. In the Noordik 2013 study, an outcome measure that was presented in the study design was not reported as an outcome. In Volker 2015, the inclusion criteria reported in the protocol changed from depressive disorders to common mental disorders as a result of a new sponsor after years of inclusion. Also, more outcomes were added.

Other potential sources of bias

We identified other sources of bias in four studies. In Birney 2016, we identified a potential conflict of interest, as the study's principal investigator may have had financial benefit from sales of the intervention. In Kaldo 2018, an unplanned subgroup analysis was conducted and in Mackenzie 2014 the work outcomes were unplanned. In Volker 2015, one occupational physician from the control condition was replaced by another occupational health physician, who then was allocated to the intervention condition.

Effects of interventions

See: **Summary of findings 1** Work-directed plus clinical intervention compared to care as usual in depressed people, medium-term follow-up; **Summary of findings 2** Work-directed intervention compared to care as usual in depressed people, medium-term follow-up; **Summary of findings 3** Psychological intervention compared to care as usual in depressed people, medium-term follow-up; **Summary of findings 4** Improved care compared to care as usual in depressed people, medium-term follow-up

Below we present the results for our primary outcome, sickness absence, for each of the comparisons. We present our secondary outcomes, depressive symptoms and work functioning, for each of the work-directed interventions as well.

1. Work-directed interventions

1.1 Work-directed plus clinical intervention compared to care as usual (medium-term follow-up)

Eleven studies examined the effect of combining a work-directed intervention with a clinical intervention in comparison to various care as usual conditions (Finnes 2017; Geraedts 2014; Hees 2013; Kaldo 2018; Lerner 2012; Lerner 2015; Lerner 2020; Reme 2015; Schene 2006; Vlasveld 2013; Volker 2015). Nine of these studies reported on sickness absence days, while two (Kaldo 2018; Reme 2015) reported on whether workers were off work (yes or no). The pooled results of nine studies revealed that work-directed interventions combined with a clinical intervention led to fewer sickness absence days in comparison with care as usual (SMD -0.25, 95% CI -0.38 to -0.12; Analysis 1.1). A sensitivity analysis showed that removing the study with a higher risk of bias (Lerner 2015) did not substantially change this outcome (SMD -0.24, 95% CI -0.41 to -0.08; Analysis 19.1).

No difference in being off work was found between work-directed and clinical interventions compared to care as usual (RR 1.08, 95% CI 0.64 to 1.83; Analysis 1.2).

Eight studies (Finnes 2017; Geraedts 2014; Hees 2013; Lerner 2012; Lerner 2015; Lerner 2020; Schene 2006; Vlasveld 2013) also reported on depressive symptoms. The summarised results of these studies showed that a work-directed interventions combined with a clinical intervention led to lower levels of depressive symptoms (SMD -0.25, 95% CI -0.49 to -0.01; Analysis 1.3).

Work functioning outcomes were reported in five studies (Finnes 2017; Hees 2013; Kaldo 2018; Lerner 2012; Lerner 2020). The summarised results of these studies did not show a difference in work functioning between the two conditions (SMD -0.19, 95% CI -0.43 to 0.06; Analysis 1.4)

As this comparison comprised various usual care conditions, we present the subgroup results of the primary outcome for each condition separately below.

1.1.1 Work-directed plus clinical interventions compared to care as usual (psychiatric clinical management)

Two studies examined the effectiveness of a work-directed intervention combined with a clinical intervention (psychiatric clinical management) in comparison to a psychiatric clinical management alone (Hees 2013; Schene 2006). The summarised effect of the two studies which added occupational therapy to psychiatric clinical management on sickness absence days was not statistically significant (SMD -0.30, 95% CI -0.61 to 0.01). The combined results of these two studies showed no difference between the interventions in effect on depressive symptoms (SMD -0.08, 95% CI -0.66 to 0.50), nor did the Hees 2013 study find an effect on work functioning (SMD -0.09, 95% CI -0.48 to 0.29).

1.1.2 Work-directed plus clinical interventions compared to care as usual (primary care)

Five studies compared the effect of a work-directed plus clinical intervention with care as usual consisting of access to primary care, of which four (Finnes 2017; Lerner 2012; Lerner 2015; Lerner 2020) measured sickness absence days and one (Kaldo 2018) investigated the number of workers on sickness absence. The pooled effect of the four studies showed that a work-directed plus clinical intervention reduced the number of sickness absence days (SMD -0.32, 95% CI -0.56 to -0.07) but the Kaldo 2018 study did not find a difference in number of workers off work (RR 1.73, 95% CI 0.70 to 4.24) between the two conditions.

1.1.3 Work-directed plus clinical interventions compared to care as usual (work-directed interventions)

Three studies (Reme 2015; Vlasveld 2013; Volker 2015) compared the effect of a work-directed plus clinical intervention with care as usual consisting of work-directed interventions, such as usual occupational healthcare or employment services. Two of these

studies (Vlasveld 2013; Volker 2015) measured sickness absence days and one (Reme 2015 examined the number of workers off work. The pooled effect of the two studies showed no difference on sickness absence days (SMD -0.20, 95% CI -0.44 to 0.04), and the Reme 2015 study did not find a difference in number of workers off work (RR 0.94, 95% CI 0.83 to 1.06) between the two conditions.

1.1.4 Work-directed plus clinical interventions compared to care as usual (no intervention)

One study (Geraedts 2014) compared the effect of a work-directed intervention with no intervention and found no effect on sickness absence days between the two conditions (SMD 0.02, 95% CI -0.33 to 0.37).

1.2 Work-directed plus clinical intervention compared to care as usual (long-term follow-up)

Two studies also reported long-term effects of (Hees 2013; Schene 2006) of their work-directed clinical intervention. These two studies found that in the long term, a work-directed intervention plus clinical intervention did not reduce sickness absence days in comparison with care as usual (SMD -0.19; 95% CI: -0.49 to 0.12; Analysis 2.1). However, one of the two studies (Hees 2013) found that the work-directed intervention reduced depressive symptoms in the long term (SMD -0.63; 95% CI -1.02 to -0.24;Analysis 2.2). That same study did not find an effect on work functioning (SMD -0.25, 95% CI -0.63 to 0.14;Analysis 2.3).

1.3 Work-directed plus clinical intervention compared to psychological intervention (medium-term follow-up)

One study (Finnes 2017) compared the effect of a work-directed intervention (facilitating a dialogue between the participant and the supervisor following a protocol) combined with a clinical intervention (Acceptance and Commitment Therapy) with the clinical intervention alone. This study did not find differences between the two intervention in terms of sickness absence days (SMD 0.04, 95% CI -0.47 to 0.56;Analysis 3.1), depressive symptoms(SMD -0.15, 95% CI -0.69 to 0.39; Analysis 3.2), nor work functioning (SMD -0.08, 95% CI -0.63 to 0.48;Analysis 3.3).

1.4 Work-directed plus clinical intervention compared to work-directed intervention (medium-term follow-up)

The Finnes 2017 study also compared the effect of a work-directed intervention (facilitating a dialogue between the participant and the supervisor following a protocol) combined with a clinical intervention (Acceptance and Commitment Therapy) to the work-directed intervention alone.This study did not find differences between the two intervention in terms of sickness absence days (SMD -0.10, 95% CI -0.65 to 0.45;Analysis 4.1), depressive symptoms (SMD -0.37, 95% CI -0.98 to 0.23;Analysis 4.2), nor work functioning (SMD 0.32, 95% CI -0.30 to 0.94;Analysis 4.3).

1.5 Work-directed intervention compared to care as usual (medium-term follow-up)

Four studies (Finnes 2017; Hellstrom 2017; Noordik 2013; Reme 2019) compared a work-directed intervention with care as usual. Two of these (Finnes 2017; Noordik 2013) reported sickness absence days, while two other (Hellstrom 2017; Reme 2019) reported the number of workers off work. Of the Reme 2019 study only the long-term result were retrieved from the authors (see comparison 1.6). In three studies (Hellstrom 2017; Noordik 2013; Reme 2019), care as usual was work-directed, such as regular occupational healthcare or employment services. In the Finnes 2017 study, a work-directed intervention was compared to primary care as the care as usual condition.

The combined effect of the Finnes 2017 and Noordik 2013 studies showed workers in the care as usual condition had less sickness absence days compared to workers who received the work-directed intervention (SMD 0.39, 95% CI 0.04 to 0.74;Analysis 5.1). The Hellstrom 2017 study did not find an effect on the number of workers off work (RR 0.93, 95% CI 0.77 to 1.11;Analysis 5.2).

The results of all four studies combined showed no effect on depressive symptoms (SMD -0.10, 95% CI -0.30 to 0.10; Analysis 5.3). Finnes 2017 was the only study in this comparison that reported work functioning, and did not find a difference between the work-directed intervention and usual care(SMD -0.32, 95% CI -0.90 to 0.26;Analysis 5.4).

1.6 Work-directed intervention compared to care as usual (long-term)

Of the studies that compared a work-directed intervention with care as usual, two studies (Hellstrom 2017; Reme 2019) reported long-term outcomes. The combined results of these studies found no effect on the number of workers off work (RR 1.00, 95% CI 0.82 to 1.22; Analysis 6.1), and the Hellstrom 2017 study did not find a difference in depressive symptoms (SMD 0.18, 95% CI -0.13 to 0.49; Analysis 6.2).

2. Clinical interventions

2.1 Psychological intervention compared to care as usual (short-term follow-up)

One study (Birney 2016) reported the short-term outcomes of a psychological intervention, unguided Internet-delivered therapy and did not find a difference in sickness absence days between the two conditions (SMD -0.05, 95% CI -0.28 to 0.17; Analysis 7.1).

2.2 Psychological intervention versus care as usual (medium-term follow-up)

Nine studies (Bee 2010; Beiwinkel 2017; Birney 2016; Eriksson 2017; Finnes 2017; Hollinghurst 2010; Mackenzie 2014; McCrone 2004; Phillips 2014) compared a psychological intervention, either face-to-face, or an E-mental health intervention with or without guidance from a care provider, with care as usual. These psychological interventions led to a smaller number of sickness absence days (SMD -0.15, 95% CI -0.28 to -0.03; Analysis 8.1). A sensitivity analysis showed that studies with a higher and lower risk of bias did not differ in their effect on days of sickness absence (test for subgroup differences: $Chi^2 = 0.00$, df = 1 (P = 1.00); Analysis 20.1).

All these studies, except for Mackenzie 2014, also reported on depressive symptoms. The pooled results of these studies showed that the psychological interventions reduced depressive symptoms (SMD -0.30, 95% CI -0.45 to -0.15; Analysis 8.2).

Only the Finnes 2017 study reported on work functioning and found no difference in this outcome between the psychological intervention and care as usual (SMD 0.05, 95% CI -0.46 to 0.57).

2.3 Psychological intervention compared to other psychological intervention (medium-term follow-up)

One study (Knekt 2013) with three treatment arms evaluated the effect of alternative psychological interventions. Two study arms assessed psychodynamic therapy, where one study arm examined short-term and the other long-term therapy. Both were compared with solution-focused therapy, but did not lead to fewer sickness absence days (SMD 0.70, 95% CI -0.19 to 1.59; Analysis 9.1).

One other study (Wormgoor 2020) compared two alternative psychological interventions and found no difference in the risk of being off work between coping focussed therapy and short-term psychotherapy in the first year of follow-up (RR 1.83, 95% CI 1.00 to 3.37; Analysis 9.2).

The Knekt 2013 study also reported on depressive symptoms, but the inconsistency (I^2) in this meta-analysis was 99.2%, we therefore refrained from pooling the results of the two psychodynamic therapy conditions for this outcome. Workers receiving short-term psychodynamic therapy had less depressive symptoms than workers who received solution-focused therapy (SMD -1.19, 95% CI -1.58 to -0.81) but workers who received long-term psychodynamic therapy had more depressive symptoms than workers receiving solution-focused therapy (SMD 2.04, 95% CI 1.62 to 2.45; Analysis 9.4).

The Knekt 2013 study also reported on work functioning, but the inconsistency (I^2) in this meta-analysis was 97.5%. We therefore also refrained from pooling the results of the two psychodynamic therapy conditions for this outcome. Workers receiving short-term psychodynamic therapy had fewer work functioning problems than workers who received solution-focused therapy (SMD -0.66, 95% CI -1.03 to -0.30) but workers who received long-term psychodynamic therapy had more work functioning problems than workers receiving solution-focused therapy (SMD 1.00, 95% CI 0.63 to 1.36; Analysis 9.3).

2.4 Psychological intervention compared to other psychological intervention (long-term follow-up)

The Knekt 2013 study also had long-term results (five-year follow up). We refrained from statistically pooling the results due to high inconsistency (I^2 = 96.2%). The separate analyses showed that long-term (SMD -4.61, 95% CI -5.84 to -3.39) and short-term (SMD -0.91, 95% CI -1.62 to -0.19) psychodynamic psychotherapy reduced sickness absence days more than solution-focused therapy in the long term (Analysis 10.1).

The Knekt 2013 study also reported on the long-term depressive symptoms, but the inconsistency (I^2) in this meta-analysis was 92.4%, we therefore refrained from pooling the results of the two psychodynamic therapy conditions for this outcome. Workers receiving short-term psychodynamic therapy had less depressive symptoms than workers who received solution-focused therapy (SMD -0.91, 95% CI -1.62 to -0.19) and the same was found for workers who received long-term psychodynamic therapy (SMD -4.61, 95% CI -5.84 to -3.39). See Analysis 10.3.

The Knekt 2013 study also reported work functioning in the long term, the pooled results of both psychodynamic therapy conditions showed that workers receiving this therapy did not have better work functioning compared to workers receiving solution-focused therapy (SMD -0.26, 95% CI -0.52 to 0.01; Analysis 10.4).

The Wormgoor 2020 study also reported the outcome after two years and found no difference in the risk of being off work between coping focused therapy and short-term psychotherapy at that time (RR 1.14, 95% CI 0.61 to 2.11); Analysis 10.2). The Wormgoor 2020 study also found no difference in depressive symptoms (SMD -0.32, 95% CI -0.66 to 0.01; Analysis 10.3)

2.5 Psychological intervention with antidepressant compared to antidepressant (medium-term follow-up)

Two studies (Burnand 2002; Sarfati 2016) compared the effect of a psychological intervention combined with antidepressants to that antidepressant alone. The Burnand 2002 study combined psychodynamic therapy with TCA medication and the Sarfati 2016 study combined telephone-administered CBT with a SSRI. The pooled results show no difference in the number of sickness absence days (SMD -0.38, 95% CI -0.99 to 0.24; Analysis 11.1).

Both studies also reported on depressive symptoms and on work functioning. The pooled results showed no difference in depressive symptoms (SMD -0.19, 95% CI -0.50 to 0.12; Analysis 11.2) nor work functioning problems (SMD -0.24, 95% CI -0.68 to 0.20; Analysis 11.3) between the two conditions.

2.6 Any antidepressant medication versus placebo

One study compared a TCA or MAO with placebo (Agosti 1991). That study found no difference in sickness absence days between the antidepressant medication and placebo, the effect may even have been in favour of the placebo condition (SMD 0.48; 95% CI -0.05 to 1.00) but this was not statistically significant (Analysis 12.1).

Measured with the work functioning subscale of the LIFE interview, Agosti 1991 did find a statistically significant positive effect in favour of antidepressant medication (SMD -0.58; 95% CI -1.11 to -0.05; Analysis 12.2).

2.7 Antidepressant medication compared to any other antidepressant medication (medium-term follow-up)

2.7.1 SSRI versus SNRI

Three studies compared a SSRI with SNRI in depressed workers (Fernandez 2005; Romeo 2004; Wade 2008). In the meta-analysis, the inconsistency of results between these three studies (I^2) was 83% and so we refrained from pooling the results. The results of the single studies were highly inconsistent. We found no difference in sickness absence between a SSRI and SNRI in the Fernandez 2005 study (SMD -0.03; 95% CI -0.37 to 0.31) as well as in the Romeo 2004 study (SMD 0.28; 95% CI -0.13 to 0.69). The Wade 2008 study revealed evidence of an effect on sickness absence favouring a SSRI (SMD -0.57; 95% CI -0.88 to -0.26; Analysis 13.1). Measured with the Sheehan disability scale, this study also reported a favourable effect on work functioning (difference of 2.4; 95% CI 0.4 to 4.1) but the reported data did not allow for inclusion in the meta-analysis.

2.7.2 SSRI versus TCA

Miller 1998 was the only study comparing an SSRI with TCA medication in depressed workers. This study found no difference between a SSRI and TCA in reducing sickness absence days (SMD 0.08; 95% CI -0.08 to 0.25; Analysis 13.1).

The Miller 1998 study measured work functioning using the SAS work composite (Wells 1989). A higher score on this measure reflects a higher level of impairment. The study reported no

significant difference on work functioning between the groups (difference of -0.08; 95% CI -0.24 to 0.09; Analysis 13.3).

2.7.3 SSRI versus SSRI

One study (Fantino 2007) compared one SSRI with another SSRI. This study found evidence of a greater reduction in sickness absence days with escitalopram compared to citalopram (SMD -0.31; 95% CI -0.54 to -0.07; Analysis 13.1.). No difference in depressive symptoms were found between the two interventions (SMD -0.23, 95% CI -0.47 to 0.00; Analysis 13.2).

2.8 Improved care compared to care as usual (medium-term follow-up)

Eight studies (Björkelund 2018; Kendrick 2005; Knapstad 2020; Meuldijk 2015; Schoenbaum 2001; Simon 1998; Wang 2007; Wikberg 2017), one of which had three study arms (Kendrick 2005), examined the effects of improved care management compared with care as usual. For the Meuldijk 2015 study, no data for the depressed subgroup be retrieved from the authors. The pooled results of the seven studies that measured sickness absence days, showed that care management did not lead to fewer sickness absence days (SMD -0.05, 95% CI -0.16 to 0.06); Analysis 14.1), however the sensitivity analysis revealed a statistically significant difference between the studies with a lower and higher risk of bias (test for subgroup differences: Chi^2 = 6.00, df = 1 (P = 0.01). The studies with a lower risk of bias (Simon 1998; Wang 2007) did show fewer sickness absence days in the care management condition (SMD -0.20, 95% CI -0.35 to -0.05; Analysis 21.1). A further sensitivity analysis showed that the results for all studies pooled together minus the one with a cluster-randomised design (Björkelund 2018) was similar to the overall meta-analysis (Analysis 22.1).

The Knapstad 2020 study investigated the numbers of workers off work and found no difference between the improved care and the care as usual condition (RR 0.97, 95% CI 0.77 to 1.21; Analysis 14.2).

Data on the depressive symptoms were available for six studies (Björkelund 2018; Kendrick 2005; Knapstad 2020; Simon 1998; Wang 2007; Wikberg 2017). The pooled results of these studies showed that improved care management led to fewer depressive symptoms (SMD -0.21, 95% CI -0.35 to -0.07; Analysis 14.3).

One study (Wang 2007) also reported on work functioning. This study found more work functioning problems in the care management condition compared to care as usual (SMD 0.50, 95% CI 0.34 to 0.66; Analysis 14.4).

2.9 Improved care compared to care as usual (long-term follow-up)

One study (Schoenbaum 2001) reported the long-term outcomes of improved care management. This study showed no difference in the number of workers who were at were off work (RR 1.06, 95% CI 0.90 to 1.23; Analysis 15.1). However workers receiving improved care had fewer depressive symptoms compared to workers receiving care as usual (RR 0.89, 95% CI 0.81 to 0.98; Analysis 15.2).

2.10 Exercise intervention compared to care as usual or relaxation (medium-term follow-up)

Three studies (Kaldo 2018; Krogh 2009; Krogh 2012) examined the effect of an exercise intervention, either compared with care as usual (Kaldo 2018) or compared with relaxation (Krogh 2009; Krogh 2012). The Krogh 2009 study had two study arms.

The Krogh 2009 and the Krogh 2012 both reported the sickness absence days, but the inconsistency (I^2) in this meta-analysis was 90.2%, we therefore refrained from pooling the results for this outcome. The results for days of sickness absence are therefore presented for the two subgroups separately.

2.10.1 Strength exercise versus relaxation

The Krogh 2009 found that supervised strength exercise led to fewer sickness absence days compared to relaxation (SMD -1.11; 95% CI -1.68 to -0.54), but no difference in depressive symptoms (SMD 0.15, 95% CI -0.39 to 0.68). See Analysis 16.1.

2.10.2 Aerobic exercise versus relaxation or stretching

The pooled effect of two studies (Krogh 2009; Krogh 2012) showed that aerobic exercise did not lead to fewer sickness absence days than relaxation or stretching in reducing sickness absence (SMD -0.06; 95% CI -0.36 to 0.24; Analysis 16.1).

Both studies also reported on depressive symptoms and also found no differences between the two conditions for this outcome (SMD -0.06, 95% CI -0.36 to 0.24).

The Kaldo 2018 study compared an exercise intervention with care as usual (primary care) and reported on the number of workers who were off work. No difference in being at work was found between the two conditions (RR 0.38, 95% CI 0.02 to 8.62; Analysis 16.2).

The Kaldo 2018 study also reported on work functioning and found no difference in work functioning problems between the two conditions (SMD 0.00, 95% CI -0.18 to 0.18; Analysis 16.4).

2.11 Art therapy compared to care as usual (medium-term follow-up)

One study (Blomdahl 2018) compared art therapy with care as usual. This study showed no difference in days of sickness absence between the two conditions (SMD -0.13, 95% CI -0.58 to 0.31; Analysis 17.1).

This study also reported on depressive symptoms, and found no difference in depressive symptoms between the two conditions (SMD -0.43, 95% CI -0.88 to 0.02; Analysis 17.2).

2.12 Diet compared to social support (medium-term follow-up)

One study (Chatterton 2018) compared a diet intervention with a social support intervention. This study showed no difference in sickness absence days between the two conditions (SMD -0.30, 95% CI -0.78 to 0.18; Analysis 18.1).

This study also reported on depressive symptoms, and found lower levels of depressive symptoms in workers who received the diet intervention (SMD -4.91, 95% CI -5.99 to -3.83; Analysis 18.2).

DISCUSSION

Summary of main results

We included 45 studies, of which 40 RCTs and five cluster-RCTs, in the review, evaluating a total of 17 work-directed interventions and 32 clinical interventions. Within these broad categories, the type of intervention varied from one study to another, which limited

the number of studies in each predefined comparison. We present 'Summary of findings' tables for the comparisons with more than two included studies. Summary of findings 1 presents the GRADE assessment of the certainty of the evidence per comparison.

Work-directed and a clinical intervention compared with care as usual

There is moderate-certainty evidence that a combination of a work-directed and a clinical intervention probably reduces the number of sickness absence days in the medium term to a small degree (4 to 12 months; SMD -0.25) more than care as usual but this does not lead to a greater number of people at work in the intervention group at the end of a one-year follow-up or beyond. The SMD of -0.25 translates back to 0.5 fewer sickness absence days in the past two weeks (CI -0.7 to -0.2 days) or to 25 fewer sickness absence days during one year (CI -37.5 to -11.8). See Summary of findings 1; Table 2.

There was high-certainty evidence that a combination of a work-directed and a clinical intervention does not lead to fewer people being off work at the end of follow-up at medium term follow-up, while we found low-certainty evidence that a combination of a work-directed and a clinical intervention may reduce depressive outcomes (SMD -0.25) and low-certainty evidence of no effect on work functioning outcomes. See Summary of findings 1.

Work-directed compared with care as usual

There is low-certainty evidence, based on two studies, that a specific work-directed intervention alone may increase the number of sickness absence days compared with work-directed care as usual (SMD 0.39, 95% CI 0.04 to 0.74) within one year follow-up. This SMD translates back to an increase of 0.7 sickness absence days in two weeks (CI 0.1 to 1.3 days) or an increase of 38 days in 12 months (CI 3.9 to 73 days). This review also found moderate-certainty evidence based on four studies that there is probably no effect of work-directed interventions on depressive symptoms within the first year of follow-up (SMD - 0.10, 95% -0.30 CI to 0.10; 4 studies, moderate-certainty evidence) and beyond (SMD 0.18, 95% CI -0.13 to 0.49; 1 study, low-certainty evidence) and low-certainty evidence of no effect on work functioning outcomes (SMD - 0.32 95% CI - 0.90 to 0.26; 1 study) within the first year of follow up. There further is moderate-certainty evidence the intervention does not lead to a lower or greater number of people at work in the intervention group at the end of a one-year follow-up (RR 0.93, 95% CI 0.77 to 1.11; 1 study, moderate-certainty evidence) or beyond (RR 1.00, 95% CI 0.82 to 1.22; 2 studies, moderate-certainty evidence). See Summary of findings 2 and Table 3

Psychological interventions compared with care as usual

One study in this review compared a psychological intervention to care as usual in the short-term (Birney 2016) and found low-certainty evidence that there may be no difference in number of sickness absence days. The SMD -0.05 of translates back to a reduction of 0.1 days per 2 weeks (CI -0.5 to 0.3 days) or a reduction of 4.9 days in 12 months (-27.6 to 16.8 days). See Table 4.

We found low-certainty evidence, based on nine studies, that a psychological intervention either face-to-face, or an E-mental health intervention with or without guidance from a care provider may reduce the number of sickness absence days compared with care as usual at medium term follow-up (SMD -0.15, 95% CI -0.28 to -0.03). This SMD translates back to a reduction of 0.3 days per 2 weeks (CI -0.5 to -0.1) or a reduction of 14.7 days in 12 months (-27.6 to -3.0).

All these studies, except for Mackenzie 2014, also reported on depressive symptoms, leading to low-certainty evidence that psychological interventions may reduce depressive symptoms (SMD -0.30, 95% CI -0.45 to -0.15).

See Summary of findings 3.

Improved care compared with care as usual

We found that care management did not lead to fewer sickness absence days in the medium term in seven studies. The SMD translates back to a reduction of 0.1 days per 2 weeks (CI -0.3 to 0.1) or a reduction of 5.9 days in 12 months (-14.8 to 3.9). However, a sensitivity analysis revealed that the studies with a lower risk of bias (Simon 1998 and Wang 2007) probably lead to fewer sickness absence days in the care management condition (SMD -0.20, 95% CI -0.35 to -0.05; moderate-certainty evidence). The SMD of this sensitivity analysis translates back to a reduction of 0.4 days per 2 weeks (CI -0.6 to -0.1) or a reduction of 19.7 days in 12 months (CI -34.5 to -4.9).

We found moderate-certainty evidence based on seven studies that improved care management probably leads to fewer depressive symptoms in the medium term (SMD -0.21, 95% CI -0.35 to -0.07). Summary of findings 4.

For the long term, we found moderate-certainty evidence based on one study that care management probably does not reduce the number of sickness absence days, nor the depressive symptoms. See Table 5.

Antidepressant medication

With regard to antidepressant medication, this review found highly inconsistent results regarding the effect of SSRIs compared with other medications on sickness absence days(four studies). Compared with SNRI medication (three studies), one single study found that SSRI reduced sickness absence (Wade 2008), no difference in effect on sickness absence was found in another (Fernandez 2005), and a non-significant difference in effect on sickness absence was found in the last (Romeo 2004). One single study found that a SSRI did not reduce sickness absence more than TCA medication (Miller 1998). One study (Fantino 2007) compared one SSRI with another SSRI. This study found that escitalopram reduced sickness absence more than citalopram (SMD -0.31). One study compared a TCA or MAO with placebo (Agosti 1991). This study found that the antidepressant medication did not reduce sickness absence more than placebo.

Overall completeness and applicability of evidence

The studies included in this review have been conducted in Europe, the United States of America, Australia and Canada only. Therefore, the generalisability of our findings to other parts of the world remains unclear. In line with our inclusion criteria, the included studies cover a range of clinical states. In 34 studies, a clinical diagnosis, most often a major depressive disorder according to the DSM-IV or III, was used as an inclusion criterion, while in 11 studies included patients based on their symptom severity as measured by a questionnaire.

Moreover, study setting is likely to be a source of clinical heterogeneity. Studies were conducted in various settings, with most (28) taking place in primary care and outpatient settings. In five studies, patients were recruited in a workplace setting, and in another five in occupational health care. In many instances, the occupation of the participants was not reported even though it is conceivable that the effect of interventions partly depends on the specific work situation. A lack of studies on work-related factors that may be predictors for work outcomes in depressed workers has already been pointed out (Lagerveld 2010). A meta-analysis of prognostic studies did not identify work-related prognostic factors for return to work in depressed workers (Ervasti 2017). A study by de Vries 2015 did find that work characteristics (work pace and workload, decision latitude, autonomy, relations with supervisor and job insecurity) were predictive of impaired work functioning after clinical recovery from a major depressive disorder. Moreover, studies have shown that work-related factors are predictors of the clinical outcome for depression. For instance, Wang 2012 found long working hours to be associated with persistence of the depressive disorder over time. Therefore, we cannot assess the potential impact of work situations on the effectiveness of the included interventions.

In this updated review, we were able to include studies on work-directed interventions as well as clinical interventions. While it is important to assess the effects of clinical interventions on occupational health, we are aware that the primary reason to choose between one or another clinical intervention is clinical effectiveness. However, in line with the emerging paradigm of value-based medicine, it is central to care to offer interventions to patients providing the greatest patient value (Brown 2013). As being able to work may be one of the factors on which patient preference is based, so that assessing occupational health outcomes for clinical interventions is a key aspect. Moreover, from the point of view of patient preference, work functioning may be as important as sickness absence. However, in most included studies this outcome was not measured. Evaluating the effect of interventions on work functioning would further enable us to assess the patient value of these interventions.

In contrast to the first version of this review, we were able to include studies in most of the predefined comparisons. However, the number of studies within some of the comparison was small, and within some of the comparisons, the interventions were too dissimilar to pool the results. Another consequence of the low number of studies per comparison is that we were unable to perform subgroup analyses for participant and intervention characteristics and other sensitivity analyses which impedes the generalisation of the results.

The clinical relevance of the observed effects can best be evaluated by looking at the absolute differences in sickness absence days. It should, however, be noted that these differences vary from one study to another. Part of the explanation is that the outcome measure 'sickness absence days' is by definition partly determined by the length of follow up. The relevance of reductions in sickness absence days depends on the perspective of the stakeholder. A reduction in sickness absence of one day may not be relevant from the worker's point of view but can be relevant for stakeholders who bear the costs of the lost productivity, such as employers or insurance companies.

Quality of the evidence

Of the included studies, 40 were RCTs and five were cluster-RCTs. The total number of participants in the analyses was 12,109. The number of participants in the smallest study subgroup was lower than 50 in 20 study arms, between 50 and 99 in 22 study arms, between 100 and 199 in 23 study arms, and 200 or more in 18. In some cases, the low number of participants was due to our need to focus only on a subgroup of the study population, disregarding participants with no or other mental disorders.

In the three cluster-RCTs, allocation concealment was not adequate, probably indicative of the non-feasibility of allocation concealment in this type of design due to all participants in one cluster (for example in a practice or with a healthcare provider) being automatically assigned to the same study arm. In seven further studies, information on random sequence generation or allocation concealment could not be retrieved, leading to a judgment of unclear risk of bias. In 35 studies, the allocation concealment was adequate and therefore posed a low risk of bias.

We found a high risk of performance bias in 10 included studies. In work-directed interventions and in clinical interventions,such as psychological or exercise interventions, blinding of the participant or care provider is not feasible. However, the risk of performance bias also depends on how desirable the intervention is compared with the control group, according to either care providers or participants. One study evaluating a psychological intervention in addition to medication managed to compose two evenly desirable psychological interventions by ensuring an equal number of supportive, instead of therapeutic, sessions.

In this review, we chose to assess detection and attrition bias separately for sickness absence and depressive symptoms. We felt that not being blind to allocation may bias a self-report assessment of depressive symptoms more than the reporting of a more factual outcome such as the days absent from work in a given period. In addition, sickness absence may be retrieved from employee attendance records while depression is measured with a self-report questionnaire. In those instances, the lack of blinding of outcome assessment (detection bias) cannot influence the sickness absence but may well bias the depressive outcome. Nonetheless, the risk of detection bias for the outcome sickness absence was still high in 27 studies, which led to downgrading of the level of evidence in many comparisons.

Potential biases in the review process

This review included studies with a study population of both workers and non-workers. This means that subgroups of the original sample were used for measuring the effect on sickness absence. These studies did not usually present all data for workers separately, but their sickness absence reports were by definition based on the workers in the study population. Some studies included participants with mental disorders other than depression. We included the studies in this review if the authors were willing to provide data for the depressed subgroup.

Subgroup analyses in individual studies may lead to biased results (Freemantle 2001), one reason being that testing many subgroups increases the likelihood of finding a statistically significant result by chance alone. We therefore predefined the subgroups and did not test multiple potential subgroups in the hope of finding

a statistically significant finding. However, we acknowledge that a lack of power leading to statistically non-significant findings may have occurred in our review. We, therefore, refrained from describing non-significant findings with wide confidence intervals as evidence of no effect. In the future, we may have to reconsider our approach of selecting depressed workers only. The distinction between with depressive and other disorders, such as anxiety disorder may become less relevant as new insights are emerging based on the network perspective on psychopathology (Fried 2017). This approach involves investigating the importance and connections of individual symptoms rather than disorders. If future clinical treatments will use the most central symptom rather than a diagnostic category as a starting point for treatment, a review such as this may have to be organised around broader categories of psychological problems, such as common mental disorders, or around selected central depressive symptoms.

This review evaluates the effectiveness of a range of interventions aiming to reduce sickness absence in depressed workers rather than one specific intervention. While we believe this is appropriate for a complex and multifactorial outcome such as sickness absence, the categorisation of interventions under the comparisons has been challenging. This categorisation is likely to influence the results as it determines, for each intervention, with which other interventions the results will be pooled and to which other interventions it will be compared. The way interventions are categorised entails a potential bias in the review process. In the 2020 update, we have re-organised the comparisons. One change was that we now distinguish between care as usual (a study arm where patients are treated without a specific intervention protocol) and an alternative intervention (an intervention that was protocolised, regardless of whether that intervention constitutes the regular care in that setting).

Another methodological issue concerns the handling of sickness absence data. For calculating standardized mean differences (SMDs) we considered sickness absence as a construct for which different instruments could be used, as long as they provided information on absenteeism. This meant that as long as we reported SMDs we could incorporate studies with different time spans (and therefore with a different maximum of sickness absence days during follow up) and scales that differed in the maximum score. Also, this enabled us to compare studies from various countries as we know that days of sickness absence tend to be calculated differently in different countries (for instance due to differences in whether calendar days or only work days are included as absenteeism days). Moreover, we transformed reports of days worked into days of sickness absence by extracting the days worked from the days that should have been worked ('the scale maximum'). This is analogous to transforming the scores of a scale in which a high score indicates a good outcome into a scale where a high score indicates a bad outcome. However, for this transformation we had to make inferences about the mean number of hours and the number of hours a day an employee would work in a specific country. This transformation, along with the different time spans, impedes the translation of standardized mean differences into overall mean of days of sickness absence.

Another issue regarding sickness absence data is that we accepted both self-report and administrative databases as sources of data on sickness absence. Administrative databases are sometimes considered the gold standard. Agreement between the two sources has been reported to be good (Ferrie 2005; Severens 2000) but also limited (Pole 2006; van Poppel 2002). A meta-analysis by Johns 2015 showed that in self-report, workers tend to underestimate their days of sickness absence with a mean of two days. And a more recent study (Thorsen 2018) found that workers with fewer than 10 sickness absence days under-reported and those with more than 30 days over-reported the sickness absence days. As we included RCTs, the relative difference between the intervention was not affected by under- or overestimation in studies using self-report. However, this does further complicate the translation of the effect estimates we found into days of sickness absence. In summary, caution is recommended when interpreting sickness absence data in meta-analyses as this is as not all methodological issues have been thoroughly investigated.

Agreements and disagreements with other studies or reviews

Furlan 2012 searched the literature until 2010 and concluded that the evidence was of insufficient quality to determine which interventions are effective and are of value for the management of depression in the workplace. This conclusion was similar to the first published version of this review (Nieuwenhuijsen 2008). Our updated review has markedly different conclusions due to the inclusion of a substantially greater number of studies. In the current and the last update we were able to conclude that, at least with moderate certainty, work-directed and clinical interventions combined are effective in reducing sickness absence. We also found low-certainty evidence that psychological interventions are able to reduce sickness absence days better than care as usual.

Our finding that a work-directed intervention alone does not reduce sickness absence is in line with the conclusion of a Cochrane systematic review (Vogel 2017) that return-to-work coordination programmes do not improve return-to-work outcomes in musculoskeletal and mental health problems. Our results further show that a combination of a work-directed and a clinical intervention does have the potential to reduce sickness absence. One possible explanation for this is that the integration of the clinical and the work-directed elements of an intervention is key in improving work outcomes.

Our finding that medication interventions show highly inconsistent effects on sickness absence has not changed with this new update. This is partly in line with findings of a systematic review (Lee 2018) which found that antidepressant medication had a positive effect on workplace functioning, but not on sickness absence. As we only included studies that measured our primary outcome, sickness absence, they included different studies in their review. Our review aims to completely cover the evidence for the effect on our primary outcome, and therefore our findings with regard to the secondary outcomes should be interpreted only in relation to the findings on our primary outcome.

AUTHORS' CONCLUSIONS

Implications for practice

A combination of a work-directed and a clinical intervention probably reduces the number of sickness absence days but not the number of people at work at the end of follow up. Specific work-directed intervention may not be more effective than usual work-directed care alone. We further found that psychological

interventions may reduce the number of sickness absence days compared to care as usual. These interventions were either provided face-to-face, or online, with or without guidance from a care provider. Improving the management of clinical care probably also leads to fewer days of sickness absence and, probably reduces the depressive symptoms. The effects of one antidepressant medication compared with another were inconsistent without a clear pattern.

Implications for research

More research is needed on combined work-directed interventions combined with clinical interventions. Such interventions probably reduce sickness absence to a small degree but it is unclear which type and combination of work-directed and clinical intervention is the most effective and what is the mechanism by which this intervention apparently works. For example, it is unclear if it is most important to add work-directed intervention to clinical treatment such as adding overcoming barriers for return-to-work to existing treatment such as CBT or the other way around to add clinical intervention to existing return-to-work services.

Most studies currently compare the intervention to care as usual. This makes comparison across studies difficult as the content of care as usual is often not specified. In future studies care as usual should be specified and it should be reported how this differs from the intervention.

Future studies should also elucidate the apparent discrepancy between the reduction in sickness absence and the lack of effect on the proportion of persons at work at end of follow-up.

Studies should focus solely on patients with depression because it is unclear if mechanisms of action differ between patient groups with various mental health problems. In addition, it has been shown that well-powered studies with depressed persons are possible. More of these studies are needed. This will also prevent the decrease in validity of the studies when analysing only subgroups of patients with depression.

To facilitate the synthesis of evidence from various intervention studies, the occupational health field should work towards standardising and validating measures of sickness absence that preferably should be measured in an objective way for example based on registry data.

For future reviews including sickness absence as an outcome measure, it is advisable to report standardised mean differences instead of means as this takes into account the differences in measurement methods.

ACKNOWLEDGEMENTS

We are grateful to José Luis Fernandez, Martin Keller, Tony Kendrick, Paul Knekt, Paul McCrone, Judith Proudfoot, Renee Romeo, Kathryn Rost, Aart Schene, Gregory Simon, Alan Wade, Penny Bee, Clément Francois, Nicholas Moore, and Ken Wells for kindly providing further information about their studies. We thank the CCDAN group for their support with the first version of this review. We thank Melissa Raven and Berend Terluin for their peer review of the updated 2020 review. We thank Janet Wale for copy editing the text of the updated 2014 review and Hacsi Horvath for copy editing the text of the updated 2020 review. In the 2020 update Joost Daams kindly conducted the search update. To conclude, we are very grateful for the help of the Cochrane Work Review Group.

We would like to acknowledge the contribution of Hiske Hees and Christina van der Feltz-Cornelis as authors of the previous version of this review

REFERENCES

References to studies included in this review

Agosti 1991 *{published data only}*

Agosti V, Stewart JW, Quitkin FM. Life satisfaction and psychosocial functioning in chronic depression: effect of acute treatment with antidepressants. *Journal of Affective Disorders* 1991;**23**(1):35-41.

Bee 2010 *{published data only}*

Bee PE, Bower P, Gilbody S, Lovell K. Improving health and productivity of depressed workers: a pilot randomized controlled trial of telephone cognitive behavioral therapy delivery in workplace settings. *General Hospital Psychiatry* 2010;**32**:337-40.

Beiwinkel 2017 *{published data only}*

Beiwinkel T, Eising T, Telle N, Siegmund-Schultze E, Rossler W. Effectiveness of a web-based intervention in reducing depression and sickness absence: randomized controlled trial. *Journal of Medical Internet Research* 2017;**19**(6):87-100. [DOI: 10.2196/jmir.6546]

Birney 2016 *{published data only}*

Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. *JMIR mHealth and uHealth* 2016;**4**(1):e8. [DOI: https://dx.doi.org/10.2196/mhealth.4231]

Björkelund 2018 *{published data only}*

Bjorkelund C, Svenningsson I, Hange D, Udo C, Petersson E, Ariai N, Nejati S, et al. Clinical effectiveness of care managers in collaborative care for patients with depression in Swedish primary health care: a pragmatic cluster randomized controlled trial. *BMC Family Practice* 2018;**19**(1):28. [DOI: https://dx.doi.org/10.1186/s12875-018-0711-z]

Blomdahl 2018 *{published data only}*

Blomdahl C, Guregård S, Rusner M, Wijk H. A manual-based phenomenological art therapy for individuals diagnosed with moderate to severe depression (PATd): a randomized controlled study. *Psychiatric Rehabilitation Journal* 2018;**41**(3):169-182. [DOI: 10.1037/prj0000300]

Burnand 2002 *{published data only}*

* Burnand Y, Andreoli A, Kolatte E, Venturini A, Rosset N. Psychodynamic psychotherapy and clomipramine in the treatment of major depression. *Psychiatric Services* 2002;**53**(5):585-90.

Chatterton 2018 *{published data only}*

Chatterton ML, Mihalopoulos C, O'Neil A, Itsiopoulos C, Opie R, Castle D, et al. Economic evaluation of a dietary intervention for adults with major depression (the "SMILES" trial).. *BMC Public Health* 2018;**18**(1):N.PAG-N.PAG. [DOI: 10.1186/s12889-018-5504-8]

Eriksson 2017 *{published data only}*

Eriksson MCM, Kivi M, Hange D, Petersson EL, Ariai N, Häggblad P, et al. Long-term effects of Internet-delivered cognitive behavioral therapy for depression in primary care–the PRIM-NET controlled trial. *Scandinavian Journal of Primary Health Care* 2017;**35**(2):126-136. [DOI: 10.1080/02813432.2017.1333299]

Fantino 2007 *{published data only}*

Fantino B, Moore N, Verdoux H, Auray J. Cost-effectiveness of escitalopram vs. citalopram in major depressive disorder. *Journal of Psychopharmacology* 2007;**22**:107-15.

Fernandez 2005 *{published data only}*

* Fernandez JL, Montgomery S, Francois C. Evaluation of the cost effectiveness of escitalopram versus venlafaxine XR in major depressive disorder. *Pharmacoeconomics* 2005;**23**(2):155-67.

Finnes 2017 *{published data only}*

Finnes A, Enebrink P, Sampaio F, Sorjonen K, Dahl J, Ghaderi A, et al. Cost-effectiveness of acceptance and commitment therapy and a workplace intervention for employees on sickness absence due to mental disorders. *Journal of Occupational and Environmental Medicine* 2017;**59**(12):1211-1220. [DOI: 10.1097/JOM.0000000000001156]

Finnes A, Ghaderi A, Dahl J, Nager A, Enebrink P. Randomized controlled trial of acceptance and commitment therapy and a workplace intervention for sickness absence due to mental disorders. *Journal of occupational health psychology* 2017 (epub);**24**(1):198-212. [DOI: 10.1037/ocp0000097]

Geraedts 2014 *{published data only}*

Geraedts AS, Kleiboer AM, Twisk J, Wiezer NM, van Mechelen W, Cuijpers P, et al. Long-term results of a Web-based guided self-help intervention for employees with depressive symptoms: Randomized controlled trial. *Journal of medical internet research* 2014;**16**(7):14-28. [DOI: http://dx.doi.org/10.2196/jmir.3539]

Geraedts AS, Kleiboer AM, Wiezer NM, Mechelen W van, Cuijpers P. Web-based guided self-help for employees with depressive symptoms (Happy@Work): design of a randomized controlled trial. *BMC Psychiatry* 2013;**13**:1-10.

Hees 2013 *{published data only}*

Hees HL, Vries de G, Koeter MW, Schene AH. Adjuvant occupational therapy improves long-term depression recovery and return-to-work in good health in sick-listed employees with major depression: results of a randomised controlled trial. *Occupational and Environmental Medicine* 2013;**70**:252-60.

Hellstrom 2017 *{published data only}*

Hellstrome L, Bech P, Hjorthoj C, Nordentoft M, Lindschou J, Falgaard Eplov L. Effect on return to work or education of Individual Placement and Support modified for people with mood and anxiety disorders: results of a randomised clinical trial. *Occupational and Environmental Medicine*

2017;**74**(10):717-725. [DOI: https://dx.doi.org/10.1136/oemed-2016-104248]

Hellstrom L, Bech P, Nordentoft M, Lindschou J, Eplov LF. The effect of IPS-modified, an early intervention for people with mood and anxiety disorders: study protocol for a randomized clinical superiority trial. *Trials* 2013;**14**:1-10. [DOI: 10.1186/1745-6215-14-442]

Hollinghurst 2010 *{published data only}*

Hollinghurst S, Peters TJ, Kaur S, Wiles N, Lewis G. Cost-effectiveness of therapist-delivered online cognitive-behavioural therapy for depression: randomized controlled trial. *British Journal of Psychiatry* 2010;**197**:297-304.

Kaldo 2018 *{published data only}*

Kaldo V, Lundin A, Hallgren M, Kraepelien M, Strid C, Ekblom O, et al. Effects of internet-based cognitive behavioural therapy and physical exercise on sick leave and employment in primary care patients with depression: two subgroup analyses. *Occupational and Environmental Medicine* 2018;**75**(1):52-58. [DOI: https://dx.doi.org/10.1136/oemed-2017-104326]

Kendrick 2005 *{published data only}*

Kendrick T, Simons L, Mynors-Wallis L, Gray A, Lathlean J, Pickering R, et al. A trial of problem-solving by community mental health nurses for anxiety, depression and life difficulties among general practice patients. The CPN-GP study. *Health Technology Assessment* 2005;**9**(37):1-104.

Knapstad 2020 *{published data only}*

Knapstad M, Lervik LV, Saether SMM, Aaro LE, Smith ORF. Effectiveness of prompt mental health care, the Norwegian version of improving access to psychological therapies: a randomized controlled trial. *Psychotherapy and Psychosomatics* 2020;**89**(2):90-105. [DOI: http://dx.doi.org/10.1159/000504453]

Knekt 2013 *{published data only}*

Knekt P, Lindfors O, Sares-Jäske L, Virtala E, Härkänen T. Randomized trial on the effectiveness of long- and short-term psychotherapy on psychiatric symptoms and working ability during a 5-year follow-up. *Nordic Journal of Psychiatry* 2013;**67**:59-68.

Krogh 2009 *{published data only}*

Krogh J, Saltin B, Gluud C, Nordentoft M. The DEMO trial: a randomized, parallel-group, observer-blinded clinical trial of strength versus aerobic versus relaxation training for patients with mild to moderate depression. *Journal of Clinical Psychiatry* 2009;**70**:790-800.

Krogh 2012 *{published data only}*

Krogh J, Videbech P, Thomsen C, Gluud C, Nordentoft M. DEMO-II trial. Aerobic exercise versus stretching exercise in patients with major depression - a randomized clinical trial. *PLOS ONE* 2012;**7**(10):e48316.

Lerner 2012 *{published data only}*

Lerner D, Adler A, Hermann RC, Hong Chang MS, Ludman EJ, Greenhill A, et al. Impact of a work-focused intervention on the productivity and symptoms of employees with depression. *Journal of Occupational and Environmental Medicine* 2012;**54**:128-35.

Lerner 2015 *{published data only}*

Adler DA, Lerner D, Visco ZL, Greenhill A, Chang H, Cymerman E, et al. Improving work outcomes of dysthymia (persistent depressive disorder) in an employed population. *General Hospital Psychiatry* 2015;**37**(4):352-359. [DOI: 10.1016/j.genhosppsych.2015.04.001]

Lerner D, Adler DA, Rogers WH, Chang H, Greenhill A, Cymerman E, et al. A randomized clinical trial of a telephone depression intervention to reduce employee presenteeism and absenteeism. *Psychiatric Services* 2015;**66**(6):570-577. [DOI: 10.1176/appi.ps.201400350]

Lerner 2020 *{published data only}*

Lerner D, Adler DA, Rogers WH, Ingram E, Oslin DW. Effect of Adding a Work-Focused Intervention to Integrated Care for Depression in the Veterans Health Administration: A Randomized Clinical Trial. *JAMA Network Open.* 2020;**3**(2):e200075-e200075. [DOI: http://dx.doi.org/10.1001/jamanetworkopen.2020.0075]

Mackenzie 2014 *{published data only}*

Mackenzie A, Harvey S, Mewton L, Andrews G. Occupational impact of internet-delivered cognitive behaviour therapy for depression and anxiety: reanalysis of data from five Australian randomised controlled trials. *Medical Journal of Australia* 2014;**201**(7):417-419.

McCrone 2004 *{published data only}*

McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. *British Journal of Psychiatry* 2004;**185**:55-62.

Meuldijk 2015 *{published data only}*

Meuldijk D, Carlier IV, Vliet IM, Hemert AM, Zitman FG, Akker-van MME. Economic evaluation of concise cognitive behavioural therapy and/or pharmacotherapy for depressive and anxiety Disorders. *Journal of Mental Health Policy and Economics* 2015;**18**(4):175-183.

Miller 1998 *{published data only}*

Miller IW, Keitner GI, Schatzberg AF, Klein DN, Thase ME, Rush AJ, et al. The treatment of chronic depression, part 3: psychosocial functioning before and after treatment with sertraline or imipramine. *Journal of Clinical Psychiatry* 1998;**59**(11):608-19.

Noordik 2013 *{published data only}*

Noordik E, Klink JJ van der, Geskus RB, de Boer MR, van Dijk FJ, Nieuwenhuijsen K. Effectiveness of an exposure-based return-to-work program for workers on sick leave due to common mental disorders: a cluster-randomized controlled trial. *Scandinavian Journal of Work, Environment & Health* 2013;**39**:144-54.

Phillips 2014 *{published data only}*

Phillips R, Schneider J, Molosankwe I, Leese M, Foroushani PS, Grime P, et al. Randomized controlled trial of computerized cognitive behavioural therapy for depressive symptoms: effectiveness and costs of a workplace intervention. *Psychological Medicine* 2014;**44**(4):741-752. [DOI: 10.1017/S0033291713001323]

Reme 2015 *{published data only}*

Reme SE, Grasdal AL, Lovvik C, Lie SA, Overland S. Work-focused cognitive-behavioural therapy and individual job support to increase work participation in common mental disorders: a randomised controlled multicentre trial. *Occupational and Environmental Medicine* 2015;**72**(10):745-52. [DOI: https://dx.doi.org/10.1136/oemed-2014-102700]

Reme 2019 *{published data only}*

Reme SE, Monstad K, Fyhn T, Sveinsdottir V, Løvvik C, Lie SA, et al. A randomized controlled multicenter trial of individual placement and support for patients with moderate-to-severe mental illness. *Scandinavian Journal of Work, Environment and Health* 2019;**45**(1):33-41. [DOI: 10.5271/sjweh.3753]

Romeo 2004 *{published data only}*

Romeo R, Patel A, Knapp M, Thomas C. The cost-effectiveness of mirtazapine versus paroxetine in treating people with depression in primary care. *International Clinical Psychopharmacology* 2004;**19**(3):125-34.

Rost 2004 *{published data only}*

Rost K, Smith JL, Dickinson M. The effect of improving primary care depression management on employee absenteeism and productivity. A randomized trial. *Medical Care* 2004;**42**(12):1202-10.

Sarfati 2016 *{published data only}*

Sarfati D, Stewart K, Woo C, Parikh SV, Yatham LN, Lam RW. The effect of remission status on work functioning in employed patients treated for major depressive disorder. *Annals of Clinical Psychiatry* 2016;**28**(4):e8-e13.

Schene 2006 *{published data only}*

Schene AH, Koeter MWJ, Kikkert MJ, Swinkels JA, Mc Crone P. Adjuvant occupational therapy for work-related major depression works: randomized trial including economic evaluation. *Psychological Medicine* 2006;**2007**(37):3.

Schoenbaum 2001 *{published data only}*

Schoenbaum M, Unutzer J, Sherbourne C, Duan N, Rubenstein LV, Miranda J, et al. Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. *JAMA* 2001;**286**(11):1325-30.

Simon 1998 *{published data only}*

Simon GE, Katon W, Rutter C, Von Korff M, Lin E, Robinson P, et al. Impact of improved depression treatment in primary care on daily functioning and disability. *Psychological Medicine* 1998;**28**(3):693-701.

Vlasveld 2013 *{published data only}*

Vlasveld MC, van der Feltz-Cornelis CM, Adèr HJ, Anema JR, Hoedeman R, van Mechelen W, et al. Collaborative care for sick-listed workers with major depressive disorder: a randomised controlled trial from the Netherlands Depression Initiative aimed at return to work and depressive symptoms. *Occupational and Environmental Medicine* 2013;**70**:223-30.

Volker 2015 *{published data only}*

Volker D, Zijlstra-Vlasveld MC, Anema JR, Beekman AT, Brouwers EP, Emons WH, et al. Effectiveness of a blended web-based intervention on return to work for sick-listed employees with common mental disorders: results of a cluster randomized controlled trial. *Journal of medical internet research* 2015;**17**(5):e116. [DOI: 10.2196/jmir.4097]

Wade 2008 *{published data only}*

Wade AG, Fernandez J, Francois C, Hansen K, Danchenko N, Despiegel N. Escitalopram and duloxetine in major depressive disorder - A pharmacoeconomic comparison using UK cost data. *Pharmacoeconomics* 2008;**26**:969-81.

Wang 2007 *{published data only}*

Wang PS, Simon SE, Avorn J, Azocar F, Ludman EJ, McCulloch J, et al. Telephone screening, outreach, and care management for depressed workers and impact on clinical and work productivity outcomes. *JAMA* 2007;**298**(12):1401-11.

Wikberg 2017 *{published data only}*

Wikberg C, Westman J, Petersson EL, Larsson ME, André M, Eggertsen R, et al. Use of a self-rating scale to monitor depression severity in recurrent GP consultations in primary care - does it really make a difference? A randomised controlled study. *BMC Family Practice* 2017;**18**(1):6. [DOI: 10.1186/s12875-016-0578-9]

Wormgoor 2020 *{published data only}*

Wormgoor MEA, Indahl A, Andersen E, Egeland J. Effectiveness of briefer coping-focused psychotherapy for common mental complaints on work-participation and mental health: a pragmatic randomized trial with 2-year follow-up. *Journal of Occupational Rehabilitation* 2020;**30**(1):22-39. [DOI: http://dx.doi.org/10.1007/s10926-019-09841-6]

References to studies excluded from this review

Aasdahl 2017 *{published data only}*

Aasdahl L, Pape K, Vasseljen O, Johnsen R, Gismervik S, Jensen C, et al. Effects of Inpatient Multicomponent Occupational Rehabilitation versus Less Comprehensive Outpatient Rehabilitation on Somatic and Mental Health: Secondary Outcomes of a Randomized Clinical Trial. *Journal of Occupational Rehabilitation* 2017;**27**(3):456-466. [DOI: 10.1007/s10926-016-9679-5]

Aasdahl 2018 *{published data only}*

Aasdahl L, Pape K, Vasseljen O, Johnsen R, Gismervik S, Halsteinli V, et al. Effect of inpatient multicomponent occupational rehabilitation versus less comprehensive outpatient rehabilitation on sickness absence in persons

with musculoskeletal- or mental health disorders: A randomized clinical trial. *Journal of Occupational Rehabilitation* 2018;**28**(1):170-9.

Aasvik 2017 *{published data only}*

Aasvik JK, Woodhouse A, Stiles TC, Jacobsen HB, Landmark T, Glette M, et al. Effectiveness of working memory training among subjects currently on sick leave due to complex symptoms. *Frontiers in Psychology* 2017;**7**(eCollection 2016):2016.02003.

Aelfers 2013 *{published data only}*

Aelfers E, Bosma H, Houkes I, van Eijk JT. Effectiveness of a minimal psychological intervention to reduce mild to moderate depression and chronic fatigue in a working population: the design of a randomized trial. *BMC Public Health* 2013;**13**:129.

Ahola 2012 *{published data only}*

Ahola, K, Vuori J, Toppinen-Tanner S, Mutanen P, Honkonen T. Resrcource-enhancing group intervention against depression at workplace: who benefits? A randomized controlled study with a 7-month follow-up. *Occupational and Environmental Medicine* 2012;**69**:870-6.

Alexopoulos 2011 *{published data only}*

Alexopoulos GS, Raue PJ, Kiosses DN, Mackin RS, Kanellopulos D, McCulloch C, et al. Problem-solving therapy and supportive therapy in older adults with major depression and executive dysfunction. *Archives of General Psychiatry* 2011;**68**(1):33-41.

Amore 2001 *{published data only}*

Amore M, Jori MC, AMISERT Investigators. Faster response on amisulpride 50 mg versus sertraline 50-100 mg in patients with dysthymia or double depression: a randomized, double-blind, parallel group study. *International Clinical Psychopharmacology* 2001;**16**:317-24.

Arends 2014 *{published data only}*

Arends I, Klink JJ, Rhenen W, Boer MR, Bültmann U. Prevention of recurrent sickness absence in workers with common mental disorders: results of a cluster-randomised controlled trial. *Occupational and environmental medicine* 2014;**71**(1):21-29. [DOI: 10.1136/oemed-2013-101412]

Bakker 2007 *{published data only}*

Bakker IM, Terluin B, van Marwijk HWJ, Windt DAWM, van der Rijmen F, van Mechelen W, et al. A cluster-randomized trial evaluating an intervention for patients with stress-related mental disorders and sick leave in primary care. *PLOS Clinical Trials* 2007;**26**:1-9.

Barbui 2009 *{published data only}*

Barbui C, Ostuzzi G, Cipriani A. SSRI plus supportive care more effective than supportive care alone for mild to moderate depression. *Evidence Based Mental Health* 2009;**12**(4):109.

Bech 2000 *{published data only}*

Bech P, Cialdella P, Haugh MC, Birkett MA, Hours A, Boissel JP, et al. Meta-analysis of randomized controlled trials of fluoxetine v. placebo and tricyclic antidepressants in the short-term treatment of major depression. *British Journal of Psychiatry* 2000;**176**:421-8.

Becker 1998 *{published data only}*

Becker DR, Drake RE, Bond GR, Xie H, Dain BJ, Harrison K. Job terminations among persons with severe mental illness. *Community Mental Health Journal* 1998;**34**:71-82.

Bejerholm 2015 *{published data only}*

Bejerholm U, Areberg C, Hofgren C, Sandlund M, Rinaldi M. Individual placement and support in Sweden-a randomized controlled trial. *Nordic Journal of Psychiatry* 2015;**69**(1):57-66. [DOI: 10.3109/08039488.2014.929739]

Bejerholm 2017 *{published data only}*

Bejerholm U, Larsson ME, Johanson S. Supported employment adapted for people with affective disorders-A randomized controlled trial. *Journal of Affective Disorders* 2017;**207**:212-220. [DOI: http://dx.doi.org/10.1016/j.jad.2016.08.028]

Beurden 2013 *{published data only}*

Beurden KM, Brouwers EP, Joosen MC, Terluin B, Klink JJ, Weeghel J. Effectiveness of guideline-based care by occupational physicians on the return-to-work of workers with common mental disorders: design of a cluster-randomized controlled trial. *BMC Public Health* 2013;**13**:1-8.

Blonk 2007 *{published data only}*

Blonk RWB, Brenninkmeijer V, Lagerveld SE, Houtman ILD. Return to work: a comparison of two cognitive behavioural interventions in cases of work-related psychological complaints among the self-employed. *Work & Stress* 2006;**20**:129-44.

Boyer 1998 *{published data only}*

Boyer P, Danion JM, Bisserbe JC, Hotton JM, Troy S. Clinical and economic comparison of sertraline and fluoxetine in the treatment of depression. A 6-month double-blind study in a primary-care setting in France. *Pharmacoeconomics* 1998;**13**(1 Pt 2):157-69.

Brandes 2011 *{published data only}*

Brandes VMT. Efficacy of auditory stimulation programs for the treatment of depression, dysthymia and symptoms of burnout - RCT results. In: European Psychiatry conference. 2011.

Brouwers 2007 *{published data only}*

Brouwers EPM, Bruijne MC de, Terluin B, Tiemens BG, Verhaak PFM. Cost-effectiveness of an activating intervention by social workers for patients with minor mental disorders on sick leave: a randomized controlled trial. *The European Journal of Public Health* 2007;**17**(2):214-20.

Carlin 2010 *{published data only}*

Carlin E. The effect of a motivational interviewing style in cognitive therapy for depression [Dissertation]. http://gradworks.umi.com/34/53/3453319.html 2010.

Castillo-Pérez 2010 *{published data only}*

Castillo-Pérez S, Gómez-Pérez V, Calvillo Velasco M, Pérez-Campos E, Mayoral M. Effects of music therapy on depression

compared with psychotherapy. *The Arts in Psychotherapy* 2010;**37**:387-90.

Dalgaard 2014 *{published data only}*

Dalgaard L, Eskildsen A, Carstensen O, Willert MV, Andersen JH, Glasscock DJ. Changes in self-reported sleep and cognitive failures: a randomized controlled trial of a stress management intervention. *Scandinavian Journal of Work, Environment & Health* 2014;**40**(6):569-581. [DOI: 10.5271/sjweh.3460]

Dalgaard 2017 *{published data only}*

Dalgaard VL, Andersen LPS, Andersen JH, Willert MV, Carstensen O, Glasscock DJ. Work-focused cognitive behavioral intervention for psychological complaints in patients on sick leave due to work-related stress: results from a randomized controlled trial. *Journal of Negative Results in Biomedicine* 2017;**16**(1):13. [DOI: 10.1186/s12952-017-0078-z]

Dalgaard 2017a *{published data only}*

Dalgaard VL, Aschbacher K, Andersen JH, Glasscock DJ, Willert MV, Carstensen O, Biering K. Return to work after work-related stress: a randomized controlled trial of a work-focused cognitive behavioral intervention. *Scandinavian Journal of Work, Environment and Health, Supplement* 2017;**43**(5):436-446. [DOI: 10.5271/sjweh.3655]

Danielsson 2019 *{published data only}*

Danielsson L, Waern M, Hensing G, Holmgren K. Work-directed rehabilitation or physical activity to support work ability and mental health in common mental disorders: a pilot randomized controlled trial. *Clinical Rehabilitation* 2020 (2019 epub);**34**(2):179-181.

Dean 2014 *{published data only}*

Dean OM, Maes M, Ashton M, Berk L, Kanchanatawan B, Sughondhabirom A, et al. Protocol and rationale-the efficacy of minocycline as an adjunctive treatment for major depressive disorder: A double blind, randomised, placebo controlled trial. *Clinical Psychopharmacology and Neuroscience* 2014;**12**(3):180-188. [DOI: http://dx.doi.org/10.9758/cpn.2014.12.3.180]

Dean 2017 *{published data only}*

Dean OM, Kanchanatawan B, Ashton M, Mohebbi M, Ng CH, Maes M, Berk L, et al. Adjunctive minocycline treatment for major depressive disorder: A proof of concept trial. *Australian & New Zealand Journal of Psychiatry* 2017;**51**(8):829-840. [DOI: 10.1177/0004867417709357]

deVries 2015 *{published data only}*

de Vries G, Schene AH. International handbook of occupational therapy interventions, 2nd ed.. Amsterdam: Adult Psychiatry, Other departments, 2015. [PERSITENT IDENTIFIER: https://www.persistent-identifier.nl/urn:nbn:nl:ui:29-dbda4ac0-0cfa-48f4-9ee3-47f5d070bdfd]

Dick 1985 *{published data only}*

Dick P, Cameron L, Cohen D, Barlow M. Day and full time psychiatric treatment. *British Journal of Psychiatry* 1985;**147**:246-9.

Dunlop 2011 *{published data only}*

Dunlop BW, Reddy S, Yang L, Lubaczewski S, Focht K, Guico-Pabia CJ. Symptomatic and functional improvement in employed depressed patients. *Journal of Clinical Psychopharmacology* 2011;**31**:569-76.

Ebert 2014 *{published data only}*

Ebert DD, Lehr D, Boß L, Riper H, Cuijpers P, Andersson G, et al. Efficacy of an internet-based problem-solving training for teachers: Results of a randomized controlled trial. *Scandinavian Journal of Work, Environment & Health* 2014;**40**(6):582-596. [DOI: http://dx.doi.org/10.5271/sjweh.3449]

Ebert 2014a *{published data only}*

Ebert DD, Lehr D, Smit F, Zarski AC, Riper H, Heber E, et al. Efficacy and cost-effectiveness of minimal guided and unguided internet-based mobile supported stress-management in employees with occupational stress: a three-armed randomised controlled trial. *BMC Public Health* 2014;**14**:807. [DOI: 10.1186/1471-2458-14-807]

Eisendrath 2014 *{published data only}*

Eisendrath SJ, Gillung EP, Delucchi KL, Chartier M, Mathalon DH, Sullivan JC, et al. Mindfulness-based cognitive therapy (MBCT) versus the health-enhancement program (HEP) for adults with treatment-resistant depression: a randomized control trial study protocol. *BMC Complementary and Alternative Medicine* 2014;**14**:95. [DOI: https://dx.doi.org/10.1186/1472-6882-14-95]

Eklund 2012 *{published data only}*

Eklund M, Wästberg BA, Erlandsson L. Work outcomes and their predictors in the Redesigning Daily Occupations (ReDO) rehabilitation programme for women with stress-related disorders. *Australian Occupational Therapy Journal* 2013;**60**:85-92.

Endicott 2014 *{published data only}*

Endicott J, Lam RW, Hsu MA, Fayyad R, Boucher M, Guico-Pabia CJ. Improvements in quality of life with desvenlafaxine 50 mg/d vs placebo in employed adults with major depressive disorder. *Journal of Affective Disorders* 2014;**166**:307-314. [DOI: 10.1016/j.jad.2014.05.011]

Erkkilä 2011 *{published data only}*

Erkkilä J, Punkanen M, Fachner J, Ala-Ruona E, Pöntiö I, Tervaniemi M, et al. Individual music therapy for depression: randomized controlled trial. *British Journal of Psychiatry* 2011;**199**:132-39.

Evans 2016 *{published data only}*

Evans VC, Alamian G, McLeod J, Woo C, Yatham LN, Lam RW. The Effects of Newer Antidepressants on Occupational Impairment in Major Depressive Disorder: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *CNS drugs* 2016;**30**(5):405-17. [DOI: https://dx.doi.org/10.1007/s40263-016-0334-7]

Finley 2003 *{published data only}*

Finley PR, Rens HR, Pont JT, Gess SL, Louie CL, Bull SA, et al. Impact of a collaborative care model on depression in a primary

care setting: a randomized controlled trial. *Pharmocotherapy* 2003;**23**:1175-85.

Folke 2012 *{published data only}*

Folke F, Parling T, Melin L. Acceptance and commitment therapy for depression: a preliminary randomized clinical trial for unemployed on long-term sick leave. *Cognitive and Behavioural Practice* 2012;**19**:583-94.

Forman 2012 *{published data only}*

Forman EM, Shaw JA, Goetter EM, Herbert JD, Park JA, Yuen EK. Long-term follow up of a randomized controlled trial comparing acceptance and commitment therapy and standard cognitive behaviour therapy for anxiety and depression. *Behavior Therapy* 2012;**43**:801-11.

Fournier 2015 *{published data only}*

Fournier JC, DeRubeis RJ, Amsterdam J, Shelton RC, Hollon SD. Gains in employment status following antidepressant medication or cognitive therapy for depression. *The British journal of Psychiatry : the Journal of Mental Science* 2015;**206**(4):332-8. [DOI: https://dx.doi.org/10.1192/bjp.bp.113.133694]

Furukawa 2012 *{published data only}*

Furukawa TA, Horikoshi M, Kawakami N, Kadota M, Sasaki M, Sekiya Y, et al. Telephone cognitive-behavioural therapy for subthreshold depression and presenteeism in workplace: a randomized controlled trial. *PLoS One* 2012;**7**:1-9.

Gournay 1995 *{published data only}*

Gournay K, Brooking J. The community psychiatric nurse in primary care: an economic analysis. *Journal of Advanced Nursing* 1995;**22**(4):769-78.

Gunnarson 2018 *{published data only}*

Gunnarsson AB, Wagman P, Hedin K, Håkansson C. Treatment of depression and/or anxiety - outcomes of a randomised controlled trial of the tree theme method® versus regular occupational therapy. *BMC Psychology* 2018;**6**(1):25.

Hackett 1987 *{published data only}*

Hackett GI, Boddie HG, Harrison P. Chronic muscle contraction: the importance of depression and anxiety. *Journal of the Royal Society of Medicine* 1987;**80**:689-91.

Han 2015 *{published data only}*

Han C, Wang SM, Kwak KP, Won WY, Lee H, Chang CM, et al. Aripiprazole augmentation versus antidepressant switching for patients with major depressive disorder: A 6-week, randomized, rater-blinded,prospective study. *Journal of Psychiatric Research* 2015;**66-67**:84-94. [DOI: 10.1016/j.jpsychires.2015.04.020]

Heer 2013 *{published data only}*

Heer EW de, Dekker J, Van Eck-Sluijs JF van der, Beekman ATF, Marwijk HWJ van, Holwerda TJ, et al. Effectiveness and cost-effectiveness of transmural collaborative care with consultation letter (TCCCL) and duloxetine for major depressive disorder (MDD) and (sub)chronic pain in collaboration with primary care: design of a randomized placebo-controlled muli-Centre trial: TCC: PAINDIP. *BMC Psychiatry* 2013;**13**:1-14.

Hirani 2010 *{published data only}*

Hirani SS, Karmaliani R, McFarlane J, Asad N, Madhani F, Shehzad S. Testing a community derived intervention to promote women's health: preliminary results of a 3-arm randomized controlled trial in Karachi, Pakistan. *Southern Online Journal of Nursing Research* 2010;**10**(3):n.a.

Hobart 2019 *{published data only}*

Hobart M, Zhang P, Weiss C, Meehan S R, Eriksson H. Adjunctive brexpiprazole and functioning in major depressive disorder: a pooled analysis of six randomized studies using the sheehan disability scale. *Int J Neuropsychopharmcol* 2019;**22**(3):173-9.

Hollon 2016 *{published data only}*

Hollon SD, DeRubeis RJ, Fawcett J, Amsterdam JD, Shelton RC, Zajecka J, et al. Notice of Retraction and Replacement. Hollon et al. Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. JAMA Psychiatry. 2014;71(10):1157-1164. *JAMA Psychiatry* 2016;**73**(6):639-640. [DOI: 10.1001/jamapsychiatry.2016.0756]

Hordern 1964 *{published data only}*

Hordern A, Burt CG, Gordon WF, Holt NF. Amitriptyline in depressive states: Six-month treatment results. *British Journal of Psychiatry* 1964;**110**:641-7.

Jansson 2015 *{published data only}*

Jansson I, Gunnarsson AB, Björklund A, Brudin L, Perseius KI. Problem-based self-care groups versus cognitive behavioural therapy for persons on sick leave due to common mental disorders: a randomised controlled study. *Journal of Occupational Rehabilitation* 2015;**25**(1):127-140. [DOI: 10.1007/s10926-014-9530-9]

Johansson 2019 *{published data only}*

Johansson O, Bjarehed J, Andersson G, Carlbring P, Lundh LG. Effectiveness of guided internet-delivered cognitive behavior therapy for depression in routine psychiatry: A randomized controlled trial. *Internet Interventions* 2019;**17 (no pagination)**(100247):2019.100247.

Kennedy 2016 *{published data only}*

Kennedy SH, Avedisova A, Belaïdi C, Picarel-Blanchot F, de Bodinat C. Sustained efficacy of agomelatine 10 mg, 25 mg, and 25-50 mg on depressive symptoms and functional outcomes in patients with major depressive disorder. A placebo-controlled study over 6 months. *European neuropsychopharmacology* 2016;**26**(2):378-389. [DOI: 10.1016/j.euroneuro.2015.09.006]

Kennedy 2019 *{published data only}*

Kennedy S H, Heun R, Avedisova A S, Ahokas A, Olivier V, Picarel-Blanchot F, et al. Functional outcome in patients with major depressive disorder treated by agomelatine 25-50 mg as compared to placebo. *European Neuropsychopharmacology* 2019;**29 (Supplement 1)**:S233-4.

Knekt 2011 *{published data only}*

Knekt P, Lindfors O, Laaksonen MA, Renlund C, Haaramo P, Härkänen, et al. Quasi-experimental study on the effectiveness

of psychoanalysis, long-term and short-term psychotherapy on psychiatric symptoms, work ability and functional capacity during a 5-year follow-up. *Journal of Affective Disorders* 2011;**132**:37-47.

Knekt 2016 *{published data only}*

Knekt P, Virtala E, Härkänen T, Vaarama M, Lehtonen J, Lindfors O. The outcome of short- and long-term psychotherapy 10 years after start of treatment. *Psychological Medicine* 2016;**46**(6):1175-88. [DOI: https://dx.doi.org/10.1017/S0033291715002718]

Kojima 2010 *{published data only}*

Kojima R, Fujisawa D, Tajima M, Shibaoka M, Kakinuma M, Shima S, et al. Efficacy of cognitive behavioral therapy training using brief e-mail sessions in the workplace: a controlled clinical trial. *Industrial Health* 2010;**48**:495-502.

Kooistra 2014 *{published data only}*

Kooistra LC, Wiersma JE, Ruwaard J, van Oppen P, Smit F, Lokkerbol J, et al. Blended vs. Face-to-face cognitive behavioural treatment for major depression in specialized mental health care: Study protocol of a randomized controlled cost-effectiveness trial. *BMC Psychiatry* 2014;**14**(1):1-11. [DOI: 10.1186/s12888-014-0290-z]

Kroenke 2001 *{published data only}*

Kroenke K, West SL, Swindle R, Gilsenan A, Eckert GJ, Dolor R, et al. Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care - a randomized trial. *JAMA* 2001;**286**:2947-55.

Kuhs 1996 *{published data only}*

Kuhs H, Färber D, Borgstädt S, Mrosek S, Tölle R. Amitriptyline in combination with repeated late deep deprivation versus amitriptyline alone in major depression. A randomized study. *Journal of Affective Disorders* 1996;**37**:31-41.

Lagerveld 2012 *{published data only}*

Lagerveld SE, Blonk RWB, Brenninkmeijer V, Wijngaards-de Meij L, Schaufeli WB. Work-focused treatment of common mental disorders and return to work: a comparative outcome study. *Journal of Occupational Psychology* 2012;**17**:220-34.

Lam 2012 *{published data only}*

Lam RW, McLeod J, Woo C, Michalak EE, Yatham LN, Bond DJ. A systematic review of the treatment effects of antidepressants on occupational functioning. In: Presented at the International Society for Affective Disorders. UK, 2012.

Lexis 2011 *{published data only}*

Lexis MAS, Jansen NWH, Huibers MJH, Amelsvoort LGPM van, Berkouwer A, Tjin A, Ton G, et al. Prevention of long-term sickness absence and major depression in high-risk employees: a randomised controlled trial. *Occupational and Environmental Medicine* 2011;**68**:400-7.

Löbner 2018 *{published data only}*

Löbner M, Pabst A, Stein J, Dorow M, Matschinger H, Luppa M, et al. Computerized cognitive behavior therapy for patients with mild to moderately severe depression in primary care: a pragmatic cluster randomized controlled trial (@ktiv). *Journal of Affective Disorders* 2018;**238**:317-326.

Maljanen 2016 *{published data only}*

Maljanen T, Knekt P, Lindfors O, Virtala E, Tillman P, Härkänen T. The cost-effectiveness of short-term and long-term psychotherapy in the treatment of depressive and anxiety disorders during a 5-year follow-up. *Journal of Affective Disorders* 2016;**190**:254-263. [DOI: 10.1016/j.jad.2015.09.065]

Martinez 2011 *{published data only}*

Martinez JM, Katon W, Greist JH, Kroenke K, Thase ME, Meyers AL, et al. A pragmatic 12-week, randomized trial of duloxetine versus generic selective serotonin-reuptake inhibitors in the treatment of adult outpatients in a moderate-to-severe depressive episode. *International Clinical Psychopharmacology* 2012;**27**:17-26.

Meyer 2009 *{published data only}*

Meyer B, Berger T, Caspar F, Beevers CG, Andersson G, Weiss M. Effectiveness of a novel integrative online treatment for depression (Deprexis): randomized controlled trial. *Journal of Medical Internet Research* 2009;**11**:e15.

Mino 2006 *{published data only}*

Mino Y, Babazono A, Tsuda T, Yasuda N. Can stress management at the workplace prevent depression? A randomized controlled trial. *Psychotherapy and Psychosomatics* 2006;**75**:177-82.

Morgan 2011 *{published data only}*

Morgan AJ, Form AF, Mackinnon AJ. Protocol for a randomized controlled trial investigating self-help email messages for sub-treshold depression: the Mood Memos study. *Trials* 2011;**12**:1-9.

Mundt 2001 *{published data only}*

Mundt JC, Clarke GN, Burroughs D, Brenneman DO, Griest JH. Effectiveness of antidepressant pharmacotherapy: the impact of medication compliance and patient education. *Depression and Anxiety* 2001;**13**:1-10.

Oakes 2012 *{published data only}*

Oakes TMM, Meyers AL, Marangell LB, Ahl J, Prakash A, Thase ME, Kornstein SG. Assessment of depressive symptoms and functional outcomes in patients with major depressive disorder treated with duloxetine versus placebo. *Human Psychopharmacology: Clinical and Experimental* 2012;**27**:47-56.

Reavley 2018 *{published data only}*

Reavley NJ, Morgan AJ, Fischer JA, Kitchener B, Bovopoulos N, Jorm AF. Effectiveness of eLearning and blended modes of delivery of Mental Health First Aid training in the workplace: randomised controlled trial. *BMC Psychiatry* 2018;**18**(1):312.

Salminen 2008 *{published data only}*

Salminen JK, Karlsson H, Hietala J, Kajander J, Aalto S, Markkula J, et al. Short-term psychodynamic psychotherapy and fluoxetine in major depressive disorder: a randomized comparative study. *Psychotherapy and Psychosomatics* 2008;**77**:351-7.

Saloheimo 2016 *{published data only}*

Saloheimo HP, Markowitz J, Saloheimo TH, Laitinen JJ, Sundell J, Huttunen MO, et al. Psychotherapy effectiveness for major depression: a randomized trial in a Finnish community. *BMC Psychiatry* 2016;**16**:131. [DOI: 10.1186/s12888-016-0838-1]

Salomonsson 2017 *{published data only}*

Salomonsson S, Santoft F, Lindsäter E, Ejeby K, Ljótsson B, Öst LG, et al. Cognitive-behavioural therapy and return-to-work intervention for patients on sick leave due to common mental disorders: A randomised controlled trial. *Occupational and Environmental Medicine* 2017;**74**(12):905-912. [DOI: 10.1136/oemed-2017-104342]

Sandahl 2011 *{published data only}*

Sandahl C, Lundberg U, Lindgren A, Rylander G, Herlofsen J, Nygren ÅN. Two forms of group therapy and individual treatment of work-related depression: a one-year follow-up study. *International Journal of Group Psychotherapy* 2011;**61**:539-55.

Schmitt 2008 *{published data only}*

Schmitt AB, Volz HP, Wiedemann K, Fritze J, Loeschmann PA. Therapy effects on working ability by venlafaxine compared to serotonin reuptake inhibitors in the treatment of major depressive disorder of different severity. *Gesundheitsökonomie & Qualitätsmanagement* 2008;**13**:269-75.

Schoenbaum 2002 *{published data only}*

Schoenbaum M, Unutzer J, McCaffrey D, Duan N, Sherbourne C, Wells KB. The effects of primary care depression treatment on patients' clinical status and employment. *Health Services Research* 2002;**37**(5):1145-58.

Shawyer 2016 *{published data only}*

Shawyer F, Enticott JC, Özmen M, Inder B, Meadows GN. Mindfulness-based cognitive therapy for recurrent major depression: A 'best buy' for health care? *The Australian and New Zealand Journal of Psychiatry* 2016;**50**(10):1001-13. [DOI: https://dx.doi.org/10.1177/0004867416642847]

Simon 2000 *{published data only}*

Simon GE, Revicki D, Heiligenstein J, Grothaus L, Von Korff M, Katon WJ, et al. Recovery from depression, work productivity, and health care costs among primary care patients. *General Hospital Psychiatry* 2000;**22**(3):153-62.

Sir 2005 *{published data only}*

Sir A, D'Souza RF, Uguz S, George T, Psych MRC, Vahip S, et al. Randomized trial of sertraline versus venlafaxine XR in major depression: efficacy and discontinuation symptoms. *Journal of Clinical Psychiatry* 2005;**66**:1312-20.

Soares 2019 *{published data only}*

Soares CN, Zhang M, Boucher M. Categorical improvement in functional impairment in depressed patients treated with desvenlafaxine. *CNS Spectrums* 2019;**24**(3):322-32.

Stant 2009 *{published data only}*

Stant AD, TenVergert EM, Kluiter H, Conradi HJ, Smit A, Ormel J. Cost-effectiveness of a psychoeducational relapse prevention program for depression in primary care. *The Journal of Mental Health Policy and Economics* 2009;**12**:195-204.

Twamley 2019 *{published data only}*

Twamley EW, Thomas KR, Burton CZ, Vella L, Jeste DV, Heaton RK, et al. Compensatory cognitive training for people with severe mental illnesses in supported employment: A randomized controlled trial. *Schizophrenia Research* 2019;**203**:41-8.

Warmerdam 2007 *{published data only}*

Warmerdam L, Straten A van, Cuijpers P. Internet-based treatment for adults with depressive symptoms. *BMC Psychiatry* 2007;**7**:1-7.

Wells 2000 *{published data only}*

Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unutzer, J et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA* 2000;**283**(2):212-20.

Winter 2015 *{published data only}*

Winter L, Kraft J, Boss K, Kahl KG. [Return to Work: A Workplace Focused Module to be Integrated in Cognitive Behavioral Therapy]. *Ruckkehr ins Erwerbsleben: Ein Arbeitsplatz-bezogenes Modul zur Integration in die kognitiv-behaviorale Therapie bei psychischen Erkrankungen.* 2015;**65**(8):321-6. [DOI: https://dx.doi.org/10.1055/s-0035-1545312]

Wisenthal 2018 *{published data only}*

Wisenthal A, Krupa T, Kirsh BH, Lysaght R. Cognitive work hardening for return to work following depression: An intervention study. *Canadian Journal of Occupational Therapy / Revue Canadienne D'Ergothérapie* 2018;**85**(1):21-32.

Zambori 2002 *{published data only}*

Zambori J, Szadoczky E, Rozsa S, Furedi J. Cost-outcome of anxiety treatment intervention in primary care in Hungary. *The Journal of Mental Health Policy and Economics* 2002;**5**(3):115-20.

Zeeuw 2010 *{published data only}*

Zeeuw ELEJ de, Tak ECPM, Dusseldorp E, Hendriksen IJM. Workplace exercise intervention to prevent depression: a pilot randomized controlled trial. *Mental Health and Physical Activity* 2010;**3**:72-7.

Zwerenz 2015 *{published data only}*

Zwerenz R, Becker J, Knickenberg RJ, Hagen K, Dreier M, Wölfling K, et al. Enhancing inpatient psychotherapeutic treatment with online self-help: Study protocol for a randomized controlled trial. *Trials* 2015;**16**:98. [DOI: 10.1186/s13063-015-0620-6]

Zwerenz 2017 *{published data only}*

Zwerenz R, Becker J, Knickenberg RJ, Siepmann M, Hagen K, Beutel ME. Online Self-Help as an Add-On to Inpatient Psychotherapy: Efficacy of a New Blended Treatment Approach. *Psychotherapy and Psychosomatics* 2017;**86**(6):341-350. [DOI: 10.1159/000481177]

Zwerenz 2017a *{published data only}*

Zwerenz R, Becker J, Gerzymisch K, Siepmann M, Holme M, Kiwus U, et al. Evaluation of a transdiagnostic psychodynamic online intervention to support return to work: a randomized controlled trial. *PloS One* 2017;**12**(5):e0176513.. [DOI: 10.1371/journal.pone.0176513]

References to ongoing studies

Deady 2018 *{published data only}*

Deady M, Johnston D A, Glozier N, Milne D, Choi I, Mackinnon A, et al. A smartphone application for treating depressive symptoms: Study protocol for a randomised controlled trial. *BMC Psychiatry* 2018;**18**(1):166.

Imamura 2018 *{published data only}*

Imamura K, Tran TTT, Nguyen HT, Kuribayashi K, Sakuraya A, Nguyen AQ, et al. Effects of two types of smartphone-based stress management programs on depression and anxiety among hospital nurses in Vietnam: a protocol for three-arm randomized controlled trial. *BMJ Open* 2019 (2018 epub);**9**(4):e025138.

Kouvonen 2019 *{published data only}*

Kouvonen A, Manty M, Harkko J, Sumanen H, Konttinen H, Lahti J, et al. Effectiveness of internet-delivered cognitive behavioural therapy in reducing sickness absence among young employees with depressive symptoms: Study protocol for a large-scale pragmatic randomised controlled trial. *BMJ Open* 2019;**9**(10):e032119.

Poulsen 2017 *{published data only}*

Poulsen R, Hoff A, Fisker J, Hjorthøj C, Eplov LF. Integrated mental health care and vocational rehabilitation to improve return to work rates for people on sick leave because of depression and anxiety (the Danish IBBIS trial): study protocol for a randomized controlled trial. *Trials* 2017;**18**:578.

Additional references

Abma 2012

Abma FI, Amick BC 3rd, Brouwer S, van der Klink JJ, Bültmann U. The cross-cultural adaptation of the work role functioning questionnaire to Dutch.. *Work* 2012;**43**(2):203-210. [DOI: doi:10.3233/WOR-2012-1362]

Adler 2006

Adler DA, McLaughlin TJ, Rogers WH, Chang H, Lapitsky L, Lerner D. Job performance deficits due to depression. *American Journal of Psychiatry* 2006;**163**(9):1569-76.

APA 1994

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). 4th edition. Washington, DC: American Psychiatric Association, 1994.

APA 2010

American Psychiatric Association (APA). Practice guideline for the treatment of patients with major depressive disorder. 3rd edition. American Journal of Psychiatry 2010:152.

APA 2013

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Arlington, VA: American Psychiatric Association, 2013.

Beck 1987

Beck AT, Steer RA. BDI: Beck Depression Inventory Manual. New York, NY: Psychological Corporation, 1987.

Bowling 1995

Bowling A. What things are important in people's lives? A survey of the public's judgements to inform scales of health related quality of life. *Social Science and Medicine* 1995;**41**(10):1447-62.

Brown 2013

Brown MM, Brown GC. Update on value-based medicine. *Current Opinion in Ophthalmology* 2013;**24**(3):183-9.

Burton 2004

Burton WN, Pransky G, Conti DJ, Chen CY, Edington DW. The association of medical conditions and presenteeism. *Journal of Occupational and Environmental Medicine* 2004;**46**(6 Suppl):S38-S45.

Cooney 2013

Cooney GM, Dwan K, Greig CA, Lawlor DA, Rimer J, Waugh FR, et al. Exercise for depression. *Cochrane Database of Systematic Reviews* 2013;**Sep 12;9:CD004366**:doi: 10.1002/14651858.CD004366.pub6.

de Vries 2015

de Vries G, Koeter MW, Nieuwenhuijsen K, Hees HL, Schene AH. Predictors of impaired work functioning in employees with major depression in remission. *Journal of Affective Disorders* 2015;**185**:180-7.

Dewa 2000

Dewa CS, Lin E. Chronic physical illness, psychiatric disorder and disability in the workplace. *Social Science and Medicine* 2000;**51**(1):41-50.

Donner 2002

Donner A, Klar N. Issues in the meta-analysis of cluster randomized trials. *Statistics in Medicine* 2002;**21**:2971-80.

Endicott 1997

Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. *Psychopharmacology Bulletin* 1997;**33**(1):13-6.

Ervasti 2017

Ervasti J, Joensuu M, Pentti J, Oksanen T, Ahola K, Vahtera J, et al. Prognostic factors for return to work after depression-related work disability: A systematic review and meta-analysis. *Journal of psychiatric research* 2017;**95**:28–36.

Evans-Lacko 2016

Evans-Lacko S, Knapp M. Global patterns of workplace productivity for people with depression: absenteeism and presenteeism costs across eight diverse countries. *Social Psychiatry and Psychiatric Epidemiology* 2016;**51**(11):1525-1537.

Ferrari 2013

Ferrari AJ, Charlson FJ, Norman RE, Flaxman AD, Patten SB, Vos T, et al. The epidemiological modelling of major depressive disorder: application for the Global Burden of Disease Study 2010. *PLoS One* 2013;**8**(7):e69637. doi: 10.1371/journal.pone.0069637.

Ferrie 2005

Ferrie JE, Kivimaki M, Head J, Shipley MJ, Vahtera J, Marmot MG. A comparison of self-reported sickness absence with absences recorded in employers' registers: evidence from the Whitehall II study. *Occupational and Environmental Medicine* 2005;**62**(2):74-9.

Freemantle 2001

Freemantle N. Interpreting the results of secondary end points and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? *BMJ* 2001;**322**(7292):989-91.

Fried 2017

Fried EI, van Borkulo CD, Cramer AO, Boschloo L, Schoevers RA, Borsboom D. Mental disorders as networks of problems: a review of recent insights. *Soc Psychiatry Psychiatr Epidemiol.* 2017;**52**(1):1-10. [DOI: doi:10.1007/s00127-016-1319-z]

Furlan 2012

Furlan AD, Gnam WH, Carnide N, Irvin E, Amick BC 3rd, DeRango K, et al. Systematic review of intervention practices for depression in the workplace. *Journal of Occupational Rehabilitation* 2012;**22**(3):312-21.

GRADEpro 2008 [Computer program]

GRADE working group GRADEpro. Brozek J, Oxman A, Schünemann H, Version 3.2 for Windows. GRADE working group, 2008.

Greenberg 2015

Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry.* 2015;**76**(2):155-62.

Haafkens 2006

Haafkens J, Moerman C, Schuring M, van Dijk F. Searching bibliographic databases for literature on chronic disease and work participation. *Occupational Medicine* 2006;**56**(1):39-45.

Hamilton 1967

Hamilton M. Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology* 1967;**6**(4):278-96.

Haslam 2005

Haslam C, Atkinson S, Brown S, Haslam RA. Perceptions of the impact of depression and anxiety and the medication for these conditions on safety in the workplace. *Occupational and Environmental Medicine* 2005;**62**(8):538-45.

Hees 2013b

Hees HL, Koeter MW, Schene AH. Longitudinal relationship between depressive symptoms and work outcomes in clinically treated patients with long-term sickness absence related to major depressive disorder. *Journal of Affective Disorders* 2013;**148**(2-3):272-7.

Higgins 2011

Higgins JPT, Green S (editors) The Cochrane Collaboration, 2011. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0, 2011. Available from www.cochrane-handbook.org, 2011.

Hirschfeld 2000

Hirschfeld RM, Montgomery SA, Keller MB, Kasper S, Schatzberg AF, Moller HJ, et al. Social functioning in depression: a review. *Journal of Clinical Psychiatry* 2000;**61**(4):268-75.

Ilmarinen 2005

Ilmarinen J, Tuomi K, Seitsamo J (eds). New dimensions of work ability.. *International congress series* 2005;**1280**:3-7.

Johns 2015

Johns G, Miraglia M. The reliability, validity, and accuracy of self-reported absenteeism from work: a meta-analysis. *Journal of Occupational Health Psychology* 2015;**20**(1):1-14..

Kessler 2006

Kessler RC, Akiskal HS, Ames M, Birnbaum H, Greenberg P, Hirschfeld RM, et al. Prevalence and effects of mood disorders on work performance in a nationally representative sample of U.S. workers. *American Journal of Psychiatry* 2006;**163**(9):1561-8.

Lagerveld 2010

Lagerveld SE, Bültmann U, Franche RL, van Dijk FJ, Vlasveld MC, van der Feltz-Cornelis CM, et al. Factors associated with work participation and work functioning in depressed workers: a systematic review. *Journal of Occupational Rehabilitation* 2010;**20**(3):275-92.

Lee 2018

Lee Y, Rosenblat JD, Lee J, Carmona NE, Subramaniapillai M, Shekotikhina M, et al. Efficacy of antidepressants on measures of workplace functioning in major depressive disorder: A systematic review. *Journal of Affective Disorders* 2018;**227**:406–415.

Lerner 2008

Lerner D, Henke RM. What does research tell us about depression, job performance, and work productivity? *Journal of Occupational and Environmental Medicine* 2008;**50**(4):401-10.

Montgomery 1979

Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* 1979;**134**:382–9.

Murray 2012

Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**(9859):2197-223.

NICE 2010

National Collaborating Centre for Mental Health. http://www.nice.org.uk/nicemedia/live/12329/45896/45896.pdf. Leicester & London: The British Psychological Society and The Royal College of Psychiatrists, 2010.

Nieuwenhuijsen 2010

Nieuwenhuijsen K, Franche RL, van Dijk FJH. Work functioning measurement: tools for occupational mental health research. *Journal of Occupational and Environmental Medicine* 2010;**52**(8):778-90.

Pole 2006

Pole JD, Franche RL, Hogg-Johnson S, Vidmar M, Krause N. Duration of work disability: a comparison of self-report and administrative data. *American Journal of Industrial Medicine* 2006;**49**(5):394-401.

RevMan 2014 [Computer program]

The Nordic Cochrane Centre. The Cochrane Collaboration Review Manager (RevMan). Version 5.2. Copenhagen: The Nordic Cochrane Centre. The Cochrane Collaboration, 2012.

Severens 2000

Severens JL, Mulder J, Laheij RJ, Verbeek AL. Precision and accuracy in measuring absence from work as a basis for calculating productivity costs in The Netherlands. *Social Science and Medicine* 2000;**51**(2):243-9.

Sheehan 1996

Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. *International Clinical Psychopharmacology* 1996;**11 Suppl 3**:89-95.

Spitzer 1979

Spitzer RL, Endicott J, Williams JB. Research diagnostic criteria. *Archive of General Psychiatry* 1979;**36**(12):1381-3.

Suzuki 2004

Suzuki K, Ohida T, Kaneita Y, Yokoyama E, Miyake T, Harano S, et al. Mental health status, shift work, and occupational accidents among hospital nurses in Japan. *Journal of Occupational Health* 2004;**46**(6):448–54.

Thorsen 2018

Thorsen SV, Flyvholm MA, Bültmann U. Self-reported or register-based? A comparison of sickness absence data among 8110 public and private employees in Denmark. *Scandinavian Journal of Work, Environment & Health* 2018;**44**(6):631-638. [DOI: doi:10.5271/sjweh.3747]

van Poppel 2002

van Poppel MN, de Vet HC, Koes BW, Smid T, Bouter LM. Measuring sick leave: a comparison of self-reported data on sick leave and data from company records. *Occupational Medicine* 2002;**52**(8):485-90.

Verbeek 2005

Verbeek J, Salmi J, Pasternack I, Jauhiainen M, Laamanen I, Schaafsma F, et al. A search strategy for occupational health intervention studies. *Occupational and Environmental Medicine* 2005;**62**(10):682-7.

Vogel 2017

Vogel N, Schandelmaier S, Zumbrunn T, Ebrahim S, de Boer WE, Busse JW, et al. Return-to-work coordination programmes for improving return to work in workers on sick leave. *Cochrane Database Syst Rev.* 2017;**3**(3):doi:10.1002/14651858.CD011618.pub2.

Wang 2012

Wang JL, Patten SB, Currie S, Sareen J, Schmitz N. Predictors of 1-year outcomes of major depressive disorder among individuals with a lifetime diagnosis: a population-based study. *Psychological Medicine* 2012;**42**(2):327-34.

Wells 1989

Wells KB, Stewart A, Hays RD, Burnam MA, Rogers W, Daniels M, et al. The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. *JAMA* 1989;**262**(7):914-9.

WHO 1992

World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders. Geneva: World Health Organization, 1992.

References to other published versions of this review

Nieuwenhuijsen 2008

Nieuwenhuijsen K, Bültmann U, Neumeyer-Gromen A, Verhoeven AC, Verbeek JH, Feltz-Cornelis CM. Interventions to improve occupational health in depressed people. *Cochrane Database of Systematic Reviews* 2008, Issue 2. Art. No: CD006237. [DOI: 10.1002/14651858.CD006237.pub2]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Agosti 1991

Study characteristics	
Methods	**Study Design**:Double-blind randomised trial

Agosti 1991 *(Continued)*

	Number of trial arms: 4 (3 treatment and one placebo). **Study grouping:** Parallel group **Recruitment**: unclear. Follow up: 6 weeks. **Lost to follow up**: 29.5%
Participants	**Participants**: 61 were randomised (T1: 38, C: 23). **Inclusion criteria:** - DSM-III diagnosis of depressive disorder - mood reactivity (i.e. significant lifting of mood in response to positive environmental events) - onset prior to age 21 yrs - rated by experienced clinician to be depressed for most or virtually all of the time through adulthood **Baseline characteristics:** Mean age: 35 yrs (SD 8.9) Female: 52% Single: 57% Married: 23% Divorced or separated: 19.6% Working: 70% **Setting:** Outpatients in New York, USA
Interventions	T1: Treatment with increasing dose of either TCA or MAO - 60 to 90 mg/day of phenelzine (T1a) - 200 to 3000 mg/day of imipramine (T1b) - 40 mg/day of L-deprenyl (T1c) Duration: 6 weeks. C: 4 to 6 placebo pills/day. Duration: 6 weeks
Outcomes	Sickness absence: 1) hours worked in past week (baseline and at 6 weeks). From the LIFE scale. Depressive symptoms: 1) CGI (measured but not reported!) 2) HAM-D (measured but not reported!) Work functioning: 1) work functioning of the LIFE scale (psychosocial functioning part) (baseline and at 6 weeks)
Notes	Country: US

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Random sequence generation not reported "Following baseline evaluation, patients were treated with single-blind placebo for 1-2 weeks, those who were still depressed were randomly assigned to 6 weeks of treatment with increasing doses of one of four agents in a double blind design."

Agosti 1991 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	Allocation concealment not reported "Following baseline evaluation, patients were treated with single-blind placebo for 1-2 weeks, those who were still depressed were randomly assigned to 6 weeks of treatment with increasing doses of one of four agents in a double blind design."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	A double blind design was used "Following baseline evaluation, patients were treated with single-blind placebo for 1-2 weeks, those who were still depressed were randomly assigned to 6 weeks of treatment with increasing doses of one of four agents in a double blind design."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Patients reported sick leave in an interview, but were blinded to treatment allocation "Sick leave was assessed by the LIFE. The LIFE is a semi-structured interview which tracks episodes of psychiatric illness. The portion of the LIFE which we used assessed the psychosocial functioning during the week in five areas; employment..etc. The LIFE was administered to the patient by the treating physician."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	Depressive symptoms were determined by personnel, were blinded to treatment allocation "Clinical outcome was determined by the treating psychiatrist on the basis of Clinical Global Improvement."
Incomplete outcome data (attrition bias) Depressive symptoms	Unclear risk	Outcome not reported
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up is considered to be high: T1: 28.9%; T2: 30.4%, even though the proportion of incomplete data was comparable in both groups
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None Identified

Bee 2010

Study characteristics	
Methods	**Study design**: RCT **Recruitment:** over 10 months, human resources mailed all potential participants a study information pack. **Follow up:** 3 months. **Lost to follow up**: overall 40%, subgroup depressed workers: 0%
Participants	**Baseline:** 53 were randomised (T1: 26; T2: 27). Subgroup of depressed workers: 12. For the subgroup of depressed workers:

Bee 2010 *(Continued)*

	mean age: 50.9 (SD 10.04) male: 58% **Inclusion criteria**: employees of a large communications company absent from work with mild to moderate mental health difficulties for 8 to 90 days authorised by general practitioner certificate **Exclusion criteria:** severe or complex disorders (psychosis, comorbid personality disorder), degenerative cognitive disorders, substance misuse or active self-harm **Setting:** large communications company.
Interventions	T1: Telephone CBT, delivered over 12 weeks by one of two registered graduate mental health workers. Participants worked with therapists through regular phone calls to identify and challenge negative thoughts, develop self-care skills and complete workbook exercises emphasizing behavioural activation. Therapists received 12 h of didactic instruction and role play and weekly supervision from a senior CBT therapist. T2: Usual care, primary and occupational health services.
Outcomes	**Sickness absence** 1) self-reported actual working hours (HPQ) in last four weeks **Depressive symptoms** 1) depression, assessed by the HADS
Notes	Country: UK

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Personal communication: "Yes there was a random component in the sequence generation – and the sequence was held by an independent trial units."
Allocation concealment (selection bias)	Low risk	"Randomization was conducted centrally by an independent service, with minimization on age, gender and illness severity". "[...] internal validity was heightened trough allocation concealment via central randomisation [..]"
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Due to the nature of the intervention, the participants could not be blinded and the control condition (usual care) was deemed less desirable,
Blinding of outcome assessment (detection bias) Sick Leave	High risk	The actual working hours were assessed by the participants themselves. As they were aware of the allocation status, risk of detection bias is considered to be high
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression is assessed by the HADS, which is a self-reported instrument. As the participants were aware of their allocation status, risk of detection bias is considered to be high
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Personal communication: "For the subgroup of depressed workers, there is no loss to follow up."
Incomplete outcome data (attrition bias)	Low risk	Personal communication: "For the subgroup of depressed workers, there is no loss to follow up."

Bee 2010 *(Continued)*
Sick Leave

Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Beiwinkel 2017

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** The researchers recruited members from Kaufmännische Krankenkasse (KKH), a statutory health insurance company with about 1.8 million members nationwide. First, to identify participants who were at high risk for sick leave due to depression, insurance members were screened for previous diagnosis of depression (ICD codes F32.0, F32.1, F33.0, F33.1, andF34.1), previous sickness absence due to depression, and current sickness absence. Second, the study team sent an invitation letter to all positively screened insurance members along with study information, the informed consent form, and a 6-digit code to login into the platform. **Follow-up:** 12 and 24 weeks
Participants	**Baseline characteristics** Clinical Intervention: psychological: I-CBT, I-guided • *Number of participants randomised*: 100 • *Gender*: 34% male • *Marital status*: 24% single, 56% married/partner, 15% divorced/separated, 5% widowed • *Occupation*: 22% executive position • *Sick leave status*: 25.60 ± 2.03 in past 90 days and average sick leave was 25.60 ± 2.03 in past 90 days • *Age*: 47.01 ± 10.36 CAU-info: Waiting list plus psycho-education • *Number of participants randomised*: 80 • *Gender*: 29% male • *Marital status*: 22.5% single, 57.5% married/partner, 17.5% divorced/separated, 2.5% widowed • *Occupation*: 18% executive position • *Sick leave status*: 27.69 ± 2.37 in past 90 days and average sick leave was 27.69 ± 2.37 in past 90 days • *Age*: 48.66 ± 11.59 Overall • *Number of participants randomised*: 180 • *Gender*: 32% male, 68% female • *Marital status*: 23% single, 57% married/partner, 16% divorced/separated, 4% widowed • *Occupation*: 20% executive position • *Sick leave status*: not reported • *Age*:47.74

Beiwinkel 2017 *(Continued)*

Inclusion criteria: Insurance members in the insurers database that after screening were adults with a previous episode of mild to moderate depression(International Classification of Disease codes F32.0, F32.1,F33.0, F33.1) or dysthymia (F34.1)

Exclusion criteria: Participants with a score of ≥ 20 on the Patient Health Questionnaire (PHQ-9), indicating severe depression, were excluded. A second exclusion criterion was suicidality as measured by one item on the presence of suicidal thoughts

Pretreatment differences: PHQ-9, Gender, relationship status, education, employment, executive position, previous depression, depression medication, being in psychotherapy were tested for group differences, none of these were statistically significant. The authors further state " No clinically relevant differences in terms of any baseline characteristics were found, and we concluded that randomisation was successful."

Setting: Participants were recruited through a statutory health insurance company

Interventions

Intervention characteristics

Clinical Intervention: psychological: I-CBT, I-guided

- *Content*: "The intervention's psychological approach includes cognitive-behavioural therapy, mindfulness training, and systemic counselling. During the development process, current research evidence on the respective therapies was used as the basis, and special emphasis was placed on a "person-based" approach,focusing on the perspectives of the people who would use the intervention. Cognitive-behavioural therapy is the most extensively researched psychological treatment approach inWeb-based interventions. From cognitive-behavioural therapy, the intervention used elements of cognitive restructuring, with an emphasis on dealing with negative moods and automatic thoughts, as well as exercises for behavioural activation. Mindfulness training has been used increasingly in psychotherapy over the past years. It was shown to be effective for depressive symptoms and can be adapted to online formats. The intervention module on mindfulness engages the user in exercises to observe the self and to practice mindfulness in daily situations. Systemic counselling is a therapeutic approach that highlights the social context surrounding the individual and its resources. Specifically, systemic questioning technique sand instructions were employed to make use of the participants' social support. Systemic principles were presented in specific weekly sessions, while homework exercises on systemic therapy encouraged the participants to adopt a systemic viewpoint and behaviour change in their everyday interactions."
- *Duration, frequency, length*: 12 weeks, 12 times, 30-45 minutes each
- *Communication means*: Web-based, email, telephone
- *Providers*: Therapist contact upon request, that is, psychologists (bachelor level or higher) trained in the intervention approach provided feedback via email or telephone

CAU-info: Waiting list plus psychoeducation

- *Content*: Wait-list plus psycho-education condition. Participants had access to text-based information on the nature of depression and its symptoms and treatment. The psycho-education content was developed by a team of trained psychologists (bachelor degree or higher) and was based upon scientific literature on depression. This type of control condition was chosen because more active control groups are considered to be more methodologically valid. The control group did not have access to therapist guidance. Participants were eligible to access the intervention after study completion, if they requested access
- *Duration, frequency, length*: 12 weeks, supposed to use each week
- *Communication means*: Web-based
- *Providers*: Not applicable

Outcomes

Sickness absence

Days lost in 90 days following randomisation

- **Outcome type**: Continuous outcome

Depressive symptoms

Beiwinkel 2017 *(Continued)*

	BDI-II • **Outcome type**: Continuous outcome
Notes	**Country** *: Germany* **Outcomes** Absenteeism: Three sickness absence measures were constructed. First, the number of persons who were absent at least once, second, absence frequency as the number of times a person was absent during the 90 day period irrelevant of duration, and third, absence duration as the total number of absence days during the 90 day period. Sickness absence data was not diagnosis-specific. From the 90 days examined at each time point, participants in the intervention group were absent from work 26 days at baseline and 25 days at post-assessment. In the control group, participants were absent 28 days at baseline and 24 days at post-assessment. I have calculated the SD based on the mean diff of 1 and P value of test difference provided by authors (P = 0.88). SD is then 41.62. (based on Cochrane handbook 4.2.5. (8.5.2.4) for differences in means Depression: PHQ data only available at T1 (12 weeks) (BDI was measures at 24 weeks as well) **Outcomes** The numbers are based on random imputation for the BDI. For sickness absence less data were available but the researchers did not use imputation here for unclear reasons. The follow-up time is unclear. The article states post-assessment but there have been assessments at 12 weeks and 24 weeks. It is unclear what 24 weeks would stand for because the waiting list group could have the active intervention now. Follow-up time not stated in protocol.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "We used a computerized block randomisation procedure (allocation ratio 1:1, block size 10)." Judgement comment: Computer generated
Allocation concealment (selection bias)	Low risk	Quote: "The researcher conducting the randomisation had no information about the participants apart from their 6-digit codes and did not participate in the enrolment and assignment of the participants to study groups, which was handled by two different researchers."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Comment: Participants were not blinded to allocation. 'The control group was a wait-list plus psycho-education condition." The control group did not have access to therapist guidance. Participants were eligible to access the intervention after study completion, if they requested access.
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Comment: Information on work absenteeism was retrieved from health insurance records. In the German health care system, such standardized health data is collected routinely. Its primary purpose is cost reimbursement and quality assurance, but it can be made available for secondary analysis. Due to the routine data collection, health insurance records are assumed to have high ecological validity.
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: Outcomes were self-assessed depression
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Quote:"Baseline characteristics between participants who completed the post-assessments and those who were lost to follow-up were tested for differences. Older participants (PHQ at T1: P=.02, BDI at T2: P=.03) and participants with higher education (PHQ at T1: P=.03, BDI at T2: P=.04) were more likely to complete the post-assessment on the primary outcome and the follow-up assessment. Participants who were not in psychotherapy during study

Beiwinkel 2017 *(Continued)*

		enrolment were more likely to complete post-assessment on one of the primary outcomes" Comment: There was more than 20% attrition
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Comment: There was less than 10% attrition
Selective reporting (reporting bias)	Low risk	Quote: "The study was registered retrospectively on February 1, 2013, under the International Standard Randomized Controlled Trial Number ISRCTN02446836; http://www.controlled-trials.com/ISRCTN02446836." Judgement comment: All outcomes relevant for this review" improvement of depressive symptoms or symptoms of adjustment disorder (weekly screening by the PHQ-9; pre-post measurement with two follow-ups by the BDI-II)Secondary outcome measures1. Reduction of sick days (routine data analysis) "were published.In the protocol, EQ-%D, ASF and SCL-14 were also listed as secondary outcomes, but these were not published.The trial was set up to also include adjustment disorder. Current incapacity to work certificate was an inclusion criterion in the protocol, but is not mentioned in the publication.
Other bias	Low risk	No other bias detected

Birney 2016

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** Outreach was conducted via the Chestnut EAP call centre, print ads, online postings and ads, email listservs, and flyers. All interested potential participants were directed to an informational website that described the broad characteristics of the study's purpose, activities, and compensation, concluding with an online screening survey. **Follow-up:** 6 weeks and 10 weeks
Participants	**Baseline characteristics** Clinical: psychological: I-CBT, I-guided • *Number of participants randomised*: 150 • *Gender*: 25% male • *Marital status*: Married/living with partner 78 (52.0%) Divorced 22 (14.7%) Widowed 3 (2.0%) Separated 5 (3.3%) Single 42 (28.0%) • *Occupation*: Full time 84 (56.0%) Part time 53 (35.3%) Self-employed 13 (8.7%) • *Sick leave status*: 100% • *Age*: 40.6 ±11.5 No intervention or Care as Usual • *Number of participants randomised*: 150 • *Gender*: 21% male

Birney 2016 *(Continued)*

- *Marital status*: Married/living with partner 72 (48.0%) Divorced 23 (15.3%) Widowed 2 (1.3%) Separated 5 (3.3%) Single 47 (31.3%)
- *Occupation*: Full time 92 (61.3%); Part time 46 (30.7%); Self-employed 12 (8.0%)
- *Sick leave status*: 100%
- *Age*: 40.7 ± 11.2

Overall

- *Number of participants randomised*: 300
- *Gender:* 23% male
- *Marital status*: Married/living with a partner 50%
- *Occupation*: Full-time 87%
- *Sick leave status*: 10%
- *Age*:40.6 ± 11.4

Inclusion criteria: (1) 18 years or older, (2) mild-to-moderate depressive symptoms as measured by the Patient Health Questionnaire-9 (PHQ-9)(score of 10-19), (3) not currently suicidal or meeting criteria for bipolar or schizo-affective disorder, (4) employed at least part time, (5) English speaking, and (6) have access to a high-speed Internet connection.

Exclusion criteria: None specified

Pretreatment: No substantial differences between intervention and control group

Setting: Occupational as part of employee assistance programmes

Interventions

Intervention characteristics

Clinical: Psychological intervention: I-CBT, I-guided

- *Content*: The Mood Hacker responsive mobile Web app was designed to educate users about depression and the benefits of CBT-based strategies to improve mood self-management and to activate(1) daily mood and activity monitoring, (2) increased engagement in positive behavioural activities, (3) decreased negative thinking and increased positive thinking, (4) increased practice of gratitude, mindfulness, and strength-based cognitions and behaviours, and (5) daily practice of these skills to improve depression symptoms and increase resilience to future mood disturbances
- *Duration, frequency, length*: 6 weeks, daily, up to users
- *Communication means*: Mobile Web
- *Providers*: Development of the MoodHacker app was undertaken by a multidisciplinary team of researchers and developers at ORCAS;input was incorporated from experts with extensive experience in CBT-based self-management interventions for adults with depression and the benefits of positive psychology

Care as Usual-info

- *Content*: Alternative care participants received an email with links to vetted online information about depression from Help Guide, the Mayo Clinic, Mental Health America, and the National Institute of Mental Health; they were encouraged to browse these sites on their own schedule for 6 weeks. The educational links were emailed after the baseline assessment. Participants in the alternative care group were then given access to the MoodHacker program after the 10-week assessment.
- *Duration, frequency, length*: 6 weeks, supposedly daily and up to users
- *Communication means*: Internet sites
- *Providers*: Provided by researchers who also evaluated the programme

Outcomes

Sickness absence

Days lost in past two weeks due to health reasons

- **Outcome type**: Continuous Outcome

Depressive symptoms

Birney 2016 *(Continued)*

	Patient Health Questionnaire Depression Score • **Outcome type**: Continuous Outcome
Notes	Country: US

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "After screening into the study, agreeing to the online informed consent, and submitting the baseline assessment, participants were blocked on race/ethnicity and randomised within block into either (1) treatment intervention group (n = 150), which used the MoodHacker intervention for 6 weeks, or (2) alternative care group (n = 150), which received links to six websites with information about depression." Judgement comment: "To enhance sample representativeness in each experimental condition, qualified participants were blocked on race/ethnicity and then randomly assigned within each race/ethnicity block to condition—treatment or alternative care—using the random number function in our subject database. Unclear sequence generation, unclear how block randomization was conducted
Allocation concealment (selection bias)	Low risk	Quote: "Emails indicating group assignment and linking participants to the online informed consent form were auto-generated in the database and sent to participants by a research assistant. Upon" Judgement comment: Automated mails sent to participants leave little room for change
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Comment: participants and personnel were not blinded. Unlikely that this will have led to different behaviour that had an effect on the outcome
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Quote:"All other research team members were blinded and, aside from crisis calls, no research team members had direct interaction with subjects after randomisation." Comment:Both the outcome 'absenteeism' and depressive symptoms were measured using self-report. Therefore, the risk of bias due to a lack of blinding of the participants (in this study the outcome assessors) is high. "Productivity loss due to work absence was assessed using the two-item WLQ Work Absence Module," "Depressive symptomatology was assessed at each assessment point using the self-reported PHQ-9 "
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Quote:"All other research team members were blinded and, aside from crisis calls, no research team members had direct interaction with subjects after randomisation." Comment:Both the outcome 'absenteeism' and depressive symptoms were measured using self-report. Therefore, the risk of bias due to a lack of blinding of the participants (in this study the outcome assessors) is high. "Productivity loss due to work absence was assessed using the two-item WLQ Work Absence Module," "Depressive symptomatology was assessed at each assessment point using the self-reported PHQ-9 "
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Quote: "Prior to conducting these analyses, we employed the single imputation procedure available in SPSS, version 21.0 (IBM Corp) to account for missing data."

Birney 2016 *(Continued)*

		Comment: Only 10/150 in intervention group and 5/150 in the control group missing. Imputation made complete ITT analysis possible. Attrition did not differ by condition
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Quote: "Prior to conducting these analyses, we employed the single imputation procedure available in SPSS, version 21.0 (IBM Corp) to account for missing data." Comment: Only 10/150 in intervention group and 5/150 in the control group missing. Imputation made complete ITT analysis possible. Attrition did not differ by condition
Selective reporting (reporting bias)	Unclear risk	Judgement comment: All outcomes listed in the trial register were reported on. However, the trial registration was conducted in 2015, while participants were recruited in 2012-2013.
Other bias	High risk	Judgement comment: Conflicts of Interest Amelia Birney was the study Principal Investigator. She is employed as a Behavioural Scientist at ORCAS, a health innovation and technology company that creates self-management programs to improve physical and emotional well-being. Softwaredevelopment was funded with a Small Business Innovation Research grant, which was designed to stimulate research and product development. Thus, improved versions of MoodHacker are being marketed. Ms Gunn and Mr Russell are no longer employed by ORCAS; they will derive no financial benefit from sales of the MoodHacker app or from publication of this research. MsBirney and Dr Ary remain employees of ORCAS with some potential for financial benefit from sales of the MoodHacker app.

Björkelund 2018

Study characteristics	
Methods	**Study design:** Cluster randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** Clusters were recruited through all primary care centres in a region in Sweden; patients were recruited through the health centres by asking all with a probably new diagnosis of depression to participate **Follow-up:** 3 months and 6 months
Participants	**Baseline characteristics** Improved Care • *Number of participants randomised*: 192 • *Number with sick leave: 89* • *Gender*: 32% male • *Marital status*: 67% cohabiting • *Occupation*: 73% working • *Sick leave status*: on sick leave 51% • *Age*: 40.8 ± 15.0 Care as Usual

Björkelund 2018 *(Continued)*

- *Number of participants randomised*: 184
- *Number with sick leave: 99*
- *Gender*: 26% male
- *Marital status*: 68% cohabiting
- *Occupation*: 66% working
- *Sick leave status*: on sick leave 55%
- *Age*: 41.6 ± 15.4

Overall

- *Number of participants randomised*: 376
- *Gender*: 71% female
- *Marital status*: 67% cohabiting
- *Occupation*: 69% working
- *Sick leave status*: 42% sick leave
- *Age*: 41

Inclusion criteria: Patients attending 23 different urban and rural PCCs,aged≥18 years, diagnosed with a new (1 month) mild or moderate (according to Montgomery-Åsberg Depression Rating Scale-Self assessment (MADRS-S), depression (ICD-10 diagnosis F32, F33)

Exclusion criteria: Diagnosed with bipolar disorder, psychosis, addiction, or cognitive impairment, not speaking/understanding Swedish

Pretreatment: There were no statistically significant differences between participants in the intervention and control patient groups at baseline concerning age, gender,lifestyle, education, occupation, sick leave, depression symptom scores (MADRS-S and BDI), or QoL.

Setting: Primary care centres (PCCs) in Sweden.

Interventions

Intervention characteristics

Improved Care: Enhanced Care for depression

- *Content*: Care as usual plus the intervention. Creating an individual care plan. Person-centered communication around depressive symptoms based on the patient's current depression symptom assessment with a self assessment instrument in connection with the regular telephone call, as well as behavioural activation. The care manager had direct and regular contact with the General Practitioner (GP), therapist, or other PCC personnel who were involved in the care of the patient. The care manager did not include any type of psycho-therapy in her/his care of the patient, but supported the patient and increased the accessibility and continuity of the PCC's care for the patient, coupled with organizational changes that would facilitate care for the patient with depression.
- *Duration, frequency, length*: 12 weeks, 6-8 times, 15-30 minutes; initial visit one hour
- *Communication means*: Initial face to face; follow-up telephone
- *Providers*: Special care manager, nurse 25% working time

Care as Usual- General Practice

- *Content*: Participants at the control PCCs received care as usual (CAU) according to standard protocol and procedures.The Swedish National Guidelines for Depression and Anxiety Disorders recommend high accessibility and continuity, early next appointment, guided self-help,cognitive behaviour therapy (CBT) (face-to-face or Internet delivered), interpersonal therapy, and/or anti-depressants first and second steps in a stepped care model.
- *Duration, frequency, length*: 12 weeks
- *Communication means*: Not specified.
- *Providers*: General Practitioner

Outcomes

Sickness absence

Sick leave days

Björkelund 2018 *(Continued)*

- **Outcome type**: Continuous outcome

Depressive symptoms

Beck Depression Inventory II

- **Outcome type**: Continuous outcome
- **Scale**: BDI II
- **Range**: 0-27
- **Direction**: Lower is better

Notes	Country: Sweden Authors provided extra information for 6 months sick leave: Number of patients, mean number of days on sick leave from 0 to 6 months: Intervention: n = 89, m = 99.9. SD 68.9 Control n = 99, m = 93.5 SD 65.6 BDI end-scores extracted from figure in article

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "Each stratum was allocated into six blocks consisting of two health care centers, in which one was randomly assigned to implement the care manager function." Judgement comment: No information on sequence generation
Allocation concealment (selection bias)	Unclear risk	Judgement comment: Allocation concealment at the level of health care centre is not detailed in the publication nor the trial registration. At the level of individuals, allocation concealment is not applicable as all individuals within the health centre received the same care.
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Comment: Participants were not blind to the intervention. It is unlikely that they changed there behaviour because they knew they were in the intervention group. The same holds for the providers.
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Quote:The 3 months assessment at the intervention PCCs was carried out by research personnel unknown to the patient, as it could have been a possible source of bias if the assessment was made by the local care manager. At the control PCCs, the 3 months assessment was administered by a specially trained research nurse. Comment: Sick leave data are self-reported by patients
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Quote:The 3 months assessment at the intervention PCCs was carried out by research personnel unknown to the patient, as it could have been a possible source of bias if the assessment was made by the local care manager. At the control PCCs, the 3 months assessment was administered by a specially trained research nurse. Comment: Depression data are self-reported by patients
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Comment: It is not completely clear, but the attrition of follow-up was" Follow-up data: 147 of the 199 in the intervention group (lost = 23 %) 152 of the 184 in the intervention group (lost = 17 %) total: 299 of the 376 (20%) As electronic patient records were used to gather missing data, risk of bias is low. The attrition rate was low, and through access to the electronic patient records,

Björkelund 2018 *(Continued)*

		complementary data especially concerning medication and sick certification were also collected.
Incomplete outcome data (attrition bias) Sick Leave	Low risk	The same holds for sick leave data
Selective reporting (reporting bias)	High risk	Judgement comment: All outcome measures reported in the trial register are reported in the publication, however the trial was retrospectively registered. "Trial registration: Identifier: NCT02378272. February 2, 2015. Retrospectively registered."Moreover, in the trial registration the MADRS-S is not mentioned, only the BDI-II. In the publication MADRS-S is mentioned as primary outcome alongside the BDI-II. In the results, the MADRS-S outcome shows statistically sign results and the BDI does not."There was a substantial reduction of depression scores both in intervention and control groups, but the reduction was significantly greater in the intervention group compared to control group when measured with MADRS-S, and the difference still progressed during the period 4-6 months, although the care manager intervention was terminated at 3 months."Depression score reduction measured by BDI-II did not reach significance. Mean depression score measured by BDI-II was 0.44 lower (95% CI [-1.62; 2.50], P = 0.67) at 3 months, and 1.96 lower (95% CI [-0.19; 4.11], P = 0.07) at 6 months"
Other bias	Low risk	Judgement comment: No other sources of bias detected

Blomdahl 2018

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Follow-up:** 13 weeks **Number of trial arms:** 2 **Recruitment:** Participants were recruited consecutively from 2 general care clinics and 2 specialist psychiatric outpatient care clinics.
Participants	**Baseline characteristics** Art Therapy + Treatment as usual • *Age*: 18-25:5 (11.6%)/ 26-35:9 (20.9%)/ 36-45:14 (32.6%)/ 46-55:13 (30.2%)/ 56-65:2 (4.7%) • *Gender*: Men 10 (23.3%) • *Marital status*: Single 12 (27.9%)/ Single parent 2 (4.7%)/ Cohabiting (spouses, partners) 12 (27.9%)/ Partners with children 12 (27.9%)/ Collective 0 (.0%)/ Live-apart 1 (2.3%)/ Other 4 (9.3%) • *Occupation*: Employment On a temporary basis 2 (5.1%)/ With conditional tenure 18 (46.2%)/ Other forms 12 (30.8%)/ Not applicable 7 (17.9%)/ (n = 4 not reported) • *Sick leave status*: Not reported • *Number of participants randomised*: N = 43 Treatment as usual • *Age*: 18-25:4 (11.1%)/ 26-35:11 (30.6%)/ 36-45:6 (16.7%)/ 46-55:10 (27.8%)/ 56-65:5 (13.9%) • *Gender*: Men 13 (36.1%)

Blomdahl 2018 *(Continued)*

- *Marital status*: Single 11 (31.4%)/ Single parent 4 (11.4%) Cohabiting (spouses, partners) 9 (25.7%)/ Partners with children 7 (20.0%)/ Collective 2 (5.7%) / Live-apart 1 (2.9%)/ Other 1 (2.9%)
- *Occupation*: Employment On a temporary basis 3 (10.7%)/ With conditional tenure 12 (42.9%)/ Other forms 11 (39.3%)/ Not applicable 2 (7.1%)/ (n = 8 not reported)
- *Sick leave status*: Not reported
- *Number of participants randomised*: N = 36

Overall

- *Age*: 18-25:9 (11.4%)/ 26-35:20 (25.3%)/ 36-45:20 (25.3%)/ 46-55:23 (29.1%)/ 56-65:7 (8.8%)
- *Gender*: Men 23 (29.1%)
- *Marital status*: Single 23 (29.1%)/ Single parent 6 (7.6%)/ Cohabiting (spouses, partners): 19 (24.1%)/ Partners with children 19 (24.1%)/ Collective 2 (2.5%)/ Live-apart 2 (2.5%)/ Other 5 (6.3%)
- *Occupation*: Employment on a temporary basis 5 (7.5%)/ With conditional tenure 30 (44.8%)/ Other forms 23 (34.3%)/ Not applicable 9 (13.4%)/ (n = 12 not reported)
- *Sick leave status*: Not reported
- *Number of participants randomised*: N = 79

Inclusion criteria: Eligible participants were adults, aged 18 or over, and outpatients who actively sought help for depression. Inclusion criteria were moderate to severe depression without psychotic symptoms.The participants underwent a clinical interview with a registered psychotherapist before definite inclusion, and further assessment with MADRS-S to ensure that they fulfilled the inclusion criteria.

Exclusion criteria: Exclusion criteria were recent traumatic events needing trauma treatment, bipolar syndrome, ongoing addiction, psychosis, and cognitive disability.

Pretreatment: There were no significant between-group differences regarding diagnosis, numbers of depression occasions, comorbidity, gender, age, forms of social life, children at home, and different aspects of employment, education at baseline or other characteristics at baseline.

Setting: General and specialist (psychiatric) care in Sweden (Region Västra Götaland in western Sweden). The health care units were located both in city and in rural areas.

Interventions

Intervention characteristics

Art Therapy + Treatment as usual

- *Content*: I. Goal-setting Exercise: Body scan art task: Description of the current situation. II. Here and now art task: Mindful exploration of art media, awareness of bodily and emotional responses elicited by sensory stimulation III. Breathing anchors art task: Body image before and after the mindfulness practice; Raise awareness and explore how breathing affects body experience. IV. Breathing-space art task: Drawing analogue pictures, explore and raise the awareness of emotional reactions. V. Body scan art task: Color and emotions In-depth exploration of emotions and state of mind. VI. Inner and outer attention art task: stressful, pleasant event pictures, enhance awareness for reactions to stressful situations and find strategies to cope with reactions. VII. One thing at a time art task: Graphic life-line, awareness of behaviour patterns and strategies. VIII. Breathing exercise art task: Roles awareness of behaviour patterns and roles. IX. Body scan art task: Description of the current situation, evaluation of treatment and process; the patient's interpretation of meanings are the focus. X. Review of all images art task: Mandala
- *Duration, frequency, length*: 10 weeks,1 hour session per week
- *Communication means*: Face-to-face
- *Providers*: Experienced occupational therapists who applied a manual, which consisted of detailed guidelines based on phenomenological art therapy

Treatment as usual

- *Content*: TAU consisted of acupuncture, cognitive-behavioural therapy, electroconvulsive therapy, interpersonal therapy, occupational therapy, pharmacological therapy, physiotherapy, psychodynamic therapy,and supportive therapy

Blomdahl 2018 *(Continued)*

- *Duration, frequency, length*: For CBT average 10 sessions. Varying for the other therapies between 0 and 10
- *Communication means*: Face-to-face
- *Providers*: Various providers for TAU: The participants' regular therapists or physicians planned and performed TAU.

Outcomes	**Sickness absence** *Sick leave percentage during follow-up* • **Outcome type**: Continuous outcome **Depressive symptoms** *MADRS* • **Outcome type**: Continuous outcome
Notes	**Country:** Sweden

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "A simple randomisation procedure was carried out using a computer-generated list of random numbers."
Allocation concealment (selection bias)	Low risk	Quote: "The result of the randomisation was stored in an opaque sealed envelope marked with an ID-code. The first author performed the randomisation procedure before any contact was established with participants. Each participant was then contacted either by phone or by mail for an appointment with the research assistant (a registered psychotherapist). The research assistant described the procedure in detail to the participant before obtaining written informed consent. The assessment started with an interview to confirm the participant's diagnosis and level of suicide risk, after which the assessment was completed with the self-assessment questionnaires. The research assistant then in- formed participants about their treatment allocation."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Quote: "There were no significant differences between PATd/TAU- group and TAU-group in relation to the frequencies of the TAU therapies that the participants received (see Table 2)." Judgement comment: Unlikely that the participants would change their behaviour based on the intervention
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Outcomes were self-reported and patients judged to report beneficial outcomes after the intervention
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Outcomes were self-reported and patients judged to report beneficial outcomes after the intervention
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Of the total group that was randomised (85), 21 were lost to follow up. Percentage lost to follow-up: 25%
Incomplete outcome data (attrition bias) Sick Leave	High risk	Of the total group that was randomised (85), 21 were lost to follow up. Percentage lost to follow-up: 25%

Blomdahl 2018 *(Continued)*

Selective reporting (reporting bias)	Unclear risk	Judgement comment: 'following a research protocol.' We did not find the info.
Other bias	Low risk	Judgement comment: No other sources of bias detected

Burnand 2002

Study characteristics	
Methods	**Study design**: RCT, random assignment stratified by presence of personality disorder, past major depressive syndrome and gender; two conditions. **Recruitment**: screening by nurse and psychiatrist of consecutive patients referred for acute outpatient treatment. **Follow up:** 10 weeks. Lost to follow up: 22%
Participants	**Baseline:** 95 were randomised (T1: 35; C: 39); Age: T1: 36 (SD 9.5); C: 36.7 (SD 10.4) Female: T1: 66%; C: 56% Stable employment: T1: 71%; C: 82% **Inclusion criteria:** age 20 to 65 years, new episode of care, MDD DSM-IV (SCID) + HDRS at least 20; **Exclusion criteria:**: bipolar disorder, psychotic symptoms, severe substance dependence, organic disorder, mental retardation, history of severe intolerance to clomipramine, poor command of French language **Setting**: outpatient community mental health centre in Switzerland;
Interventions	T1: Psychodynamic psychotherapy: individual sessions by nurse + clomipramine: 25 mg first day, gradually increasing to 125 mg on fifth day (dosage adjustment allowed). Refusal or severe side effects: 20 to 40 mg citalopram per day. Duration: 10-week program, frequency psychotherapy sessions not fixed, duration of clomipramine 10 weeks C: Supportive care: individual sessions: empathic listening, guidance and support. + clomipramine: 25 mg first day, gradually increasing to 125 mg by fifth day (dosage adjustment allowed). Refusal or severe side effects: 20 to 40 mg citalopram per day. Duration supportive care: not fixed, duration clomipramine 10 weeks
Outcomes	Sickness absence 1) number of days of sick leave in 10 weeks Depressive symptoms: 1) full remission (at most 7 HDRS) (at 10 weeks) 2) severity of depression (HDRS score; GAS) (at 10 weeks) Work functioning: 1) “adjustment to work” subscale of the modified Health-Sickness Rating Scale (HSRS).
Notes	**Country:** Switzerland
Risk of bias	

Bias	Authors' judgement	Support for judgement

Burnand 2002 *(Continued)*

Random sequence generation (selection bias)	Unclear risk	Randomisation procedure not reported
Allocation concealment (selection bias)	Unclear risk	Randomisation procedure not reported
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	No blinding but risk of performance bias low as both treatments can be considered equally desirable for patients "Both treatments involved the same clomipramine protocol and intensive nursing in a specialized milieu. In addition, the amount of structured psychodynamic psychotherapy provided during combined treatment was comparable to the amount of supportive care provided during treatment with clomipramine alone."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Outcome assessor for sick leave was blinded, but (non-blinded) patients had to report the number of sick leave days to them "The psychologists who made the assessments of hospitalizations, number of days of sick leave, and GAS scores were blinded to each patient's treatment assignment."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	"The individuals who rated the presence and severity of major depression and HSRS scores at ten weeks were not blinded to treatment assignment."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up is high: 22%. Risk of attrition bias due to follow up losses is therefore considered to be high, although multiple analyses were used to study the effect on the findings and the authors conclude otherwise: "Twenty-one patients (12 in the experimental and nine in the control group, or 22 percent) were excluded from the analysis--four who did not return for treatment (three in the experimental group and one in the control group), three who dropped out against medical advice (two in the experimental group and one in the control group), and 14 who were discharged because they had exclusion characteristics that were not detected at entry, including severe alcohol or drug dependence (five in each group) and adverse effects (two in each group). These patients were not significantly different from the other patients in terms of the main outcome variables at intake. The 74 patients who completed the study were not significantly different from the 21 who were withdrawn or from the group of 95 as a whole. To control for intent to treat, the analyses were repeated with all 95 patients who had been randomly assigned to treatment." "This finding was unchanged when we repeated the analyses and controlled for age, gender, initial severity of depression, GAS score at intake, compliance and intent to treat"
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up is high: 22%. Risk of attrition bias due to follow up losses is therefore considered to be high, although multiple analyses were used to study the effect on the findings and the authors conclude otherwise: "Twenty-one patients (12 in the experimental and nine in the control group, or 22 percent) were excluded from the analysis--four who did not return for treatment (three in the experimental group and one in the control group), three who dropped out against medical advice (two in the experimental group and one in the control group), and 14 who were discharged because they had exclusion characteristics that were not detected at entry, including severe alcohol or drug dependence (five in each group) and adverse effects (two in each group). These patients were not significantly different from the other patients in terms of the main outcome variables at intake. The 74 patients who completed the study were not significantly different from the 21 who were with-

Burnand 2002 *(Continued)*

		drawn or from the group of 95 as a whole. To control for intent to treat, the analyses were repeated with all 95 patients who had been randomly assigned to treatment." "This finding was unchanged when we repeated the analyses and controlled for age, gender, initial severity of depression, GAS score at intake, compliance and intent to treat"
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None Identified

Chatterton 2018

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Follow-up:** 12 weeks **Number of trial arms:** 2 **Recruitment:** Community-based recruitment strategies were used to identify study participants, including flyers in medical waiting rooms, pharmacies and university campuses; newsletters; and contact with potential referral sources (e.g. general practitioners, private psychiatrists and local psychiatric inpatient units). Media interviews and advertisements in social media (e.g. Twitter, Facebook), Google, local newspapers and radio stations were also employed as recruitment strategies.
Participants	**Baseline characteristics** Diet • *Age*: 37.5 (10.7) • *Gender (% female)*: 61.8 (21) • *Marital status*: not reported • *Occupation (income > $ 80 000)*: 25.0 (8) • *Sick leave status*: not reported • *Number of participants randomised*: 33 Social Support • *Age*: 43.1 (14.6) • *Gender (% female)*: 81.8 (27) • *Marital status*: not reported • *Occupation (income > $ 80 000)*: 21.2 (7) • *Sick leave status*: not reported • *Number of participants randomised*: 34 Overall • *Age*: 40.3 (13.1) • *Gender (% female)*: 71.6 (48) • *Marital status*: not reported • *Occupation (income > $ 80 000)*: 23.1 (15)

Chatterton 2018 *(Continued)*

- *Sick leave status*: not reported
- *Number of participants randomised*: 67

Inclusion criteria: Eligibility criteria included participants who were at screening: aged 18 or over and could provide informed consent;successfully fulfilled the DSM-IV-TR diagnostic criteria for a major depressive episode; scored 18 or over on the Montgomery–Åsberg Depression Rating Scale and scored 75 or less, out of a possible score of 104, on a Dietary Screening Tool (DST) modified for Australian food products. The DST was completed to confirm 'poor' dietary quality, before enrolment. This screening tool was used to reflect usual daily or weekly intake of specified foods. Broadly defined, participants had to report a poor (low) intake of dietary fibre, lean proteins and fruit and vegetables, and a high intake of sweets, processed meats and salty snacks. If participants were on antidepressant therapy or undergoing psychotherapy, they were required to be on the same treatment for at least 2 weeks prior to randomization. Participants had to be readily available for a 12-week period and have the ability to eat foods as prescribed, without religious,medical, socio-cultural or political factors precluding participation or adherence to the diet.

Exclusion criteria: Participants were ineligible if they had: (1) a concurrent diagnosis of bipolar I or II disorder; (2) two or more failed trials of antidepressant therapy for the current MDE; (3) known or suspected clinically unstable systemic medical disorder; (4) pregnancy; (5) commencement of new psychotherapy or pharmacotherapy within the preceding2 weeks; (6) severe food allergies, intolerances or aversions; (7) current participation in an intervention targeting diet or exercise; (8) a primary clinical diagnosis of a personality disorder and/or a current substance use disorder.

Pretreatment: The dietary group had significantly lower scores on the dietary screening tool and the ModiMedDiet score than the social support control group at baseline, primarily due to lower intakes of fruit and higher intakes of extras. Otherwise, groups were well matched on characteristics

Setting: Participants were recruited from two sites: Barwon Health in Geelong and St. Vincent'sHealth in Melbourne (Victoria, Australia)

Interventions

Intervention characteristics

Diet

- *Content*: The dietary intervention comprised personalised dietary advice and nutritional counselling support, including motivational interviewing, goal setting and mindful eating, from a clinical dietician in order to support optimal adherence to the recommended diet. This comprised the 'ModiMedDiet', developed by RO and CI, which was based on the Australian Dietary guidelines and the Dietary Guidelines for Adults in Greece and is concordant with our previous dietary recommendations for the prevention of depression. This was provided in adjunction to regular clinical therapy.
- *Duration, frequency, length*: Seven individual 1 hour dietary support sessions; the first four sessions occurred weekly and the remaining three sessions occurred every 2 weeks.
- *Communication means*: face-to-face
- *Providers*: Intervention delivered by an Accredited Practising Dietician

Social Support

- *Content*: Befriending consists of trained personnel discussing neutral topics of interest to the participant, such as sport,news or music, or in cases where participants found the conversation difficult, engaging in alternate activities such as cards or board games, with the intention of keeping the participant engaged and positive. This is done without engaging in techniques specifically used in the major models of psychotherapy. This was provided in adjunction to regular clinical therapy.
- *Duration, frequency, length*: Seven individual 1 hour sessions the first four sessions occurred weekly and the remaining three sessions occurred every 2 weeks.
- *Communication means*: face-to-face
- *Providers*: Research assistants (RAs) in this trial completed manual-guided training and also participated in role-playing training exercises to ensure consistent delivery of the protocol.

Outcomes

Sickness absence

Days lost from paid work in past 3 months

Chatterton 2018 *(Continued)*

- **Outcome type**: Continuous Outcome

Depressive symptoms

MADRS

- **Outcome type**: Continuous Outcome
- **Scale**: 6-point scale
- **Range**: 0-60
- **Direction**: Lower is better
- **Data value**: Endpoint

Notes	**Country**: Australia Communication from study authors: Investigators asked participants the number of sickness absence days taken in the past month. They multiplied by 3 to get results for 3 months.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "The randomisation sequence was computer generated by an independent person (OD) using a 2 × 2 block design."
Allocation concealment (selection bias)	Low risk	Quote: "mental health assessments were blind to participants' group allocations, and the randomisation schedule and coding of group allocations were not, at any time, accessible to the research assistants conducting the assessments, or to the biostatistician (SC). At the conclusion of the baseline appointment, the dietician/befriender would meet privately with the participant and inform them of their group allocation in order to maintain blinding of the research assistants." Quote: "The sequence was saved to a password-protected spreadsheet, and groups were coded A and B. The randomisation allocation was managed by the trial dieticians or 'befrienders', in order to ensure that the research assistants responsible for"
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Even though blinding was not possible. The control strategy was designed to make it highly unlikely that patients changed their behaviour. However, as pointed out in a published critique of the study: Molendijk et al. (2018) 'The SMILES trial: do undisclosed recruitment practices explain the remarkably large effect?,' the dietary intervention was positively advertised during recruitment, leading to a high risk of performance bias.
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	"The research assistants remained blind to condition for the final assessment of the outcomes"
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	"The research assistants remained blind to condition for the final assessment of the outcomes"
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	"16.4% lost to follow-up. But attrition was accounted for in the analyses. To test departures from missing at random (MAR), a weighted sensitivity analysis using the Selection Model Approach was applied to the main outcome findings [43, 44]. Briefly, once data had been imputed under MAR (n = 5), parameter estimates from each imputed dataset were reweighted to allow for the data to be missing not at random (MNAR). The chosen constant values used to add to the imputed missing data to account for MNAR were multiplications of standard error (i.e. 1.6) for main outcome comparison under MAR assumptions. To eval-

Chatterton 2018 *(Continued)*

		uate the robustness of our findings, different degrees of departure from the MAR assuming plausible values ranging from 10*SE to –8*SE were considered"
Incomplete outcome data (attrition bias) Sick Leave	Low risk	See depressive symptoms.
Selective reporting (reporting bias)	Low risk	Judgement comment: A trial protocol was published before start of the study. Australia and New Zealand Clinical Trials Register (ANZCTR): ACTRN12612000251820. Registered on 29 February 2012. Absenteeism was not included in the protocol, but it was reported in the design paper in BMC Psychiatry.
Other bias	Low risk	Judgement comment: No other sources of bias detected

Eriksson 2017

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** A total of 90 patients were enrolled between March 2010 and March 2013 at 16 primary care centers (PCCs) located in the south-west region of Sweden. All patients were assessed by a psychologist/psychotherapist (therapist) and randomised to either Internet-delivered cognitive behavioural therapy (ICBT) or treatment as usual (TAU).. **Follow-up:** 3, 6 and 12 months
Participants	**Baseline characteristics** Clinical: psychological: I-CBT, I-guided • *Number of participants randomised*: 44 • *Number with sick leave: 41* • *Gender*: 33% men • *Marital status*: 60% married • *Occupation*: 80% employed • *Sick leave status*: 48% last year • *Age*: 37.1 ± 12.8 Care as Usual-GP • *Number of participants randomised*: 46 • *Number with sick leave: 29* • *Gender*: 26% men • *Marital status*: 61% married • *Occupation*: 78% employed • *Sick leave status*: 40% last year • *Age*: 35.1 ± 9.9 **Inclusion criteria:** men and women aged 18 years and older with symptoms of depression who attended the study PCCs were recruited by the GPs and nurses. Patients positive to ICBT as a treatment option, who had not recently (last month) started or changed possible antidepressant medication were

Eriksson 2017 *(Continued)*

	asked about their willingness to participate in the study; assessed by a psychologist/psychotherapist (therapist), patients had to meet diagnostic criteria for depression according to DSM-IV (assessed via MINI), have a MADRS-S score below 35, and have access to a computer with speakers or headphones. **Exclusioin criteria:** severe depression (according to a MADRS-S score 35), a principal diagnosis of anxiety (assessed by the therapist), psychosis, bipolar disorder or hypomanic episode, antisocial personality disorder, substance dependence or alcohol abuse (all of the above assessed by therapists using MINI), medium or high suicide risk (defined as MADRS-S question 9 > 3p and/or MINI Part B–Suicide > 9p, or previous suicide attempt); other severe mental disorder, cognitive disability or communication difficulties that would prevent participation in the ICBT program (only available in Swedish) **Pretreatment:** no significant differences in age, gender, socioeconomic status, medication, severity of depression, quality of life or psychological distress except for use of sedatives (n = 5 [ICBT] versus n = 0 [TAU]; P = 0.049) **Setting:** Primary Care Centers - GP
Interventions	**Intervention characteristics** Psychological intervention: I-CBT • *Content*: CBT program Depressionshjälpen © consisting of behavioural activation and components of acceptance and commitment therapy. Web page with 7 modules and therapist involvement with telephone calls. All patients in the study could receive usual care • *Duration, frequency, length*: 8 to 12 weeks, frequency unclear, length unclear • *Communication means*: Internet, email and telephone • *Providers*: Licensed psychologists or psychotherapists with training in CBT Care as usual • *Content*: Scheduled contacts with GPs, nurses and other personnel at the PCC, face-to-face-psychotherapy, antidepressants, sick leave certification and combinations of these treatments • *Duration, frequency, length*: no restrictions • *Communication means*: Face to face • *Providers*: general practitioner/ GP, nurse, primary care psychologist/ psychotherapist, other personnel
Outcomes	**Sickness absence** *Sick leave during follow-up* • **Outcome type**: Continuous outcome **Depressive symptoms** *Beck Depression Inventory II* • **Outcome type**: Continuous outcome
Notes	**Country**: Sweden Additional information received from the authors: Means of number of sick leave days and standard deviations for 0-12 months for the intervention and control group: Intervention - Internet ICBT : mean 45.8 days, SD 99.3, Control –CAU: mean 49.4 days, SD 92.8.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	The randomisation was computer generated by the randomisation unit.

Eriksson 2017 *(Continued)*

Allocation concealment (selection bias)	Low risk	Quote: "The patients were consecutively randomised to either ICBT or TAU by an independent research unit at the University of Gothenburg/Sahlgrenska University Hospital. The randomisation process was performed with all study patients as one group, which concealed the allocation to be from both the PCC personnel and the researchers." Judgement comment: The authors communicated: The unit set up a telephone service office hours (8-12, 13-16) where the primary care research nurse could call. The center put up a computer routine so that the person who answered the phone calls from the research nurses at the different primary care centers could log in to the computer and get the treatment option for the patient. A confirmation letter was also sent for every patient to the research leader.
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Comment: It is unlikely that participants behaved differently knowing that they belonged to the intervention group as the control condition was equally desirable (scheduled contacts with GPs, nurses and other personnel at the PCC, face-to-face-psychotherapy, antidepressants, sick leave certification and combinations of these treatments).
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Comment: Self-reported outcomes but no blinding
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: Self-reported outcomes but no blinding
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Out of 90 participants, 22 were lost to follow-up =24%. Also 8 control participants crossed over to the intervention group and were counted there as intervention participants
Incomplete outcome data (attrition bias) Sick Leave	High risk	Out of 90 participants, 22 were lost to follow-up =24%. Also 8 control participants crossed over to the intervention group and were counted there as intervention participants
Selective reporting (reporting bias)	Low risk	Judgement comment: Study reported in two articles Eriksson 2017 and Hange 2017 but they report different results of the trial. The authors communicated that this was due to analysing only working patients
Other bias	Low risk	No other biases detected

Fantino 2007

Study characteristics	
Methods	**Study design** RCT. Recruitment: patients were recruited by psychiatrists or by general practitioners. **Follow up:** 8 weeks. **Lost to follow up**: 8.1%
Participants	**Inclusion criteria:** all patients fulfilling the DSM-IV criteria for MDD and having a baseline MADRS total score of at least 30 were eligible for the study. **Exclusion criteria**: patients meeting DSM-IV for primary diagnoses for any axis I disorder other than MDD or those with a history of mania, bipolar disorder, schizophrenia or other psychotic disorder, ob-

Fantino 2007 *(Continued)*

	sessive-compulsive disorder, cognitive disorder including mental retardation or personality disorder, patients who met the DMS-IV criteria for substance abuse or dependence within the past 12 months, or used a depot antipsychotic within 6 months before study inclusion or any antipsychotic or anticonvulsant medications within 2 weeks before the first administration of study medication **Baseline characteristics**: 280 were randomised (T1: 138; T2: 142). Setting: outpatient; general or psychiatric practices in France. Male: T1: 28.3%; T2: 38.0% Age: T1: 44.1 (SD 10.9); T2: 46.2 (SD 11.1) Family situation: T1: 23.9% single; T2: 16.2% single T1: 49.3% married, living with partner; T2: 50.7% married living with partner T1: 26.8% separated, divorced, widowed; T2: 33.1% separated, divorced, widowed Occupational status: T1: 35.5% unemployed; T2: 29.6% unemployed T1: 64.5% employed; T2: 70.4% T1: 4.5% craftsman, tradesman; T2: 7.0% craftsman, tradesman T1: 9.0% manager; T2: 12.0% manager T1: 21.3% technician; T2: 30.0% technician T1: 9.0% workman; T2: 4.0% workman
Interventions	T1: Escitalopram (SSRI) 10 mg daily during the first week, 20 mg per day for the remaining 7 weeks T2: Citalopram (SSRI) 20 mg/day daily during the first week, 40 mg per day for the remaining 7 weeks All study medications were provided in identical blister packs of identical capsules administered as one capsule per day, regardless of dose or treatment group. No adjustment of dosage was allowed
Outcomes	**Sickness absence** 1) days of sick leave for the 2-month pre-study period and for the 8-week study period (percentage of patients and mean consumption of those patients) **Depressive symptoms** 1) depression severity, assessed by the Montgomery-Asberg Depression Scale (MADRS) 2) remission, defined as the total score MADRS of ≤ 12 3) MADRS-S, the self-reported version of MADRS
Notes	**Country:** France

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Personal communication: "Allocation was random. This includes random allocation using equal block sizes."

Fantino 2007 *(Continued)*

Allocation concealment (selection bias)	Low risk	Personal communication: "Allocation was concealed. Investigators allotted patients to a treatment defined by the patient inclusion number. All treatments were prepared and identical, the only difference being the treatment number, corresponding to the allocation table, which was kept by the person who prepared the treatments. The investigators were not aware of the nature of the treatments."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Trial is double-blind: "Those meeting the eligibility criteria were randomly assigned to receive double-blind, fixed doses of either escitalopram 20 mg daily or citalopram 40 mg daily during 8 weeks, with equal block randomisation at baseline." "All study medications were provided in identical blister packs of identical capsules administered as one capsule per day, regardless of dose or treatment group." Personal communication: "The psychiatrist or GP both included the patient, dispensed the study medication, and did the assessments. Patient and investigator were both blind to the treatment, which were identical in aspect. Since this was not placebo-controlled, both comparators were active and quite similar, differing only be the presence of 20 mg R-citalopram in the 40 mg citalopram.This actually reduces the risk of unblinding by recognizable drug effects or side-effects."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	"A standardized form was used by trained investigators to record healthcare services and days of sick leave for the 2-month pre-study period and for the 8-week study period." Since the investigators were blinded, the risk of bias is considered to be low
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	The MADSR was done by investigators who are trained or confirmed in the proper use of the MADSR scores and who were blinded for the allocation status. The MADSR-S is a self-reported version, but patients were also blinded for treatment allocation
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Loss to follow up is considered to be low. T1: 4.3%; T2: 10.6%
Incomplete outcome data (attrition bias) Sick Leave	Low risk	No missing sick leave data: "Valid resource utilization information corresponding to the pre study and study periods was thus available for 280 patients."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Fernandez 2005

Study characteristics	
Methods	**Study design**: Randomised, double-blind, flexible-dose, multinational, clinical trial with a one-week run-in period with no treatment. After randomization: two treatment arms **Recruitment**: patient were asked to participate by GP. Follow up: 8 weeks. Lost to follow up: 16%
Participants	**Inclusion criteria: P**atients in primary care, age 18 to 85 yrs, DSM-IV diagnosis of MDD (current or first), Minimal MADRS score of 18. **Exclusion criteria**: History of mania or any bipolar disorder, schizophrenia or any psychotic disorder, Currently suffering from obsessive-compulsive disorder, eating disorder, mental retardation, any per-

Fernandez 2005 *(Continued)*

	vasive development disorder, or cognitive disorder (DSM-IV criteria), MADRS of at least 5 on item 10 (suicidal thoughts), Alcohol or drug abuse problems within the previous 12 months, Having had treatment with: antipsychotics, antidepressants, psychotropics (except zolpidem or stable low doses of benzodiazepines for insomnia), serotonin receptor antagonists, lithium, carbamazepine, valproate, or valpromide, ECT, treatment with CBT or psychotherapy, Being pregnant or breastfeeding, Medications likely to interfere **Baseline characteristics**: 293 were randomised (T1: 148; T2: 145). **Setting:** primary care at 44 sites in 8 European countries. with the study Mean age T1: 48.4; T2: 46.5 Sex: T1: 75.4% female; T2: 71.2% female Married or cohabiting: T1: 61.9%; T2: 56% Employed: T1: 51.5%; T2: 60% Long-term sickness absence: T1: 11.1%; T2: 11.2% Higher education: T1: 9.5%. T2: 11.2%
Interventions	T1: Escitalopram (SSRI): initial 10 mg/day. At week 2 or 4 dose could be increased to 20 mg/day at the investigator's discretion if patient's response was unsatisfactory. After 8 weeks of treatment, 1 week run-out period. Patients on 20 mg/day were down-tapered to 10 mg for the first 4 days and placebo the last 3. Patients on lower dose received 7 days of placebo T2: Venlafaxine XR (SNRI), initially 75 mg/day. At week 2 or 4 dose could be increased to 150 mg/day at the investigator's discretion if patient's response was unsatisfactorily. After 8 weeks of treatment, 1-week run-out period. Patients on 150 mg/day were down-tapered to 75 mg for the first 4 days and placebo the last 3. Patients on lower dose received 7 days of placebo
Outcomes	**Sickness absence** 1) % of patients on sick leave and average length of sick leave per week (3 months prior baseline and during 8 weeks of study) 2) personal communication; days of sick leave during 8 weeks of study, for workers only **Depressive symptoms** 1) MADRS (at 8 weeks) 2) HAM-D (at 8 weeks)
Notes	**Country**: UK

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Personal communication with first author: "Patients who met the selection criteria at the baseline visit were assigned to 8 weeks of double-blind treatment according to a computer-generated randomization list."
Allocation concealment (selection bias)	Low risk	Personal communication with first author: "The details of the randomization series were unknown to any of the investigators and were contained in a set of sealed opaque envelopes."

Fernandez 2005 *(Continued)*

Blinding of participants and personnel (performance bias) Sick Leave	Low risk	An economic evaluation was conducted alongside a double-blind,multinational, randomised clinical trial. Personal communication with first author: "This means that both investigator and patient were blinded regarding allocation to treatment."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	"Data at baseline consisted of self-reported patient questionnaires recording use of healthcare services and days of sick leave ...." Personal communication with first author: "Patients were blinded regarding allocation to treatment."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	"Depressive symptoms were assessed by trained raters." Personal communication with first author: "Outcome assessors were blinded for the allocation of patients."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow-up depression data is 15%, which we consider high and no appropriate method has been used to account for attrition. "Efficacy analyses were conducted on the intention-to-treat (ITT) population, which included all randomised patients who took at least 1 dose of double-blind study medication and who had at least 1 valid post-baseline assessment of the MADRS total score. The ITT population thus comprised 146 patients in the escitalopram group and 142 patients in the venlafaxine group. A total of 249 patients (of 293) completed the study."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Lost to follow-up sick leave data is 16%, which we consider high and no appropriate method has been used to account for attrition. "Data at baseline consisted of self-reported patient questionnaires recording use of healthcare services and days of sick leave. Of the 293 patients in the trial, valid cost information in the 3-month pre-study period was available for 251 patients; for 22 patients in the escitalopram arm and 20 patients in the venlafaxine arm, either the physician or patient did not fill in the resource use questionnaire. Of the 251 evaluable patients, 126 received escitalopram and 125 received venlafaxine. Of these, 245 patients reported valid cost information for the 8-week duration of the trial (four escitalopram and two venlafaxine patients were lost relative to the pre-study period). "Given the very low rate of attrition in the sample during the trial, patients with missing data were unlikely to represent serious bias to the results of the present analysis. As a result, no attempt was made to impute missing data."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None Identified

Finnes 2017

Study characteristics	
Methods	**Study design:** Randomized controlled trial **Number of trial arms:** 4 **Recruitment:** Through the Swedish Social Securance Agency

Finnes 2017 *(Continued)*

	Follow-up: 3 months and 9 months
Participants	**Baseline characteristics** Work-directed intervention combined with clinical intervention:WDI plus ACT • *Age*: 47.2 ± 9.2 • *Gender*: 21.6 % male • *Marital status*: not reported • *Occupation*: not reported • *Sick leave status*: 149.4 ± 102.2 • *Number of participants randomised*: 90 • *Numbers in subgroup of depressed participants only*: 27 Care as usual • *Age*: 46.9 ± 9.5 • *Gender*: 25 % male • *Marital status*: - • *Occupation*: - • *Sick leave status*: 143.2 ± 100.1 • *Number of participants randomised*: 89 • *Numbers in subgroup of depressed participants only*: 31 Psychological intervention: ACT • *Age*: 46.0 ± 8.2 • *Gender*: 19.1% male • *Marital status*: - • *Occupation*: - • *Sick leave status*: 139.6 ± 87.4 • *Number of participants randomised*: 90 • *Numbers in subgroup of depressed participants only*: 32 Work-directed intervention: WDI • *Age*: 44.9 ± 8.6 • *Gender*: 26.7 % male • *Marital status*: - • *Occupation*: - • *Sick leave status*: 154.4 ± 107.5 • *Number of participants randomised*: 90 • *Numbers in subgroup of depressed participants only*: 24 **Inclusion criteria:** Participants from Stockholm County,Sweden, of working age holding a current employment status of at least 50% (working at least 20 hr per week) and a current SA status between 25% and 100% for the past 1 to 12 months; diagnostic criteria of an anxiety disorder, depression, or stress-related ill-health as defined by the diagnostic criteria for exhaustion disorder (according to the ICD-10, Diagnostic Groups F32, F33, F43.8) **Exclusion criteria:** Active suicide ideation, severe depression, history of bipolar disorder or psychosis, substance abuse or dependence,unemployment or self-employment, and insufficient comprehension of the Swedish language. **Pretreatment:** "There were no significant pretreatment differences between the groups on the sociodemographic variables or on the pretreatment outcome measures" Diagnostic groups were also similar between interventions **Setting:** A large employer in Sweden in the public sector

Finnes 2017 *(Continued)*

	Depression Subgroup Analysis: Authors re-analysed the data for the subgroup of depressed participants only.
Interventions	**Intervention Characteristics** Work-directed intervention combined with clinical intervention: WDI plus ACT • *Content*: Combined ACT and WDI intervention with two different therapists • *Duration, frequency, length*: Nine intervention sessions/meetings over three months • *Communication means*: Face-to-face • *Providers*: see separate interventions Care as Usual • *Content*: Normal treatment:medical doctor plus psychologist, social worker, physical therapist or nurse • *Duration, frequency, length*: 2.7 ± 1.4 appointments with doctor • *Communication means*: Face-to-face • *Providers*: see content Psychological intervention • *Content*: Acceptance and commitment Therapy: six core processes in the ACT-model:acceptance, mindfulness, defusion, self as context, values, and committed action. The second part of the intervention focused on increasing behaviour repertoire in a valued direction, involving discriminating between rule-governed (e.g. must do, should do) and avoidant behaviours on one side and those driven by positive reinforcement (want to do)on the other side. • *Duration, frequency, length*: Six sessions over three months • *Communication means*: Face-to-face • *Providers*: Licensed clinical psychologists with training in ACT with weekly clinical supervision Work-directed intervention • *Content*: The Work Directed Intervention aims at the facilitation of dialogue between the participant and the workplace through a series of steps involving the participant and the nearest supervisor. (a) the participant-interview including six open questions regarding the participant's perception of causes of SA and factors that may facilitate RTW; (b) the supervisor interview carried out at the workplace (c) the convergence dialogue meeting between the participant and the supervisor, consisting of an analysis of the two former inter-views aiming at agreeing on a rehabilitation plan. • *Duration, frequency, length*: Three meetings over three months • *Communication means*: Face-to-face • *Providers*: Licensed clinical psychologists, behavioural therapist, psychiatric nurse with weekly clinical supervision
Outcomes	**Sickness absence** *Sick leave days during follow-up period* • **Outcome type**: Continuous Outcome **Depressive symptoms** *HADS* • **Outcome type**: Continuous Outcome *Work functioning: work ability index* • **Outcome type**: Continuous Outcome
Notes	**Country**: Sweden

Finnes 2017 *(Continued)*

Additional information received from authors: data were re-analysed for the subgroup that scored above the clinical cut-off score for depression on the HADS only

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Judgement comment: No information on sequence generation
Allocation concealment (selection bias)	Low risk	Quote: "Following baseline measurements, a blinded administrator made random allocation in blocks of eight, each block containing two possibilities of each condition."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Comment: Unlikely that participants would have behaved differently as a result of the intervention
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Comment: Objective outcome sick leave from register
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: Self-report
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Comment: Loss to follow-up about 10% across different groups
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Comment: Loss to follow-up about 10% across different groups
Selective reporting (reporting bias)	Unclear risk	Judgement comment: Protocol published. Not all secondary outcomes mentioned in protocol reported. Follow-up times in report pre, post, 3 mo, 9 mo different from protocol 6, 12, 24 and 60 months
Other bias	Low risk	Judgement comment: No other sources detected

Geraedts 2014

Study characteristics

Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** participants were recruited via 6 different companies in the Netherlands,2 banking companies, 2 research institutes, 1 security company, and 1 university, through banners and digital pamphlets on the company's Intranet and via posters. Employees who showed interest in the study received an information leaflet and an informed consent form via email. After participants gave informed consent, they received a link to an online screening questionnaire via email.

Geraedts 2014 *(Continued)*

	Follow-up: 12 months
Participants	**Baseline characteristics** Clinical intervention: psychological intervention: I-CBT plus Problem Solving Treatment (PST), I-guided • *Age*: 43 ± 8.9 • *Gender*: 33.6% male • *Marital status*: 74.1 in relationship • *Occupation*: all are workers; education low: n = 11 (9.5%), middle: n = 31 (26.7%), high: n = 74 (63.8%); working hours, mean (SD): 33.7 (4.8) • *Sick leave status*: all not on sick leave • *Number of participants randomised*: 116 No intervention • *Age*: 43.8 ± 9.6 • *Gender*: 41.7% male • *Marital status*: 78.3 in relationship • *Occupation*: all are workers; education low: n =5 (4.3%), middle: n = 37 (32.2%), high: n = 73 (63.5%); working hours, mean (SD): 34.0 (5.3) • *Sick leave status*: all not on sick leave • *Number of participants randomised*: 115 Overall • *Age*: years mean (SD): 43.4 (9.2) • *Gender*: females: n = 144/231 (62.3%) • *Marital status*: relationship: n = 176/231 (76.2%) • *Occupation*: all are workers, education low n = 16 (6.9%), middle: n = 68 (29.4%), high: n = 147 (63.6%); working hours mean (SD): 33.9 (5.0) • *Sick leave status*: all not on sick leave • *Number of participants randomised*:231 **Inclusion criteria:** Employees with elevated depressive symptoms as measured by a score of 16 or higher on the Center for Epidemiologic Studies Depression scale (CES-D) who were not on sick leave (at the time they completed the baseline questionnaire). Access to the Internet and an email address were required **Exclusion criteria:** Using medication for depressive symptoms for less than 1 month or if they had a legal labor dispute with the employer **Pretreatment:** At baseline there were statistically significantly more men in the control group (41.7%) against 33.6% in the intervention group. Otherwise, there were no relevant differences. **Setting:** RCT comparing a web-based, guided self-help intervention with care as usual for workers recruited from different companies in the Netherlands
Interventions	**Intervention characteristics** Psychological intervention: I-CBT plus PST, I-guided • *Content*: Happy@Work: a brief Web-based intervention delivered with minimal guidance; consists of 2 evidence-based treatments: problem-solving treatment (PST) and cognitive therapy (CT) [and a guideline for employees to help them to prevent work-related stress; Web-based lessons has a different theme, but always follows the same structure: information about the theme, examples, and assignments. Themes of the lessons are introduction of problem solving (lesson 1), problem-solving methods (lesson 2), changing cognitions (lesson 3), dealing with work-related problems (lesson 4), social support (lesson 5), and relapse prevention (lesson 6). Happy@Work is a tunnelled intervention, which means that participants can start with a new lesson after they have received feedback on their

Geraedts 2014 *(Continued)*

assignments from a coach.Participants were viewed as treatment completers if they had followed at least the basic information and assignments of PST and CT (completion of lessons 1-3).

- *Duration, frequency, length*: Six weeks, weekly with an option of 1 week extra time in case of delay,
- *Communication means*: Web-based intervention with minimal guidance
- *Providers*: participants receive feedback on assignments from a coach; coaches were trained Master's-level students in clinical psychology. All coaches used a protocol-treatment Manual; to ensure treatment fidelity, all feedback was reviewed by a supervisor (AG) before it was placed on the website.

No intervention

- *Content*: Care as usual group received an email with the randomization outcome only and were advised to consult their (occupational) physician or a psychologist if they wanted treatment for their depressive symptoms. Participants could seek any additional treatment they wanted
- *Duration, frequency, length*: Six weeks
- *Communication means*: none
- *Providers*: none from study point of view

Outcomes

Sickness absence

Sick leave in past 2 months at 6 months follow/up

- **Outcome type**: Continuous outcome

Depressive symptoms

CES-D

- **Outcome type**: Continuous outcome

Notes

Country: the Netherlands

Work performance was measured but this is different from work ability and we did not use it

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "random blocks containing 4, 6, or 8 allocations. An independent researcher made the allocation schedule with a computerized random number generator"
Allocation concealment (selection bias)	Low risk	Quote: "The participants were informed about randomisation outcome via email."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Comment: Unlikely that the participants would change their behaviour based on the intervention
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Comment: unblinded and self-reported outcomes
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: unblinded and self-reported outcomes
Incomplete outcome data (attrition bias)	Unclear risk	Quote: Missing data were handled by multiple imputation via data augmentation. Data augmentation is an iterative Markov chain Monte Carlo method to

Geraedts 2014 *(Continued)*

Depressive symptoms		generate the imputed values assuming a multivariate normal distribution. Five imputations were used in all analyses and reported in the effectiveness analyses. Comment: Loss to follow-up was 60/116 (52%) in intervention group and 65/115 (56%) in the control group at 12 mo follow-up. Even though acceptable imputation used, it is unclear if this repairs the enormous loss to follow-up.
Incomplete outcome data (attrition bias) Sick Leave	Unclear risk	Quote: Missing data were handled by multiple imputation via data augmentation. Data augmentation is an iterative Markov chain Monte Carlo method to generate the imputed values assuming a multivariate normal distribution. Five imputations were used in all analyses and reported in the effectiveness analyses. Comment: Loss to follow-up was 60/116 (52%) in intervention group and 65/115 (56%) in the control group at 12 mo follow-up. Even though acceptable imputation used, it is unclear if this repairs the enormous loss to follow-up.
Selective reporting (reporting bias)	Low risk	Judgement comment: More secondary outcomes stated in protocol than reported: quality of life, social support and locus of control. Unlikely that this has introduced bias
Other bias	Low risk	Judgement comment: No other sources of bias detected

Hees 2013

Study characteristics	
Methods	**Study design:** Two armed RCT. **Recruitment**: Between December 2007 and October 2009, participants were referred by occupational physicians from several occupational health services. **Follow up:** 18 months. Lost to follow up: 13.7%
Participants	**Inclusion criteria**: Age 18 to 65, DSM-IV diagnosis of MDD, Absent from work at least 25% of their contract hours due to their depression. In addition, the duration of the depression had to be at least 3 months or the duration of their sickness absence had to be at least 8 weeks. Finally, there had to be a relation between the depressive disorder en the work situation, that is, work was one of the determinants of depressive disorder and contributed substantially (> 25%), or the depressive symptoms reduced productivity or hindered RTW. **Exclusion criteria:** severe alcohol or drug dependence, bipolar disorder, psychotic disorder, depression with psychotic characteristics, indication of inpatient treatment **Baseline characteristics** 117 were randomised (T1: 39; T2: 78); Age: T1: 41.5 (SD 9.6); T2: 43.8 (SD 9.0) Male: T1: 41%; T2: 53% Education (years): T1: 13.9 (SD 3.7); T2: 13.5 (SD 3.1) Martial status: T1: 59% married or living together; T2: 58% married or living together; T1: 23% single; T2: 28% single; T1: 18% divorced or widowed; T2: 14% divorced or widowed Contract (number of hours): T1: 32.7 (SD 5.8); T2: 35.0 (SD 5.0)

Hees 2013 *(Continued)*

	Absenteeism (number of hours): T1: 27.1 (SD 8.8); T2: 27.6 (SD 10.0) Duration of absenteeism (months): T1: 3.8 (IQR 2.0 - 6.5); T2: 5.0 (IQR 2.8 - 5.0) Occupational sector: financial or insurance: T1: 54%; T2: 58%; Health care: T1: 18%; T2: 9%; Other: T1: 28%; T2: 33% Work experience (years): T1: 14.1 (SD 9.6); T2: 15.9 (SD 11.0) **Setting**: Outpatient; Department of Psychiatry, Academic Medical Center
Interventions	T1: Treatment as usual: treatment by psychiatric residents in an outpatient university clinic according to a treatment protocol consistent with the APA guidelines. 19 visits consisted of clinical management, including psycho education, supportive therapy and cognitive behavioural interventions. Therapies were supervised on a weekly basis by an experienced senior psychiatrist specialised in depression. If needed, participants received pharmacotherapy according to a protocolised algorithm. If the participant's condition deteriorated and outpatient treatment was no longer deemed adequate, he or she was referred to day treatment or inpatient treatment T2: Adjuvant occupational therapy: consisted of 18 sessions (nine individual sessions, eight group sessions and a meeting with the employer), and was conducted by two experienced occupational therapists who had received extensive training in the intervention protocol. During the intervention, the occupational therapist frequently communicated with the occupational physician and the resident treating psychiatric. Employees were recruited to work at least 2 hours per week when starting OT, so that employees were able to directly practise the things learned (e.g. new coping strategies) during therapy
Outcomes	**Sickness absence** 1) work participation, defined in: a) average number of hours of absenteeism over each 6-month period and b) duration of sick leave due to depression in calender days from the start of treatment until partial (or full) RTW. Time until partial or full RTW was operationalised as the duration of sick leave due to depression in calendar days from the start of treatment until partial (or full) RTW. Partial RTW was defined as working an increment of at least 5 hours (compared with hours worked at baseline), for at least 4 weeks without partial or full recurrence. Full RTW was defined as working the full number of contract hours in own or other work for at least 4 weeks, without partial or full recurrence **Depressive symptoms** 1) severity of depression, assessed by the Hamilton Rating Scale for depression (HRSD) 2) depression remission, defined as having HRSD ≤ 7 3) severity of depression, assessed by the Questionnaire Inventory of Depressive Symptoms Self-Report (StIDS-SR) **Functioning:** 1) at work functioning: weekly self-report records of work efficiency on a scale 1-0 and 3 sub scales of WLQ: Output, time, mental-interpersonal 2) health-related functioning, 3 subscales of MOS-SF 36: role limitations due to emotional problems, mental health, role limitations due to physical problems
Notes	**Country**: The Netherlands

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomization was conducted by an independent research assistant, using software based on a minimization randomisation procedure."

Hees 2013 *(Continued)*

Allocation concealment (selection bias)	Low risk	"Randomization was conducted by an independent research assistant, using software based on a minimization randomisation procedure."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	"Due to the nature of the intervention, neither patients nor therapists could be blinded to the patient's allocation status." However, it is unlikely that patients or providers will have changed their behaviour based on knowledge of the intervention
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sickness absence data are measured by the use of self-report. As patients are not blinded for the allocation status, risk of bias is high
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	"Study assessment were conducted by a psychiatrist and a researcher who where blind to group allocation." As the HRSD is a clinician-rated instrument, there is a low risk of bias for the HRSD outcome
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Lost to follow up: T1: 15.4%; T2: 12.8% but appropriate imputation methods have been used. "To take potential biased outcomes caused by selective loss to follow up into account, we used multiple imputation (five imputed datasets), which, assuming missing at random for missing values, gives unbiased results with correct SEs."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Lost to follow up: T1: 15.4%; T2: 12.8% but appropriate imputation methods have been used. "To take potential biased outcomes caused by selective loss to follow up into account, we used multiple imputation (five imputed datasets), which, assuming missing at random for missing values, gives unbiased results with correct SEs."
Selective reporting (reporting bias)	Low risk	The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way
Other bias	Low risk	None identified

Hellstrom 2017

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** Via mental health centres and psychiatrists **Follow-up:** 12 and 24 months **Subgroup participants with depression:** Authors reanalyzed the data for the subgroup of participants with depression only: At 12 months follow-up there were 113 participants with depression in the intervention group and 113 in the control group. At 24 months follow-up these numbers were 113 and 112 respectively. For depression score the were 87 and 77 at 12 months and 87 and 73 at 24 months follow-up.
Participants	**Baseline characteristics** Work-directed intervention: Individual Placement and Support (IPS)

Hellstrom 2017 *(Continued)*

- *Age*: 34 ± 10
- *Gender*: 29% male
- *Marital status*: 38% married
- *Occupation*: no info
- *Sick leave status*: 54% sickness benefit
- *Number of participants randomised*: 162
- *Number with depression: 113*

Care as Usual

- *Age*: 36 ± 11
- *Gender*: 35% male
- *Marital status*: 36% married
- *Occupation*: no info
- *Sick leave status*: 61% sickness benefit
- *Number of participants randomised*: 164
- *Number with depression: 113*

Inclusion criteria: (1) age 18–60 years; (2) diagnosis of affective disorder (International Classification of Diseases, 10th Revision (ICD-10): F30-39) or anxiety (ICD-10: F40-41); (3) no contact with mental health services for more than the past 3 years; (4) employed or enrolled in education at some time during the past 3 years (this criterion was changed during the trial from originally 2 years); (5) motivated to return to work or education; (6) not ready to return to work within 3 months after inclusion (equal to match group 2 or 311; used by the job centres in Denmark to estimate how far from the labour market people are. Match group 2: able to participate in pre-vocational training but not able to work and be off public benefits within 3 months. Match group 3: severe long-term problems; cannot work or participate in pre-vocational training); (7) able to read and understand Danish and(8) give informed consent.

Exclusion criteria: (1) somatic co-morbidity causing reduced ability to work; (2) primary large-scale alcohol or substance abuse and(3) legal guardian or forensic psychiatric arrangements

Pretreatment: The two groups were comparable at baseline

Setting: Participants were recruited from mental health centres (inpatients and outpatients) and private practising psychiatrists within the Capital Region of Denmark, from 1 October 2011 until February 2013 (inclusion period was extended with 5 months)

Interventions

Intervention characteristics

Work-directed intervention: IPS

- *Content*: Briefly, the intervention consisted of mentor support and career counselling, providing five basic services: individualised mentor support based on psychiatric knowledge; coordination of services provided; career counselling; impartial help to clarify private economy; and contact with employers to help participants obtain jobs and keep them. Focus was on competitive employment and support was time unlimited
- *Duration, frequency, length*: The number and duration of contacts depended on the individual needs; most met with their mentor once a week for 1 to 1 ½ hours.
- *Communication means*: Face to face
- *Providers*: Mentors had a minimum of 10 years' experience from mental health services, as nurses, social workers or occupational therapists. Career counsellors had many years of experience from workplace career counselling or human resources in the private sector. Mentors and career counsellors worked closely together.Newly appointed mentors and career counsellors had a 2-week introduction to working routines and the IPS-MA method. Team members received monthly supervision provided by a trained psychologists

Care as usual: WD

Hellstrom 2017 *(Continued)*

- *Content*: Services as offered by the job centres in Denmark, for instance, courses, company internship programs, wage subsidy jobs, skill development and guidance, mentor support or gradual return to employment.
- *Duration, frequency, length*: Not measured
- *Communication means*: Not measured
- *Providers*: See above

Outcomes	**Sickness absence** Number returned to work • **Outcome type**: Dichotomous outcome **Depressive symptoms** *HAM depression score* • **Outcome type**: Continuous outcome
Notes	**Country:** Denmark Return to work numbers and HAM depression score provided by the authors for the subgroup that was diagnosed with depression

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "computer-generated allocation sequence with varying block sizes of 4, 6 and 8,"
Allocation concealment (selection bias)	Low risk	Quote: "concealed from the investigators."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Quote: It was not possible to blind participants, mentors, career counsellors or care providers. Outcome assessors and research team were blinded to allocation throughout the trial period, data collection and statistical analysis. Self-reported online surveys were answered using an identification number enabling the research team to remain blinded. The randomization code was broken when all analyses were completed, and two conclusions had been drawn. Comment: Unblinded, but unlikely that the knowledge of the intervention will have changed the behaviour of patients or providers.
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Comment: Objective outcome from administrative database
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: Subjective self-reported outcome
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Quote: All analyses were conducted according to the intention-to- treat (ITT) principles. According to the protocol, 10 we would use mixed model with repeated measurements to handle missing data, but we chose to use multiple imputations, since we believed that this would give us a better estimate of the missing values. 28 Predictions were based on variables with full information indicative of missing values; 100 imputations were made. If more than 50% was missing, we chose to report results based on the actual data, but com-

Hellstrom 2017 *(Continued)*

		pared these with results based on multiple imputations, both being prone to bias, results did not differ. We had complete data on all register data.
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Quote: All analyses were conducted according to the intention-to- treat (ITT) principles. According to the protocol, 10 we would use mixed model with repeated measurements to handle missing data, but we chose to use multiple imputations, since we believed that this would give us a better estimate of the missing values. 28 Predictions were based on variables with full information indicative of missing values; 100 imputations were made. If more than 50% was missing, we chose to report results based on the actual data, but compared these with results based on multiple imputations, both being prone to bias, results did not differ. We had complete data on all register data.
Selective reporting (reporting bias)	Low risk	Judgement comment: Protocol published. No differences with adjusted protocol
Other bias	Low risk	Judgement comment: No other sources of bias detected No other sources of bias detected

Hollinghurst 2010

Study characteristics	
Methods	**Study design**: RCT. **Recruitment**: patients were recruited from 55 general practices in Bristol, London, and Warwickshire between October 2005 and February 2008. **Follow up**: 8 months. Lost to follow up: 53% for sickness absence and 29% for clinical outcomes
Participants	**Inclusion criteria:** patients between 18 and 75 who where identified in primary care as having a new episode of depression which was defined as being diagnosed within the 4 weeks preceding referral. Depression was defined as a score of 14 or more on the BDI12 and an ICD-10 diagnosis of depression using the CIS-R. **Exclusion criteria:** patients treated for depression in the 3 months before the present episode, patients with a history of bipolar disorder, psychotic disorder, alcohol or substance misuse, and those already receiving psychotherapy **Baseline characteristics** 297 were randomised (T1: 149; T2: 148). Setting: patients between who where identified in primary care as having a new episode of depression Female: T1: 69%; T2: 67% Age: T1: 35.6 (SD 11.9); T2: 34.4 (SD 11.3) Marital status: T1: 34% married; T2: 39% married T1: 50% single; T2: 47% single T1: 16% separated or divorced or widowed; T2: 15% separated or divorced or widowed Employment status: T1: 65% employed; T2: 56% employed T1: 15% student; T2: 24% student

Hollinghurst 2010 *(Continued)*

	T1: 20% not in employment; T2: 20% not in employment Highest educational level: T1: 65% A level or above; T2: 63% A level or above T1: 32% other; T2: 33% other T1: 3% no educational qualifications; T2: 4% no educational qualifications **Setting**: primary care
Interventions	T1: Online CBT in addition to usual care: participants receiving online CBT were offered up to ten sessions of 55 minutes, to be completed within 4 months from the date of randomization when possible. Each participant was assigned their own therapist (psychologist) for the duration of the study. Participants and therapists typed free text into the computer, with messages sent instantaneously, using only this means of communication T2: Usual care from GP while on a 8-month waiting list for online CBT: participants on the waiting list were not to receive psychotherapy during the study follow-up period. Those on the waiting list who had still an eligible Beck Depression Inventory (BDI) score after 8 months were offered the intervention at that time
Outcomes	**Sickness absence** 1) the number of working days lost because of depression (time off work) over 8 months **Depressive symptoms** 1) depression severity, assessed by the BDI 2) recovery, defined as a score of less than 10 on the BDI
Notes	**Country:**UK

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomization was by means of a computer-generated code, implemented by an individual who was not involved in the recruitment process, and communicated to the participant within 48 h of the baseline interview."
Allocation concealment (selection bias)	Low risk	"Randomization was by means of a computer-generated code, implemented by an individual who was not involved in the recruitment process, and communicated to the participant within 48 h of the baseline interview." "The allocation was concealed in advance from participants, researchers involved in recruitment, and therapists."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Comment: unblinded and the control condition (usual care and waiting list) was deemed less desirable,
Blinding of outcome assessment (detection bias) Sick Leave	High risk	The number of working days lost because of depression was recorded in a diary by the participants themselves. As participants were aware of their intervention status, risk of bias high
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	The BDI is a self-report inventory. As participants were aware of their intervention status, risk of bias high

Hollinghurst 2010 *(Continued)*

Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up is high: T1: 27%; T2: 32% even though appropriate method has been used to account for these missing data: "Fourth, a sensitivity analysis investigated the effect of missing data with multiple imputation by chained equation methods in Stata." "Analyses imputing missing values suggested that differences in attrition between the groups did not introduce any noticeable bias."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up is high: T1: 50%; T2: 55% even though appropriate method has been used to account for this missing data: "we imputed missing observations of cost and QALYs using the multiple imputation by chained equation procedure in Stata release 10." "We acknowledge that more complete data would have been available if we had used questionnaires completed face to face or data from practice records. However, the results of the imputation suggest that any information lost is unlikely to have a major influence on the results or conclusions."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Kaldo 2018

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 3 **Recruitment:** Patients (unclear which) were screened in primary health care centers and further assessed if they scored 10 or above on the PHQ-9 **Follow-up:** 3 and 12 months
Participants	**Baseline characteristics** Psychological intervention: I-CBT, I-guided • *Age*: 42.1% (18-34 year) • *Gender*: 34.2% male • *Marital status*: not reported • *Occupation*: not reported • *Sick leave status*: 7.3% long-term sick leave • *Number of participants randomised*: 317 Care as usual • *Age*: 40% (18-34 year) • *Gender*: 45% male • *Marital status*: not reported • *Occupation*: not reported • *Sick leave status*: 5.7% long-term sick leave • *Number of participants randomised*: 312 Clinical intervention: Exercise

Kaldo 2018 *(Continued)*

- *Age*: 25% (18-34 year)
- *Gender*: 40% male
- *Marital status*: no reported
- *Occupation*: not reported
- *Sick leave status*: 10.4% long-term sick leave
- *Number of participants randomised*: 316

Inclusion criteria: For the original trial inclusion:10 or above on the Patient Health Questionnaire-9 at an initial screening for depression, age (≥18 years), no severe somatic illness, no primary alcohol or drug use disorder, and no psychiatric diagnosis requiring specialist treatment. In this secondary analysis, the trial is restricted to those who were employed at baseline and 3 months and 12 months follow-up.

Exclusion criteria: none reported

Pretreatment: 10% more women, 7% more employed, 8% less anti-depressant use in the control group in the sample restricted to employed persons; exercise group younger

Setting: Primary Care

Interventions

Intervention characteristics

Clinical Intervention: Psychological intervention, I-CBT, I-guided

- *Content*: The treatment was based on self-help text modules, each based on established CBT principles and presenting information on a specific problem area, useful methods to handle it and an online homework report. four of the modules aimed at managing problems related to work and sick leave: social insurance agency—participants on sick leave could receive this module, which included information about the sick leave process and homework assignments about initiating better communication with the authority; returning to work—participants on sick leave also could receive this module about the transition of going back to work, including home-work assignments about keeping contact with the employer; handling problems at work
- *Duration, frequency, length*: 12 weeks, weekly assignments, length variable
- *Communication means*: Internet, telephone
- *Providers*: Active support from a therapist: a clinical psychologist or last-year psychology student under supervision

Care as usual

- *Content*: Primary care standard treatment for depression determined by the patient's general practitioner. It could for example include antidepressants, counselling with a CBT focus conducted for about 1 hour and group-based interventions. Twenty-five percent of patients in this group received no recorded treatment
- *Duration, frequency, length*: Variable
- *Communication means*: Variable
- *Providers*:

Clinical intervention, Exercise, Aerobic exercise

- *Content*: Within the PE arm, patients were randomly allocated to one of three levels of exercise: 'light' (yoga or similar), 'moderate' (inter-mediate level aerobics) and 'vigorous' (higher intensity aerobics). In this study, all three groups were analyzed together.
- *Duration, frequency, length*: 12 weeks, 3 times per week, 60 minutes
- *Communication means*: Face-to-face, individual guidance
- *Providers*: Qualified personal trainer:, ICB

Outcomes

Sickness absence

On sick leave: Long-term sick leave past month

Kaldo 2018 *(Continued)*

- **Outcome type**: Dichotomous outcome

The authors reported only the percentage of employed participants on long-term sick leave and the number of events. We recalculated the number of participants based on these numbers and used this as input in RevMan data-tables. The authors also reported on estimated number of days absent per month, but deemed these estimates to be too imprecise to capture effects over time.

Work functioning: Work ability

- **Outcome type**: Continuous outcome
- **Scale**: work ability index
- **Range**: 0 to 10
- **Direction**: Higher is better

Notes	**Country:** Sweden

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "patients)" Quote: "trial research organisation where a computer program generated the group allocation, with the size of the randomization blocks (36 patients) being unknown to the"
Allocation concealment (selection bias)	Low risk	Quote: "When the research nurse had entered all assessment data into the study database and confirmed that all inclusion criteria were met, the allocation for the new patient was revealed via the user interface of the database."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Unblinded, but unlikely that patients or providers changed their behaviour based on knowledge of the intervention
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Comment: Patients' self-reported and unblinded to intervention
Blinding of outcome assessment (detection bias) Depressive symptoms	Unclear risk	Depressive symptoms not assessed
Incomplete outcome data (attrition bias) Depressive symptoms	Unclear risk	Depressive symptoms not assessed
Incomplete outcome data (attrition bias) Sick Leave	High risk	Comment: Large attrition: PE 30%, ICBT 25%, TAU 37%
Selective reporting (reporting bias)	High risk	Judgement comment: Protocol retrospectively published according to trials register (article says 'preregistered'), no mention of work outcomes; work outcomes constructed after first data-analysis
Other bias	High risk	Judgement comment: Subgroup analysis of those employed only that were not prespecified

Kendrick 2005

Study characteristics	
Methods	**Study design:** RCT, randomization on the level of patients stratified for referring GP; 3 conditions. **Recruitment:** general practices referred patients to the study. CMHNs were employed by local NHS trusts. Follow up: 26 weeks. **Lost to follow up**: 26%
Participants	**Inclusion criteria:** age: 18-65; new episode of anxiety, depression or reaction to life difficulties; minimum duration symptoms: 4 weeks; maximum duration symptoms: 6 months; GHQ-12 score at least 3 **Exclusion criteria:** patient already in contact with psychiatric services; Patient already receiving psychological treatment; Severe mental illness such as schizophrenia, manic-depressive psychosis; severe substance misuse, dementia or severe depression with active suicidal ideas; housebound patients; patients without the spoken and written language skills necessary to participate; seriously ill and terminally ill patients; temporary residents **Baseline characteristics** 247 randomised (T1: 90; T2: 79; T3: 78) Mean age: T1: 35.8 (SD 10.92); T2: 34.2 (SD 11.33); T3: 34.9 (SD 11.77) Female: T1: 72%; T2: 70%;T3: 69% Married or cohabiting: T1: 60%; T2: 58%; T3: 48% Fulltime or part-time employed: T1: 66%; T2: 75%; T3: 69% **Setting:** community mental health, UK.
Interventions	Improved care by additional community health nurses T1: CMHN problem-solving treatment 1. explanation of treatment and rationale 2. clarification and definition of problems 3. choice of achievable goals 4. generations of alternative solutions 5. selection of preferred solution 6. clarification of necessary steps to implement solution 7. evaluation of progress; Initial 1-hour session + 5 follow-up sessions of 30-45 minutes. T2: Generic CMHN; nurses were asked to use whatever treatment they were experienced in giving; initial 1-hour session + 5 follow-up sessions of 30 to 45 minutes. Range 0 to 8 sessions T3: GP care: usual care, but asked not to refer patients to a psychological therapist during the study period unless absolutely necessary
Outcomes	**Sickness absence** 1) number of days off paid work **Depressive symptoms** 1) CIS-R 2) HADS-D **Work functioning** 1) SAS, however, sub-scale "work outside the home" not separately reported
Notes	**Country:** UK Personal communication: data for depressed sub-sample was provided. In our analyses, the two CMHN subgroups were combined, leaving two study arms.
Risk of bias	

Kendrick 2005 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The telephone randomization service at the university of York was contracted."
Allocation concealment (selection bias)	Low risk	"Remote central randomization was provided by telephone" "Randomisation sequences were in block sizes of either three or six, to prevent practitioners from guessing to which arm the next referral would be."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	High risk for the comparison with the GP usual care group (T3) as this treatment cannot be considered equally desirable as T1 and T1 for patients and patients were not blinded. "Table 16: n = 50 received their preferred treatment; n = 114 did not receive their preferred treatment; n = 83 reported no preference"
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "Number of days off paid work was captured by a resource-use questionnaire filled out by patients." "Patients were reminded not to reveal their allocation at the follow-up assessments."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression symptom score (CIS-R and HADS-D) were measured by self-report and patients were not blinded. "The computerised version of the CIS-R, which is self-complete, was used in this study." "Patients were reminded not to reveal their allocation at the follow-up assessments."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up is considered to be high (26%). Risk of attrition bias due to follow-up losses is therefore considered to be high, although sensitivity analyses were conducted and the authors conclude otherwise; "sensitivity analyses were conducted to see whether the result changed depending on what assumptions were made about the missing data". "Table 12 shows that the main findings are not particularly sensitive to the different assumptions about missing data that were investigated." It was harder to retain patients in the GP care (thus higher loss to follow up in that group): "Although the overall follow-up rates were good, there was a lower follow-up rate in the GP arm. It is difficult to tell whether this biased the findings in a particular direction. Follow-up rates were better among those patients who received the treatment they preferred, so it is likely that there were more disaffected patients in the GP care arm. However, it is not known whether those who dropped out remained more symptomatic than those who were followed up. Failing to receive their treatment of preference was not associated with a worse outcome on the CIS-R among those who were followed up. The sensitivity analyses suggest that CMHN care, whether generic care or specific PST, is unlikely to be more effective than GP care, unless one believes the LOCF analysis and makes the extreme assumption that all the dropouts remained as symptomatic as they were at the time of last assessment."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up for sick leave data is considered to be high (26%). Risk of attrition bias due to follow-up losses is therefore considered to be high, although sensitivity analyses were conducted and the authors conclude otherwise; "cost results from this analysis were validated by substituting where possible data from the GP case notes in place of imputed values for missing data, and repeating the analysis. Overall, the results did not change significantly." "36% had at least one resource item missing over the 6-month follow up. Therefore, complete resource use data were available for 159 (64%) of the

Kendrick 2005 *(Continued)*

		patients. The results presented here are based mainly on the 184 patients for whom complete CIS-R data were available over the 6-month period. To achieve this sample, 25 (14%) of the patients who had CIS-R data but not resource-use information had to be imputed. The results were then compared with those obtained using data from GP notes where available instead of imputation, and those obtained using only the 159 patients with complete resource-use data. After imputing missing values for the 25 patients with missing resource-use data, the numbers of patients included in the economic analysis in each group were as follows: 51 patients in GP care (28%), 62 patients in generic CMHN care (34%) and 71 patients in PS CMHN care (38%)."
Selective reporting (reporting bias)	Low risk	No indication for selective reporting could be identified. However, in the design study, the comparisons of T1 with T2 was not pre-specified
Other bias	Low risk	None identified

Knapstad 2020

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Follow-up:** At 3 and 6 months after start of the intervention **Number of trial arms:** 2 **Recruitment:** Information about the trial was provided on the municipality web pages, in local newspapers, and on local radio. All GPs in the catchment areas were informed through an information letter from the National Institute of Public Health and directly by the service providers at local GP association meetings. All clients contacting PMHC in Sandnes and Kristiansand, both GP- and self-referred, got an appointment for individual assessment at the PMHC clinic. In this detailed screening and assessment, one of the therapists conducted a clinical interview with the client. The therapist identified the relevance and severity of the mental health problems, the available client resources, and motivation for treat-ment. The client received information about the study and the treatment methodology within PMHC. To minimize the nocebo effect, comprehensive information about the rationale for randomisation was provided. The therapist then reviewed all information and decided on inclusion/exclusion in consultation with the client.
Participants	**Baseline characteristics** Enhanced care • *Age*: 34.6 (SD 11.8) • *Gender Female*: Female 65.7% • *Marital status (having a partner)*: Having a partner 55.1% • *Occupation (having regular work)*: N.R. • *Sick leave status (functional status)*: N.R. • *Number of participants randomised*: 463 • *Number depressed participants only randomised*: 417 Care as Usual • *Age*: 35.3 (SD 13.1) • *Gender Female*: Female 68.4% • *Marital status (having a partner)*: Having a partner 58.9% • *Occupation (having regular work)*: N.R.

Knapstad 2020 *(Continued)*

- *Sick leave status (functional status)*: N.R.
- *Number of participants randomised*: 219
- *Number depressed participants only randomised*: 199

Overall

- *Age*: 34.8 (SD12.2)
- *Gender Female*: Female 66.5%
- *Marital status (having a partner)*: Having a partner 56.3%
- *Occupation (having regular work)*: N.R.
- *Sick leave status (functional status)*: N.R.
- *Number of participants randomised*: 774
- *Number depressed participants only randomised*: 616

Inclusion criteria: PHQ-9/GAD-7 scores above cutoff level; being 18 years of age or above and a resident in one of the pilot site municipalities; basic verbal and oral Norwegian proficiency

Exclusion criteria: Entitled to secondary care services due to eating disorder, suicide risk, bipolar disorder, severe depression, invaliding anxiety, psychotic symptoms, severe substance abuse, personality disorder, two or more previous treatment attempts without effect, or serious physical health problem as prime problem

Pretreatment: Not tested, authors: "As displayed in Table 2, baseline demographic and clinical characteristics were generally similar across the two treatment groups".

Setting: General Practice and Municipal communities

Interventions

Intervention characteristics

Enhanced care (Prompt Mental Health Care, PMHC)

- *Content*: CBT. Access to care within 48 hrs. Choice (of therapist and client combined) of either group-based psycho-education, guided self-help, individual treatment or combination. Most start with four-sessions psychoeducational course.
- *Duration, frequency, length*: Variable, median of 5 sessions. 77.% received 4 sessions or more. The median number of treatment sessions was lowest for guided self-help (1.5, IQR = 1–5), medium for group-based psychoeducation (4, IQR = 3–4), and high-est for individual CBT (7, IQR = 4–10) and mixed treatment (9, IQR = 7–12).
- *Communication means*: Face-to-face and paper/Internet self-help
- *Providers*: PMHCtherapists. Each had minimum of 3 years of relevant higher education and had completed an additional mandatory 1-year training in CBT including an IAPT-based curriculum, adjusted to the Norwegian context. All therapists had individual treatment responsibilities. Clinical psychologist supervised.

Care as usual

- *Content*: ll ordinary services available to the target population. In the two included municipalities, this usually included follow-up by the GP, or alternatively by pri-vate psychologists or occupational health services. After randomisation, the TAU group received a response letter in which they were encouraged to contact the GP for further follow-up as well as references to publicly available self-help resources (internet, books).
- *Duration, frequency, length*: 59% received help, 45% of alle patients in TAU recieved four sessions or more.
- *Communication means*: Face-to-face and self-help
- *Providers*: Not pre-specified, but of the 59% receiving help, 47% had sessions with a General practitioner and 39% with a psychologist/psychiatrist.

Outcomes

At work

- **Outcome type**: Dichotomous Outcome

Knapstad 2020 *(Continued)*

- **Reporting**: Fully reported
- **Direction**: Higher is better

Depressive symptoms

Patient Health Questionnaire Depression Score

- **Outcome type**: Continuous Outcome

Notes	**Country:** Norway

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Quote: "The participants were randomized on a 70: 30 ratio (PMHC versus usual GP care [TAU]) with simple randomization within each of the two sites with no further constraints. A computerized random number generator was used for group assignment."
Allocation concealment (selection bias)	Low risk	Quote: "Full allocation concealment was achieved by using the web-based central allocation ap- plication that was integrated in the data portal from the Norwegian Social Science Data Services."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Quote: Participants were subsequently informed about their allocation – PMHC clients by their assigned therapist and TAU clients through a standardized letter that was sent by mail by the project coordinator. Because of the nature of the intervention, participants and therapists could not be blinded to treatment. Comment: The TAU condition, ordinary services available to target population, cannot be considered considered equally desirable as the intervention, prompt mental health care (PMHC) Personnel could not be blinded. Unclear if this would have led to change of behaviour
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Quote: Work participation was assessed by means of two questions, one multi-response item about current work status and one multi-response item about sources of income. Based on these two questions, it was determined whether participants were in full- or part-time regular work without receiving benefits or not (coded as a binary variable). Commetn: self-report and unblinded
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: All outcomes were self-reported and could be influenced by the knowledge or participants belonging to the intervention group
Incomplete outcome data (attrition bias) Depressive symptoms	Unclear risk	Comment: Intervention group: 73% data complete (27% lost to follow up) TAU: 69% data complete (31% lost to follow up)
Incomplete outcome data (attrition bias) Sick Leave	Unclear risk	Quote: Altogether slightly more outcome data were available in the PMHC than the TAU group (data available at 3- and/or 6-month follow-up for 73 vs. 67%, respectively). Comment: Intervention group: 73% data complete (27% lost to follow up) TAU: 69% data complete (31% lost to follow up)
Selective reporting (reporting bias)	Low risk	Quote: " Statement of Ethics:The trial protocol was approved by the Regional ethics commit- tee for Western Norway (REK-vest No. 2015/885) and the trial is

Knapstad 2020 *(Continued)*

		registered at ClinicalTrials.gov (NCT03238872). No changes were made to the primary or secondary outcomes after trial approval. All participants gave their written consent" Judgement comment: The outcomes listed in the trial protocol were reported in the publication
Other bias	Low risk	Judgement comment: No other sources of bias perceived

Knekt 2013

Study characteristics	
Methods	**Study design:** RCT. **Recruitment**: a total of 459 eligible outpatients were referred to the Helsinki Psychotherapy Study from psychiatric services in the Helsinki region from June 1994 to June 2000. Follow up: 5 years. **Lost to follow up**: 19% (for all participants over five years), lost to follow up for the subgroup of people with depressive disorder: 51% (over five years)
Participants	**Inclusion criteria:**: 20 to 45 years of age and suffered from a longstanding (> 1 year) disorder causing dysfunction in work ability. They were also required to meet DSM-IV criteria for anxiety or mood disorders **Exclusion criteria:** psychotic disorder or severe personality disorder, adjustment disorder, substance-related disorder, organic brain disease or other diagnosed severe organic disease, and mental retardation. Individuals treated with psychotherapy within the previous 2 years and psychiatric health employees were also excluded **Baseline characteristics** 326 were randomised (T: 97; T2: 101; T3: 128). Subgroup of people with depressive disorder: 161. Age: T1: 33.6 (SD 7.2); T2: 32.1 (SD 7.0); T3: 31.6 (SD 6.6) Male: T1: 25.8%; T2: 25.7%; T3: 21.1% Employed or student: T1: 83.2%; T2: 85.1%; T3: 75.4% Academic education: T: 28.9%; T2: 19.8%; T3: 75.4% **Setting:** outpatient.
Interventions	T1: Solution-focused therapy: is a brief, focal, transference-based therapeutic approach which helps patients by exploring and working through specific intrapsychic and interpersonal conflicts. The therapy included one session every second or third week, with a limit of 12 sessions, over no more than 8 months T2: Short-term psychodynamic psychotherapy: is characterized by the exploration of a focus, which can be identified by both the therapist and the patient. This consists of material from current and past interpersonal and intrapsychic conflicts and the application of confrontation, clarification, and interpretation in a process in which the therapist is active in creating the alliance and ensuring the time-limited focus. The therapy was scheduled for 20 weekly treatment sessions over 5 to 6 months T3: Long-term psychodynamic psychotherapy: is an open-ended, intensive, transference-based therapeutic approach which helps patients by exploring and working through a broad area of intrapsychic and interpersonal conflicts. The therapy is characterized by a framework in which the central elements are exploration of unconscious conflicts, developmental deficits, and distortions of intrapsychic structures. Confrontation, clarification and interpretation are major elements, as well as the therapist's ac-

Knekt 2013 *(Continued)*

	tions in ensuring alliance and working through the therapeutic relationship to attain conflict resolution and greater self-awareness. Therapy includes both expressive and supportive elements, the use of which depends on patient needs. The frequency of sessions was 2 to 3 times a week, and the duration of the therapy was up to 3 years
Outcomes	**Sickness absence** 1) number of sick-leave days during last 3 months Depressive symptoms: 1) depressive symptoms assessed by the Beck Depression Inventory (BDI) 2) depressive symptoms assessed by the Hamilton Depression Rating Scale (HDRS) **Work Functioning:** 1) the work subscale (SAS-work) of the social adjustment scale (SAS-SR)
Notes	Country: Finland

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Concealed assignment codes were given sequentially to patients in consecutively numbered envelopes."
Allocation concealment (selection bias)	Low risk	"The patients who fulfilled the selection criteria at baseline were randomized into solution-focused therapy, short-term psychodynamic psychotherapy or long-term psychodynamic psychotherapy or long-term psychodynamic psychotherapy in a 1:1:1.3 ratio using a central computerized randomization schedule. Concealed assignment codes were given sequentially to patients in consecutively numbered envelopes."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Due to the nature of the intervention, the participants and personnel could not be blinded, however it is unlikely that this would have changed their behaviour.
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and the patients were not blinded for their allocation status. Outcome is likely to be influenced by this lack of blinding. "The number of sick leave days from work during the past 3 months were collected by single-item questions included in a follow-up questionnaire developed in the project." "Unavoidable weaknesses in a study like this are [...] the lack of blindness of assessments."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	The BDI is a self-report inventory and patient were not blinded for their allocation status. Outcome is likely to be influenced by this lack of blinding. The HDRS is a clinician-administered scale but clinicians were also not blinded: "raters were not blinded since they were provided with information on the treatment group at the five interview sessions during the 3-year follow up."
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Loss to follow up is 19% and missing values were replaced by multiple imputation; this did not alter the results. "Analyses based on multiple imputation and taking into account the need for treatment at the time of dropout did not, however, notably alter the results, suggesting that the results presented are unbiased (data not shown)."
Incomplete outcome data (attrition bias)	High risk	Loss to follow up is considered to be high: 39% at one year and 52% at five years

Knekt 2013 *(Continued)*
Sick Leave

Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Krogh 2009

Study characteristics	
Methods	Randomised pragmatic trial. Recruitment: between January 2005 and July 2006. Follow-up: 12 months. Lost to follow up: 17% at 4 months and 22% at 12 months
Participants	**Study design**: **Inclusion criteria:** age 18-55 years, referred by a medical doctor or psychologist, meeting ICD-10 criteria for unipolar depression, living in the Greater Copenhagen catchment area, able to read and understand informed consent. **Exclusion criteria:** being engaged in regular sports activity for more than 1 hour per week, ongoing alcohol or substance abuse judged to be at risk of suicide, poor Danish language skills, having a medical condition that contraindicated physical exercise, or had been on sickness leave for than 24 consecutive months **Baseline characteristics** 165 were randomised (T1:55; T2:55; T3:55) Age: T1: 41.9 (SD 8.7); T2: 38.1 (SD 9.0); T3: 36.7 (SD 8.7) Female: T1: 81.8%; T2: 78.2%; T3: 61.8% Ethnicity: T1: 90.9% Caucasian; T2: 92.7% Caucasian; T3: 90.9% Caucasian Occupational status: T1: 41.8% unemployed; 40% full time work; 14.5% part-time work; 3.6% < 20 hrs/wk T2: 54.5% unemployed; 32.7% full time work; 10.9% part-time work; 1.8% < 20 hrs/wk T3: 36.4% unemployed; 41.8% full time work; 18.2% part-time work; 3.6% > 20 hrs/wk **Setting**: outpatient
Interventions	T1: Supervised strength training. Designed to increase muscular strength, initially with 12 repetitions of 50% of repetition maximum 2 or 3 times per exercise. As the patients progressed, the numbers of repetitions were reduced to 10 and 8, with an increase of RM to 75%. The training was a circuit-training program with 6 exercise on machines involving large muscle groups. As a supplement to this, free weights and sandbags were used for exercising the calf muscles, the arm abductors, the triceps muscles, and the hip abductors. All patients were scheduled to meet twice per week during a 4-month period for a total of 32 sessions T2: Aerobic training. Designed to increase fitness as measured by maximal oxygen uptake. The program involved 10 different aerobic exercises using large muscle groups. Machines were used for cycling, running, stepping, abdominal exercises, and rowing. Additional exercises were sliding movements on small carpets, trampoline, step bench, jump rope, and Ski Fitter. During the first 8 sessions, each exercise was done twice for 2 minutes with a 2-minute rest at an intensity level of 70% of maximal heart rate. This gradually increased to a level at which exercise was done for 3 minutes with a 1-minute rest at

Krogh 2009 *(Continued)*

	an intensity level of 89% during the last 8 sessions. All patients were scheduled to meet twice per week during a 4-month period for a total of 32 sessions T3: Relaxation training. Designed to avoid muscular contractions or stimulation of the cardiovascular system, and the patients did not engage in activities perceived higher than 12 on the Borg Scale. The first 20 to 30 minutes were used for exercises on mattresses or Bobath Balls or back massage using a Ball Stick Ball. This was followed by light balance exercises for 10 to 20 minutes and by relaxation exercises with alternating muscle contraction and relaxation in different muscle groups while lying down for 20 to 30 minutes
Outcomes	**Sickness absence** 1) self-reported percentage of days absent from work during the last 10 working days at 4 and 12 months 2) off work: a) % on sick leave b) % unemployed **Depressive symptoms** 1) severity of depression, assessed by the Hamilton Rating Scale for Depression (HAM-D17) 2) remission, defined as not fulfilling the ICD-10 criteria for depression and having a HAM-D17 < 8 3) severity of depression, assessed by the Beck Depression Inventory (BDI)
Notes	Country: Denmark

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomization was carried out by the CTU using computerized restricted randomization with a block size of 6. The block size and thus the allocation sequence were unknown to the DEMO trial staff." "The strengths of our trial were the centralized randomization, which provided adequate generation of the allocation sequence and adequate allocation concealment"
Allocation concealment (selection bias)	Low risk	"Randomization was centralized and stratified according to medicine status." "The strengths of our trial were the centralized randomization, which provided adequate generation of the allocation sequence and adequate allocation concealment"
Blinding of participants and personnel (performance bias) Sick Leave	High risk	"The same 2 physiotherapists were used throughout the trial period. The type and number of exercise interventions were distributed evenly between the two, and thus the physiotherapists were not blinded to allocation". "And the patients were instructed not to reveal their group assignment." "The lack of blinding of treatment allocation for patients and psychotherapists could lead to collateral interventions, possibly confounding our results." As the relaxation condition was not equally desirable to patients as the other two groups, the risk of performance bias is considered high
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Absenteeism measured by self-report. As patients were aware of their allocation status, risk of bias high
Blinding of outcome assessment (detection bias)	Low risk	For HAM-D17: "The assessor was blinded to intervention group, and the patients were instructed not to reveal their group assignment. After assessment

Krogh 2009 *(Continued)*

Depressive symptoms		the assessor was requested to guess which group the patient has been assigned to, making it possible to examine whether the blinding was successful [..] This indicated that the blinding of the assessors was successful"
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up at endpoint was high: 22% (36/165) and skewed. Risk of attrition bias was therefore considered high although an appropriate method was used to deal with missing values in the analyses and the authors conclude otherwise. "Analysis of age, sex, HAM-D17, or absence from work during the last 10 working days at entry did not suggest any significant differences between missing participants and participants included in the analysis at either 4 months or 12 months." "It is then plausible to consider the missing data as 'missing at random,' making the mixed effect model a plausible approach to estimate the effect, based on the total sample with missing cases included." "This approach uses data from all included patients (intention-to-treat), handles entry differences, and is able to handle missing data (restricted maximum likelihood procedure) with higher precision and power compared to more traditional methods such as the last observation carried forward." "There was skewed attrition, and the follow-up assessment was significantly later than 4 months in the control group."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up at endpoint was high: 22% (36/165) and skewed. Risk of attrition bias was therefore considered high although an appropriate method was used to deal with missing values in the analyses and the authors conclude otherwise. "Analysis of age, sex, HAM-D17, or absence from work during the last 10 working days at entry did not suggest any significant differences between missing participants and participants included in the analysis at either 4 months or 12 months." "It is then plausible to consider the missing data as 'missing at random,' making the mixed effect model a plausible approach to estimate the effect, based on the total sample with missing cases included." "This approach uses data from all included patients (intention-to-treat), handles entry differences, and is able to handle missing data (restricted maximum likelihood procedure) with higher precision and power compared to more traditional methods such as the last observation carried forward." "There was skewed attrition, and the follow-up assessment was significantly later than 4 months in the control group."
Selective reporting (reporting bias)	High risk	In the study protocol, no report was made regarding the third treatment group (relaxation)
Other bias	Low risk	None identified

Krogh 2012

Study characteristics	
Methods	**Study design:** A single-centre, two-armed, parallel-group, observer-blinded randomised clinical superiority trial. **Recruitment:** between September 2008 and April 2011, participants were referred to trial site from various clinical settings. **Follow up:** 3 months. Lost to follow up: 13%
Participants	**Inclusion criteria:**: men and women between 18 and 60 years of age, referred from a clinical setting by a physician or a psychologist, a diagnose of major depression (DSM-IV) based on the Danish version

Krogh 2012 *(Continued)*

	of the Mini International Neuropsychiatric Interview, score above 12 on the HAM-D17 and living in the Greater Copenhagen catchments area, able to comprehend and sign the informed consent statement. **Exclusion criteria:** current drugs abuse, any antidepressant medication within the last two months, current psychotherapeutic treatment, contraindications to physical exercise, more than 1 hour or recreational exercise per week, suicidal behaviour according to the 17-item Hamilton depression rating scale (HAM-D17 item 3 > 2), pregnancy, current/previous psychotic or manic symptoms, or lack of informed consent **Baseline characteristics** 115 were randomised (T1: 56; T2: 59). Age: T1: 39.7 (SD11.3); T2: 43.4 (SD 11.2) Female: T1: 71.4%; T2: 62.7% Occupational status: T1: 35.7% unemployed; T2: 45.7% T1: 35.7% sickness leave; T2: 30.5% sickness leave T1: 74.3% job attendance, last 10 days; T2: 73.8% job attendance, last 10 days **Setting**: outpatient
Interventions	T1: Aerobic training group: designed to increase fitness as measured by maximal oxygen uptake. After initial 10 minutes of general low-intensity warm-up, the participants did 30 minutes of aerobic exercise on a stationary cycle ergometer followed by five minutes low-intensity cool down period. During the initial four weeks, the aim was to work out at intensity levels corresponding to at least 65% to their maximal capacity, progressing to 70% and 80% during the second and third month, respectively. The participants carried a pulse monitor during exercise to guide and document intensity levels T2: Stretching exercise group: designed as an attention control group with the purpose of providing the same level of social interaction and contact with health care professionals as in the aerobic exercise group. This was done in order to assess the potential antidepressant effect of aerobic exercise in it self, and not the effect of aerobic exercise plus social interaction. This stretching exercise group performed low intensity exercise, which we did not expect to contain any antidepressant effect per se. The initial 10 minutes were low-intensity warm-up on a stationary bike, then a 20 minutes program of stretching, followed by 15 minutes of various low intensity exercises such as throwing and catching balls Both groups were scheduled to meet three times per week for three months for a total of 36 sessions
Outcomes	**Sickness absence** 1) the number of days spent on the job within the last ten working days, expressed as a percentage 2) off work: employment status or sick leave at the time of the interview **Depressive symptoms** 1) depression severity, assessed by the HAM-D17 2) core depression items, assessed by HAM-D6 3) remission, defined as not fulfilling the DSM-IV criteria for major depression and a HAM-D17 score below 8 4) self-reported depression, assessed by the Beck Depression Inventory (BDI)
Notes	Country: Denmark
Risk of bias	

Krogh 2012 *(Continued)*

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The randomization was centralized and carried out by the Copenhagen Trial Unit (CTU) using a computerized randomization sequence with alternating block sizes unknown to the investigators."
Allocation concealment (selection bias)	Low risk	"The randomization was centralized and carried out by the Copenhagen Trial Unit (CTU) using a computerized randomization sequence with alternating block sizes unknown to the investigators."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	"Prior to the first training session of the participant, the trial psychotherapist would contact the CTU by phone for participant allocation." "Neither participants nor the physiotherapist conducting the intervention were blinded to the allocation." However, since the interventions would be similar to participants and providers, it is unlikely that they would have behaved differently because they knew being the intervention group
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	The outcome assessors were all blinded to participant allocation "Prior to the follow up interview, participants were instructed not to reveal their allocation to the outcome assessors. The statistical analysis and preparation of the first draft was carried out blinded to group assignment."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	The outcome assessors were all blinded to participant allocation. The HAM-D17 is a structured interviewer based questionnaire, so risk of bias low (this does not apply to the BDI as this is a self-report instrument)
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Lost to follow up: T1: 16.1%; T2: 10.2% but appropriate method has been used to account for these missing data: "All continuous outcome measures were analyzed using a repeated measurement linear mixed effect model with an unstructured variance matrix [..] The mixed effects function is able to handle missing continuous data using a likelihood estimation of missing data."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Lost to follow up: T1: 16.1%; T2: 10.2% but appropriate method has been used to account for these missing data: " "All continuous outcome measures were analyzed using a repeated measurement linear mixed effect model with an unstructured variance matrix [..] The mixed effects function is able to handle missing continuous data using a likelihood estimation of missing data."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None reported

Lerner 2012

Study characteristics	
Methods	**Study design**: RCT. **Recruitment**: 6 months. **Follow up:** 4 months. **Lost to follow up:** 8.9%

Lerner 2012 *(Continued)*

Participants	**Inclusion criteria:** ages 18 to 62 years and employed 15 hours per week or more and fulfilled the criteria for current MDD and/or dysthymia, a WLQ productivity loss of at least 5% in the past 2 weeks (this score is consistent with an impaired ability to work approximately 20% of the time over 2 weeks). **Exclusion criteria:**: planning to retire within 2 years, receiving work disability benefits, active alcoholism or drugs-abuse based on the five-item CAGE, pregnant or 6 months postpartum, schizophrenia or bipolar disorder, non-English speaking and/or reading, and/or diagnosed with one or more of 12 medical conditions that have symptoms that potentially interfere with working (e.g. angina, congestive hart failure, stroke, diabetes, chronic obstructive lung disease) **Baseline characteristics** 79 were randomised (T1:52; T2:27); Comorbidity: T1: 80.8%; T2: 71.1% Age: T1: 45.5 (SD 9.8); T2: 45.9 (SD 8.6) Male: T1: 23.1%; T2: 18.5% Ethinicity: T1: 100% white; 96.3% white Marital status: T1: 47.1% married; T2: 48.1% married **Setting**: workplace
Interventions	T1: Work and Health Initiative (WHI) intervention. Provided over the phone by EAP counsellors trained in its methods. The program lasts for 8 weeks with 1-hour visits occurring every 2 weeks. This multi component work-focused programs consists of: 1) work coaching and modification, 2) care coordination, 3) cognitive-behavioural strategies. In the WHI, the counsellor and employee co-create a care plan for dealing with each functional problem and review specific assignments and progress at each session. A motivational enhancement approach is utilized to promote and solidify change. In both groups: electronic feedback on depression and advise to seek care T2: Usual care. Primary care, specialty care, behavioural health programs, and/or standard EAP services. In both groups: electronic feedback on depression and advise to seek care
Outcomes	**Sickness absence** Sick leave; Days lost in past two weeks due to health reasons (WLQ Work Absence Module) Outcome type: Continuous Outcome **Depressive symptoms**; Patient Health Questionnaire Depression Score Outcome type: Continuous Outcome **Work functioning:** Work Limitations Questionnaire Outcome type: Continuous Outcome
Notes	**Country:** US

Risk of bias

Bias	Authors' judgement	Support for judgement

Lerner 2012 *(Continued)*

Random sequence generation (selection bias)	Low risk	"Employees were allocated by electronic randomization."
Allocation concealment (selection bias)	Low risk	Web-based randomisation
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Participants received information about the RCT and were aware of the treatment condition to which they were randomised. Seven counsellors volunteered to conduct the WHI intervention. However, it is unlikely that they will have changed behaviour because they knew they were in the intervention
Blinding of outcome assessment (detection bias) Sick Leave	High risk	The WLQ Work absence module is a self-report measure. As participants were aware of their allocation status, risk of bias high
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	The PHQ-9 relies on patient self-report. As participants were aware of their allocation status, risk of bias high
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	"Five (9.6%) employees in the WHI treatment group and 2 (7.4%) of the usual group did not complete the follow-up questionnaire and were considered dropouts." "Sensitivity analyses including the seven employees that were lost to follow-up confirmed the results."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	"Five (9.6%) employees in the WHI treatment group and 2 (7.4%) of the usual group did not complete the follow-up questionnaire and were considered dropouts." "Sensitivity analyses including the seven employees that were lost to follow-up confirmed the results."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Lerner 2015

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** Two: Work-focused intervention versus care as usual **Recruitment:** Eligibility screening on a privacy-protected study Web site was offered at 24 sites: 13 private-sector employers, six public-sector employers, and five organizations serving employed populations (for example, employee benefits organizations). Screening was voluntary, anonymous, available during the workday and after work hours, and open to employees (and in some cases dependents) **Follow-up:** 4 months
Participants	**Baseline characteristics** Work-directed intervention combined with clinical intervention • *Age*: 54.6 +- 6.1 • *Gender*: 69% female

Lerner 2015 *(Continued)*

- *Marital status*: 52% married
- *Occupation*: 72% white collar
- *Sick leave status*: 1.5 +/- 2.1 days missed in past two weeks
- *Number of participants randomised*: 217

No intervention or Care as Usual

- *Age*: 54.8 +/- 6.1
- *Gender*: 75% female
- *Marital status*: 58% married
- *Occupation*: 69% white collar
- *Sick leave status*: 1.6 +/- 2.3 days missed in past two weeks
- *Number of participants randomised*: 214

Inclusion criteria: Eligible individuals were age 45 or older and employed; met criteria for major depressive disorder, persistent depressive disorder (formerly dysthymia), or both (double depression); and had work limitations. Major depression required a Patient Health Questionnaire–9 (PHQ-9) score of five of nine symptoms at qualifying levels. Persistent depressive disorder required a score of at least two of six symptoms on the Primary Care Screener for Affective Disorder. Work limitations were signified by an at-work productivity loss score of ≥5% on the Work Limitations Questionnaire (WLQ).

Exclusion criteria: Psychosis, bipolar disorder, current alcohol abuse or dependence, inability to speak English, and severe physical limitations (a physical component score of ≤35 on the 12-item Short-Form Health Survey)

Pretreatment: The WFI and usual-care groups were similar at baseline (Table 1), except that the proportion of married individuals was larger in the usual-care group (58% versus 46%, P = .01) as was the mean number of baseline comorbid general medical conditions (3.2 versus 2.7, p<.01).

Setting: 13 private-sector employers, six public-sector employers, and five organizations serving employed populations (for example, employee benefits organizations).

Interventions

Intervention characteristics

Work-directed intervention combined with clinical intervention

- *Content*: Three integrated modalities: "I) Care coordination addressing barriers to functional improvement related to a misalignment of goals and expectations among the individual with depression, his or her regular provider, and the counselor; psychoeducation (filling in gaps in knowledge of depression and treatment and their impact in work); motivational enhancement (promoting active engagement in care); counselor promotes three-way participant-provider-counselor communication by assessing depression symptom severity and work limitations monthly and sharing results.II) Work-focused cognitive-behavioural therapy (CBT) strategy: With counselor guidance and a workbook, participants learn to identify the thoughts,feelings, and behaviours that are eroding work functioning and respond by using more effective coping strategies. III) Work coaching and modification: addresses barriers to functioning resulting from imbalances between the characteristics of the worker and those of the job and work environment. Using a semi-structured interview approach,the counselor obtains information about the participant's work limitations (reported on the WLQ) and work life (job demands, ability to control the work, and availability of workplace supports and stressors). A customized plan is developed that guides the participant to change specific work behaviours, work processes, or environmental conditions, to begin using compensatory strategies, or both. With methods culled from disability management,vocational rehabilitation, supported employment, and management, the plan is designed to be self-administered and not require employer approval.In each session, the homework and results are discussed. Finally, the counselor and participant co-create a self-care plan to reinforce continued use of helpful CBT and work strategies."
- *Duration, frequency, length*: Eight 50-minute telephone sessions every two weeks (four months total)
- *Communication means*: Telephone
- *Providers*: Masters-level counselors with EAP experience; The 11 counselors were employed by Optum EAP, Eden Prairie, Minnesota. Study personnel provided the counselors with 2.5 days of in-person WFI training

Lerner 2015 *(Continued)*

	No intervention or Care as Usual • *Content*: Each usual-care enrollee was advised to contact a health care provider (for example,primary care physician, psychiatrist, or behavioural health specialist) and, when applicable,an employer-sponsored employee assistance program (EAP). The study provided no direct care to the usual-care group. All study participants were shown Web links to depression information and care resources, including care offered through their affiliated study site. Most sites offered EAPs and insurance coverage (medical, behavioural, and pharmacy).During the study, participants were not restricted from using other services. • *Duration, frequency, length*: not specified • *Communication means*: Web links no personal contact • *Providers*: not specified
Outcomes	**Sickness absence** *Days lost in past two weeks due to health reasons* • **Outcome type**: Continuous Outcome **Depressive symptoms** *Patient Health Questionnaire Depression Score* • **Outcome type**: Continuous Outcome
Notes	Country: US

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "Randomization to the WFI or usual-care group occurred next, with use of an automated 1:1 scheme with random permutations of six consecutive enrollees." Judgement comment: In 2010 to 2013, eligible, consenting employed adults with depression were randomly assigned to the WFI or usual-care groups. Not clear how randomisation was achieved. Ask authors for clarification
Allocation concealment (selection bias)	Unclear risk	Judgement comment: Randomisation to the WFI or usual-care group occurred next, with use of an automated 1:1 scheme with random permutations of six consecutive enrollees. Unclear how this was concealed
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Quote: The study provided no direct care to the usual-care group. During the study, participants were not restricted from using other services. Participants could not be blinded to group assignment. Precautions to minimize bias included prohibiting the WFI counselors from providing care to any members of the usual care group and not informing study participants which questions specifically measured the study's endpoints. Comment: Unblinded to intervention but unlikely that they will have changed their behaviour because of this knowledge
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Comment: Unblinded and based on self-report
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: Unblinded and based on self-report

Lerner 2015 *(Continued)*

Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Quote: To assess the robustness of model results, six sensitivity analyses were performed, including a reanalysis using the last-observation-carried-forward (LOCF) method instead of mixed effects models and multiple versions of the original mixed-effects models. The original mixed-effects models were modified to assess the impact on results of including participants with missing follow-up WLQ data (including those unemployed at follow-up who did not complete the WLQ).... Sensitivity analyses of at-work productivity loss and depression symptom severity results supported the findings. [Results of sensitivity analyses are presented in the online data supplement.] LOCF models comparing the difference in outcome change between the groups yielded slightly smaller, significant effect sizes. For at-work productivity loss, the effect size changed from –.72 in the original model to –.60 in the LOCF model. For depression symptom severity, the parallel change in effect size was –.60 to –.48.
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Quote: Next, assigning participants with missing WLQ follow-up data the maximum at-work productivity loss score reduced that variable's effect size from –.72 (original model) to –.62 (new model) (statistical significance was maintained). Adding the days from baseline to follow-up survey completion yielded at-work productivity and depression severity results that were similar to those obtained in the original models.
Selective reporting (reporting bias)	Low risk	Judgement comment: All outcomes that were described in the protocol were measured and reported
Other bias	Low risk	Judgement comment: No other sources of bias

Lerner 2020

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Follow-up:** 4 and 8 months **Number of trial arms:** 2 **Recruitment:** The IC program referred veterans to the study. Research assistants conducted the pretrialscreening, involving a health record review and telephone interview. Potentially eligible, interested patients were invited to come in for informed consent and to complete a further eligibility assessment, which also served as the study baseline measurement.
Participants	**Baseline characteristics** Work-directed plus Clinical Care • *Age*: 46.5 (11.6) • *Gender: men (%)*: 122 (87.8) • *Marital status*: not reported • *Occupation*: not reported • *Sick leave status: off work due to illness (%)*: 0 • *Number of participants randomised*: 139 Care as Usual (Primary Care) • *Age*: 44.8 (11.6) • *Gender: men (%)*: 96 (84.2)

Lerner 2020 *(Continued)*

- *Marital status*: N.R.
- *Occupation*: N.R.
- *Sick leave status: off work due to illness (%)*: 4 (3.5%)
- *Number of participants randomised*: 114

Overall

- *Age*: 45.7 (11.6)
- *Gender: men (%)*: 218 (86.2
- *Marital status*: N.R.
- *Occupation*: N.R.
- *Sick leave status: off work due to illness (%)*: 4 (1.6%)
- *Number of participants randomised*: 253

Inclusion criteria: Age 18 years or older,Worked at least 15 hours per week in jobs they had occupied for at least 6 monthsWork limitations resulting in at least 5% at-work productivity loss based on validated questionnaire assessment. Current major depressive disorder or persistent depressive disorder confirmed by diagnostic interview

Exclusion criteria: Inability to speak or read English, Planned maternity leave, History of bipolar disorder or psychosis

Pretreatment: At baseline and after attrition the treatment groups were not significantly different on any variable.

Setting: A large Veteran Health Administration primary mental health care medical center and 2 smaller satellite sites.

Interventions

Intervention characteristics

Work-directed plus Clinical Care

- *Content*: Integrated Care (IC): see care as usual.The BWAW intervention: Be Well at Work targets 3 areas of coping. 1 Coping with depression treatment: motivational enhancement and psychoeducational strategies to enhance engagement in BWAW and treatment by regular healthcare provider. 2. work-focused cognitive-behavioural therapy strategy training based on a self-help manual. This component is aimed at increasing the patient's ability to identify maladaptive patterns of thinking, feeling, and behaving that affect work and substitute more adaptive coping behaviours. 3 identifying and addressing workplace barriers to effective functioning and potential work-appropriate coping strategies. Patients are guided to make small specific changes in how they perform work tasks (eg, new time management techniques), work routines (eg, collaborating vs staying isolated), and/or the work environment (eg, rearranging the workspace). Some patients may be guided to develop compensatory work strategies (eg, using memory aids). The BWAW intervention culminates with the development of a customized self-care plan. At the booster session, self-care progress is reviewed, and if necessary, the plan is adjusted.
- *Duration, frequency, length*: Eight 50-minute telephone visits occurring biweekly for 4 months, and 1 booster session approximately 4 months later.
- *Communication means*: Integrated Care (IC) sessions by telephone or in-person. Be well at Work (BWAW) provided by telephone
- *Providers*: For Integrated care, (IC) see Care as Usual. In addition: Two doctoral-level psychologists who were not providing IC provided BWAW counseling under thesupervision of its developers (a psychiatrist and a workplace health specialist). Initially,counselors received an intensive 2.5-day training session, followed by weekly telephone supervisioninvolving in-depth case reviews.

Care as Usual (Primary Care)

- *Content*: Integrated Care (IC): Promotion of adherence to prescribed antidepressants as well as activities to increase positive social interactions, healthy living, and selfesteem. Continuous assessments to check whether specialized service were needed.

Lerner 2020 *(Continued)*

- *Duration, frequency, length*: Duration: 4 months Frequency IC: (from protocol):Typically, one or two sessions per month. Duration IC: 30 minute sessions
- *Communication means*: Integrated Care (IC) sessions by telephone or in-person.
- *Providers*: Practitioners in the IC team included psychologists, nurses, and social workers, supervised by a VHA psychiatrist, who supported the use of brief evidence-based psychotherapy and pharmacotherapy

Outcomes	**Sickness absence** *Leave of absence* • **Outcome type**: Dichotomous Outcome • **Direction**: Lower is better **Depressive symptoms** *Patient Health Questionnaire Depression Score* • **Outcome type**: Continuous Outcome **Work functioning** *At work productivity loss (WLQ)* • **Outcome type**: Continuous Outcome • **Direction**: Lower is better
Notes	**Outcomes** At T2 no SDs provided. We took SDs from T1. Work productivity loss is mean percentage decrease. **Country:** US

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Judgement comment: "If all eligibility criteria were met, the patient was assigned by simple randomization to 1 of the treatment groups. In this automated process, a random uniform number was calculated using theMath.random function in Java Math Library random function software version 8 (Oracle TechnologyNetwork) and the patient was assigned to IC plus BWAW half of the time."
Allocation concealment (selection bias)	Low risk	Quote: "If all eligibility criteria were met, the patient was assigned by simple randomization to 1 of the treatment groups. In this automated process," Judgement comment: Apparently allocation concealed
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	No blinding, but risk of performance bias considered low as the treatment of interest (Integrated care + be well at work module) cannot be considered less desirable as Treatment as usual for patients (Integrated care).
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Judgement comment: The research assistants were blinded for allocation, but sick leave was measured with self-report. "In this single-blind study, research assistants collecting study data were blinded to treatment group assignment."" The WLQ Time Loss Module measures the number of hours of work missed in the past 2 weeks owing to a health-related issue."
Blinding of outcome assessment (detection bias)	High risk	Judgement comment: Research assistants were blinded, but depressive symptoms were measured with self-report instruments. "In this single-blind study,

Lerner 2020 *(Continued)*

Depressive symptoms		research assistants collecting study data were blinded to treatment group assignment.""Potentially eligible, interested patients were invited to come in for informed consent and to complete a further eligibility assessment, which also served as the study baseline measurement (T0). For this initial assessment, the research assistant administered standardized questions to assess depression (Patient HealthQuestionnaire [PHQ-9];"
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Judgement comment: Lost to follow-up was between 10-20% (17.8%) but the analyses show that lost to follow-up was not related to any of the (outcome) measures. "The T1 follow-up survey was completed by 96 patients (83.5%) in the IC group and 115 patients (82.7%) in the IC plus BWAW group, and the T2 follow-up survey was completed by 97 patients (84.3%) in the IC group and 111 patients (79.9%) in the IC plus BWAW group (study attrition rate, 17.8%)." "At baseline and after attrition, the treatment groups were not significantly different on any variable."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Judgement comment: Lost to follow-up was between 10-20% (17.8%) but the analyses show that lost to follow -up was not related to any of the (outcome) measures."The T1 follow-up survey was completed by 96 patients (83.5%) in the IC group and 115 patients (82.7%) in the IC plus BWAW group, and the T2 follow-up survey was completed by 97 patients (84.3%) in the IC group and 111 patients (79.9%) in the IC plus BWAW group (study attrition rate,17.8%).""At baseline (Table 1 and Table 2) and after attrition (eTable 1 and eTable 2 in Supplement 2),the treatment groups were not significantly different on any variable."
Selective reporting (reporting bias)	Low risk	Judgement comment: All outcomes are either reported or not measured at follow-up.As the authors report in their suppl. 'deviations from protocol': The SF-12 Veterans Health Survey (VR-12) was going to be administered at baseline and both follow-ups, but it was only administered at baseline (to reduce respondent burden). '
Other bias	Low risk	Judgement comment: No other sources of bias detected

Mackenzie 2014

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** Via the website thiswayup.org.au participants with depression were recruited for two trials reported by Perini 2009 and Titov 2010. Later, the trials were restricted to participants that worked and work outcomes were analysed **Follow-up:** 11 weeks
Participants	**Baseline characteristics** Clinical: Psychological, I-CBT, I-guided • *Age*: not reported • *Gender*: not reported • *Marital status*: not reported • *Occupation*: not reported

Mackenzie 2014 *(Continued)*

- *Sick leave status*: 0.86 (0.18) mean (SE) work loss days in previous week before treatment
- *Number of participants randomised*: 58

No intervention or Care as Usual

- *Age*: not reported
- *Gender*: not reported
- *Marital status*: not reported
- *Occupation*: not reported
- *Sick leave status*: 0.84 (0.24) mean (SE) work loss days in previous week before treatment
- *Number of participants randomised*: 37

Inclusion criteria: i) DSM-IV criteria for depressive disorder, as determined by the Mini International Neuropsychiatric Interview Version 5.0.0 ii) aged 18 years or over iii) no previous history of a psychotic disorder for drug or alcohol misuse iv) not actively suicidal, as determined by a risk assessment

Exclusion criteria: i) unemployed and casually employed individuals ii) participants with missing baseline scores on the Sheehan Disability Scale or diagnosis-specific questionnaires

Pretreatment: In the trials from which the participants came there were already baseline differences between groups. In Titov 2010 there are more men, older and married persons in the control group. In Perini 2009 there are 30% more participants in the intervention group, but the control participants are more often male and less often married

Setting: Experiment with Internet-based mental health treatment in an academic institution

Interventions

Intervention characteristics

Clinical, psychological: I-CBT, I-guided

- *Content*: Principles and techniques of CBT described in the Sadness © programme: behavioural activation, cognitive restructuring, problem solving, assertiveness skills
- *Duration, frequency, length*: six online lessons with home work, to be completed wiithin 11 weeks
- *Communication means*: Internet, email
- *Providers*: Therapist supervised and supported the programme and communicated with participants
- *Name used by researchers*: iCBT

Care as Usual-WL

- *Content*: Wait list control group
- *Duration, frequency, length*: not reported
- *Communication means*: not reported
- *Providers*: not reported
- *Name used by researchers*: waiting list period

Outcomes

Sickness absence

Work loss days in the previous week

- **Outcome type**: Continuous Outcome

Notes

Country: Australia

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Judgement comment: In both Titov 2010 and Perini 2009: "were randomized via a true randomization process (www.random.org) "

Mackenzie 2014 *(Continued)*

Allocation concealment (selection bias)	Unclear risk	Judgement comment: No reporting of allocation concealment
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Comment: Patients could not be blinded and outcome was self-reported subjective; as the controls were on the waiting list, which was less desirable to patients, this may have changed their behaviour.
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Comment: Patients were not blinded and all outcomes were self-assessed and subjective.
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: Patients were not blinded and all outcomes were self-assessed and subjective.
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Comment: Both trials had selective loss to follow up in the intervention groups of more than 20%. Missing values were replaced by baseline values.
Incomplete outcome data (attrition bias) Sick Leave	High risk	Comment: Both trials had selective loss to follow up in the intervention groups of more than 20%. Missing values were replaced by baseline values.
Selective reporting (reporting bias)	High risk	Judgement comment: Retrospectively registered protocol; work participation not mentioned as an outcome.
Other bias	High risk	Judgement comment: Two trials (Perini 2009 and Titov 2010) were combined and the subgroup of working participants was analysed for work outcomes, which was an unplanned outcome

McCrone 2004

Study characteristics	
Methods	**Study design**: RCT, 2 conditions. **Recruitment:** by screening in the GP waiting rooms and of GP referrals using the GHQ-12. Score at least 4: seen by GP who administered inclusion and exclusion criteria. **Follow up**: 6 months. **Lost to follow up** at 6 months: T1: 27%; T2: 24%
Participants	**Inclusion criteria:** GP patients aged 18 to 75 years; diagnosis (ICD): depression, mixed anxiety/depression or anxiety disorder. CIS-R score at least 12 **Exclusion criteria:** active suicidal ideas, Psychotic disorder, organic mental disorder or alcohol or drug dependence. Having taken medication for anxiety or depression continuously for at least 6 months immediately prior to entry; unable to read or write; unable to attend 8 sessions at practice **Baseline characteristics** 274 were randomised (T1: 146; T2: 128). Mean age: T1: 43.6 (SD 14.3); T2: 43.4 (SD 13.7) Female: T1: 73% T2: 75% Married or cohabiting: T1: 54%; T2: 52%

McCrone 2004 *(Continued)*

	Employed: T1: 66%; T2: 58% **Setting**: Primary care
Interventions	T1: Computerised CBT: interactive, multimedia. Feedback to patient and GP after each session. 15 minute introductory video, 8 x 50 minute sessions of CBT, with homework projects between sessions T2: TAU: General practitioner care as usual: no constraints. Could include medication, discussion of problems with GP, practical or social help, referral to counsellor, practice nurse, mental health professional, or further physical examination
Outcomes	**Sickness absence** 1) Number of days of absence from work (certified by GP) during 8 months **Depressive symptoms** 1) BDI **Work functioning** 1) Work and Social Adjustment Scale
Notes	Country: UK

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The random allocation schedule was generated at the Institute of Psychiatry. An individual unit of randomization was used."
Allocation concealment (selection bias)	Low risk	"Random allocation schedule was generated at the Institute of Psychiatry, before the study commenced and away from GP practices. Cards in sealed and numbered envelopes were used. Only to be opened by practice nurse who ran study. Integrity was checked by the first author on her regular visits to the practices."
Blinding of participants and personnel (performance bias) Sick Leave	High risk	No blinding, risk of performance bias considered high as the treatment of interest (T1) cannot be considered equally desirable as Treatment as usual (T2) for patients. "Patients randomized to 'Beating the Blues' (T1) also received pharmacotherapy, if prescribed by their GP, and/or general GP support and practical/social help", offered as part of treatment as usual, with the exception of any face-to-face counselling or psychological intervention. We did not constrain the interventions received by patients allocated to treatment as usual (T2)." Moreover, patients in the Treatment as Usual (T2) group were found to attend other health care professionals more often. "Large differences were observed for the proportion of patients attending accident and emergency or outpatient departments, and having contacts with community psychiatric nurses, counsellors and other therapists. Greater use was made by the TAU group for all these services."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	No blinding of outcome assessors was reported. Sick leave was based on the sick leave certificates of the GP, who was also the treatment provider of treatment as usual. "We recorded the number of days of absence from work during the baseline and follow-up periods on the basis of an issue of a certificate by the general practitioner."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	No blinding of patients was reported and depressive symptoms were measured by self-report

McCrone 2004 *(Continued)*

		"Depressive symptoms were measured with self-report and participants were not blinded to treatment allocation."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up was relatively high (> 20%) for the depression outcome From Figure 2 of the publication on depression outcome (Proudfoot et al 2004): Loss to follow up: T1: 27%; T2: 24%
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Sick leave data were part of the cost data, and a high percentage of the cost data were complete at follow up. "A total of 274 patients were randomised into two groups (BtB, n = 146; TAU, n = 128), with cost data available for both baseline and follow-up periods for 261 (95%) patients (138 BtB, 123 TAU)."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Meuldijk 2015

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Follow-up:** 12 months **Number of trial arms:** 2 **Recruitment:** Patients were recruited from outpatient mental health care clinics. From protocol: "First all referred patients are globally screened by an experienced psychiatrist for the presence of depression and/or anxiety disorders as current, main problem. This global screening is based on written information provided by the GP containing an interpretation of the patient's current health status and referral for further mental health care; this step does not require face-to-face contact with the patient. Subsequently, the potentially eligible patients are invited for a first ROM assessment. Prior to this first ROM assessment, the psychiatric research nurse conducting the ROM assessment invites the patients to participate in the study. Those who agree to participate are asked to provide written informed consent."
Participants	**Baseline characteristics** Concise Care • *Age*: 36.0 (12.0) • *Gender*: 58 (62%) female • *Marital status*: 51 (58%) married • *Occupation*: 41 (47%) employed • *Sick leave status*: not reported • *Number of participants randomised*: 93 • *Number depressed and randomised*: 34 Care as Usual • *Age*: 37.0 (11.98) • *Gender*: 53 (60%) female • *Marital status*: 42 (47%) married • *Occupation*: 49 (55%) employed

Meuldijk 2015 *(Continued)*

- *Sick leave status*: not reported
- *Number of participants randomised*: 89
- *Number depressed and randomised*: 36

Overall

- *Age*: 36.5 (12.3)
- *Gender*: 111 (61%) female
- *Marital status*: 93 (53%) married
- *Occupation*: 90 (51%) employed
- *Sick leave status*: not reported
- *Number of participants randomised*: 182
- *Number depressed and randomised*: 70

Included criteria: Eligible participants:Patients referred to the mental health clinic by their general practitioner, Age: 18–65 yearsMeeting the DSM IV-TR criteria for a primary current diagnosis of anxiety disorder and/or depression, established using the Mini-International Neuropsychiatric Interview-Plus.

Excluded criteria: Excluded were patients with suicidal or homicidal risk, severe social dysfunction, delusions, hallucinations and/or suffering from bipolar or psychotic disorders. Other co-morbidity with psychiatric disorders was allowed. Insufficient mastery of Dutch was areas on for exclusion.

Pretreatment: None reported "Randomization reasonably balanced the treatment groups with respect to the baseline characteristics."

Setting: Outpatient Mental Health Clinics (MHCs). These clinics were part of Rivierduinen (RD), a secondary Regional MentalHealth Provider (RHMP) in the province of South-Holland, the Netherlands.

Interventions

Intervention characteristics

Concise Care

- *Content*: In contrast, concise care started within one week after the baseline measurement and had to be given within 7 weeks thereafter. Concise care was initially described as 4 to maximum 7 individual 45-min psychotherapy sessions, depending on the treatment protocol. The pharmacotherapy protocol for depressive and/anxiety disorders in concise care was confined to a maximum of 4 sessions within7 weeks. Moreover, therapists' treatment choice in both standard and concise care followed the principles of shared decision-making. Contrary to standard care, treatment goals and procedures in concise care are clearly established and mutually agreed on, prior to initiating treatment. In addition, treatment success of concise care was evaluated at the end of treatment. When either the patient or therapists convinced that the clinical effects are insufficient or patients are insufficiently helped by the initial treatments in concise care, ‘stepping up’ or continuation of (additional) standard treatment, in line with stepped-care principles, was possible. Pharmacotherapy in concise care was also evaluated after 7 weeks, and continued when necessary according to the (inter) national clinical guidelines. After implementation changes to the treatment protocols were made at the recommendation of the MHCs; these included extending the treatment duration of concise care to a maximum of 7 sessions in 7–9 weeks. This was to allow treatment continuation of concise care in case of cancelled or missed sessions by therapists or patients.
- *Duration, frequency, length*: Duration psychotherapy: 7 weeks, with 4 to maximum 7 individual 45-min psychotherapy sessions, depending on the treatment protocol.Duration pharmacotherapy: maximum of 4 sessions within 7 weeks. At the end the clinical effect was evaluated and a continuation of care was a possibility when needed.
- *Communication means*: Face-to-face
- *Providers*: Therapists, experienced psychiatrists and psychologists, providing concise care received a 2 h instruction in the core elements of the intensified psychotherapy and/or pharmacotherapy, as described in the protocols.

Care as Usual

- *Content*: Choice between pharmacotherapy with a selective serotonin reuptake inhibitor, cognitive behavioral therapy and, in case of a posttraumatic stress disorder, Eye Movement Desensitization

Meuldijk 2015 *(Continued)*

	and Reprocessing therapy. A combination of pharmacotherapy and psychotherapy was also possible. Therapists' treatment choice in both standard and concise care followed the principles of shared decision-making. • *Duration, frequency, length*: In standard care the number of sessions, start and duration of treatment is variable and treatment could continue during the entire study period of 1 year. On average, psychotherapy is provided in 3–6 months on a weekly basis, but in practice once every 2 to 3 weeks, pharmacotherapy for 1 year or longer • *Communication means*: Face-to-face • *Providers*: Therapists, experienced psychiatrists and psychologists, in the standard condition did not get additional training
Outcomes	**Sickness absence** *Days lost* • **Outcome type**: Continuous Outcome Data on depressed subgroup not provided.
Notes	**Country:** the Netherlands

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Comment: Protocol states: " Random allocation was generated by using a variable blocked design developed by an independent researcher from the Department of Medical Statistics & BioInformatics, LUMC and derived by computer"
Allocation concealment (selection bias)	Low risk	"Participants and clinicians are informed about the outcome of the randomization; the psychiatric test nurses (assessors)involved in the ROM assessment in the study, are kept blinded to the randomization condition throughout the entire study.Randomization and the subsequent assignment of participants to the intervention will be performed by the researcher(D.M.), whom is not an assessor."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	The control condition cannot be considered less desirable: "In both concise and standard care, a choice could be made between pharmacotherapy with a selective serotonin reuptake inhibitor (SSRI; Cognitive Behavioral Therapy (CBT) and, in case of a posttraumatic stress disorder, Eye Movement Desensitization and Reprocessing- therapy (EMDR)."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave data were collected by a self-report questionnaire and patients were not blinded to treatment allocation. "Patients and therapists were informed about the outcome [of the randomization]; the psychiatric test nurses responsible for the ROM assessments were not. "Absence from work [....] are measured in the second part of the Tic-P. Work absenteeism is measured by two questions related to short- and long-term absence from work."
Blinding of outcome assessment (detection bias) Depressive symptoms	Unclear risk	No depressive symptoms measured
Incomplete outcome data (attrition bias) Depressive symptoms	Unclear risk	No depressive symptoms measured

Meuldijk 2015 *(Continued)*

Incomplete outcome data (attrition bias) Sick Leave	High risk	Comment: Full economic data (of which sick leave was part) was only available for 22% of the participants
Selective reporting (reporting bias)	Low risk	Comment: All outcomes that were described in the protocol were measured and reported
Other bias	Low risk	Comment: No other sources of bias detected

Miller 1998

Study characteristics	
Methods	**Study design:** RCT. multi-centre, 2 conditions. **Recruitment:** referrals from physicians or mental health professionals, media advertising, and word of mouth. **Follow up:** 12 weeks. **Lost to follow up**: 2%
Participants	**Inclusion criteria:** age 21 to 65 years; Diagnosis of chronic MDD with two or less cumulative depression-free months and who had not met DSM-II-R criteria for dysthymia within 2 months of the onset of current MD episode OR of concurrent MD episode superimposed on antecedent DSM-III-R dysthymia; Premenopausal women: adequate contraception **Exclusion criteria:** organic mental syndrome, current or lifetime diagnosis of bipolar disorder or cyclothymia, schizophrenia, other psychotic disorders, obsessive-compulsive disorder, antisocial, schizotypical or severe borderline personality disorder; Principal DSM-III-R diagnosis of panic disorder, generalized anxiety disorder or PTSD within the past 6 months; DSM-II-R defined anorexia or bulimia nervosa within the past year; Drug or alcohol abuse or dependence within the past 6 months; Patients deemed at immediate suicide risk/ medical contraindications to antidepressants; Significant general medical disorder;Concomitant therapy with any psychotropic drug (except chloral hydrate or temazepam); Failure of adequate trial of sertraline or imipramine; Treatment with MOA-inhibitors within 3 weeks; Any depot neuroleptic within 6 months'; Fluoxetine within 1 month; Regular daily neuroleptic, anxiolytic, or antidepressant medication within 2 weeks; ECT within 3 months **Baseline characteristics** 635 were randomised: (T1: 426; T2: 209). Mean age: 41.1 (SD 10.1) Female: 63% Married: 38% Employed: 71% **Setting**: 12 outpatient centres
Interventions	T1: sertraline (SSRI). Week 1-3: 50 mg/day, then weekly titration in 50 mg/day increments (max 200 mg/day). 12 weeks, visits every week for the first 6 weeks and every 2 weeks for last 6 weeks. Before this, 1 week placebo run-in T2: Imipramine (TCA). Week 1: 50 mg/day, week 2: 100 mg/day, week 3: 150 mg/day. Then weekly titration 50 mg/day increments with a max of 300 mg/day by week 6. 12 weeks, visits every week for the first 6 weeks and every 2 weeks for last 6 weeks. Before this, 1 week placebo run-in
Outcomes	**Sickness absence**

Miller 1998 *(Continued)*

	1) hours worked per week (12 weeks) 2) off work: employed (yes or no) **Depressive symptoms:** 1) full remission, both CGI-I (=sub scale CGI) score of 1 or2 AND total HAM-D score of 7 (or less) at last visit 2) satisfactory therapeutic response, at last visit: both CGI-I (=sub scale CGI) score of 1 or 2 AND total HAM-D score of 15 or less AND HAM-D-score reduction of at least 50% since baseline AND final GSI-S (= subscale CGI) score of 3 or less 3) 24-HAM-D 4) MADRS 5) BDI **Work functioning:** 1) SAS work composite 2) LIFE work functioning
Notes	**Country:**US

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Not reported
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	"A novel statistical method was employed for unblinding patients who experienced recurrence or clinically significant worsening of symptoms." "In consultation with FDA personnel, the sponsor's statistician monitored the ability of each investigator to guess the treatment assignment of their patients still in the study. When breaking the blind for any patient, the statistician (R.J.M.) examined the effect of unblinding on our ability to guess the treatment assignment for the remaining patients at that site. If any of these probabilities exceeded 75%, the site agreed to refer all subsequent relapsers to a third party for treatment."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Sick leave was assessed by the LIFE interview. Interviewers were blind to treatment condition. "Finally, it should be noted that while blind to treatment condition, patients and interviewers were not blind to the fact that patients were receiving active medication nor were they blind to the time of assessment (baseline, week 4, endpoint)."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	Depressive symptoms were measured with the 24 HAM-D (clinician-rated). Interviewers were blind to treatment condition. "Finally, it should be noted that while blind to treatment condition, patients and interviewers were not blind to the fact that patients were receiving active medication nor were they blind to the time of assessment (baseline, week 4, endpoint)."
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	For depressive symptoms, ITT rates of remission could be calculated for 623 (of the 635) patients, which is 98%. "See Figure 1, Keller et al, 1998."
Incomplete outcome data (attrition bias) Sick Leave	Unclear risk	Completeness of sick leave data not reported. "Sample sizes [on psychosocial variables] vary due to sporadic missing data."

Miller 1998 *(Continued)*

Selective reporting (reporting bias)	Unclear risk	No indication for selective reporting could be identified. The design was published in a paper by Rush et al, albeit concurrently with the publications on the outcome
Other bias	Low risk	None identified

Noordik 2013

Study characteristics	
Methods	**Study design:** Two-armed cluster randomised trial. **Recruitment:** Recruitment of workers started in November 2006 and ended in December 2007. Workers eligible according to the OP were invited to participate. **Follow up**: 12 months. **Lost to follow up** main outcome: 10.6% for all participants and 11% for depressed subgroup
Participants	**Inclusion criteria:** workers who were on sick leave due to CMD between 2 and 8 weeks. CMD were defined as stress-related, adjustment, anxiety or depressive disorders. Stress-related disorders were classified according to the Dutch guidelines for OP (19). Anxiety, depressive, and adjustment disorders were classified by the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) **Exclusion criteria:** workers with a primary somatic disorder according to the OP and those who were not able to speak Dutch **Baseline characteristics** 160 were randomised (T1: 75; T2: 85). Subgroup of depressed workers: 37 (T1: 18; T2:19). Mean age: T1: 44.9 (SD 9.8); T2: 45.9 (SD 9.8) Female: T1: 75.7%; T2: 66.7% Educational level: Low: T1: 8.7%; T2: 17.9% Middle: T1: 24.6%; T2: 23.1% High: 66.7%; T2: 59.0% **Setting:** Occupational healthcare.
Interventions	This study was conducted in the Netherlands, where most of the workers on sick leave due to CMD visit an OP. The OP offers RTW interventions to these workers according to the evidence-based (Dutch) guidelines T1: Exposure based return to work intervention (RTW-E): In the RTW-E program, workers received CAU and were gradually exposed in vivo to more demanding work situations structured by a hierarchy of tasks evoking increasing levels of anxiety, stress, or anger. The RTW-E program provided workers with several homework assignments aimed at preparing, executing, and evaluating an exposure-based RTW plan T2: Care as usual (CAU): aims to help workers regain control and rebuild social and occupational contacts and activities, according to the OP practice guidelines for CMD. The OP can support this process by using recommended methods such as stress inoculation training, cognitive restructuring, graded activity, and time contingency during the RTW

Noordik 2013 *(Continued)*

Outcomes	**Sickness absence** 1) the time-to-full RTW, calculated as the number of calendar days from the first day of sick leave to the first day of full RTW. Full RTW was defined as the total number of contracted working hours per week lasting ≥28 calendar days without a recurrence of sick leave **Depressive symptoms** 1) symptoms of depression, assessed by the Four-Dimensional Symptom Questionnaire (4DSQ)
Notes	Country: the Netherlands

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"We performed a restricted randomization with blocks of four OPs." After randomization researcher KN informed EN about the allocation of every OP and saved the randomization file." Personal communication: "The randomization followed a schedule generation by randomization software."
Allocation concealment (selection bias)	High risk	"The validity of the results of this study may have been limited due to a selection bias because of the absence of allocation for each OP. As a result, the potential for the selective inclusion of workers was rather high." "However, we could not prevent some OP from including zero workers, which could have introduced selection bias."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Blinding of participants and researchers, but not of personnel was ensured: "The workers were blind to the differences in RTW-E and CAU." "The researchers were blind to the allocation and outcome measurement." However, it is unlikely that this will have changed their behaviour.
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Sick leave was assessed by workers' diaries. As workers are blinded to allocation status, risk of detection bias for sick leave is considered to be low
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression is assessed by the 4DSQ, a self-report questionnaire. As the participants were blinded to allocation status, risk of detection bias for depressive symptoms is considered to be low
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow-up for depression for the subgroup of depressed workers: 52%
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up of sick leave data for the subgroup of depressed workers was: 11%. No appropriate method was used to take selective attrition into account
Selective reporting (reporting bias)	High risk	Not all (secondary) outcomes measures announced in the design paper were reported in the effect study, of which the data on the HADS-depression subscale
Other bias	Low risk	None identified

Phillips 2014

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** Occupational health sections of three large employers agreed to participate in the trial, by directing their staff to the website and promoting the opportunity internally. The workplaces were thus a convenience sample, which spanned, as it turned out,the transport, health and communications sectors. **Follow-up:** 6 weeks and 12 weeks
Participants	**Baseline characteristics** Clinical, Psychological intervention, I-CBT, Unguided • *Age*: 42.2+-9.6 • *Gender*: 43% male • *Marital status*: 69% married • *Occupation*: 49% manager/prof • *Sick days in past 6 months*: 29.3 +- 41 • *Number of participants randomised*: 318 No intervention or Care as Usual • *Age*: 42.7+-9.6 • *Gender*: 50% male • *Marital status*: 61% married • *Occupation*: 50% manager/prof • *Sick days in past 6 months*: 27.5+-37.6 • *Number of participants randomised*: 319 **Inclusion criteria:** Aged over 18 years and: on the PHQ-9 score 2 or more or five of the nine items, including 2 or more on item 1 (little interest in doing things) or item 2 (feeling hopeless); at least one of the items made it difficult to work, take care of things at home, or get along with other people.All participants were required to give a telephone number as a condition of joining the study. **Exclusion criteria:** - medical history or treatment for brain injury, stroke, bipolar disorder- receiving CBT **Pretreatment:** More males were randomised to control than to MoodGYM **Setting:** Occupational health
Interventions	**Intervention characteristics** Psychological intervention, I-CBT, Unguided • *Content*: MoodGYM is a freely available CBT course developed at Australia National University (ANU) which allows participants to proceed at their own pace • *Duration, frequency, length*: Five weeks, weekly, 1 h-long modules • *Communication means*: Internet, assisted by weekly telephone calls • *Providers*: not reported Care as Usual- information • *Content*: Control participants were directed to websites judged from a previous review of self-help in mental health to be reliable sources of information about mental health problems

Phillips 2014 *(Continued)*

- *Duration, frequency, length*: Probably similar
- *Communication means*: Internet
- *Providers*: UK government, NHS

Outcomes	**Sickness absence** *Number of sick leave days in the past six months* • **Outcome type**: Continuous Outcome **Depressive symptoms** *Patient Health Questionnaire Depression Score* • **Outcome type**: Continuous Outcome
Notes	**Country**: UK Authors provided sick leave data and a full report to the commissioner the BOHRF

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "parallel randomized controlled trial" Judgement comment: Information received from authors: Once potential participants had completed the screening questions, if eligible for inclusion in the trial, they were given a study ID, allocated through the website, and they were then invited to join the trial. A list was produced by the Nottingham Clinical Trials Unit to allow simple (unrestricted) randomisation.[i] If participants consented, they were randomised and sent to the portal designers to be incorporated in its pathway. In this way the randomisation status of participants was concealed from their employers and from the research team until the study was completed.
Allocation concealment (selection bias)	Low risk	Judgement comment: See previous domain
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Comment: Unblinded but unlikely that there is a deviation from the intended intervention because of knowledge of the intervention
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Comment: Unblinded but unlikely that there is a deviation from the intended intervention because of knowledge of the intervention
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: From authors: The telephone interviewers were not 'blind' to the status of the participants, but they only recorded service use measures, not the main outcome. However all outcomes subjective self-report.
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Comment: From authors: The telephone interviewers were not 'blind' to the status of the participants, but they only recorded service use measures, not the main outcome. However all outcomes subjective self-report.
Incomplete outcome data (attrition bias) Sick Leave	High risk	Comment: More than half of the participants were lost to follow-up. For missing items in questionnaires means were imputed

Phillips 2014 *(Continued)*

Selective reporting (reporting bias)	Low risk	Judgement comment: The protocol says that also WLQ will be measured but this is not mentioned or reported in the article. Moreover protocol retrospectively registered. Information received from authors: I really cannot recall what might have happened to the WLQ
Other bias	Low risk	Judgement comment: No other sources of bias detected

Reme 2015

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** Unclear, probably through social insurance system **Follow-up:** **Subgroup participants with depression:**
Participants	**Baseline characteristics** Work-directed intervention combined with clinical intervention • *Age*: 17% >50 y • *Gender*: 30% male • *Marital status*: 30% married • *Occupation*: 7.3% unemployed • *Sick leave status*: 40.2% • *Number of participants randomised*: *626* • *Number with depression: 319* Care as Usual • *Age*: 21% > 50 y • *Gender*: 35% male • *Marital status*: 33% married • *Occupation*: 8.7% unemployed • *Sick leave status*: 37.3% • *Number of participants randomised*: *549* • *Number with depression: 314* **Inclusion criteria:** People: aged 18–60 years who were struggling with work participation attributable to common mental disorders and expressed a motivation to RTW/stay at work. Clinical psychologist assessed the presence of common mental disorders. **Exclusion criteria:** other reason than common mental disorder as the primary cause of problems with work participation; for example no motivation to participate in working life, severe psychiatric disorder or high suicide risk, ongoing substance abuse, or pregnancy, inability to read Norwegian, or engagement in psychotherapy elsewhere **Pretreatment:** **Setting:** National Insurance Scheme of Norway

Reme 2015 *(Continued)*

Interventions	**Intervention characteristics** Work-directed intervention combined with clinical intervention (CBT) • *Content*: The AWaC (At Work and Coping) programme combines individual CBT and job support. CBT was characterised by 'cognitive work-coping', and focused on managing mental health problems as they relate to work situations.. The individual job support was based on the'Individual Placement and Support (IPS)' approach, and was offered to those in need of individual job support (primarily participants on long-term disability) to facilitate workplace adaptations or identification of appropriate employment. IPS consisted of -eligibility based on consumer choice, focus on competitive employment, integration of mental health and employment services, attention to client preferences, work incentives planning, rapid job search, systematic job development and individualised job support. • *Duration, frequency, length*: up to 15 sessions CBT, IPS based on need (32% ended up receiving IPS). • *Communication means*: Face-to-face • *Providers*: Mini-teams of therapists and employment specialists. All therapists were monitored, videotaped and scored according to the Cognitive Therapy Adherence and Competence Scale. The employment specialists were required to have relevant qualifications and broad experience with supported employment, and extensive knowledge regarding the IPS principles and the job market in the team's region Care as Usual-WD • *Content*: A letter with information and encouragement to use available services and self-help resources. Employment and health care services for the control group were not restricted (beyond ruling out the AWaC), they could well be followed up by other psychologists and/or participate in other employment schemes initiated by NAV. • *Duration, frequency, length*: Not reported • *Communication means*: Face-to-face • *Providers*: GP, national insurance office (NAV), other health professionals
Outcomes	**Sickness absence** *At work at 18 months follow-up (maintained or work participation* • **Outcome type**: Dichotomous Outcome **Depressive symptoms** *Depression HADS at 12 months follow-up* • **Outcome type**: Continuous Outcome
Notes	**Country: Norway** Authors provided a re-analysis of the data for participants with a score of 8 or higher on the HADS.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "computer-generated randomisation list stratified by centre."
Allocation concealment (selection bias)	Low risk	Quote: "The allocation code, including details of block size (10), was not revealed to the researchers or the clinicians until recruitment and data collection were complete."
Blinding of participants and personnel (performance bias)	Low risk	Comment: Unblinded but unlikely that this will have led to different behaviour

Reme 2015 *(Continued)*

Sick Leave		
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Comment: Participants could not be blinded but outcome objective register-based
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: Unblinded and self-report
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Quote: For the secondary outcomes (mental health), we per- formed analyses with inverse probability weights 22 to account for possible attrition bias. Analyses adhered to the 'intention-to-treat' principle.
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Comment: Data on the main outcome measure, work participation, were complete for all participants.
Selective reporting (reporting bias)	Low risk	Judgement comment: Prospectively published protocol; all outcomes reported
Other bias	Low risk	Judgement comment: No other sources of bias detected

Reme 2019

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Follow-up:** 18 months for sick leave and 12 months for depressive symptoms **Number of trial arms:** Two **Recruitment:** Participants were recruited to any of the six IPS centers from regional primary and secondary mental healthcare settings while they were undergoing treatment for moderate-to-severe mental illness
Participants	**Baseline characteristics** IPS • *Age*: 34.92 (10.78) • *Gender*: Female: n = 114; 49.8% • *Marital status*: Married/ cohabiting: 43; 19.2% • *Occupation*: not provided • *Sick leave status*: 4.8% • *Number of participants randomised*: 229 (N = 85 in the depressed group) No intervention or Care as Usual • *Age*: 34.92 (10.78) • *Gender*: Female: n = 85; 47% • *Marital status*: Married/ cohabiting: 45 ;25.3% • *Occupation*: n.a. • *Sick leave status*: 7.2%

Reme 2019 *(Continued)*

- *Number of participants randomised*: 181 (N = 53 in the Care as Usual group)

Overall

- *Age*:
- *Gender*:
- *Marital status*:
- *Occupation*:
- *Sick leave status*:
- *Number of participants randomised*: 410 (N = 138 in depressed subgroup)

Inclusion criteria: At least one diagnosed psychiatric disorder (for this review only data of depressed subgroup were used). Currently out of the labor market but with an expressed desire to work.

Exclusion criteria: Insufficient Norwegian language skills to answer the questionnaires

Pretreatment: The only statistical significant difference between groups was the number with bipolar disorder which was higher in the intervention group, however the proportion of persons with a specific diagnosis was similar between groups

Setting: The participants were recruited through their treatment provider but the IPS was conducted in specialized units related to the employment office.

Interventions

Intervention characteristics

IPS

- *Content*: IPS, followed a structured and manualized approach focused on competitive employment. IPS is a structured approach based on eight principles; competitive employment, eligibility based on client choice, integration of rehabilitation and mental health services, attention to client preferences, personalized benefits counseling, rapid job search (starting within one month), systematic job development, and time unlimited and individualized support.
- *Duration, frequency, length*: Not provided
- *Communication means*: Face-to-face
- *Providers*: Employment specialist

No intervention or Care as Usual

- *Content*: At the Norwegian Labor and Welfare Administrations (NAV) participants received a high quality version of treatment as usual (TAU). This involves being offered a prioritized spot in a vocational rehabilitation scheme, primarilyWork with assistance (AB) and/or Traineeship in a sheltered business (APS). AB involves assistance by a personal facilitator, and includes finding suitable work, negotiating wage and employment conditions, modified duties, and follow-up at the work place. APS involves testing of work capability within a sheltered environment doing tasks that are modified to individual skills and challenges, with follow-up as necessary by an advisor. Participants in this group may also be offered additional interventions based on the individual needs, as they normally would in TAU.
- *Duration, frequency, length*: Not specified.
- *Communication means*: face-to-face
- *Providers*: Not reported

Outcomes

Sickness absence

Employed

- **Outcome type**: DichotomousOutcome
- **Reporting**: Fully reported
- **Unit of measure**: percentage at work
- **Direction**: Higher is better
- **Data value**: Endpoint

Reme 2019 *(Continued)*

	Depressive symptoms *HADS* • **Outcome type**: ContinuousOutcome • **Direction**: Lower is better • **Data value**: Endpoint
Notes	**Country:** Norway

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Comment: Protocol states: Using computer-generated randomization lists, the participant will be allocated to one of the two groups.
Allocation concealment (selection bias)	Low risk	Quote: "After randomization, participants were informed about group allocation. Participants in the intervention group were given a date for their first session, and participants in the control group were referred to the Norwegian Labor and Welfare Administration (NAV) for a prioritized spot in a vocational rehabilitation scheme." Judgement comment: The protocol states:' When the participant has filled out the baseline-questionnaires, the person conducting the introductory interview contacts the research technician at Uni Research Health by email, stating the participants ID-number, gender, and year of birth.'
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Participants were aware that they were in the intervention group and could have changed their behaviour.
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Sick leave was used from an administrative database and thus an objective outcome
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Participants reported depressive symptoms which can have changed due to the knowledge of being part of an intervention.
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	More than 20% for depressive symptoms
Incomplete outcome data (attrition bias) Sick Leave	Low risk	No loss to follow.up.
Selective reporting (reporting bias)	Low risk	Comment: Not all secondary outcomes announced in the protocol reported: fatigue, drug abuse, social support but these are outside our scope
Other bias	Low risk	Comment: No other sources of bias detected

Romeo 2004

Study characteristics	
Methods	**Study design:** RCT. multicenter, 2 conditions. **Recruitment:** from general practitioners' practices. **Follow up**: 24 weeks. **Lost to follow up:** T1: 6%; T2: 14%
Participants	**Inclusion criteria**: > 18 years old; Depressive episode according to DSM-IV checklist; 17-HAM-D score > 18 **Exclusion criteria**: schizophrenia, Bipolar, suicidal, illicit drug abuse or alcohol dependence; Treatment with any other psychotropic drug within 1 week before entry, or mirtazapine or paroxetine during the present episode, or treatment within 5 weeks before entry with fluoxetine, or any other antidepressant within 2 weeks before entry; renal, hepatic, respiratory, cardiovascular, or cerebrovascular disease; pregnancy or lactating, or no contraception **Baseline characteristics** 177 were randomised: (T1:93; T2:84) Age: T1: 40 (SD 14.3); T2: 40 (SD 11.7) Female: T1: 75%; T2: 71% Fulltime or part-time employed: T1: 48%; T2: 58% **Setting:** primary care, outpatients
Interventions	T1: Mirtazapine (TCA): 30 tto 45 mg/day oral Week 1 - 4 30 mg/day Week 5 - 24: optional increase to 45 mg/day (discretion of the investigator) T2: Paroxetine (SSRI): 20-30 mg/day oral Week 1 - 4: 20 mg/day Week 5 - 26 optional increase to 30 mg/day (discretion of the investigator)
Outcomes	**Sickness absence** 1) total mean days lost due to illness in 24 weeks **Depressive symptoms** 1) primary: change from baseline on 17-HAM-D; Secondary: 17-HAM-D responder rates (= at least 50% change from baseline to endpoint); 17 HAM-D remitter rates (= % with score of 8 or less on two assessments after the first score of 8 or less)
Notes	Country: UK

Risk of bias		
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	A randomisation list was used that was prepared in advance "Randomization was performed according to centrally prepared randomization lists."
Allocation concealment (selection bias)	Low risk	"Randomization was performed according to centrally prepared randomization lists." Personal communication: "The person assessing eligibility for inclusion was blind to allocation concealment."

Romeo 2004 *(Continued)*

Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Double-blind study design. Personal communication: "Medication was dispensed by the GP who was blinded to treatment allocation."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Double-blind study design. Sick leave was assessed by questionnaires filled out by patients, who were blinded to treatment allocation
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	Double-blind study design. Personal communication: "Outcomes were assessed by trained research nurses who were blind to treatment allocation."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Lost to follow-up: T1: 6%; T2: 14% and no appropriate imputation methods have been used "Six excluded mirtazapine patients, four were lost to follow-up, one dropped out early, and one refused participation in the study. Of the 14 excluded paroxetine patients, five were lost to follow-up, four were early drop outs, two did not participate any further, one discontinued due to the lack of efficacy, one was hospitalized as a results of a concomitant disease and one did not fulfil the selection criteria." "The high attrition rate observed in our study should be taken in to account when interpreting efficacy results due to possible influence on overall efficacy results."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Lost to follow-up: T1: 6%; T2: 14% and no appropriate imputation methods have been used. "Six excluded mirtazapine patients, four were lost to follow-up, one dropped out early, and one refused participation in the study. Of the 14 excluded paroxetine patients, five were lost to follow-up, four were early drop outs, two did not participate any further, one discontinued due to the lack of efficacy, one was hospitalized as a results of a concomitant disease and one did not fulfil the selection criteria." "The high attrition rate observed in our study should be taken in to account when interpreting efficacy results due to possible influence on overall efficacy results."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Rost 2004

Study characteristics	
Methods	**Study design:** RCT, randomisation on the level of practice, 12 practices were randomised. **Recruitment:** Trained administrative staff recruited patients who made routine-length visits to physicians. They asked eligible (see inclusion) patients to participate in 2 min first stage depression screener. Patients who screened positive and did not meet exclusion criteria were immediately invited to complete 5 min second stage screener. If they screened positive, they were asked to participate in study. **Follow up:** 24 months. **Lost to follow up**: 27%

Rost 2004 *(Continued)*

Participants	**Inclusion criteria:** Age > 18, sufficient literacy in English and cognitive function to complete surveys requiring 6-months recall, acces to telephone; Positive first screen: 2 weeks or more depressed or loss of interest in past year AND 1 week or more of this in last month; Second screen: 5 or more of 9 criteria for major depression in past 2 weeks on Inventory to diagnose depression. **Exclusion criteria:**pregnant, breastfeeding or < 3 months postpartum; Acute life-threatening physical conditions; Pos screeners who reported that symptoms started after loss of a loved one; pos screeners who did not intend to receive ongoing care in the clinic in the next year; Second stage screener: self-report lifetime mania, use of lithium or current alcohol dependence **Baseline characteristics** 326 employed persons were randomised: (T1: 158; T2: 168). Age: T1: 37.9 (SD 10.9); T2: 40.2 (SD 10.3) Female: T1: 84.2%; T2: 85.7% Married: T1: 45%; T2: 51% Employed: 100% **Setting:** Community primary care practices
Interventions	T1: Enhanced care. Primary care team was trained to provide high quality depression treatment. After enrolment, patients were evaluated for depression by physician and asked to return within one week to nurse care manager. Subsequent visit: education about treatment, addressing treatment barriers, checklist for physician's review, scheduling of next appointment in one week. This continued for 5-7 weeks. Then patients were monitored (symptoms and treatment adherence) for one year. Physicians reviewed patients monthly based on report of nurses to see whether guideline recommendations were followed. Medication algorithm of guideline: initially SSRI or secondary amine tricyclic. Switch drug classes when response failure T2: Usual Care. Regular Primary physicians care
Outcomes	**Sickness absence** 1) total number of work hours lost due to illness or doctor visits over past 4 weeks **Depressive symptoms** 1) depression severity: CES-D (adapted) **Work functioning:** 1) subjective rating on 0 to 10 scale of productivity at work
Notes	Country: US

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	"The practices were stratified and matched into six pairs." "Within each pair, one practice was randomized to the 'enhanced' care condition and the other practice delivered usual care to study participants."
Allocation concealment (selection bias)	High risk	Personal communication: "The allocation of the practice was known to the administrative staff who screened patients."
Blinding of participants and personnel (performance bias)	Low risk	Personal communication: "The allocation of the practice was known to patients eligible to participate. However, these patients did not know that there was another arm of the study that other practices participated in."

Rost 2004 *(Continued)*

Sick Leave		
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "We measured absenteeism at baseline, 6, 12, 18, and 24 months by calculating lost work hours from employee reports of how many full workdays and part workdays they missed due to illness or doctor visits, reflecting that employee reports demonstrate high agreement with employer records of absenteeism."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression was measured by self-report (CESD-D) and patients were not blinded to treatment allocation
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up at endpoint is considered to be high (27%). Risk of attrition bias was therefore deemed high although analyses accounted sufficiently for missing data according to authors: "Because analysis of missing data patterns produced no evidence of non ignorable missingness, we present unweighted models, noting that weighted models produce closely comparable results."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Loss to follow up at endpoint is considered to be high (27%). Risk of attrition bias was therefore deemed high although although analyses accounted sufficiently for missing data according to authors: "Because analysis of missing data patterns produced no evidence of non ignorable missingness, we present unweighted models, noting that weighted models produce closely comparable results."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Sarfati 2016

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** Through clinics and advertisements **Follow-up:** 12 weeks
Participants	**Baseline characteristics** Psychological treatment (T-CBT) plus antidepressant • *Age*: 42.3 ± 10.4 • *Gender*: 44% male • *Marital status*: 42% married • *Occupation*: Production 25% • *Sick leave status*: 9.3 ± 16.1 hours work missed in past two weeks • *Number of participants randomised*: 52

Sarfati 2016 *(Continued)*

Antidepressant with medication reminder telephone calls

- *Age*: 44.2 ± 9.9
- *Gender*: 47% male
- *Marital status*: 35% married
- *Occupation*: Production 24%
- *Sick leave status*: 9.5 ± 12.6 hours missed/2 weeks
- *Number of participants randomised*: 53

Inclusion criteria: (a) male and female out-patients aged 19–65 years (b) diagnosis of major depressive disorder by DSM-IV criteria, as confirmed using the Mini International Neuropsychiatric Interview (c) current paid employment of more than 15 h/week (d) score of 19 or higher on the Montgomery–Asberg Depression Rating Scale (MADRS), indicating at least moderate depression, at both screening and baseline (e) competency to give informed consent.

Exclusion criteria: (a) off work on short- or long-term disability (b) pregnant or lactating women, and sexually active women of child-bearing potential who were not using medically accepted means of contraception (c) serious suicidal risk as judged by the clinician (d) unstable medical conditions (e) diagnoses of organic mental disorders, substance misuse/dependence, including alcohol, active within the past year schizophrenia or other psychotic disorders; primary diagnosis of panic disorder, generalised anxiety disorder, obsessive–compulsive disorder, or post-traumatic stress disorder;bipolar disorder; eating disorders(f) use of antidepressants or psychotropic drugs within 7 days of baseline visit (14 days for monoamine oxidase inhibitors,5 weeks for fluoxetine) (g) treatment-resistance in the current episode, as defined by failure (lack of clinically significant response) of two or more antidepressants at therapeutic doses for at least 6 weeks (h) previous use of escitalopram or CBT for depression (i) use of any additional treatment for depression during the study.

Pretreatment: No significant or relevant differences in baseline age, marital status, education, job type , income, length of current episode, depression rating, employment absence and productivity scale, health and work performance scale.

Setting: Clinical

Interventions	**Intervention characteristics** Psychological treatment plus antidepressant • *Content*: SSRI (Escitalopram) 10–20 mg/day and a telephone-administered CBT programme that is based on a published manual and validated in an RCT in primary care was used. The initial session focused on motivation enhancement exercises, whereas subsequent sessions emphasised identifying, challenging and distancing from negative thoughts,and the final session focused on a personal care plan and self-management skills. There was no systematic consideration of work-related issues in this programme. • *Duration, frequency, length*: For 8 weeks, weekly, 30-40 minutes • *Communication means*: Telephone • *Providers*: The CBT providers were PhD- or Master's degree-level experienced therapists who received formal training by the developers of the treatment manual and fidelity was monitored by inspection of therapist task check lists for each session and review of random audio-taped sessions. • *Name*: Telephone-administeredcognitive–behavioural therapy plus Escitalopram Antidepressant with medication reminders • *Content*: SSRI (Escitalopram) 10–20 mg/day and adherence-reminder telephone calls. • *Duration, frequency, length*: A research coordinator provided a 10-minute structured telephone call weekly for 8 weeks, with enquiry aboutet progress and reminders to take medication properly. • *Communication means*: Telephone • *Providers*: "A research coordinator" • *Name*: Adherence-reminder telephone calls.
Outcomes	**Sickness absence**

Sarfati 2016 *(Continued)*

Sick Leave: Days lost in past two weeks due to health reasons

- **Outcome type**: Continuous Outcome
- **Unit of measure**: hours missed recalculated into days missed

Depressive symptoms:

Patient Health Questionnaire Depression Score

- **Outcome type**: ContinuousOutcome

Work functioning: Work ability, Sheehan Disability Scale

- **Outcome type**: ContinuousOutcome
- **Range**: 0-10

Notes	**Country:** Canada

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "central computerised randomisation process was generated by an independent statistician, stratified for site and conducted in random blocks of 4 or 8." Judgement comment: Adequate
Allocation concealment (selection bias)	Low risk	Quote: Concealment of allocation was accomplished using an automated online system that revealed the treatment allocation only after the unique participant number was entered.
Blinding of participants and personnel (performance bias) Sick Leave	High risk	Comment: unblinded and control condition (SSRI + adherence telephone calls) cannot be considered considered equally desirable as the intervention (SSRI + telephone administered CBT).
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Comment: self-report and unblinded
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	Quote: Within 2 days of each study visit, participants were rated using the MADRS over the telephone by trained independent evaluators, masked to treatment assignment and adverse events, using a structured interview guide.
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Quote: Analysis was conducted based on a modified intent-to-treat (mITT) sample comprising randomised patients who had at least one valid post-randomisation assessment. Missing data were imputed using last observation carried forward (LOCF). An observed-case completer analysis was also conducted on the sample of participants with data at the primary week 12 end-point. Comment: Authors used LOCF but inappropriate in situation that could also evolve in a negative direction In the intervention group 12 were not available for analysis and in the control group 7. We used the completer data in our analysis because the LOCF could led to too favourable results
Incomplete outcome data (attrition bias)	High risk	The same as for depressive symptoms

Sarfati 2016 *(Continued)*
Sick Leave

Selective reporting (reporting bias)	High risk	Judgement comment: The protocol states that there will be a 6 months assessment, which is not reported. It is also noteworthy that the study is powered for detecting a change in depression scores, where the protocol says: "Outcome will be rigorously evaluated by assessing absenteeism and work productivity, response and remission rates, and quality of life, after acute (3 month) treatment and longer-term (6 month) follow-up"
Other bias	Low risk	Judgement comment: No other biases detected

Schene 2006

Study characteristics	
Methods	**Study design**: RCT, two conditions; **Recruitment:** Regular referrals (including from occupational physicians) **Follow up:** 42 months. **Lost to follow up** at 12 months: T1: 13%; T2: 3%; at 42 months: T1: 25%; T2: 20%
Participants	Inclusion:18 years; MDD (single episode or recurrent); BDI score >15; Work absenteeism due to depression of at least 50% of regular hours worked per week with a duration between 10 weeks and 2 years;Clinically estimated contribution of work to the onset and/or continuation of depression of > 50% of supposed causal factors Exclusion: MDD with psychotic features;history of psychosis, manic, hypomanic, or cyclothymic features; history of active drug or alcohol abuse or dependence; personality disorder according to DSM-IV **Baseline characteristics** 62 were randomised (T1:32; T2:30). Age: T1: 45.2 (SD 7.5); T2: 46.6 (SD 7.4) Female: T1: 53%; T2: 50% Married: T1: 63%; T2: 53% Mean hours employment: T1: 36.5 (SD 10.4); T2: 36.4 (SD 7.8) **Setting:** outpatient unit of Psychiatric department of Academic hospital.
Interventions	T1: Treatment as usual (TAU) following evidence-based guidelines (APA Guideline);This consisted of clinical management according to APA Guideline and antidepressants, if indicated and accepted by patients, according to our standardized stepwise drug treatment regimen or algorithm. Visits consisted of symptom assessment, psycho-education, general support and cognitive behavioural techniques, and if indicated medication prescription, dose titration and review of adverse effects. In case of any clinical significant deterioration in condition patients could be referred for partial or full-time hospitalisation within the Program. Patients were treated by three supervised senior psychiatric residents. Visits regularly took 30 minutes every 2 to 4 weeks T2: Treatment as usual + occupational therapy (TAU + OT) TAU plus occupational therapy (OT): same outpatient treatment; OT: diagnostic phase (4 weeks): occupational history, video observation in a role-played work situation, contact with occupational physician of patient's employer and written conclusions including a plan for work reintegration therapeutic phase (24 weeks): this phase had three sub-phases: preparation of work reintegration, contacting the working place and if possible starting to work. In the individual sessions these three phases were followed: further analyses of the relationship between work and depression, exploration of work problems, and support and evaluation of work resume. Specific individual issues from the group sessions were elaborated. The first half of these two-hour group sessions were spend on discussing and

Schene 2006 *(Continued)*

	exchanging individual progress. In the second half seven themes were successively discussed: being passive, stress on the work place, personal bounds and limits, powerful and powerless, perfectionism, conflicts and prevention. Patients were treated by three supervised senior psychiatric residents. + two occupational therapists diagnostic phase (4 weeks): 5 visits therapeutic phase (24 weeks): 24 weekly group sessions (8-10 patients) and 12 individual sessions (45 minutes) follow-up phase (20 weeks): 3 individual visits
Outcomes	**Sickness absence** 1) total number of hours worked during 6-month periods up to 42nd month (primary outcome) 2) proportion of patients working at least 1 hour per week 3) proportion of patients working at least 16 hours per week 4) time from T1 to partial or full return to work **Depressive symptoms** 1) % meeting DSM IV criteria at 6/42 months 2) change in BDI at 6/42 months 1) depression according to DSM-IV at 12 months 2) change in BDI-score (baseline-12 months)
Notes	**Country:** the Netherlands

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Patients who met the inclusion criteria were randomly assigned to TAU or TAU +OT in blocks of 20 by use of computer-generated cards stored as concealed assignment codes in consecutively number sealed envelopes under the responsibility of an independent research associate."
Allocation concealment (selection bias)	Low risk	"Patients who met the inclusion criteria were them randomly assigned to TAU or TAU +OT in blocks of 20 by use of computer-generated cards stored as concealed assignment codes in consecutively number sealed envelopes under the responsibility of an independent research associate."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Personal communication: "patients and clinical personnel were not blinded." It is unlikely that the knowledge of the intervention would have led to different behaviour.
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "Work resumption data were assessed by a study-specific questionnaire at T2, T3, T4 and T5."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	"Depression was assessed by the BDI, a self-report measure of severity of depressive symptoms." Patients were not blinded to treatment allocation
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Loss to follow up was high: T1: 25%; T2: 20%. Risk of attrition bias was therefore deemed high even though appropriate imputation methods have been used: "Complete T4 data were obtained on 28 (88%) of TAU patients and on 29 (97%) of TAU +OT patients. For T5 these figures were 24 (75%) for TAU and 24 (80%) for TAU + OT." "Both GEE and Proc Mixed give unbiased effect estimates taking into account missing data."
Incomplete outcome data (attrition bias)	High risk	Loss to follow up was high: T1:25%; T2: 20%. Risk of attrition bias was therefore deemed high even though appropriate imputation methods have been

Schene 2006 *(Continued)*

Sick Leave		used: "Complete T4 data were obtained on 28 (88%) of TAU patients and on 29 (97%) of TAU +OT patients. For T5 these figures were 24 (75%) for TAU and 24 (80%) for TAU +OT." "Both GEE and Proc Mixed give unbiased effect estimates taking into account missing data."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Schoenbaum 2001

Study characteristics	
Methods	**Study design:** RCT with randomisation on the level of clinic. Clinic clusters were matched based on patient demographics, clinician specialty, and distance to mental health providers. **Recruitment:** study staff screened consecutive patient visitors. **Follow up**: 24 months. **Lost to follow-u**p: T1: 15%; T2: 13%
Participants	**Inclusion criteria:** depressed, intend to use clinic for next 12 months; Probable depressive disorder:at least 2-weeks depressed mood or loss of interest in last year or persistent over year + at least 1 week depression in last 30 days **Exclusion criteria:** < 18 years, acute medical emergency, did not speak English or Spanish, no insurance or public pay arrangement that covered care delivered by mental health specialists **Baseline characteristics** 1356 were randomised (T1:913; T2: 443). Age: T1: 44.5 (SD15.5); T2: 42.2 (SD 13.9) Female: T1: 74%; T2: 69% Married: T1: 54%; T2: 55% **Setting**: Primary care
Interventions	T1: Quality improvement program (QI meds or QI therapy). Treatment type or content Quality improvement (QI) program: practices were provided with training and resources to initiate and monitor QI programs according to local practice goals and resources. For both interventions (QI-meds and QI therapy): local practice teams were trained in a 2-day workshop to provide clinician education and to supervise intervention staff. Practice nurses were trained as depression specialists, following a written protocol, to assist in initial patient assessment, education and motivation for treatment. Practice teams were given patient education pamphlets and videotapes, patient tracking forms, and clinician manuals and pocket reminder cards and were encouraged to distribute them. The materials described guideline-concordant care and described antidepressant medication and psychotherapy as equally effective. In both conditions resources were made available to obtain specific form of therapy (medication or psychotherapy) For QI-meds: nurse specialists were trained to support medication adherence through monthly telephone contacts or visits for 6 or 12 months, randomised at patient level In QI-therapy: practice therapists were trained to provide individual and group CBT, following a protocol T2: Usual care: mailing of practice guidelines to primary care professionals
Outcomes	**Sickness absence** 1) days worked during 24 months follow-up for whole sample

Schoenbaum 2001 *(Continued)*

	2) number of reported sick days for employed subsample in previous 4 weeks at each 6 months period 3) Off work: being at work after 12 months **Depressive symptoms** 1) CES-D
Notes	Country: US

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Within blocks, we used a random number table to assign clusters to usual care or QI interventions."
Allocation concealment (selection bias)	High risk	Randomisation was on the level of practice and primary care clinicians were not blinded for allocation during enrolment of patients
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Patients and personnel were not blinded: "We asked all primary care clinicians to enroll prior to their knowledge of intervention status." "Patients learned of their intervention status after enrolment." Personal communication: "Subjects in the interviews were not blinded, but may or may not have known their intervention status given the nature of interventions." Unlikely that knowledge of the intervention would have led to different behaviour.
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation "We also examined days missed from work due to illness, which patients reported for the 4 weeks preceding each follow-up study."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Depression was measured by self-report and patients were not blinded to treatment allocation. 'We assessed depressive symptoms at baseline and follow-up using a 23-item version of the Center of Epidemiologic Studies Depression (CES-D) Scale, developed by Daniel Ford. This version drops 6 items and adds others to approximate DSM-IV criteria. Items responses were summed."
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Lost to follow up for the depressive symptoms is 15% but appropriate imputation methods have been used. "The data are weighted for the probability of study enrolment and follow-up response to the characteristics of the eligible sample. We used multiple imputations for missing items at each wave."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Lost to follow for the economic survey is 15% but appropriate imputation methods have been used. "The data are weighted for the probability of study enrolment and follow-up response to the characteristics of the eligible sample. We used multiple imputations for missing items at each wave."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Simon 1998

Study characteristics	
Methods	**Study design**: RCT (consisting of 2 sub-studies) with two conditions. **Recruitment:** participating primary care physicians were asked to refer any adult outpatient initiating care for depression and willing to consider treatment with antidepressant medication. The research assistant screened for eligibility **Follow up:** 7 months. **Lost to follow up:** sub-study 1: 15%; sub-study 2: 23%
Participants	**Inclusion criteria:** diagnosis definite or probable major depression by primary care physician; Agreed to antidepressant medication; SCL-score of at least 0.75; Age 18 to 80 yrs **Exclusion criteria:** current alcohol abuse (score at least 2 CAGE questionnaire); current psychotic symptoms or serious suicidal ideation or plan; dementia; pregnancy; terminal illness; limited; command of English; plan to disenrol from insurance plan within 12 months **Baseline characteristics** 156 patients with MDD were randomised (T1: 80; T2: 76). Age: substudy1: T1: 43.2 (SD 15.4); T2: 42.3 (SD 12.7); substudy2: T1: 43.1 (SD 9.3); T2: 44.8 (SD 15.9) Female: substudy1: T1: 78%; T2: 88%; sub-study: 77%; T2: 74% Married or cohabiting: substudy1: T1: 47%; T2: 55%; substudy2: T1: 48%; T2: 32% Employed: sub-study 1: T1: 71%; T2: 63% sub-study 2: T1: 87%; T2: 74% **Setting:**Primary care
Interventions	T1: Improved Care. Multifaceted intervention. Goal: increase likelihood that treatment would be conform primary care depression guidelines Components: (1) written and videotaped patient education material (2) increased frequency of follow-up visits during first 8 weeks (3) advice to physicians regarding changes in pharmacotherapy (4) monitoring of medication side-effects, medication adherence, treatment response and follow-up visits frequency by study staff to treating physician substudy1, psychiatrist-liaison collaborative intervention: (a) co-management by consulting psychiatrist and physicians during first 6 weeks of treatment, (b) 1 week after start treatment all patients attended an extended structured visit with physician to review symptoms, barriers to adherence, side-effects, and goals for behavioural activation. (c) after 2 weeks: consultation with study psychiatrist discussing treatment response and medication (adjustment if needed), (d) week 3 physician visit, (e) week 4 psychiatrist visit (f) monthly case conferences between psychiatrist and physician substudy 2, psychologist-liaison collaborative intervention: Standardised brief psychotherapy program. Face-to-face psychiatric consultation on as-needed basis. Components psychotherapy: (a) education, skills training, and written homework (b) interventions to enhance medication adherence (c) behavioural activation and (d) brief cognitive interventions. Weekly meetings between therapists and study psychiatrists. Study clinicians communicated with physicians throughout study about progress and changes in medication psychotherapy: 4-6 visits over 6 weeks (total time 2,5 to 3,5 hour) Telephone contacts at 2, 4, 12 and 24 weeks after last face-to-face session T2: Usual primary care. Could include any service normally available including pharmacotherapy, referral to mental health service or self-referral to non-GHC services
Outcomes	**Sickness absence** 1) % unable to work due to illness 2) n of days of missed work or school out of last 90 for employed sub-sample **Depressive symptoms** 1) proportion of patients with MDD who experienced at least 50% reduction in depressive symptoms on IDS

Simon 1998 *(Continued)*

	2) SCL for employed sub-sample 3) IDS for employed sub-sample
Notes	**Country:** US Data are provided for subgroup of MDD only, both sub-studies combined

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Patients were randomly assigned using computer generated random numbers."
Allocation concealment (selection bias)	Low risk	Personal communication: "The primary care physicians or the research assistant did not know anything about the randomization status of the next patient. Randomization was performed 1-7 days after the baseline assessment by the study manager."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Personal communication: "Patient participants and their treating clinicians were not blinded – and it would not have been possible to do so." It is unlikely that knowledge of the intervention will have changed behaviour
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Sick leave was measured by self-report and patients were not blinded to treatment allocation. "One of the four assessments included questions adapted from the National Health Interview Survey regarding days of missed work or school due to health."
Blinding of outcome assessment (detection bias) Depressive symptoms	Low risk	"Folllow-up telephone interviewers were blinded to treatments assignment." "Two of the assessments included a 20-item depression scale extracted from the Hopkins Symptom Checklist or SCL and a version of the clinician-rated Inventory of Depressive Symptoms or IDS modified for telephone administration."
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Lost to follow-up is considered to be high: T1: 17%; T2: 21%. Risk of attrition bias was therefore deemed high although appropriate imputation methods have been used: "Model were estimated using generalized estimating equations (GEE) to account for multiple assessments and to allow for missing data"
Incomplete outcome data (attrition bias) Sick Leave	High risk	Lost to follow-up is considered to be high: T1: 17%; T2: 21%. Risk of attrition bias was therefore deemed high although appropriate imputation methods have been used: "Model were estimated using generalized estimating equations (GEE) to account for multiple assessments and to allow for missing data"
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Vlasveld 2013

Study characteristics

Methods	**Study design:** RCT. Recruitment: 22 months.

Vlasveld 2013 *(Continued)*

	Follow up: 12 months. **Lost to follow up**: 41.3%
Participants	**Inclusion criteria:** workers on sickness absence between 4 and 12 weeks, whose absence was diagnosed by occupational physicians (OPs) as due to mental disorder, who screened positively for depressive disorder (i.e. score ≥ 10 on 9-item 0 to 27 depression subscale of Patient Health Questionnaire), who have informed consent and who met the DSM-IV criteria for MDD and gave written informed consent **Exclusion criteria:** workers who were suicidal, psychotic or had a primary diagnosis of substance abuse or dependence, as assessed by the MINI **Baseline characteristics** 126 were randomised (T1:65; T2:61) Age: T1: 43.4 (SD 11.4); T2: 41.9 (SD 11.4) Male: T1: 45.9%; T2: 46.2% Marital status: T1: 73.3% married or cohabiting; T2: 60.0% married or cohabiting Educational level: T1: 35.0% high; T2: 36.1% high; T1 30.0% average; T2: 36.0% average; T1: 35.0% low; T2: 27.9% low Dutch nationality: T1: 91.8%; T2: 95.4% **Setting**: occupational health care
Interventions	T1: Work-directed plus clinical intervention. Provided by the Occupational Physician Care Manager (OP-CM), contained the following elements: 6 to 12 sessions of Problem Solving Therapy, manual-guided self-help, a workplace intervention and, depending on patient preference, prescription of antidepressant medication according to a treatment algorithm. In order to enhance the adherence to the treatment model, ongoing supervision and psychiatric consultation was provided to the OP-CMs. Also, a web-based tracking system was developed to support the OP-CM in monitoring treatment outcomes and in adhering to the stepped care protocol. In case of questions regarding the treatment, prescription of antidepressants, or (lack of) progress of the worker, the OP-CM was prompted by the web-based tracking system to consult the psychiatrist T2: Usual care by GP. Sick-listed workers start to visit the company's OP before the 6th week of sickness absence. The guidance of company's OP is protocolised according to the OP guidelines of the Dutch Board for Occupational Medicine. In practice, whether or not sick-listed workers will receive treatment for MDD may vary considerable. The actual care that was provided was assessed by questionnaires in both groups
Outcomes	**Sickness absence** 1) the duration until lasting, full RTW. The duration until lasting, full RTW was defined as the duration of sickness absence due to MDD in calendar days, from the day of randomisation until full RTW for at least 4 weeks without partial or full recurrence 2) the total number of sickness absence days, calculated for the entire follow up **Depressive symptoms** 1) severity of depression, assessed by the PHQ-9 2) time to first response on depressive symptoms. Response is defined as a reduction in depressive symptoms of at least 50% 3) time to first remission, defined as a score of less than 5 on the PHQ-9

Vlasveld 2013 *(Continued)*

Notes	**Country:** the Netherlands

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The randomization scheme was prepared by a computer, with blocks of four, by an independent statistician."
Allocation concealment (selection bias)	Low risk	"While assessing eligibility for the study, both the research assistant and the participant were blinded for the allocation."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Quote: "Then, the participant was informed about the computer generated allocation status by the research assistant. Next, the baseline questionnaire was sent by mail." Comment: unlikely that patients or providers changed their behaviour based on knowledge of having the intervention.
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Low risk as sickness absence data were based on registration database. "Sickness absence data were derived from the register of the occupational service 1 year after randomization."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Data about depressive symptoms were collected by a self-report questionnaire and patients were not blinded to treatment allocation
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Lost to follow up was high. "Lost to follow-up rates at 3,6, 9 and 12 months were respectively 22.2%, 28.6%, 33.3% and 41.3%." Risk of attrition bias was considered high even though an appropriate method has been described to account for this missing data: 'If there is missing data on costs and/or effects, and the additional uncertainty it introduces, multiple imputation will be used."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	No missing sickness absence data
Selective reporting (reporting bias)	Low risk	The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way
Other bias	Low risk	None identified

Volker 2015

Study characteristics

Methods	**Study design:** Cluster randomized controlled trial **Study grouping:** Parallel group **Number of trial arms:** 2 **Recruitment:** 1. Via 29 occupational health physicians in 6 regional clusters 2. via one occupational health physician in one mental health institution **Follow-up:** 12 months **Subgroup analysis of participants with depression:**

Volker 2015 *(Continued)*

Participants	**Baseline characteristics** Work-directed intervention combined with clinical intervention • *Age*: 43.4 ± 9.5 • *Gender*: 23% male • *Marital status*: 91% • *Occupation*: not reported • *Sick leave status*: 73 days median sick leave • *Number of participants randomised*:131 • *Number in subgroup with depression*:88 No intervention or Care as Usual • *Age*: 45.5 ± 10.7 • *Gender*: 46% male • *Marital status*: 62% • *Occupation*: not reported • *Sick leave status*: 70 days median sick leave • *Number of participants randomised*: 89 • *Number in subgroup with depression: 55* **Inclusion criteria:** Employees (aged ≥18 years) who were on sickness absence between 4 and 26 weeks and screened positive (score ≥10) on either the depression scale of the PHQ-9 and/or the somatization scale of the PHQ-15 and/or the GAD-7 were included. We used only the data of participants score more than 10 on PHQ-9 **Exclusion criteria:** Employees were excluded for participating in this study if they had insufficient knowledge of the Dutch language, were pregnant, or were involved in legal action against their employer. Furthermore, employees without access to the Internet were excluded **Pretreatment:** No relevant differences **Setting:** Occupational health care
Interventions	**Intervention characteristics** Work-directed intervention combined with clinical intervention • *Content*: The intervention was made up of 1. a decision aid including half a day training for occupational health physicians 2. a 16 module web-based CBT/PST intervention for sick listed employees • *Duration, frequency, length*: Half a day training for physicians; it is unclear if there was a time limit for the participants in using the web-modules • *Communication means*: Web-based for the CBT and problem solving skills. Additional face-to-face contacts with occupational health physician • *Providers*: 1. experts through web-based programme 2. occupational health physicians • *Name*: E-health module embedded in Collaborative Occupational health care" (ECO Care as Usual-WD • *Content*: No training; care as usual according to professional guideline which was not adhered to well. • *Duration, frequency, length*: No training; number of contacts with participant and occupational health physician not reported • *Communication means*: Face-to-.face contacts • *Providers*: Occupational health physicians • *Name*: Care as Usual
Outcomes	**Sickness absence**

Volker 2015 *(Continued)*

	Sick leave days at end of follow-up • **Outcome type**: Continuous Outcome
Notes	**Country:** the Netherlands Sick leave data for depressed participants only received from authors; 143 participants scored more than 10 on PHQ; 89 intervention and 55 control participants. Of these 32 (58%) returned to work at end of follow-up in the control group and 59 (67%) in the intervention group. This yielded a HR of (HR 1.3, 95% CI 0.87-2.05). Used mean days of absence and SD as input for the review. Data on depression could not be re-analysed by the author.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Quote: "At GGz Breburg, only 1 occupational physician was available. For this reason, a cluster crossover design was used at first with the first 100 employees approached as the control condition and subsequently the second 100 employees approached as the intervention condition. However, at the end of the planned control condition, the occupational physician was replaced with another occupational physician, with whom the intervention condition was conducted. Therefore, this can be considered as a pseudo-randomization design in GGz Breburg." Quote: "The clusters of occupational physicians were randomized by an independent statistician using a computer algorithm for randomization." Judgement comment: Unclear if the total procedure led to a random sequence generation
Allocation concealment (selection bias)	Low risk	Judgement comment: "The research assistants and the participants were blind to the allocation when assessing the eligibility of sick-listed employees for participating in this study.""Employees who were considered as screen-positive on any of the 3 screening instruments were contacted by a research assistant, who was blinded to group assignment, by telephone."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Comment: Unblinded but unlikely that this knowledge led to different behaviour
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Comment: Even though patients were not blinded, RoB is low as the outcome is objectively retrieved from employers' registers
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Comment: self-report and unblinded
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Quote: Sickness absence data were available for 86 employees in the control condition and for 130 employees in the intervention condition. For unknown reasons, the sickness absence data of 4 participants could not be found in the registers. These 4 participants did not differ significantly on average at baseline on sickness absence duration, depressive, somatization, or anxiety symptoms from the other participants
Incomplete outcome data (attrition bias)	High risk	Comment: Only 60% response at latest follow-up

Volker 2015 *(Continued)*
Sick Leave

Selective reporting (reporting bias)	High risk	Quote: "Netherlands Trial Register NTR2108; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2108. (Archived by WebCite at http://www.webcitation.org/6YBSnNx3P)." Quote: "Secondary outcome measures were the severity of depression, anxiety, and somatization symptoms as measured with the PHQ-9, GAD-7, and PHQ-15 in terms of response and remission. Response was defined as a 50% reduction in symptoms on the PHQ-9, GAD-7, or PHQ-15, with the restriction that the baseline score on the questionnaire on which response was evaluated was above the cut-off point of 10 (otherwise it was defined as no response). Remission was defined as a score lower than 5 on the PHQ-9, GAD-7, or PHQ-15, with the restriction that the baseline score on the questionnaire on which remission was evaluated was above the cut-off point of 10 [24-26]." Judgement comment: According to the protocol in 2009 the trial was first set up for workers with major depressive disorder. Start and closing dates of recruitment were reported as 1.11.10 and 1.10.14. As a result of a new sponsor in 2014 the focus changed to CMD. The protocol first reports: secondary outcomes: severity of depressive symptoms, as measured with the Patient Health Questionnaire (PHQ9), and Costs. As a result of the new sponsor, more outcomes are added.
Other bias	High risk	Judgement comment: There were two sources of patients: 1. clusters of occupational health physicians who were randomised to control and intervention condition. 2. one occupational health physician at a mental health institution who was assigned to the control conditions. This person was replaced by another occupational health physician who then was allocated to the intervention condition.

Wade 2008

Study characteristics	
Methods	**Study design**: A double-blind, multinational randomised study. **Recruitment:** outpatients with MDD were recruited in psychiatric and general practice settings, from September 2005 to September 2006. **Follow up:** 24 weeks. **Lost to follow up**: 23% (clinical outcome) and 24.4% (sick leave)
Participants	**Setting**: outpatients of 35 centra of psychiatric and general practice settings. **Inclusion criteria**: patients with MDD (current episode assessed with the MINI), according to the DSM IV-TR criteria, outpatient of either sex, aged 18-65 years, with a MADRS total score ≥ 26 and a CGI-S score ≥ 4 at baseline visit. Patients with a secondary current comorbid anxiety disorder (DSM-IV TR criteria) could be included in the study, expect for obsessive-compulsive disorder, post-traumatic stress disorder, or panic disorder. **Exclusion criteria**: if they met one or more of the DSM IV-TR criteria for any of the following: bipolar disorder, psychotic disorder or features, current eating disorders (anorexia nervosa, bulimia), mental retardation, any pervasive developmental disorder or cognitive disorder, or alcohol or drug abuse-related disorders within 12 months prior to baseline. In addition, patients at serious suicide risk, based on the investigator's clinical judgement, or who had a score of ≥ 5 on item 10 of the MADRS scale, were also excluded, as were those receiving formal behavioural therapy, or systematic psychotherapy, or were pregnant or breast feeding, or had a history of lactose intolerance. Patients with a history of hy-

Wade 2008 *(Continued)*

persensitivity or non-response to citalopram, or escitalopram, or duloxetine, or with increased intra-ocular pressure, or at risk of acute narrow-angle glaucoma, were also excluded. Patients were also excluded if they were taking the following psychotropic drugs within 2 weeks prior to baseline or during the study: MAOI or RIMA, SSRI (fluoxetine within 5 weeks), SNRIs, and tricyclic antidepressants, tryptophan, psychoactive herbal remedies, any drug used for augmentation of antidepressant action or any other antidepressant drugs, oral antipsychotic and anti-manic drugs (including lithium), or ECT (within 6 months), dopamine antagonists, any anxiolytics (including benzodiazepines), any anticonvulsant drug, serotonergic agonists, narcotic analgesics, cardiac glycosides, type 1c anti-arrhythmics, oral anticoagulants, cimetidine, potent inhibitors of CYP2C19, CYP1A2, or medicinal products with a narrow therapeutic index predominantly metabolised by CYP2D6.

Baseline Characteristics

295 were randomised (T1: 144; T2: 151).

Female:

T1: 73.8%; T2: 71.2%

Age:

T1: 43.3 (SD 11.6); T2: 44.5 (SD 11.0)

Marital status:

T1: 27.0% single; T2: 20.5% single

T1: 50.4% married or living as a couple; T2: 50.7% married or living as a couple

T1: 17.7% divorced or separated; T2: 25.3% divorced or separated

T1: 5.0% widowed; T2: 3.4% widowed

Level of education:

T1: 5.0% no degree or diploma; T2: 4.1% no degree or diploma

T1: 29.1% elementary school; T2: 26.0% elementary school

T1: 43.3% high school; T2: 45.2% high school

T1: 11.3% non-university degree; T2: 15.1% non university degree

T1: 11.3% university; T2: 9.6% university

Employment status:

T1: 58.9% paid employment or self-employed; T2: 60.3% paid employment or self-employed

T1: 15.6% unemployed; T2: 18.5% unemployed

T1: 5.0% student; T2: 4.8% student

T1: 6.4% non-working spouse; T2: 3.4% non-working spouse

T1: 7.8% retired; T2: 10.3% retired

T1: 6.4% other; T2: 2.7% other

Occupational status:

T1: 34.8% no data available; T2: 36.3% no data available

T1: 6.5% manager or administrator; T2: 12.9% manager or administrator

T1: 16.3% professional; T2: 15.1% professional

Wade 2008 *(Continued)*

	T1: 10.9% associate professional; T2: 10.8% associate professional T1: 8.7% clerical worker/secretary; T2: 10.8% clerical worker/secretary T1: 26.1% skilled labourer or factory worker; T2: 17.2% skilled labourer or factory worker T1: 27.2% services/sales (retail); T2: 24.7% services/sales T1: 4.3% other; T2: 8.6% other
Interventions	T1: escitalopram (SSRI), 10 mg/day for the first 2 weeks, and 20 mg/day for the rest of the period T2: duloxetine (SNRI), 60 mg/day for the 24 weeks, in accordance with the recommendations in the package insert for duloxetine in the participating countries
Outcomes	**Sickness absence** 1) percentage of patients taking sick leave 2) mean per patient sick leave duration in days **Depressive symptoms** 1) adjusted mean change in the MADRS total score 2) MADRS total score 3) HAMD-17 4) remission, defined as MADRS ≤ 12 or post hoc as HAMD-17 ≤ 7) 5) response, defined as ≥50% decrease from baseline in MADRS or (post hoc) HAMD total score **Work functioning:** 1) impairment, assessed by the Sheehan Disability Scale
Notes	**Country:** UK

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Patients who met the selection criteria at the baseline visit were assigned to 24 weeks of double-blind treatment in a 1:1 ratio of escitalopram or duloxetine treatment according to a computer-generation randomization list." "At each study centre, sequentially enrolled patients were assigned to the lowest randomization number available in blocks of 4."
Allocation concealment (selection bias)	Low risk	"The details of the randomization series were unknown to any of the investigators and were contained in a set of sealed opaque envelopes."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	"All study personnel and participants were blinded to treatment assignment for the duration of the entire study."
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Sick leave was assessed by physicians, who are blinded for allocation status
Blinding of outcome assessment (detection bias)	Low risk	The MADRS and HAMD-17 are assessed by a doctor, who were blinded for allocation status

Wade 2008 *(Continued)*

Depressive symptoms		
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Lost to follow up is considered to be high (23%). Risk of attrition bias was therefore deemed high and no appropriate method has been used to account for this missing data: "The primary endpoint was the adjusted mean change in MADRS total score from baseline to week 24, based on the intention-to-treat set (ITT), comprising all patients who took at least one valid post-baseline MADRS assessment, and using last-observation-carried-forward (LOCF) analysis."
Incomplete outcome data (attrition bias) Sick Leave	High risk	Lost to follow up is considered to be high (24.4%) Risk of attrition bias was therefore deemed high and no appropriate method has been used to account for this missing data: "In cases of premature study withdrawal, patients were assigned zero sick leave for missing assessments."
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Wang 2007

Study characteristics	
Methods	**Study design**: RCT. **Recruitment:** occurred between January 2004 and February 2005 using a 2-phase procedure. **Follow up:** 12 months. **Lost to follow up**: 12.3%
Participants	**Inclusion criteri**a: Respondents with at least moderate depression (phase 1: K-6 ≥ 9; Phase 2: QIDS-SR ≥ 8);18 years and older **Exclusion criteria**: employees with lifetime bipolar disorder, substance disorder, recent mental health specialty care or suicidally **Baseline Characteristics** 604 were randomised (T1:304; T2:300); Age: T1: 40.7 (SD 10.5); T2: 42.4 (SD 10.8) Female: T1: 70.7 %; T2: 77.0%% College graduates: T1: 38.0%; T2: 43.8% (24.6%) **Setting:** primary care
Interventions	T1: The structured telephone intervention: telephone outreach and care management program. Systematically assessed needs for treatment, facilitated entry into in-person treatment (both psychotherapy and antidepressant medication), monitored and supported treatment adherence, and (for those declining in-person treatment) provided a structured psychotherapy intervention by telephone. Intervention participants declining in-person treatment and experiencing significant depressive symptoms after 2 months were offered a structured 8-session cognitive behavioural psychotherapy program T2: Usual care. Patients were advised to consult a clinician and could receive any normally available insurance benefit or service (eg, psychotherapy or pharmacotherapy), just not the additional telephone care management components provided to those in the intervention group

Wang 2007 *(Continued)*

Outcomes	**Sickness absence** 1) actual weekly hours worked among the employed, assessed by Health and Productivity Questionnaire (HPQ), a validated self-report instrument **Depressive symptoms:** 1) depression severity, assessed by QIDS-SR **Work functioning:** 1) on-the-job performance, assessed by HPQ
Notes	**Country**: US

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Randomization was carried out by the survey research firm conducting eligibility assessments with a computerized procedure that classified respondents for eligibility and used a random number generator to assign participants to intervention or usual care."
Allocation concealment (selection bias)	Low risk	"Patient treatment allocation was concealed."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Participants were not blinded but unlikely to have changed behaviour. Quote: "Participants were advised not to offer information to their interviewers regarding their intervention status." "Respondents were told they might be invited to participate in an innovative treatment program."
Blinding of outcome assessment (detection bias) Sick Leave	High risk	HPQ is a self-report instrument. As patients were aware of their allocation status, risk of bias high
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	QID-SR is a self-report instrument. As patients were aware of their allocation status, risk of bias high
Incomplete outcome data (attrition bias) Depressive symptoms	Low risk	Lost to follow up: T1: 14.5%; T2: 10% but appropriate method has been used to account for missing data: "Multiple imputation was used to adjust for some participants not completing either 6-months (35 intervention and 22 usual care) or 12 month (44 intervention and 30 usual care) interviews." "Intervention effects on depression severity were estimated using multiple imputation linear regression with simulated standard errors."
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Lost to follow up: T1: 14.5%; T2: 10% but appropriate method has been used to account for missing data: "Multiple imputation was used to adjust for some participants not completing either 6-months (35 intervention and 22 usual care) or 12 month (44 intervention and 30 usual care) interviews." "Comparable multiple imputation regression analyses were used to estimate intervention effects on work outcomes.''
Selective reporting (reporting bias)	Unclear risk	No design paper or trial registration could be identified to assess this risk
Other bias	Low risk	None identified

Wikberg 2017

Study characteristics	
Methods	**Study design:** Cluster randomised controlled trial **Study grouping:** Parallel group **Number of trial arms:** Two **Recruitment:** All patients who fulfilled the diagnostic criteria for depressive disorder, i.e. mild to moderate (BDI score≤36), were asked if they would like to participate in the study. **Clustering:** There were 91 GPs recruited that in turn recruited 258 patients. The average cluster size was 2.8. The design effect was thus: 1.09. Outcomes adjusted for the cluster effect **Follow-up:** 3, 6 and 12 months **Subgroup analysis for working participants only:** Authors provided data on a subgroup analysis of working participants only
Participants	**Baseline characteristics** Improved care • *Age*: 125 • *Gender*: 25% men • *Marital status*: 49% single • *Occupation*: Working/studying: 81.1% • *Sick leave status*: • *Number of participants randomised*: 42.2 Care as Usual • *Age*: 133 • *Gender*: 34% men • *Marital status*: 43% single • *Occupation*: Working/studying: 80.5% • *Sick leave status*: • *Number of participants randomised*: 44.8 **Inclusion criteria:** Inclusion criteria were: written informed consent, age≥18 years, diagnosed with mild to moderate depressive disorder and either not prescribed antidepressant medication or had no changes in antidepressant medication during the preceding 2 months **Exclusion criteria:** Exclusion criteria were: lack of written informed con-sent, antidepressant medication introduced or changed during the 2 months prior to baseline, diagnosed with severe depressive disorder (BDI-II >36, confirmed by diagnostic procedure by GP), diagnosed with severe mental disorder (i.e., bipolar disorder, antisocial personality disorder, psychosis, substance use disorder, or other serious mental disorder), suicidal ideation or earlier suicide attempts, did not speak or understand Swedish,and/or had cognitive disabilities that made it difficult or impossible to complete the assessment instruments,including MADRS-S **Pretreatment:** "No significant differences were found between the participants in the intervention and TAU groups at baseline." **Setting:** The PRI-SMA study was a multicentre, controlled trial that took place at primary health care centres (PHCCs) and was randomised at the GP level.
Interventions	**Intervention characteristics**

Wikberg 2017 *(Continued)*

	Improved Care • *Content*: The intervention consisted of using a patient depression self-rating scale (MADRS-S) in recurrent monthly consultations during the 3-month intervention. Patients made 4 visits to their GPs, at which time they completed MADRS-S to monitor changes in their depressive symptoms that were then discussed in the person-centred consultation. MADRS-S was used as a supplement to, rather than as a substitute for, TAU. • *Duration, frequency, length*: 3 months, 4 visits to GP • *Communication means*: Face to face.The GPs randomised to the group that provided the intervention received four hours of guidance about how to include the results of MADRS-S in the person centred consultation. The intervention GPs also received a video CD with the same pre-recorded information. The person-centred consultations involved patients and GPs collaborating to increase patients' ability to manage their depression. The guidance therefore included a reminder to the GPs that MADRS-S was used for the sake of the patients rather than the GPs. • *Providers*: GPs at primary health care centres (PHCCs) No intervention or Care as Usual • *Content*: The GPs randomised to the group that would provide TAU were instructed to manage patients with depression the same way they usually did (but with the addition of the diagnostic procedure in the initial consultation).In general Swedish GPs are very knowledgeable about and use of person-centred consultation methods in their daily practice of the kind described in Maguire2002. • *Duration, frequency, length*: 3 months • *Communication means*: Face to face. Three months after baseline, patients in the control group were followed up in an appointment with a nurse at the PHCC. • *Providers*: GPs at primary health care centres (PHCCs)
Outcomes	**Sickness absence** *Number of days on sick leave* • **Outcome type**: Continuous Outcome **Depressive symptoms** *Beck Depression Inventory II* • **Outcome type**: ContinuousOutcome
Notes	**Country**: Sweden The authors provided extra data on sick leave and BDI, also reported in Petersson et al. Work 2018;60(1):63-73 Number of patients, mean number of days on sick leave from baseline to 12 months (data from EPR and patients): Intervention: n = 40, m = 119.2 SD 98.0 Control n = 52, m = 107.4 SD 88.5

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Judgement comment: Randomisation took place at the GP level.'All GPs took part in an information meeting about the study. All GPs also met with the study leaders when the leaders visited each participating PHCC at the time of the intervention start-up at that PHCC. Before the intervention started, the GPs at each PHCCwere randomised to either intervention treatment or TAU. All GP names were written on slips of paper and mixed in a container, and an administrative employee blinded to the aim of the trial drew names. The GP whose

Wikberg 2017 *(Continued)*

		name was first drawn was assigned to the intervention group, the GP whose name was drawn second was assigned to the control group, and so on until all names were drawn.''We randomised at the GP level. Randomisation at the patient level would have necessitated changing doctor for some patients, or GPs trained in the intervention would have provided the intervention to some patients but treatment as usual to others, increasing the possibility of contamination.'
Allocation concealment (selection bias)	Unclear risk	Judgement comment: "The GP whose name was first drawn was assigned to the intervention group, the GP whose name was drawn second was assigned to the control group, and so on until all names were drawn." This leaves unclear if the assignment was irrevocable.
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	Comment: Unblinded but unlikely that knowledge of intervention changed behaviour
Blinding of outcome assessment (detection bias) Sick Leave	High risk	Quote:A study nurse collected data from participants in the intervention and control groups during the first visit (baseline), at a follow-up visit to the PHCC at the end of the intervention (3 months after baseline), and by postal questionnaires 6 and 12 months after baseline. Comment: unblinded and self-report
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	See sick leave
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	There was a statistically significant difference between participants and drop-outs during the study concerning age (mean age 44.3 in participants, mean age 37.3 in drop-outs, P = 0.02), gender (male 14/62, 22.6%, female 16/166, 9.6%, P = 0.034), and ethnicity (born in Sweden 21/194, 10.8%, and born outside Sweden 9/32, 28.1%, P = 0.035)
Incomplete outcome data (attrition bias) Sick Leave	High risk	There was a statistically significant difference between participants and drop-outs during the study concerning age (mean age 44.3 in participants, mean age 37.3 in drop-outs, P = 0.02), gender (male 14/62, 22.6%, female 16/166, 9.6%, p = 0.034), and ethnicity (born in Sweden 21/194, 10.8%, and born outside Sweden 9/32, 28.1%, p = 0.035)
Selective reporting (reporting bias)	Unclear risk	Quote: "Trial registration: ClinicalTrials.gov Identifier: NCT01402206. Registered June 27 2011(retrospectively registered)." Judgement comment: There are considerable differences between protocol and report. Non-randomised changed to randomised. The protocol is also retrospectively registered. All protocol outcomes have been reported.
Other bias	Low risk	None detected

Wormgoor 2020

Study characteristics	
Methods	**Study design:** Randomised controlled trial **Study grouping:** Parallel group

Wormgoor 2020 *(Continued)*

Follow-up: 3, 12, 24 months

Number of trial arms: 2

Recruitment: Participants were invited to participate if 'mental complaints' was the main reason for referral to the outpatient clinic. All referrals were assessed by the clinic's psychologist coordinator.

Participants	**Baseline Characteristics** Psychological, brief coping focussed therapy • *Age*: 40.3(10.9) • *Gender; women, %*: 68% • *Marital status*: not reported • *Occupation*: not reported • *Sick leave status; % any type of sick leave*: 74.5 • *Number of participants randomised*: 143 (2 withdrawn) • *Numbers randomized with depression only*: 126 Other psychological, short-term psychotherapy • *Age*: 42.9(10.4) • *Gender; women, %*: 64% • *Marital status*: N.R. • *Occupation*: N.R. • *Sick leave status; % any type of sick leave*: 80.4 • *Number of participants randomised*: 144 (1 withdrawn) • *Numbers randomized with depression only*: 118 Overall • *Age*: • *Gender; women, %*: 66% • *Marital status*: • *Occupation*: • *Sick leave status; % any type of sick leave*: • *Number of participants randomised*: 287 • *Numbers randomized with depression only*: 244 **Included criteria:** Mental complaints was the main reason for referral.Inclusion criteria: - patients had to be employed and on or at risk of sick leave. - Sick-leave had to be < 9 months during the preceding 2 years- Age: at least 18 years - Adequate ability to communicate in Norwegian **Excluded criteria:** acute or severe pathology **Pretreatment:** Workers in the Short PST group were slightly older (42.9 vs. 40.3), all other group difference were not statistically significant. **Setting:** Outpatient rehabilitation department
Interventions	**Intervention Characteristics** Psychological, brief coping focussed therapy • *Content*: Focus on normalisation of common health complaints, redirecting patients' concerns and restoring confidence, acceptance and adaptive coping strategies with their health problems and working life. Implicit aim to enhance work participation, but not an explicit objective. Work-site contacts and visits were not incorporated.

Wormgoor 2020 *(Continued)*

- *Duration, frequency, length*: Duration: 5-hour transdiagnostic group-education. Optional: 5-day coping-course and individual coaching sessions. After that: six psychotherapy sessions. first session: 90-min intake session, subsequent sessions: 50-min psychotherapy sessions. Length: median 15 weeks
- *Communication means*: Face-to-face
- *Providers*: Group education and coping course and coaching: various health professionals (interdisciplinary team of psychologists, physicians, physiotherapists and health educators). Psychotherapy: psychotherapists

Other psychological, short-term psychotherapy

- *Content*: Besides coping of mental health and challenges concerning work participation, an emphasis on both an extensive anamnesis and possibility to establish a so-called central theme based on previous or current challenging issues such as trauma, difficult childhood conditions, and personality-related issues. Additional aims of the intervention could include reducing symptoms and problematic behaviour and an improvement of home situation, with deeper focus on cognitive maladaptive coping strategies or dynamic repetitions.
- *Duration, frequency, length*: Duration: 20 sessions first session: 90-min intake session, subsequent sessions: 50-min psychotherapy sessions. Length: Median 27 weeks
- *Communication means*: Face-to-face
- *Providers*: Ten psychotherapists were involved in the study. Client-therapist allocation was done on a random basis according to capacity. Two therapists were full-time employed with the clinic and eight therapists worked part-time during the project period.

Outcomes	**Sickness absence** *At work* • **Outcome type**: Dichotomous Outcome • **Direction**: Higher is better **Depressive symptoms** *Beck Depression Inventory II* • **Outcome type**: Continuous Outcome
Notes	**Country**: Norway

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Judgement comment: " Randomization procedure was carried out at Uni Research, Bergen, Norway. It was concealed and based on computer-generated randomization lists but stratified by gender."
Allocation concealment (selection bias)	Low risk	Judgement comment: "If the participant had completed the baseline questionnaire, the research assistant, not involved in the treatment, called the randomization unit to be informed about allocation."
Blinding of participants and personnel (performance bias) Sick Leave	Low risk	The control condition cannot be considered less desirable: "In this pragmatic RCT the objective was to compare brief psychotherapy with focus on normalization and coping (Brief-PsT) with short-term psychotherapy of standard duration with more extended focus (Short-PsT), as otherwise used in our Mental Health services. ' 'We hypothesized that in the short term, Brief-PsT could facilitate or sustain WP in a superior fashion to Short-PsT in persons who are on, or at risk of sick leave due to mental health problems. Although we expected a substantial long-term rate of clinical recovery and reduction in mental health-related symptoms in both groups, we had no specific hypothesis regarding the

Wormgoor 2020 *(Continued)*

		extent and direction of possible group differences in these clinical measurements.'
Blinding of outcome assessment (detection bias) Sick Leave	Low risk	Outcome based on registry data: "The primary outcome, short-term WP, was based on registry data from the Norwegian Labour and Welfare Administration (NAV)." ''The psychologists treating the participants were not involved in any of the research processes of this present study and were not directly involved in sickness certification of the participants."
Blinding of outcome assessment (detection bias) Depressive symptoms	High risk	Judgement comment: Outcome based on self-report data and patients were not blinded to treatment allocation:
Incomplete outcome data (attrition bias) Depressive symptoms	High risk	Judgement comment: > 20% lost to follow-up (46%)
Incomplete outcome data (attrition bias) Sick Leave	Low risk	Comment: < 10% lost to follow-up during (8%)
Selective reporting (reporting bias)	Unclear risk	Judgement comment: Personal communication with author: No protocol was published/registered.
Other bias	Low risk	Judgement comment: No other sources of bias detected

BDI = Beck Depression Inventory
CAGE = The name of which is an acronym of its four questions, is a widely used method of screening for alcoholism
CAU = Care as usual
CES-D = Center for Epidemiologic Studies Depression scale
CMD = Common mental disorders
CMHN = Community Mental Health Nursing
CIS-R = Clinical Interview Schedule-Revised
CTU = Copenhagen Trial Unit
DSM-IV = Diagnostic and Statistical Manual of Mental Disorders
4DSQ = Four-Dimensional Symptom Questionnaire
EAP = Employee Assistance Programme
ECT = Electroconvulsive therapy
FDA = Food and Drug Administration
GAS = Global Assessment Scale
GCI = Clinical Global Impression Scale
GEE = Generalized Estimating Equation
GP = General practitioner
GHQ-12 = General Health Questionnaire
HADS(-D)= Hospital Anxiety en Depression Scale
HAMD-D(17) = Hamilton Rating Scale for Depression
HDRS = Hamilton Rating Scale for Depression
HPQ = Health and Work Performance Questionnaire
HRSD = Hamilton Rating Scale for Depression
ICD-10 = International Statistical Classification of Diseases and Related Health Problems
IDS = Inventory of Depressive Symptomatology
LOCF = Last Observation Carrierd Forward
MADRS = Montgomery-Asberg Depression Scale
MAO = Monoamine oxidase
MAOI = Monoamine oxidase inhibitor
MINI = Mini International Neuropsychiatric Interview
MOS-SF 36 = Medical Outcomes Study 36-Item Short Form Health Survey
MDD = Major depressive disorder
OP = Occupational Physician

OT = Occupational therapy
PHQ = Patient Health Questionnaire
PST = Problem Solving Therapy
QI = Quality improvement
QIDS-SR = Quick Inventory of Depressive Symptomatology-self-report
RCT = Randomised controlled trial
RIMA = Reversible inhibitors of monoamine oxidase A
RTW = Return to work
RTW-E = Exposure based return to work program
SAS = Social Adjustment Scale
SCL = Symptom Checklist Score
SNRI = Selective Serotonin and Noradrenalin Reuptake Inhibitor
SSRI = Delective serotonin reuptake inhibitor
TAU = Treatment as usual
TCA = Tricyclic antidepressant
WLQ = Work Limitations Questionnaire
WHI = Work and Health Initiative

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Aasdahl 2017	No sickness absence outcome
Aasdahl 2018	No specification of depression diagnosis
Aasvik 2017	No sickness absence outcome
Aelfers 2013	Participants are people with a mild to moderate depression
Ahola 2012	Sickness absence was not measured as outcome measure
Alexopoulos 2011	No worker population and sickness absence not measured as outcome measure
Amore 2001	Sickness absence was not measured as outcome measure
Arends 2014	No diagnosis of depression
Bakker 2007	Patients suffered from mental health problems, less than 50% of these are patients with a depressive disorder
Barbui 2009	Sickness absence was not measured as outcome measure
Bech 2000	It is a meta-analysis instead of a RCT
Becker 1998	Participants were people with severe mental illness such as schizophrenia
Bejerholm 2015	No diagnosis of depression
Bejerholm 2017	No sickness absence outcome
Beurden 2013	No depressed subgroup
Blonk 2007	Patients suffered from psychological complaints, including adjustment disorders. Patients with a major depression were excluded from the study
Boyer 1998	Sickness absence was not measured as outcome measure

Study	Reason for exclusion
Brandes 2011	Sickness absence was not measured as outcome measure
Brouwers 2007	It is meta-analysis instead of a RCT
Carlin 2010	Sickness absence was not measured as outcome measure
Castillo-Pérez 2010	Sickness absence was not measured as outcome measure
Dalgaard 2014	No sickness absence outcome
Dalgaard 2017	No diagnosis of depression
Dalgaard 2017a	No diagnosis of depression
Danielsson 2019	No sickness absence outcome
Dean 2014	Protocol
Dean 2017	No sickness absence outcome
deVries 2015	paper on already included trial
Dick 1985	This study took place in an inpatient care setting
Dunlop 2011	Sickness absence was not measured as outcome measure
Ebert 2014	No diagnosis of depression
Ebert 2014a	Protocol
Eisendrath 2014	Protocol
Eklund 2012	No RCT but a matched-control design was used
Endicott 2014	No sickness absence outcome
Erkkilä 2011	Sickness absence was not measured as outcome measure
Evans 2016	Not RCT
Finley 2003	Sickness absence was not measured as outcome measure
Folke 2012	This study is done in a sample of unemployed individuals
Forman 2012	Participants were students
Fournier 2015	No sickness absence outcome
Furukawa 2012	Participants with mild depression were included in this study; people with a major depressive disorder were excluded
Gournay 1995	Participants suffered from a range of non-psychotic symptoms, data for the depressed subgroup only could not be provided
Gunnarson 2018	Participants not workers

Study	Reason for exclusion
Hackett 1987	Inclusion criterion in this study was 'clinical diagnosis of chronic muscle contraction headache'
Han 2015	No sickness absence outcome
Heer 2013	study was prematurely terminated
Hirani 2010	Sickness absence was not measured as outcome measure
Hobart 2019	No sickness absence outcome
Hollon 2016	No sickness absence outcome
Hordern 1964	This study took place in a hospital setting
Jansson 2015	No diagnosis of depression
Johansson 2019	No sickness absence outcome
Kennedy 2016	No sickness absence outcome
Kennedy 2019	No sickness absence outcome
Knekt 2011	It is quasi-experimental study
Knekt 2016	No sickness absence outcome
Kojima 2010	Sickness absence was not measured as outcome measure
Kooistra 2014	Protocol
Kroenke 2001	Sickness absence was not measured as outcome measure
Kuhs 1996	Sickness absence was not measured as outcome measure
Lagerveld 2012	Major depressive disorder was excluded in this study
Lam 2012	Sickness absence was not measured as outcome measure
Lexis 2011	The focus in this study is on relatively mild complaints
Löbner 2018	No sickness absence outcome
Maljanen 2016	paper on already included trial
Martinez 2011	Sickness absence was not measured as outcome measure
Meyer 2009	Sickness absence was not measured as outcome measure
Mino 2006	Prevention study; subjects were not depressed
Morgan 2011	Participants are people with sub-threshold depression
Mundt 2001	Sickness absence was not measured as outcome measure
Oakes 2012	Sickness absence was not measured as outcome measure

Study	Reason for exclusion
Reavley 2018	No diagnosis of depression
Salminen 2008	Sickness absence was not measured as outcome measure
Saloheimo 2016	No sickness absence outcome
Salomonsson 2017	No diagnosis of depression
Sandahl 2011	Sickness absence was not measured as outcome measure
Schmitt 2008	It is not a RCT but a review
Schoenbaum 2002	This study turned out to be a publication on the same study as Schoenbaum 2001 (which was also included)
Shawyer 2016	No sickness absence outcome
Simon 2000	Sickness absence was not measured as outcome measure
Sir 2005	Sickness absence was not measured as outcome measure
Soares 2019	No sickness absence outcome
Stant 2009	Sickness absence was not measured as outcome measure
Twamley 2019	Participants not workers
Warmerdam 2007	No sick leave was reported
Wells 2000	This trial is the basis of the economic evaluation of Schoenbaum 2001
Winter 2015	narrative review with description of a study already excluded in first publication of our review
Wisenthal 2018	Not RCT
Zambori 2002	Design was CCT instead of RCT
Zeeuw 2010	This study focuses on employees with minimal symptoms of depression
Zwerenz 2015	Protocol
Zwerenz 2017	No sickness absence outcome
Zwerenz 2017a	No diagnosis of depression

Characteristics of ongoing studies *[ordered by study ID]*

Deady 2018

Study name	Deady 2018
Methods	RCT

Deady 2018 *(Continued)*

Participants	Employees from a range of industries
Interventions	Smart phone application
Outcomes	Depressive symptoms, work functioning
Starting date	unknown
Contact information	Deady
Notes	

Imamura 2018

Study name	Jprn 2018
Methods	RCT
Participants	Nurses
Interventions	Smart phone application
Outcomes	Depressive symptoms, sickness absence, work characteristics
Starting date	1.8.2018
Contact information	nkawakami@m.u-tokyo.ac.jp
Notes	

Kouvonen 2019

Study name	Kouvonen 2019
Methods	RCT
Participants	Young adults working in the public sector
Interventions	Internet delivered face/to/face CBT
Outcomes	Sickness absence
Starting date	1.1.2019 till 1.8.2024
Contact information	anne.kouvonen@helsinki.fi
Notes	

Poulsen 2017

Study name	IBBIS ((Integrated Mental Health Care and Vocational Rehabilitation to Individuals on Sick Leave Due to Anxiety and Depression)
Methods	RCT with three arms
Participants	Patients with anxiety or depression on sick leave
Interventions	1. IBBIS mental health care and standard vocational rehabilitation 2. Integrated IBBIS mental health care and IBBIS vocational rehabilitation 3. Standard mental health care and standard vocational rehabilitation
Outcomes	Time from baseline to the event return to work within 12 months after baseline. Work is defined as having 4 consecutive weeks of working with a salary and with no concurrent vocational benefits
Starting date	April 2016
Contact information	rie.poulsen@regionh.dk
Notes	Authors promised to conduct subgroup analyses for depressed patients

ADDITIONAL TABLES

Table 1. Work functioning outcome: Risk of bias

Study	Blinding of outcome assessment (detection bias)	Incomplete outcome data: attrition bias
Agosti 1991	Low risk (blinded clinician)	High risk
Burnand 2002	High risk (self-report)	High risk
Finnes 2017	High risk (self-report)	Low risk
Hees 2013	High risk (self-report)	Low risk
Kaldo 2018	High risk (self-report)	High risk
Knekt 2013	High risk (self-report)	Low risk
Lerner 2012	High risk (self-report)	Low risk
Miller 1998	High risk (self-report)	Unclear risk
Sarfati 2016	High risk (self-report)	High risk
Wang 2007	High risk (self-report)	Low risk
Lerner 2020	High risk (self-report)	Low risk

Table 2. Work-directed plus clinical compared to care as usual in depressed people, long-term follow-up

Work-directed plus clinical compared to care as usual in depressed people, long-term follow-up

Patient or population: Depressed persons
Setting: Various: workplaces, outpatient and occupational healthcare
Intervention: Work-directed plus clinical
Comparison: Care as usual

Outcomes	Anticipated absolute effects* (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with work-directed intervention plus clinical intervention			
Days of sickness absence, long-term follow-up	SMD 0.19 lower (0.49 lower to 0.12 higher)	179 (2 RCTs)	⊕⊕⊖⊖ LOW [1,2]	The SMD translates back to -0.3 days per 2 weeks (CI -0.9 to 0.2) and -18.7 days in 12 months (-48.3 to 11.8).
Depressive symptoms, long-term follow-up	SMD 0.63 lower (1.02 lower to 0.24 lower)	117 (1 RCT)	⊕⊕⊖⊖ LOW [3]	
Work functioning, long-term follow-up	SMD 0.25 lower (0.63 lower to 0.14 higher)	117 (1 RCT)	⊕⊕⊖⊖ LOW [3]	

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RCT:** Randomised controlled trial; **SMD:** Standardised mean difference.

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]Both studies at high risk because of unblinded outcome assessment. Rated down one level due to high risk of bias.
[2]Pooled effect size includes small harms and appreciable benefits; sample size small; rated down one level due to imprecision.
[3]One study only, with small number of participants; downgraded two levels due to imprecision.

Table 3. Work-directed compared to care as usual in depressed people, long-term follow-up

Work-directed compared to care as usual in depressed people, long-term follow-up

Patient or population: Depressed persons
Setting: Workplace and occupational healthcare
Intervention: Work-directed
Comparison: Care as usual (long-term)

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with care as usual	Risk with work-directed intervention				

Table 3. Work-directed compared to care as usual in depressed people, long-term follow-up *(Continued)*

Off work	606 per 1.000	606 per 1.000 (497 to 739)	RR 1.00 (0.82 to 1.22)	363 (2 RCTs)	⊕⊕⊕⊖ MODERATE [1]
Depressive symptoms	-	SMD 0.18 higher (0.13 lower to 0.49 higher)	-	160 (1 RCT)	⊕⊕⊖⊖ LOW [2]

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RCT:** Randomised controlled trial; **RR:** Risk ratio; **SMD:** Standardised mean difference.

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]CI includes appreciable harm and benefit. Rated down one level due to imprecision.
[2]CI include appreciable harm and benefit; one study only; rated down two levels due to imprecision.

Table 4. Psychological intervention compared to care as usual in depressed people, short-term follow-up

Psychological intervention compared to care as usual in depressed people (short-term follow-up)

Patient or population: Depressed persons
Setting: Various: workplaces, primary care, insurance institute and academic hospital
Intervention: Psychological intervention
Comparison: Care as usual (short-term)

Outcomes	Anticipated absolute effects* (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with psychological intervention			
Days of sickness absence; follow-up short term	SMD 0.05 lower (0.28 lower to 0.17 higher)	300 (1 RCT)	⊕⊕⊖⊖ LOW [1 2]	The SMD translates back to -0.1 days per 2 weeks (CI -0.5 to 0.3) or -4.9 days in 12 months (-27.6 to 16.8).
Depressive symptoms	No data available			
Work functioning	No data available			

*__The risk in the intervention group__ (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RCT:** Randomised controlled trial; **SMD:** Standardised mean difference.

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect

Table 4. Psychological intervention compared to care as usual in depressed people, short-term follow-up *(Continued)*

Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]One study with high risk of bias; rated down one level.
[2]One study only with 300 participants; rated down one level.

Table 5. Improved care compared to care as usual in depressed people, long-term follow-up

Improved care compared to care as usual in depressed people

Patient or population: Depressed persons
Setting: Primary Care and community mental health
Intervention: Improved care
Comparison: Care as usual

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with care as usual	Risk with improved care				
Off work, long-term follow-up	607 per 1.000	656 per 1.000 (601 to 717)	RR 1.08 (0.99 to 1.18)	1356 (1 RCT)	⊕⊕⊕⊖ MODERATE [1]	
Depressed yes/no, long-term follow-up	614 per 1.000	546 per 1.000 (497 to 602)	RR 0.89 (0.81 to 0.98)	1356 (1 RCT)	⊕⊕⊕⊖ MODERATE [1]	
Work functioning	No data available					

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RCT:** Randomised controlled trial; **RR:** Risk ratio.

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

[1]At risk of bias because of lack of allocation concealment. Rated down one level due to high risk of bias.

WHAT'S NEW

Date	Event	Description
6 November 2020	Amended	Correction in Plain Language Summary

HISTORY

Protocol first published: Issue 4, 2006
Review first published: Issue 2, 2008

Date	Event	Description
15 October 2020	Amended	Contact details updated
23 September 2020	New citation required and conclusions have changed	The conclusions have changed, due to the new comparisons that were implemented and 23 new studies.
18 March 2020	New search has been performed	New search. New categorisation of interventions and comparisons. Inclusion of 23 new studies
4 May 2018	New search has been performed	New search conducted on 2 May 2018.
6 June 2014	New citation required and conclusions have changed	Full update. This updated review includes 12 new studies with 3440 new participants (added to the 11 studies with 2556 participants of the former version). We have modified the names of the interventions in the comparisons: we now include work-directed and clinical interventions, while in the 2008 version clinical interventions were under worker-directed interventions. In the update, we refrained from handsearching journals as this strategy did not yield additional studies in the 2008 version. We have re-assessed all studies that we originally included to be able to use the GRADE method. Two new authors have joined the review team: Babs Faber and Hiske Hees.
2 November 2008	Amended	Converted to new review format.
20 November 2007	New citation required and conclusions have changed	Substantive amendment

CONTRIBUTIONS OF AUTHORS

Original review

KN wrote the initial draft of the protocol and will write subsequent drafts of the protocol and review. She and AN designed and conducted the search strategy. AV, UB, CF, AN, and JV contributed to the draft version of the protocol and contributed to subsequent versions and revisions of the protocol and review. KN, AV, and UB included eligible studies. UB and CF conducted the quality assessment of eligible studies. KN and AN extracted the data from the original studies. KN, CF, and JV conducted the data synthesis.

Update 2014

BF adapted the search strategy and conducted the searches. BF, KN, CF, UB, and AV checked resulting studies for eligibility. BF, KN, AN, AV CF, HH, and UB conducted data extraction. BF, KN, AN, AV, CH, HH, UB, and JV assessed included studies for risk of bias. BF, KN, and JV ran the analyses. KN wrote the draft of the updated review and all others commented on this draft. JV acted as an advisor on the whole review process and several specific topics such as meaningful comparisons, GRADE, and meta-analysis.

Update 2020

JV conducted the searches. KN and JV checked resulting studies for eligibility. JV, BF, KN, AN, AV, and UB conducted data extraction and KN and JV assessed included studies for risk of bias. KN, and JV ran the analyses. KN wrote the draft of the updated review and all others commented on this draft.

DECLARATIONS OF INTEREST

Karen Nieuwenhuijsen was an author of one of the included studies: Noordik 2013.

Babs Faber: none known.

Jos Verbeek: none known.

Angela Neumeyer-Gromen: none known.

Hiske Hees (author on the 2014 update) was an author of one of the included studies: Hees 2013.

Arco Verhoeven: none known.

Christina van der Feltz-Cornelis (author on the 2008 and 2014 versions) was an author of one of the included studies: Vlasveld 2013. Her employer received an unrestricted grant from Eli Lilly for an investigator-initiated trial on depression and pain. She also received payment from Benecke for speaking at a symposium on chronic pain. She has received royalties from various publishers on her books on psychiatry.

Ute Bültmann: none known.

None of the authors assessed studies they were authors of for eligibility or risk of bias.

SOURCES OF SUPPORT

Internal sources

- Coronel Institute of Occupational Health, Netherlands

  Salary for Karen Nieuwenhuijsen (on going) and Babs Faber (update 2014)
- Trimbos Instituut - Netherlands Institute of Mental Health and Addiction, Netherlands

  Salary for Christina van der Feltz-Cornelis
- Federal Institute for Occupational Safety and Health, Germany

  Salary for Angela Neumeyer-Gromen
- Finnish Institute of Occupational Health, Finland

  Salary for Jos Verbeek
- University Medical Center Groningen, Netherlands

  Salary for Ute Bültmann
- Dutch Research Center for Insurance Medicine, Netherlands

  Support and training for authors
- Radboud University Medical Centre Nijmegen, Netherlands

  Salary of Arco Verhoeven

External sources

- KIS programme, Ministry of Social Affairs and Employment, Netherlands

  A small grant to Karen Nieuwenhuijsen to help her finish the first version of this review
- Cochrane Review Support Programme , UK

  £5,000 payment upon publication of the update (2020) review

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

In order to reflect the latest guidance available in the *Cochrane Handbook for Systematic Reviews of Interventions*, we used the GRADE approach. In the first version of the protocol and the published review, we used the Downs and Black checklist to assess quality, while in this update we used Cochrane's 'Risk of bias' tool. Also, we no longer formally tested heterogeneity but rather assessed the I^2 statistic. Furthermore, our search strategy was simplified and we no longer handsearched journals as these were indexed in MEDLINE and did not yield additional studies. Instead of searching the CCDAN registers, we now directly searched CENTRAL.

In the 2020 update, we have re-organised the comparisons. One change was that we now distinguish between care as usual (a study arm where patients are treated without a specific intervention protocol) and an alternative intervention (an intervention that was protocolised, regardless of whether that intervention constitutes the regular care in that setting). We also made a small change in the classification of the interventions where we no longer divided the work-directed interventions into subgroups. We divided the psychological interventions into a subgroup with and a group without guidance or face to face contact with a therapist.

In the 2020 update we renamed the outcome 'employment status' as being off work and treated this as a second operationalisation of sickness absence (next to days of sickness absence).

In the previous version, we had not specified the assessment of clinical heterogeneity. We added this now.

INDEX TERMS

Medical Subject Headings (MeSH)

*Absenteeism; Antidepressive Agents [therapeutic use]; Bias; Cognitive Behavioral Therapy; Depression [*therapy]; Depressive Disorder, Major [*therapy]; Muscle Stretching Exercises; *Occupational Health; Randomized Controlled Trials as Topic; Return to Work [*psychology]; Sick Leave; Work Performance

MeSH check words

Adult; Humans