February 2011

Systematic review of intervention practices for depression in the workplace

Institute for Work & Health
Research Excellence Advancing Employee Health

If you have questions about this report, please contact us at:
Institute for Work & Health
481 University Avenue, Suite 800
Toronto, Ontario M5G 2E9
info@iwh.on.ca
www.iwh.on.ca

Institute for Work & Health

Systematic review of intervention practices for depression in the workplace

Andrea D. Furlan MD, PhD[1,2,3], William H. Gnam MD, PhD[1,4], Nancy Carnide M.Sc[1,5], Emma Irvin B.A[1], Benjamin C. Amick III PhD[1, 6, 7], Kelly DeRango PhD, Robert McMaster MD[2], Kimberley Cullen M.Sc[1, 8], Tesha Slack M.Sc[1], Sandra Brouwer PhD[9], Ute Bültmann PhD[9]

[1] Institute for Work & Health, Toronto, Canada
[2] Department of Medicine, University of Toronto, Toronto, Canada
[3] Toronto Rehabilitation Institute, Toronto, Canada
[4] Centre for Addiction & Mental Health, Toronto, Canada
[5] Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
[6] University of Texas, School of Public Health, Division of Environmental & Occupational Health, Houston, TX, USA
[7] Southwest Center for Occupational & Environmental Health, Houston, TX, USA
[8] Biophysics Interdisciplinary Group, University of Guelph, Guelph, Canada
[9] University Medical Center Groningen, Department of Health Sciences, Work & Health, University of Groningen, Groningen, The Netherlands

Acknowledgements
Funding: Canadian Institutes of Health Research.
Nancy Carnide is supported by a Canadian Institutes of Health Research Vanier Canada Graduate Scholarship.
Kimberley Cullen is supported by a National Sciences and Engineering Research Council of Canada Graduate Scholarship.
The team is grateful to Quenby Mahood and Joanna Liu for their assistance with literature searches and article retrieval.
We would also like to thank all of the stakeholders who participated in our two meetings.

Please cite this report as:
Furlan AD, Gnam WH, Carnide N, Irvin E, Amick III BC, DeRango K, McMaster R, Cullen K, Slack T, Brouwer S, Bültmann B. Systematic review of intervention practices for depression in the workplace. Toronto: Institute for Work & Health; 2011.

Table of contents

Appendices

List of Tables

List of Figures

1.0 Introduction

Depression is a widespread, disabling psychiatric illness with far-reaching personal and economic consequences (Hirschfeld 2000, Stephens 2001). By the year 2020, depression is expected to impose the second largest of all illness burdens in developed economies (Murray 1997). The symptoms of depression have a substantial impact on an individual's quality of life (QOL). A major contributor to this loss in QOL is the adverse influence of depression on the ability to function in daily capacities, including social and familial capacities and the role of worker. (Adler 2006).

According to the 2002 Canadian Community Health Survey, 3.7% of the employed population in Canada aged 25 to 64 years experienced an episode of depression in the previous year, with a significantly higher prevalence in women than in men (5.1% versus 2.6%) (Statistics Canada Survey 2003). Even higher depression prevalence figures have been reported in American workers (Kessler 2005). Besides its high prevalence, the associated workplace effects of depression are extensive. In a large study of several firms in the United States, employees with depression reported significantly more health-related lost productivity time than those without depression (an average of 5.6 hours per week, versus an expected 1.5 hours per week, respectively), with 81% of the lost productivity costs explained by reduced performance while at work (Stewart 2003) – a phenomenon sometimes referred to as "presenteeism" but which we will refer to as "work functioning". These findings have been corroborated by other research tracking the performance of depressed workers at work (Adler 2006, Wang 2004). Furthermore, compared to workers with most other health conditions, those with depression have higher rates of absenteeism and short-term disability spells (Kessler 1999), as well as higher rates of job turnover (Lerner 2004).

Research conducted over the past decade has also revealed important economic consequences of workplace depression (Goetzel 2002, Lerner 2008). For instance, workers with depression utilize significantly more general health services than other workers without depression (Simon 1995). However, the most significant economic impact of depression relates to lost productivity due to absenteeism and reduced productivity while at work. Economic analyses have consistently demonstrated that the costs of lost productivity associated with depression far exceed the costs of resources utilized to treat and manage the disorder (Greenberg 1993). Furthermore, depression has been shown to be one of the most costly of common health conditions that affect the ability to work and an individual's productivity while at work (Adler 2006, Kessler 2006, Sanderson 2007, Burton 2004).

Not surprisingly, employers have increasingly expressed concern over the considerable burden of depression imposed on their employees and the impact on the workplace (Goetzel 2002). Canadian employers, in particular, have identified depression and other mental disorders as one of the principal causes of workplace absenteeism (Watson Wyatt Worldwide 2005).

If depression in many (if not most) cases does not arise primarily from work or workplace exposures, why do employers have incentives to support or implement policies, programs, or interventions targeted at depression in the workplace? The major reasons appear to be that the clinical management of depression often does not consider RTW or reducing work disability, and the management of depression generally is far from optimal. Narrative reviews of depression disability management studies have found a paucity of evidence pertaining to RTW and disability in comparison to clinical outcomes (Goldner 2004). Although depression treatment rates have increased in the past twenty years (Patten 2002), many workers with depression never receive standard clinical management of their episode (Stewart 2003). There is abundant evidence that current management practices for depression are not optimal, that depression remains under-detected, and that many workers with depression do not receive evidence-based interventions or treatment (Patten 2002, Kessler 2003).

Consequently, many employers offer general mental health benefits through Employee Assistance Programs, health promotion or wellness programs in an attempt to bridge the management gap. Yet, employer-sponsored programs that specifically target depression remain uncommon. Despite the clear necessity for these programs, there are barriers and information gaps that may prevent employers from making further investments to reduce the impact of depression in the workplace. The most significant information gap may be the paucity of readily accessible information on targeted interventions that improve workplace outcomes most directly relevant to employers, such as absenteeism and productivity. While the published literature on the clinical treatment and management of depression is voluminous, much of the evidence does not indicate whether the interventions studied could be feasibly implemented, supported, or facilitated by employers. Moreover, the evidence on the impact of interventions on relevant work-related outcomes appears to be scattered (Lerner 2008). As a result, we undertook this review to determine the range of possible evidence-based interventions or programs that could be implemented in workplaces to improve workers' depression and reduce associated productivity losses. We felt this would be beneficial for a range of stakeholders particularly the employer community.

The systematic review reported here was conducted to address these information gaps by answering the following research question: "Which intervention approaches to manage depression in the workplace have been successful and yielded value for employers in developed economies?" Systematic reviews identify, appraise, and summarize the scientific literature – and offer advantages over other forms of review due to their replicable, transparent, and scientific methods, which are designed to reduce bias. Systematic reviews also strive to be relevant and accessible to stakeholders, by involving them throughout the review process and consolidating a vast amount of information into a format more readily accessible to stakeholders. There are a few published reviews on this topic (Nieuwenhuijsen 2009, Pomaki 2010). However, a comparison of the other two systematic reviews may be found in Section 4.3.

2.0 Methods

This systematic review used a reviewing process that was developed by the Cochrane Collaboration (www.cochrane.org/training/cochrane-handbook), and adapted by the Institute for Work & Health (IWH) systematic review program.

A review team comprised of 11 researchers from Canada, United States, and Europe participated. Reviewers were identified based on their expertise in conducting epidemiologic or intervention studies related to worker productivity and sustainable return to work for individuals with mental health disorders, their experiences in conducting systematic reviews, or their clinical expertise. The backgrounds of review team members included psychiatry, epidemiology, ergonomics, kinesiology, occupational medicine, labour economics, knowledge transfer and exchange, and information science.

The basic steps of the systematic review process are listed below. The review team used a consensus process throughout the review:

1. Formulate the research question and search terms.
2. Convene a stakeholder meeting to review the research question, definitions, search terms, and relevancy criteria.
3. Conduct the literature search and pool articles with those submitted by experts, ensuring a majority of review team members' key articles have been captured by the search.
4. Conduct the review to exclude non-relevant studies.
5. Conduct the review to assess methodological quality of relevant articles.
6. Conduct the review to extract data from relevant articles that were identified for evidence synthesis.
7. Complete the evidence synthesis.
8. Convene a stakeholder meeting to review evidence synthesis and develop key messages.

The research question addressed was: "***which intervention approaches to manage depression in the workplace have been successful and yielded value for employers in developed economies?***"

To answer this question, we searched and appraised the intervention literature aimed at: 1) assisting workers with depression to stay at work and be productive members of the workforce, and 2) assisting workers with depression who are currently absent from work due to their illness to return to work (RTW).

2.1 Definition of Terms

In order to perform a well-defined literature search, we established definitions for the terms “workplace or work setting” and “depression”.

Workplace or work setting was defined as any location where a worker performs his or her assigned work.

Depression was defined as “current or remitted depression” and could be determined within the study using any one of the following methods: a screening interview or instrument (e.g., World Mental Health Composite International Diagnostic Interview), a clinician-derived diagnosis (as stated by the authors), a diagnosis established using formal standardized diagnostic criteria (i.e., fulfilling criteria from the Diagnostic and Statistical Manual of Mental Disorders, or other similar classification), or validated self-report instruments (e.g., Center for Epidemiologic Studies Depression Scale).

2.2 Inclusion and Exclusion Criteria

The P.I.C.O. (***P***opulation, ***I***ntervention, ***C***omparison, ***O***utcome) framework was used to guide the strategy for study identification.

(P) Population: Men and/or women of working age (i.e., approximately 18-65 years old) with a diagnosis of depression, as defined previously. In some instances, the study population included individuals with other mental health disorders. In this case, we only included studies in which at least 50% of the population had depression.

We excluded studies that reported on patients with either a serious mental disorder (i.e., bipolar disorder or schizophrenia) or chronic severe depression. Here we defined chronic severe depression as that which involved onset of depressive symptoms in adolescence or early adulthood that precluded patients from any meaningful labour market participation.

We also excluded studies where the primary focus was on persons with alcohol or other substance abuse or dependence disorders, depression related to pregnancy, and depression in the following populations: military personnel and veterans, seniors, the elderly population, and children. Studies focused on bereavement, burnout, and anxiety were also excluded.

(I) Intervention: We included studies evaluating *interventions or programs* that were workplace-based or that could be explicitly implemented and/or facilitated by the workplace. Such interventions or programs for workers

with depression could involve the prevention of disability, the management of depression, or the rehabilitation of workers to promote stay at work (SAW) or return to work (RTW). Examples of such interventions include:

- *Prevention of disability*: health risk management, mental health promotion, resiliency training, time/stress management, supportive human resource (HR) policies (conflict resolution, work-life balance, recognition/reward), work re-organization, supportive leadership and management/supervision, education and training, healthy workplace strategies.
- *Management of depression*: performance management, medical surveillance, employee assistance program (EAP), depression screening, assessment and referral, self-care programs, acute and chronic stress management, early RTW program (case management, practice guidelines, work accommodations such as modified work), enhanced access to mental health providers (MHPs), preferred provider networks, shared-care or independent medical evaluations (IMEs), employee satisfaction surveys.
- *Rehabilitation*: case management, practice guidelines, mental job analysis, functional capacity assessments, IMEs, task/job modification, vocational rehabilitation, preferred provider network or shared-care to increase access to MHPs, relapse prevention, and long-term disability (LTD) depression screening.

We excluded in-patient intervention programs i.e., any health or psychosocial intervention that occurred when a client was admitted to a hospital or psychiatric facility. We also excluded studies focusing entirely on drug efficacy in depression.

(C) Comparison/Control: We included any study with a comparator. This included randomized controlled trials, as well as non-randomized studies with before-and-after comparisons within the same group or comparisons between distinct non-randomized groups. We excluded studies that did not have any sort of comparison or control because in workplace studies there are usually many co-occurring interventions that could influence the outcome.

(O) Outcomes: We included studies that examined the impact of interventions on **primary outcomes** relevant to employers, such as: a) changes in productivity, b) changes in sickness absence, absenteeism, worker turnover, and long-term disability, c) changes in on-the-job performance and health-related performance, d) changes in rates of job-related accidents, and e) economic outcomes. These primary outcomes were essential to the study's inclusion into the review. **Secondary outcomes** included changes in clinical measures of depression, general

well-being, patient satisfaction, and quality of life. These outcomes were considered important, but not essential to a study's inclusion into the review. Studies reporting secondary outcomes only were excluded.

Additional criteria: The review team considered published or in-press peer-reviewed scientific articles. There were no language restrictions. Book chapters, dissertations, and conference proceedings were excluded.

2.3 Stakeholder engagement

Stakeholders from Ontario's health and safety system were invited to several meetings with the research team to provide direction and feedback on the review. The first meeting was to solicit input related to the specifics of the research question, the literature search terms, and the manner in which findings from this review would be best presented. During this feedback session, the stakeholders also received a presentation on the systematic review process. Fifteen stakeholders representing the Ontario Ministry of Health and Long-Term Care (MOHLTC), the Ontario Ministry of Government Services (MGS), insurance providers, disability management service providers, mental health organizations, mental health disorder survivors, organized labour, and employers attended this two-hour meeting. An interim meeting was held with representatives from MOHLTC only to provide an update on the progress of the review and to determine if there were any particular elements that they would like to have extracted from articles that had passed the quality appraisal stage. The final meeting was held to update the wider group of stakeholders on the preliminary findings of the review, to determine the types of messages that would emerge from the review, and to determine appropriate communication channels.

2.4 Literature Search

The literature search was based on the research question, and our definitions of work setting (or workplace) and depression. Key terms were identified and combined to search the following databases from their inception dates: MEDLINE, EMBASE, CINAHL, Central, PsycINFO and Business Source Premier (BSP).

Search terms were identified for four broad areas: work setting terms, depression terms, intervention terms, and work outcome terms. Both database-specific controlled vocabulary terms and keywords were included. As the controlled vocabulary and the ability to handle complicated multi-term searches differs across the databases searched, search terms were customized for each database, as required. The complete list of terms used in our search is reported in Table 1 (*the complete search strategies used for MEDLINE, EMBASE, and PsycINFO are included in Appendix A*).

The search categories were chosen to be exclusive within each area. The terms within each category (work setting, depression, intervention, and work outcome) were combined using a Boolean OR operator. The four main categories were then combined using a Boolean AND operator. A simplified example of this search would be: worker AND depression AND workplace intervention AND return to work. This would identify an article that describes a workplace intervention for depression among workers and evaluates RTW as an outcome.

Table 1: Search terms*

Search Term Area	List of Terms
Work Setting	Apprentice, boss, branch, company, contractor, department, employee, employer, employment, facilities, factory, firm, health services, hospital, industry, institution, isolation pay, labourer, leader, manager, office, operator, organizational, personnel, plant, retail, skilled trade, staff, supervisor, team, telecommunications, unionized, work, work environment, work site, worker, working at home, workplace
Depression	Affective disorder, affective symptoms, depression, depressive disorder, depressive symptoms, dysthymia, mood disorder, mood symptoms, seasonal affective disorder
Intervention	Access to care, accommodation, acute stress management, adjustment, advocate, affinity groups, alternate duties, assessment and referral, benefits, case management, chronic stress management, club membership, coaching, community services, contracted ombudsman services, counselling, cultural resources, depression screening, disability management program, diversity resources, employee assistance program (EAP), early intervention, education, education and training, e-learning, embrace diversity, employee satisfaction surveys, employer resource groups, engagement, enhanced access, fitness group, flexible work, functional capacity assessments, functionality, gardening, grassroots, gym membership, health and wellness, health risk management, healthy workplace strategies, horticulture, independent medical evaluations (IMEs), inviting an organization in, job control, job modification, joint labour management initiatives, long-term disability (LTD) benefits, management of individual, medical surveillance, mental health promotion, mental job analysis, mentoring, modified duties, modified work, modified work, nature, occupational health services, organizational culture, organizational policies and practices (OPPs), pastoral care, peer support, performance management, pet therapy, positive psychology, practice guidelines, prayer room, preferred provider networks, prevention, prevention for all, promoting recovery, psychological

	safety, psychosocial risk factors, organizational culture, quiet room, quiet space, reflection room, rehabilitation, reintegration, relapse prevention, resiliency training, return to work, reward, second opinion, self help, self-care programs, shared-care, shared-care, short-term disability (STD) benefits, spiritual care, spirituality, stay at work, stress management, support groups, support options (support, in general) in small business, supportive leadership, supportive management, supportive supervision, task modification, time management, training, transitional/graduated return to work, treatment support, universal access, vocational rehabilitation, wellness strategy, work environment intervention, work re-organization, workplace adjustment, workplace intervention
Work Outcome	Absenteeism, accommodation, benefit duration, cost-effectiveness, co-worker conflict, cultural shift, disability pension, employee satisfaction, engagement, job match, job turnover, labour force participation, long-term disability, lost time, lost workday, new employer, new job, presenteeism, productivity, productivity ratio, reassignment, recovery, reduced costs, reduction in complaints, reduction in harassment, reemployment, remission, resilience, return on investment, return to work, short-term disability, sick leave, sickness absence, stay at work, stigma, successful stay at work, supportive at-work solutions, talent, time on benefit, unemployment, vocational assessment, wage replacement, wellness strategy, work ability, work absence, work adaption, work adjustment, work capacity, work disability, work functioning, work impairment, work limitations, work loss, work performance, work re-entry, work reintegration, work resumption, work retention, workers compensation, work-life balance

* Terms within each category were combined using a Boolean OR operator and the four categories were combined using a Boolean AND operator

Additional steps were taken to ensure the search for relevant papers was comprehensive. We asked our stakeholder group to notify us of any articles they were aware of that should be considered in our review. Each of the review team members were also asked to examine their personal libraries for relevant articles. Finally, the reference lists of all articles that were identified as relevant to our review were hand-searched for additional potentially relevant articles.

2.5 Selection for Relevance (Level 1 & 2)

The inclusive search strategy captured many articles that were not relevant to our research question. As a result, a two-level relevance assessment was designed to identify and exclude articles that were irrelevant to the review research question as efficiently as possible, based on our inclusion/exclusion criteria. Reviewers entered responses for all levels of

the process on commercial review software, DistillerSR (Evidence Partners, Ottawa, Canada http://systematic-review.net/), allowing centralized article tracking and access.

In Level 1, reviewers read only the article title and abstract (when available) and evaluated the relevance of each article using three questions (*see Table 2*). A response of "No" to any one of the three questions led to the article's exclusion from the review. If reviewers were unsure on how to answer a question, they were instructed to mark it as "Unclear" (*see Appendix B for Level 1 Reviewer Guide*). A "Yes" response to all questions or a combination of "Yes" and "Unclear" responses would move an article forward to Level 2 relevance assessment where the full paper was obtained to definitively determine the article's relevancy.

Table 2: Level 1 relevance screening questions*

Relevance Question	**Response That Led to Exclusion****
Does the article describe:	
Population: 1. People of working age with depression?	No
Intervention: 2. An intervention to prevent further disability, manage depression or the rehabilitation of workers to promote stay at work (SAW), return to work (RTW), or reduction of job-related injuries?	No
Comparison: 3. A study with a comparison group?	No

* A complete description of each question is provided in the Level 1 Reviewer Guide in **Appendix B**

** The given response to any one question excluded the article from further review

In Level 2, full articles were obtained for all studies that passed through Level 1 (either those meeting all Level 1 criteria, or those with insufficient information to determine relevancy at Level 1). In addition to the three questions asked at Level 1, full article relevance was assessed with four additional questions (*see Table 3; see Appendix C for Level 2 Reviewer Guide*).

Table 3: Level 2 relevancy screening questions*

Relevance Question	Response That Led to Exclusion**
Does the article describe:	
Population: 1. People of working age with depression?	No
Intervention: 2. An intervention to prevent further disability, manage depression or the rehabilitation of workers to promote stay at work (SAW), return to work (RTW), or reduction of job-related injuries?	No
Comparison: 3. A study with a comparison group?	No
Outcome: 4. Primary outcome(s) that are relevant to employers?	No
Other: 5. Should this article be included for another purpose? If so, please state why. 6. Is this a review article on depression in the workplace? 7. Are there other studies listed in this reference list that should be retrieved for consideration? (if 'Yes', please include author/year/publication information)	No Yes Non-exclusion question

* A complete description of each question is provided in the Level 2 Reviewer Guide in **Appendix C**

** The given response to any one question excluded the article from further review

Each article's relevance was assessed independently by two members of the review team at each of Levels 1 and 2. Any conflicts were resolved by consensus between review partners. A third reviewer was available to assist in the decision-making process during consensus, if required.

2.6 Quality Assessment (Level 3)

Relevant articles were moved forward for methodological quality assessment at Level 3. The team developed quality assessment questions based on existing forms and pilot tested them using a relevant article. This resulted in one relevancy screening question and 18 methodological criteria questions for assessing quality, which are shown in Table 4 (*see Appendix D for Quality Assessment Reviewer Guide*).

Each article was independently reviewed by two team members. To reduce potential reviewer bias, the same two members did not review all of the same articles. Instead, each reviewer was randomly paired with other team members. Reviewer pairs were required to reach consensus on all criteria. Where reviewer pairs disagreed in their assessment, they were encouraged

to resolve their disagreement through discussion. In cases where agreement could not be reached, a third reviewer was consulted to ensure consensus was obtained. Team members did not review articles they had consulted on, authored, or co-authored.

Table 4: Level 3 quality assessment questions

Screening Question
Should this article be excluded from DE because it does not meet our inclusion criteria for the population, intervention, comparison and outcomes?
Design and Objectives
1. Is the research question clearly stated?
2. Were comparison group(s) used?
3. Was an intervention allocation method performed adequately?
Level of Recruitment
4. Was recruitment (or participation) rate reported and adequate?
5. Did the author(s) examine whether important differences existed between those who participated and those who did not?
6. Were pre-intervention (baseline) characteristics described and appropriately balanced?
7. Was loss to follow up (attrition) less than 35%?
8. Did the author(s) examine whether important differences existed between the remaining and drop-out participants after the intervention?
Intervention Characteristics
9. Was the intervention process adequately described to allow for replication?
10. Was there any potential for contamination and/or co-intervention?
Intervention Intensity
11. Was compliance with the intervention in all groups described and adequate?
Outcomes
12. Were the instruments used to assess the outcomes valid and reliable?
13. Were the outcomes described at baseline and follow-up?
14. Was the length of follow-up three months or greater?
Analysis
15. Was there adjustment for pre-intervention differences (if necessary)?
16. Were the statistical analyses appropriate?
17. Were all participants' outcomes analyzed by the groups to which they were originally allocated (intention-to-treat analysis)?
18. Was there a direct between-group comparison?

2.6.1 Determining Risk of Bias

The presence of five types of bias was assessed for each study: 1) selection bias; 2) attrition bias; 3) performance bias; 4) measurement bias; and 5) reporting bias. Responses from the 18 quality assessment criteria were grouped according to the bias they addressed to form a set of criteria used to judge the risk of each particular bias. Additional items on blinding (of intervention providers, participants, and outcome assessors), though not included in the initial quality assessment, were later examined for each study and incorporated into the other established criteria for the appropriate bias. Each criterion within a bias was then judged as either "critical" or "non-critical" for the risk of bias (*see Table 5*). A "critical criterion" was defined as a major flaw that indicated the study was particularly vulnerable to the specified bias. A "non-critical criterion" was considered to be important and suggestive of bias, but alone, insufficient to judge the study to be at high risk of the specified bias. Using these categorizations, the overall risk of each respective bias was determined as follows:

- *Low risk of [insert type] bias:* all criteria (critical and non-critical) were met
- *Moderate risk of [insert type] bias:* at least one criterion considered to be "non-critical" was unmet, but no criterion considered to be "critical" was unmet
- *High risk of [insert type] bias:* at least one criterion considered to be "critical" was unmet

Table 5: Categorization of quality assessment criteria according to the type of bias and critical/non-critical classification

Bias	Corresponding Quality Appraisal Question	Criteria to Judge Risk of Bias	Critical or Non-critical if Unmet
Selection Bias	Q4: Was recruitment (or participation) rate reported and adequate?	Participation rate >65%	Non-critical
	Q5: Did the author(s) examine whether important differences existed between those who participated and those who did not?	No differences between participants and non-participants	Critical
	Q17: Were all participants' outcomes analyzed by the groups to which they were originally allocated (intention-to-treat	Intention-to-treat analyses completed	Critical

Bias	Corresponding Quality Appraisal Question	Criteria to Judge Risk of Bias	Critical or Non-critical if Unmet
	analysis)? Q3: Was an intervention allocation method performed adequately?	Adequate allocation to treatment groups	Critical
	Q6: Were pre-intervention (baseline) characteristics described and appropriately balanced? Q15: Was there adjustment for pre-intervention differences (if necessary)?	Baseline differences balanced or if unbalanced, accounted for in the analyses	Critical
Attrition Bias	Q7: Was loss to follow up (attrition) less than 35%?	Attrition rate <35%	Critical
	Q8: Did the author(s) examine whether important differences existed between the remaining and drop-out participants after the intervention?	No differences between remaining participants and those lost to follow-up	Critical
Performance Bias	Q9: Was the intervention process adequately described to allow for replication?	Intervention process adequately described	Non-critical
	Q10: Was there any potential for contamination and/or co-intervention?	Minimal opportunity for co-interventions	Critical
	Q10: Was there any potential for contamination and/or co-intervention?	Minimal opportunity for contamination	Critical
	Q11: Was compliance with the intervention in all groups described and adequate?	Adequate compliance with the intervention	Critical
	No corresponding question on the original quality appraisal	Intervention providers blinded	See note*
	No corresponding question on the original quality appraisal	Participants blinded	See note*
Measurement Bias	Q12: Were the instruments used to assess the	Valid and reliable outcome instruments	Critical

Bias	Corresponding Quality Appraisal Question	Criteria to Judge Risk of Bias	Critical or Non-critical if Unmet
	outcomes valid and reliable?		
	Q13: Were the outcomes described at baseline and follow-up?	Outcome described at both baseline and follow-up	Non-critical
	Q14: Was the length of follow-up three months or greater?	Follow-up >3 months duration	Non-critical
	No corresponding question on the original quality appraisal	Outcome assessors blinded	Critical
	No corresponding question on the original quality appraisal	Participants blinded	See note*
Reporting Bias	Q1: Is the research question clearly stated?	Research question clearly stated	Non-critical
	Q18: Was there a direct between-group comparison?	Direct between-group comparison completed	Critical
	Q16: Were the statistical analyses appropriate?	Appropriate statistical analyses	Critical

* Not included in the determination of risk. Given the nature of the studies in this review, blinding of the intervention providers and the participants to the intervention would have been impossible. Therefore, these studies are all at risk for bias due to this lack of blinding. However, it was decided not to penalize the studies because of lack of blinding since this is an unrealistic expectation in workplace studies.

Using the risk of bias judgements made for each individual type of bias, an individual study's overall risk of bias was determined as follows:

- *Low risk of bias overall:* all five individual types of bias considered as low risk
- *Moderate risk of bias overall:* at least one type of bias considered as moderate risk, but no bias considered high risk
- *High risk of bias overall:* at least one type of bias considered as high risk

2.7 Data Extraction (Level 4)

Following quality assessment, data were extracted from each article at Level 4 in order to contribute to the synthesis of evidence used to answer the review's research question.

The team developed a standardized data extraction form based on existing forms and data extraction procedures (Franche 2004, Franche 2005, Brewer 2006, Van Eerd 2006) (*see Appendix E for Data Extraction Reviewer Guide*). Extracted data were used to build summary tables to inform evidence synthesis and to develop our overall conclusions.

Data extraction was performed independently by two reviewers. Again, reviewer pairs were rotated to reduce bias. Team members did not review articles they had consulted on, authored, or co-authored. For the articles reporting the findings of economic evaluations, we allocated the economists on the review team as one of the two reviewers in order to ensure accuracy in extracting these outcomes. Differences in data extracted between reviewers were identified and resolved by discussion. In cases where agreement could not be reached, a third reviewer was consulted to ensure consensus was obtained.

Reviewers extracted data on: year of study; jurisdiction; type of work setting; study design; source population; sample characteristics; how the presence of depression was determined; length of follow-up; intervention characteristics; outcomes of interest to this review (productivity, sickness absence, health-related, and economic measures); statistical analyses; covariates/confounders; and study findings (*see Table 6 for the complete list of data extraction questions*).

Initially, we planned to calculate the effect sizes for each article in order to evaluate the strength of associations in a uniform manner (Cooper 1994, Kristensen 2005, Cole 2005, Tompa 2007). However, this approach was abandoned early in the process due to the amount of heterogeneity in outcome measures and study methods, and the lack of data necessary to calculate effect size in some studies.

Table 6: Level 4 data extraction questions

Screening Question
1. Should this article be excluded from DE because it does not meet our inclusion criteria for the population, intervention, comparison and outcomes?
Study Design and Setting
2. State the research question/objective(s) 3. Write the last name of the first author and the year of publication

4. State the jurisdiction where the study was completed
5. Describe the source population from which the participants were recruited
6. Describe the type of setting/workplace/work setting the study was conducted in
7. List the job titles/classification of the participants that participated in the study
8. Describe how the presence of depression among potential participants was determined
9. Please clearly list the inclusion criteria described in the study for worksite characteristics.
10. Please clearly list the inclusion criteria described in the study for individual characteristics.
11. Please clearly list any other inclusion criteria described in the study.
12. Please clearly list the exclusion criteria described in the study for worksite characteristics.
13. Please clearly list the exclusion criteria described in the study for individual characteristics.
14. Please clearly list any other exclusion criteria described in the study.
15. What is the study design?
16. Was the study protocol reviewed and approved by a Research Ethics Board (REB)?

Intervention Characteristics

17. Describe the nature of the intervention for the intervention(s) group?
18. Describe the nature of the intervention for the comparison group
19. How often was the intervention applied for the intervention(s) group?
20. How often was the intervention applied for the comparison group?
21. What was the duration of the intervention for the intervention(s) group?
22. What was the duration of the intervention for the comparison group?
23. Indicate the time period between the baseline measurement and all subsequent follow-up measurements

Sample Characteristics at Baseline

24. Describe the intervention group at baseline
25. Describe the control group at baseline
26. Describe the overall (study) group at baseline

Covariate Questions

27. When were potential covariates/confounders measured?
28. Were covariates/confounders ultimately controlled for in the final analysis?

Outcomes

29. Provide a list of outcome variables used to evaluate intervention effectiveness, that are relevant to our review project
30. Were direct and indirect costs associated with the intervention measured?
31. Were any outcome measures monetized (converted into a dollar figure)?
32. Was a cost-effectiveness analysis (CEA) conducted?
33. Was a cost-benefit analysis (CBA) used?
34. If the answer to question 35 or 36 was “Yes”, was the CEA or CBA done from the point of view of society or of the employer?
35. What time frame was used for the CEA or CBA?
36. Are the results sensitive to the time frame used?
37. If the answer to question 36 was “Yes”, given the flow of benefits and the costs, did the authors calculate how long it would take to recoup the costs?

38. Was there an inflation adjustment? 39. Did the CEA or CBA perform discounting?
Statistical Analysis and Results
40. Please indicate the types of final analyses done for testing the observed effects of the intervention and provide details for which outcome in the text box 41. Describe for each outcome of interest, the observed intervention effects 42. Were additional statistical analyses conducted to increase your confidence in the observed effects? 43. Remark on the findings or enter information that is unique about the study that may not be adequately captured in the other DE questions.

2.8 Evidence Synthesis (Level 5)

The quality of evidence and strength of recommendations from this review were established in two steps, following published guidelines from the Grading of Recommendations Assessment, Development and Evaluation (short GRADE) Working Group (http://www.gradeworkinggroup.org).

2.8.1 Evaluating the Quality of the Evidence

Based on the various outcomes reported in the studies of the review, a number of sub-questions were formulated in order to answer the review's larger research question. These sub-questions were developed according to a framework of outcomes that was suggested by stakeholders at the final meeting and, as a result, may be found in the Results section (*see Table 9*). The evidence addressing each question was then aggregated and graded on six domains: study design, risk of bias, consistency of the evidence, directness or generalizability of the evidence, data precision, and economic benefit.

Following GRADE guidelines, the final grade for quality of evidence for each sub-question was categorized as follows:

- High (⊙⊙⊙⊙): further research is very unlikely to change our confidence in the estimate of effect
- Moderate (⊙⊙⊙○): further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- Low (⊙⊙○○): further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- Very Low (⊙○○○): any estimate of effect is very uncertain

The evidence available to answer each sub-question began the grading process with a 'High' grade and was then graded on the six domains in the following manner:
***Study Design*:** Studies were classified as either a randomized controlled trial (RCT) or non-randomized study (NRS). The evidence was downgraded (-2 points) if the design was not an RCT.

Risk of Bias: As described above in Section 2.6.1, we examined studies on five types of biases (selection, attrition, performance, measurement, and reporting bias). The overall risk of bias for each study was used in the GRADE synthesis. Studies were downgraded when considered moderately biased (-1 point) or highly biased (-2 points):

- Low risk of bias overall = all types of biases considered low risk
- Moderate risk of bias overall = at least one bias considered moderate risk (and the others were low)
- High risk of bias overall = at least one bias considered high risk (and the others were low or moderate)

***Consistency*:** If there was only one study examining the particular research question, then this was considered not applicable and the grade remained unchanged. When several studies attempting to answer the same research question yielded widely differing estimates of treatment effect (heterogeneity or variability in results), the strength of recommendations derived from even a rigorous randomized controlled trial was considered weaker due to the lack of consistency and the evidence was downgraded (-1 point).

***Directness or generalizability**:* Both the study populations and outcomes were evaluated for directness of evidence. The overarching research question for this review pertained to workers suffering from depression. When the evidence was derived from a similar, but not identical population of those of interest to our review question, the evidence was considered indirect and the strength of evidence was downgraded. For example, when the study sample was recruited at the workplace, the evidence was considered direct, while evidence derived from a study with a population recruited from physicians' offices where patients were presenting for routine care was considered indirect.

When the outcome was a surrogate measure of what was most relevant to our research question, the evidence was considered indirect. For example, a measure of absenteeism is judged indirect when the study assessed "work status at baseline and follow-up", as this outcome does not account for any absences that may have occurred in between follow-ups. On the

other hand, a measure of absenteeism would be judged direct when measured as “length (days) of sick leave in the past year”.

The evidence was downgraded when the population (-1 point) or the outcome (-1 point) were judged as indirect. Therefore, when both were judged as indirect, the evidence was downgraded twice (-2 point).

***Precision or sparse data**:* When only one study provided evidence to answer a particular research sub-question, the evidence was considered imprecise. When multiple studies were available, but the total sample size was small or few events occurred, yielding wide confidence intervals, the data was considered to be imprecise or sparse. In each case, the evidence was downgraded (-1 point).

- When more than one study is included, the evidence should be considered as imprecise or sparse data when the confidence intervals are sufficiently wide that the estimate is consistent with conflicting recommendations.

Economic benefit: Studies were upgraded (+1 point) if they provided evidence of a sizeable economic benefit based on evidence with low or moderate risk of bias.

Summary of steps to determining the final grade:

1. Start with study design:
   - Randomized trial = start with high grade (⊙⊙⊙⊙)
   - Non-randomized study = start with low grade (⊙⊙○○)

2. Decrease grade by one level (exceptions specified) if:
   - Moderate risk of bias (-1) or high risk of bias (-2)
   - Important inconsistency (if NA, no change in grade)
   - Uncertainty about directness: population or outcome (-1), population and outcome (-2)
   - Imprecision (or sparse data)

3. Increase grade by one point if:
   - Evidence of large economic benefit based on evidence with low or moderate risk of bias

2.8.2 Summary of Findings and Development of Key Messages

The summary of findings and key messages from this review were established following published guidelines from the Cochrane Collaboration (Chapter 11 www.cochrane.org/training/cochrane-handbook). The summary of findings provide key information concerning the grade of the evidence

and summarize the available data on all important outcomes for a given comparison. This process, which is recognized as a broader system of evaluating and presenting evidence, increases the usability of the findings for stakeholders and decision-makers (Terracciano 2010, Oxman 2006, Atkins 2005). Briefly, the summary of findings was developed in the following manner: for each intervention assessed, the findings corresponding to each primary outcome category related to work disability/sickness absence (prevention, management, recurrence), work functioning, and economic benefit were classified as positive (i.e., the intervention group was statistically significantly better than the control group at $p<0.05$), as negative (i.e. the control group was statistically significantly better than the intervention at $p<0.05$), or neutral (i.e., intervention was not statistically significantly different from the control group, or $p\geq0.05$).

The summary of findings were then used as the platform for the development of key messages and identification of relevant information for future research. Key messages for each intervention approach were extracted following the framework shown in Table 7. Note, key messages were only extracted for those studies employing an inactive control group (e.g., usual care). Key messages were not drawn from studies using an active control group due to a lack of confidence in being able to attribute the outcome effect to the intervention. Instead, the findings of these studies were used only to inform future research.

Table 7: **Translation from summary of findings to key messages**

GRADE	Consistency	Terminology for Key Messages
High	Intervention is consistently better* than inactive control	Recommendation to implement the intervention
	Intervention is consistently inferior than inactive control**	Recommendation against implementation of the intervention
Moderate or Low	Intervention is consistently better than inactive control	Practice consideration or promising practice[#]
	Intervention is consistently inferior than inactive control	No recommendation. Need for more research
Very low	Intervention is consistently better or inferior than inactive control	No recommendation. Need for more research
Any	Findings are mixed*** or contradictory****	No recommendation. Need for more research

*** Consistently better:** When all the comparisons for primary outcomes demonstrated positive findings (i.e., in favour of the intervention group)
** **Consistently inferior:** When all the comparisons for primary outcomes demonstrated negative findings (i.e., in favour of the control group)
*** **Mixed findings**: When the comparisons for primary outcomes were a mix of positive and neutral (no difference between intervention and control) findings or a mix of negative and neutral findings

****** Contradictory findings**: When the comparisons for primary outcomes were a mix of positive and negative findings
 # Practice consideration or **promising practice** refers to interventions that a group may try in collaboration with an evaluator to further assess the utility of the approach. These practices still require high quality evidence, but the evidence to date suggests there is promise in the effectiveness of the intervention.

2.9 Narrative Review of Depression Self-Report Screening Instruments Best Suited for Workplace Settings

There are numerous screening instruments used to identify depression. During the completion of this systematic review, we identified several self-report measures that were used in each study to identify workers with depression or depressive symptoms. To better inform our review findings, we decided to evaluate these screening instruments in order to assess their measurement properties and to make recommendations for their use in this population.

It is important to note that we did not set out to address this question at the outset of the review. We did not systematically search and appraise the literature for depression screening instruments. Rather, we only examined those that were used within the studies identified as part of this systematic review. Therefore, the findings of this work must be considered “narrative” in scope. Nevertheless, we feel that this work complements the findings from our systematic review and provides direction for choosing screening instruments for the diagnosis of depression within the workplace.

We were interested in evaluating screening instruments that could be used in a workplace setting. Thus, we selected instruments for further evaluation based on the following criteria: the instrument could be used for screening and to rate the severity of depression and depressive symptoms, and had to be a self-report instrument.

For each instrument identified, a literature search was conducted in MEDLINE, CINAHL, and EMBASE to identify studies evaluating the measurement properties of these instruments. Studies were chosen that evaluated the psychometric properties in a primary care setting (unless otherwise stated) in order to achieve a more reasonable comparison (since validation in general working populations or environments are not readily available). When there were many validation studies for a particular instrument, a best estimate was performed by a single reviewer to choose studies that most closely approximated our population of interest.

We developed an evaluation criteria tool to evaluate and grade each instrument (adapted from Andresen 2000). This tool included items that assessed an instrument's reliability, validity, respondent burden (time to complete instrument), cost and availability, and language availability. Each screening instrument was evaluated based on this tool and given grades (each category was assigned a maximum of three stars, and there were five categories, therefore, a potential total score could equal 15) and a narrative outline for their recommended use (*the evaluation tool can be found in Appendix F*).

3.0 Results

3.1 Literature Search and Selection for Relevance

We identified a total of 5,416 articles using the search terms listed in Table 1. After the articles from the various databases were merged and duplicate articles were removed, a total of 4,214 articles remained (*see Figure 1*).

At Level 1 relevancy screening, a total of 3,921 articles were excluded for not meeting the inclusion criteria (*refer to Table 2 for criteria*).

As a result, a total of 293 articles proceeded to Level 2 for relevancy screening. Using the exclusion criteria in Table 3, another 270 articles were excluded. In addition, two articles were not reviewed because we were unable to retrieve them despite several attempts to contact the authors and publishers. Therefore, a total of 272 articles were excluded at this level.

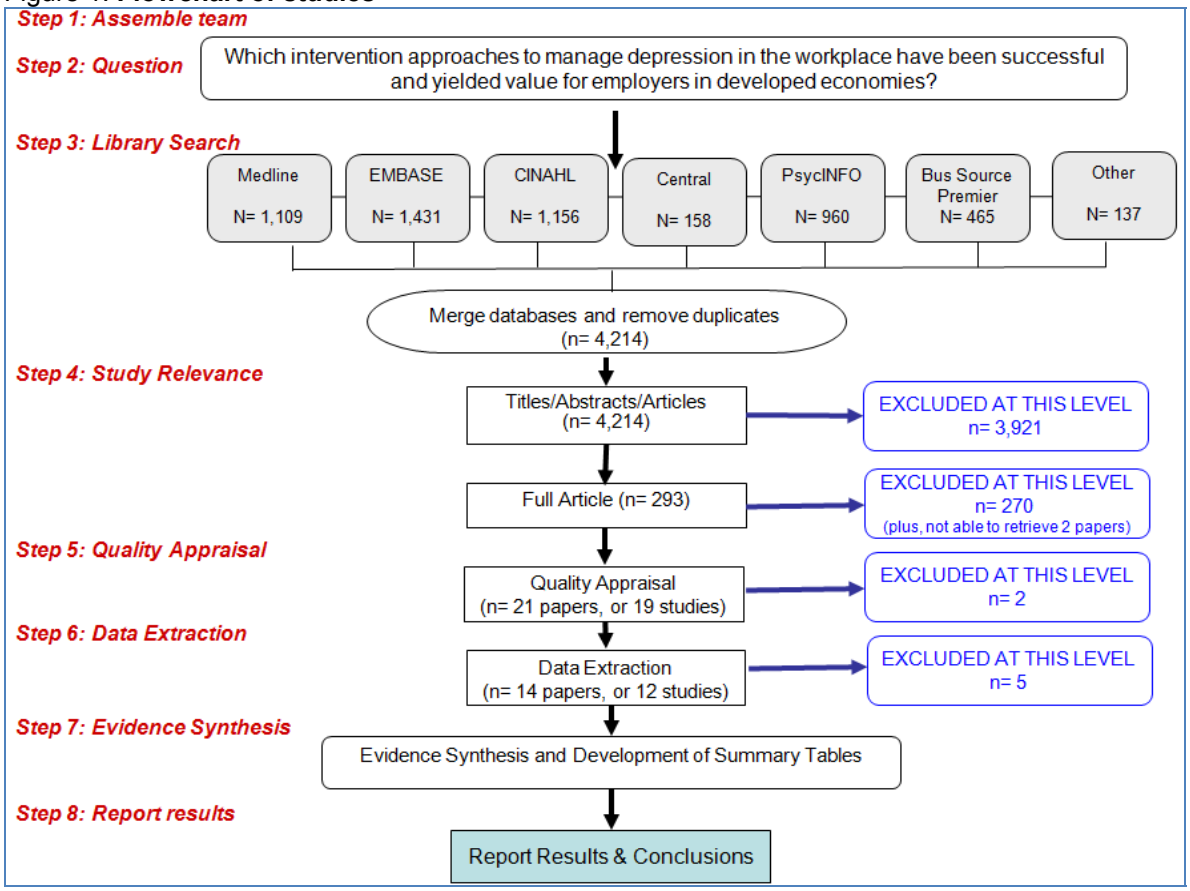
Consequently, 21 articles proceeded to Level 3 methodological quality assessment. An additional check on relevancy at this stage led to the exclusion of a further two articles as they were determined not to meet our inclusion criteria. Furthermore, four articles were paired with a corresponding article that described results from the same study. Therefore, instead of considering the eight articles separately, they were considered as four studies, each with a supplemental article.[1] There was a fifth pair of studies (Lo Sasso 2004 and Smith 2002), for which we chose to consider the articles separately because their exclusions led to slightly different underlying study samples.

At Level 4, a total of 19 articles (17 studies) were included for extraction. In a final check for relevancy, five additional articles were excluded for not meeting our inclusion criteria. Ultimately, data from 14 articles (12 studies) were extracted. Altogether, we included 24 publications which report on 12 primary studies (Blonk 2006, Dewa 2009, Kawakami 1997, Knekt 2008a and 2008b Krogh 2009, Lo Sasso 2006 *and* Rost 2004, Rebergen 2009a *and* Rebergen 2009b, Schene 2007, Schoenbaum 2001, Smith 2002, van der Feltz-Cornelis 2010, Wang 2007) and a total of 12 supplemental articles (Krogh 2007, Rost 2000, Rost 2001, Rebergen 2007, Schoenbaum 2004, Jaycox 2003, Miranda 2003, Sherbourne 2001, Wells 1999, Wells 2000,

[1] Articles grouped together included: Proudfoot 2004 with McCrone 2004, Lo Sasso 2006 with Rost 2004, Rebergen 2009a with Rebergen 2009b and Brouwers 2007 with Brouwers 2006

van der Feltz-Cornelis 2007) related to interventions for depression in the workplace.

Figure 1: **Flowchart of studies**



3.2 Study Characteristics

3.2.1 Study jurisdictions

Of the studies identified, four were conducted in The Netherlands (Blonk 2006, Rebergen 2009a, 2009b, Schene 2007, van der Feltz-Cornelis 2010), four in the USA (Lo Sasso 2006, Rost 2004, Schoenbaum 2001, Smith 2002, Wang 2007) and one each in Canada, Finland, Denmark, and Japan (Dewa, Knekt 2008a, Knekt 2008b, Krogh 2009, Kawakami 1997). The jurisdictions where the studies were conducted influenced the type of interventions tested (e.g., all three studies looking at enhanced primary care

were conducted in the USA) and the type of outcomes considered (e.g., one study conducted in The Netherlands measured the process of returning to part time work before full time work (Rebergen et al., 2009a). How work disability is operationalized across the range of jurisdictions also varies: for instance, in Ontario and other North American jurisdictions, any period of absence due to health reasons is generally referred to as a period of work disability and absences are often characterized as either short-term or long-term work disability episodes, depending on the length of absence. On the other hand, in the Netherlands, health-related work absences are generally termed sickness absence or sick leave days for which workers receive benefits to compensate for lost work time. In the Dutch system, work disability refers specifically to a state of being unable to return to work after two full years of health-related sickness absence, as determined by an insurance physician or medical advisor. At this stage, a worker may then apply for a work disability pension. Given these key differences, we maintain use of the individual terms work disability and sickness absence (or sick leave) when referring to North American and non-North American studies, respectively, throughout the remainder of the report. It should also be noted that the studies were conducted in multiple different social security systems (see Appendix G for details on the various social security systems). Moreover, even when studies were conducted in the same jurisdiction, they often drew their samples from different study populations. Ultimately, these differences across studies pose a challenge to drawing direct comparisons between studies. For more detailed information on individual study characteristics, please refer to Appendix H, Table 1.

3.2.2 General study characteristics

Of the 12 included studies, 10 were randomized controlled trials (RCTs) (Blonk et al., 2006; Knekt et al., 2008a and 2008b; Krogh et al., 2009; Lo Sasso et al., 2006 and Rost et al., 2004; Rebergen et al., 2009a and 2009b; Schene et al., 2007; Schoenbaum et al., 2001; Smith et al., 2002; van der Feltz-Cornelis et al., 2010, Wang et al., 2007) and two were non-randomized studies (NRSs) with a separate control group (Dewa et al, 2009; Kawakami et al, 1997), Studies were conducted in a variety of settings, including various workplace companies (Blonk 2006, Dewa 2009, Kawakami 1997, Wang 2007), primary care practices (Lo Sasso 2006, Rost 2004, Schoenbaum 2001, Smith 2002), occupational health services (Rebergen 2009a, 2009b, van der Feltz-Cornelis 2010), and other specialty medical clinics (Knekt 2008a, Knekt 2008b, Krogh 2009, Schene 2007). Inclusion and exclusion criteria varied greatly across studies. Most study inclusion/exclusion criteria focused on the identification of depressive disorders in the potential study population, though diverse methods were used across studies. Some studies also used inclusion/exclusion criteria designed to include only those with a specific working status (e.g., in some

studies, participants were required to be on sick leave at baseline). Additional details can be found in Appendix H, Table 1.

3.2.3 Participant Characteristics

The source population from which study participants were recruited included physicians' practices (Knekt 2008a, Knekt 2008b, Krogh 2009, Lo Sasso 2006, Rost 2004, Schene 2007, Schoenbaum 2001, Smith 2002), workplaces (Dewa 2009, Kawakami 1997, Rebergen 2009a, 2009b, van der Feltz-Cornelis 2010, Wang 2007), or administrative databases of insurance/compensation companies (Blonk 2006). Participants across the studies also represented workers with various job titles. Half of the studies had small sample sizes, where fewer than 100 individuals comprised each group (Blonk 2006, Dewa 2009, Krogh 2009, Lo Sasso 2006, Rost 2004, Schene 2007, van der Feltz-Cornelis 2010). The studies with the largest sample sizes were those conducted in the USA by Schoenbaum et al. (2001) and Wang et al. (2007).

In some studies, all participants were working at baseline (Kawakami 1997, Smith 2002, Wang 2007), while in others, all participants were on a work disability leave (Dewa 2009) or on sick leave (Blonk 2006, Rebergen 2009a, 2009b, van der Feltz-Cornelis 2010). Some studies also included a mix, where some participants were working, others were on leave, and/or were currently unemployed (Knekt 2008a, Knetk 2008b, Krogh 2009, Lo Sasso 2006, Schene 2007), One study did not report working status at baseline (Schoenbaum 2001). For specific details on study participants, including additional data on age, sex, and education status of participants, please refer to Appendix H, Table 4.

3.2.4 Interventions

Table 8 presents a summary of the interventions used in the studies, including the intervention category, the provider of the intervention, and the working status of study participants at the time of the intervention. The interventions from the 12 studies identified were classified according to the most prominent focus of change being implemented as a result of the intervention. The 12 studies identified cover a diverse range of interventions that include psychological interventions (psychotherapy) (Blonk 2006, Knekt 2008a, Knekt 2008b), enhanced primary care (Lo Sasso 2006, Rost 2004, Smith 2002, Schoenbaum 2001), psychiatric care with adjuvant occupational therapy (Schene 2007), enhanced occupational physician roles (Rebergen 2009a, Rebergen 2009b, van der Feltz-Cornelis 2010), integrated care management (Dewa 2009, Wang 2007), exercise (Krogh 2009), and a worksite intervention (Kawakami 1997).

Most intervention categories include only one study. However, there were two studies that evaluated integrated care management (Dewa 2009, Wang

2007), two studies that evaluated enhanced primary care delivered by physicians and (nurse) care managers (Lo Sasso 2006, Rost 2004, Smith 2002, Schoenbaum 2001), and two studies that examined an enhanced role of occupational physicians in the management of depressed workers (Rebergen 2009a, Rebergen 2009b, van der Feltz-Cornelis 2010). As would be expected, the diverse interventions included in the studies were delivered by various professionals, including psychologists, primary care physicians and nurse care managers, occupational physicians, occupational therapists, psychiatrists, physiotherapists, and workplace supervisors. The interventions were administered to workers of various working status at baseline, The worksite stress reduction intervention (Kawakami 1997) was administered at the worksite to employees working at baseline, none of whom were specifically identified as being depressed. One integrated care management intervention was administered exclusively to depressed employed workers who were still at work (Wang 2007), and six other interventions (three enhanced primary care trials (Lo Sasso 2006, Rost 2004, Smith 2002, Schoenbaum 2001), one trial of psychotherapy (Knekt 2008a, 2008b), one trial of psychiatric care with adjuvant occupational therapy (Schene 2007), and one exercise trial (Krogh 2009)) were administered to at least some participants who remained working. Six studies included participants who were employed, but were absent from work for varying durations (Blonk 2006, Schene 2007, Rebergen 2009a/b, van der Feltz 2010, Dewa 2009, Krogh 2009). Five studies (Knekt 2008a and 2008b, Lo Sasso 2006; Rost 2004; Smith 2002; Schoenbaum 2001, Krogh 2009) included at least some participants who were unemployed when the intervention was provided (*see Appendix H, Table 2 for details of the intervention characteristics*).

Table 8: **Characteristics of interventions**

Category of Intervention	Intervention Description	Intervention Provider	Employment and Baseline Work Status	Studies
Psychological interventions	Cognitive-behavioural therapy	Psychologists (CBT)	Self-employed, 100% on sick leave	Blonk 2006
	Brief solution- focused psychotherapy Short-term psychodynamic psychotherapy Long-term psychodynamic psychotherapy	Therapists with qualifications in the specific psychotherapy delivered	80.7% Employed, working or studying at baseline 19.3% Unemployed	Knekt 2008a/b
Enhanced primary care delivered by	Enhanced care delivered by primary care physicians and	Physicians and Nurse care managers	45.3% Unemployed 54.7%	Lo Sasso 2006; Rost

Category of Intervention	Intervention Description	Intervention Provider	Employment and Baseline Work Status	Studies
physicians and nurses	nurses		Employed, full or part-time	2004; Smith 2002
	Quality improvement program for improved psychotherapy with primary care clinicians Quality improvement program for improved access to medications with primary care clinicians	"Practice therapists" for psychotherapy Nurses for medication follow-up	63.1% Employed 36.9% Unemployed	Schoenbaum 2001
Psychiatry plus occupational therapy	Psychiatric treatment with adjuvant occupational therapy	Psychiatrists and Occupational Therapists	Employed, but reduced or no work hours for 10 weeks to 2 years. 19.4% had reduced part-time hours, 80.6% were absent	Schene 2007
Enhanced occupational physician role	Guideline-based care by occupational physician	Occupational physicians	Employed, on sick leave	Rebergen 2009a/b
	Occupational physicians with specialized training	Occupational physicians	Employed, with an absence spell of at least 6 weeks	van der Feltz-Cornelis 2010
Integrated care management	Collaborative mental health program (enhanced disability management)	Psychiatrists	Employed, 100% on short-term work disability	Dewa 2009
	Telephone screening, outreach, and care management	Masters-level mental health clinicians	Employed, working	Wang 2007
Exercise	Strength training; aerobic training; relaxation training	Physiotherapists	44.2% Unemployed 46.1% Sick leave 9.7% Working	Krogh 2009
Worksite intervention	Worksite stress reduction program	Worksite supervisors	Employed, 100% working	Kawakami 1997

3.2.5 Diagnosis of depression

A number of different methods were used to identify individuals with depression in the studies of this review. Six studies utilized self-report questionnaires (Blonk 2006, Kawakami 1997, Rebergen 2009a/2009b, Schoenbaum 2001, van der Feltz-Cornelis 2010, Wang 2007). Five studies utilized personal interviews: administered by a psychiatrist and trained staff (Schene 2007), by a psychologist and research assistant (Krogh 2009), and by administrative staff (Lo Sasso 2006, Rost 2004, Smith 2002). Telephone interviews were used to identify depression in two other studies: administered by a psychologist (Blonk 2006) and by a survey interviewer (Wang 2007). Finally, two studies did not report the method of depression identification (Dewa 2009, Knekt 2008a/2009b). See Appendix H, Table 3 for further details on the various methods used, as well as depression characteristics of the samples in each study.

3.2.6 Study Outcomes

The primary and secondary outcomes of interest to this review were determined a priori as part of the review inclusion criteria (*see Section 2.2 Inclusion and Exclusion Criteria*). Please refer to Appendix H, Table 5 for details about the specific primary and secondary outcomes reported in each study.

Primary Work-Related Outcomes

Through the process of data extraction and evidence synthesis, a classification framework was established to categorize the primary work-related outcomes reported in the studies of this review into one of three categories. This framework, informed by stakeholders at the final meeting, is described below and outlined in Table 9.

1) *Prevention and management of work disability/sickness absence*
Prevention of work disability/sickness absence refers to measures or interventions aimed at preventing depressed workers who are currently working from 1) taking intermittent leaves of absence or even vacation time to cope with their depression; and 2) taking a leave of absence from work paid for by either a publicly or privately paid disability benefit program.

Examples of outcomes contained in this category and measured in the studies of this review include:

- Number of employed workers taking days off
- Number of sick leave days
- Number of days absent from work
- Number of days or hours worked
- Job retention
- Work status (working or on disability benefits)

Management of work disability/sickness absence refers to measures or interventions aimed at reducing the burden of a work disability claim, either to a private or public insurance payer, or sickness absence benefits. The main goals of these interventions are to promote RTW, to expedite RTW, or to prevent the transition from short-term to long-term disability or from sickness absence to work disability pension among workers currently on a paid leave of absence from work due to their depression.

Examples of outcomes contained in this category and measured in the studies of this review include:

- Return to work (RTW) (yes/no)
- Time until partial RTW, until full RTW, or until any RTW

2) *Work functioning* refers to measures or interventions aimed at the maintenance and/or improvement of work performance and productivity among depressed workers who are currently working. Outcomes within this category are designed to capture changes in performance and productivity at work due to workers' ongoing management of his or her depression. These outcomes are often referred to as presenteeism measures.

Examples of outcomes contained in this category and measured in the studies of this review include:

- Work ability, as measured by the Work Ability Index (Ilmarinen 1999)
- Work performance, as measured by the Health and Work Performance Questionnaire (HPQ) (Kessler 2003)

3) *Recurrence of work disability/sickness absence* refers to measures or interventions aimed at 1) preventing a relapse or recurrence of work disability/sickness absence due to depression and 2) managing recurrent episodes of depression-related work disability/sickness absence. Such interventions are geared towards workers who have returned to work after a previous period of depression-related work disability/sickness absence.

Please note that while we included "changes in rates of work-related accidents" as an inclusion criterion in our Level 2 screening, we did not find any studies that looked specifically at this outcome. However, in our secondary outcomes section, we include the category "reduction in critical workplace incidents" to account for a study (Wang 2009) that assessed work-related accidents combined with other critical workplace incidents as a composite outcome.

Table 9: **Framework of work-related outcomes relevant to review stakeholders**

Outcome Category	**Prevention of Work Disability/Sickness Absence**	**Management of Work Disability/Sickness Absence**	**Work Functioning**	**Recurrence of Work Disability/ Sickness Absence**
Relevant Study Population	Depressed workers, currently working and not on work disability leave/sickness absence	Depressed workers currently on work disability leave/sickness absence due to their depression	Depressed workers, currently working and not on work disability leave/sickness absence	Depressed workers who are currently working, but have had a prior episode of work disability/ sickness absence due to their depression
Among this study population, is there an effective intervention to:	Promote stay at work, promote job retention, or to prevent or reduce the number of casual sick leave days taken due to depression (e.g., use of vacation days or unpaid sick days) or paid sickness absence days?	Promote a return to work, to hasten a return to work, to prevent the transition from short-term work disability leave to long-term leave, or to prevent the transition from sickness absence to work disability?	Maintain or improve a worker's functioning both in terms of productivity and performance?	Prevent or reduce recurrences of work disability leave/ sickness absence due to depression?
Outcome Measures	- Number of causal sick leave days or vacation days - Number of paid sickness absence or sick leave days - Hours worked - Job retention - Transition to work disability leave	- Return to work - Duration on work disability leave/sickness absence - Transition from short-term disability to long-term disability - Transition from sickness absence to work disability	- Productivity and performance measures (e.g., Work Ability Index, Health and Work Performance questionnaire)	-Recurrence of work disability/ sickness absence - Number of work disability/ sickness absence recurrences - Duration of a recurrent work disability leave/ sickness absence

Primary Economic Outcomes
Five of the twelve studies in this review (Dewa 2009, Lo Sasso 2006, Schoenbaum 2001, Rebergen 2009a, Schene 2007) included an economic evaluation that measured the economic outcomes of the intervention compared to the comparison intervention. The interventions that included an economic evaluation were enhanced primary care (two studies) (Lo Sasso 2006, Schoenbaum 2001), enhanced occupational physician role (Rebergen 2009a), psychiatric care enhanced by occupational therapy (Schene 2007), and system integration and care management (Dewa 2009). Studies varied according to perspective from which analyses were conducted (societal perspective versus employer perspective), the type of economic analysis conducted, and the economic outcomes reported (s*ee Appendix H, Tables 5 and 7 for details of the economic outcomes*).

Secondary Outcomes
Studies in this review reported on a number of outcomes that we have classified as secondary outcomes. Of most relevance to this review were outcomes that documented improvements in depression severity and remission of depression. Other key secondary outcomes included psychosocial work outcomes (e.g., workplace conflict) and critical workplace incidents. Additional secondary outcomes were captured (*see Appendix H, Table 5*), but do not contribute to the evidence synthesis presented further below in this report.

3.3 Risk of Bias in Included Studies

3.3.1 Summary of Bias Criteria Across Studies
The overall findings pertaining to the risk of bias criteria are presented in Table 10, with criteria that were met, unmet, or unclear denoted by striped, black, and gray circles, respectively (*see Appendix I for more details*). A bias-specific summary follows below.

Table 10: **Summary assessment of risk of bias**

Bias	Criteria to Judge Risk of Bias	Blonk 2006	Dewa 2009	Kawa-kami 1997	Knekt 2008a, 2008b	Krogh 2009	Lo Sasso 2006, Rost 2004	Reber-gen 2009a, 2009b	Schene 2007	Schoe-nbaum 2001	Smith 2002	van der Feltz-Cornelis, 2010	Wang, 2007
Selection Bias	Participation rate >65%	◍	◍	○	◍	◍	◍	○	◍	◍	◍	◍	○
	No differences between participants and non-participants	○	◍	○	○	○	●	○	○	○	●	◍	○
	Intention-to-treat analyses completed	●	◍	◍	◍	◍	◍	◍	◍	◍	◍	◍	◍
	Adequate allocation to treatment groups	○	●	●	◍	◍	○	◍	◍	○	○	◍	●
	Baseline differences balanced or if unbalanced, accounted for in the analyses	○	◍	●	◍	●	◍	◍	●	◍	◍	◍	◍
Attrition Bias	Attrition rate <35%	◍	◍	◍	◍	◍	◍	◍	◍	◍	◍	◍	◍
	No differences between remaining participants and those lost to follow-up	◍	◍	●	●	◍	○	○	○	●	○	○	●
Performance Bias	Intervention process adequately described	◍	◍	●	◍	◍	◍	◍	◍	◍	◍	◍	◍
	Minimal opportunity for co-interventions	◍	◍	◍	○	●	◍	◍	◍	◍	◍	◍	◍
	Minimal opportunity for contamination	○	◍	◍	◍	●	◍	●	●	○	◍	●	◍
	Adequate compliance with the intervention	○	○	○	◍	●	◍	○	◍	●	◍	○	◍
	Intervention providers blinded												
	Participants blinded												

Bias	Criteria to Judge Risk of Bias	Blonk 2006	Dewa 2009	Kawa-kami 1997	Knekt 2008a, 2008b	Krogh 2009	Lo Sasso 2006, Rost 2004	Reber-gen 2009a, 2009b	Schene 2007	Schoe-nbaum 2001	Smith 2002	van der Feltz-Cornelis, 2010	Wang, 2007
Measurement Bias	Valid and reliable outcome instruments	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜
	Outcome described at both baseline and follow-up	⊜	⊜	⊜	⊜	⊜	●	⊜	⊜	⊜	⊜	●	⊜
	Follow-up >3 months duration	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	●	⊜	⊜	⊜
	Outcome assessors blinded	⊜	⊜	○	●	●	⊜	⊜	⊜	⊜	⊜	⊜	⊜
	Participants blinded												
Reporting Bias	Research question clearly stated	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜
	Direct between-group comparison completed	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜	⊜
	Appropriate statistical analyses	⊜	⊜	⊜	⊜	⊜	●	⊜	⊜	●	⊜	⊜	●

Legend: ⊜ = criteria met; ○ = unclear if criteria met due to insufficient information; ● = criteria not met

3.3.1.1 Selection Bias

Of all 12 included studies, 10 were randomized controlled trials (RCTs) (Blonk et al., 2006; Knekt et al., 2008a and 2008b; Krogh et al., 2009; Lo Sasso et al., 2006 and Rost et al., 2004; Rebergen et al., 2009a and 2009b; Schene et al., 2007; Schoenbaum et al., 2001; Smith et al., 2002; van der Feltz-Cornelis et al., 2010, Wang et al., 2007). Two were non-randomized studies (NRSs) with a separate control group (Dewa et al, 2009; Kawakami et al, 1997) and were considered to have an inadequate allocation method. Of the 10 RCTs, one was judged to have performed random group allocation inadequately due to a potential lack of allocation concealment (Wang 2007). Three others described group allocation as random, but did not provide sufficient details to gauge the adequacy of the process (Blonk 2006, Lo Sasso 2006, Schoenbaum 2001). All other RCTs were considered to have performed allocation adequately.

Almost all studies analyzed the participants according to the groups to which they were originally allocated (i.e., intention-to-treat), with one exception (Blonk 2006). In this study, participants who actually received the intervention were removed from the analysis because of a misunderstanding, therefore breaking the randomized allocation.

Participation among eligible potential participants was described in nine studies (Blonk 2006, Dewa 2009, Knekt 2008a, Krogh 2009, Lo Sasso 2006, Schene 2007, Schoenbaum 2001, Smith 2002, van der Feltz-Cornelis 2010), and in all cases, the reported participation rate was judged adequate (more than 65%). However, only two of these studies examined whether important differences existed between participants and non-participants (Lo Sasso 2006, Rost 2004), finding major differences between the groups on a number of factors, including depression symptoms.

We also assessed whether baseline characteristics were described and appropriately balanced between groups, which is a consequence of appropriate intervention allocation methods. The two NRSs (Dewa 2009 and Kawakami 1997) demonstrated important baseline differences between the intervention and control groups, but in the latter study (Kawakami 1997), analyses were conducted without adjustment for these differences. One RCT (Blonk 2006) did not report baseline characteristics. Five of the 10 RCTs also showed important baseline differences (Knekt 2008a/2008b, Krogh 2009, Lo Sasso, Schene 2007, Smith 2002) and in three of these studies (Krogh 2009, Schene 2007, Smith 2002), no adjustment was made for these differences.

3.3.1.2 Attrition Bias

Losses to follow-up were not applicable in one NRS based on administrative data (Dewa 2009). All other studies had an acceptable attrition rate (lower

than 35%). Five studies (Blonk 2006, Kawakami 1997, Knekt 2008a/2008b, Krogh 2009, Wang 2007) examined whether there were any important differences between those who remained in the study and those who lost to follow-up, with important differences found in four of these studies (Kawakami 1997, Knekt 2008a, Schoenbaum 2001, and Wang 2007), and no important differences found in the remaining two studies (Blonk 2006, Krogh 2009).

3.3.1.3 Performance Bias

The interventions were described adequately to allow replication in all, but one study (Kawakami 1997). Only seven studies reported on participants' compliance with their assigned intervention (Knekt 2008a/2008b, Krogh 2009, Lo Sasso 2006, Schene 2007, Schoenbaum 2001, Smith 2002 and Wang 2007). All were judged to have adequate rates of compliance, with the exception of two studies (Krogh 2009, Schoenbaum 2001). In the former study (Krogh 2009), mean participation in the 32 sessions of strength, aerobic, and relaxation training groups was only 18 (56.2%), 16.2 (50.6%) and 10.5 (32.8%) sessions, respectively. Similarly, in the Schoenbaum et al. study (2001), only 30% of the medication group were followed for the anticipated duration, while only 40% of the therapy group actually received cognitive behavioural therapy.

One common problem with the interventions included in this review was the potential for contamination or co-interventions. We determined that in six studies (Blonk 2006, Krogh 2009, Rebergen 2009a/2009b, Schene 2007, Schoenbaum 2001, van der Feltz-Cornelis, 2010) contamination had a potential to introduce bias, while bias associated with co-interventions was possible in two studies (Knekt 2008a/2008b, Krogh 2009).

3.3.1.4 Measurement Bias

The outcome measures used in the studies were judged to be adequate, valid and reliable in eleven studies. Three studies only reported data on the main outcomes at follow-up (Lo Sasso 2006, Schoenbaum 2001, van der Feltz-Cornelis 2010). All 12 studies had a length of follow-up longer than 3 months. We also considered whether outcome assessors were blinded to participants' group status. In four studies (Blonk 2006, Dewa 2009, Rebergen 2009a/2009b, Schoenbaum 2001), this was not applicable, while in one study (Kawakami 1997), blinding was not described. Finally, measurement bias was likely in two studies (Knekt 2008a/2008b, Krogh 2009) that did not perform a blinded outcome assessment.

3.3.1.5 Reporting Bias

All 12 studies included in this review reported the findings of a direct comparison between intervention groups. All studies, except three (Lo Sasso 2006, Schoenbaum 2001, Wang 2007) were judged to have

adequate statistical analyses. In these three studies, imputation techniques were performed for missing data, though in each case, there was a significant amount of missing information, potentially increasing the risk of reporting bias in these studies.

3.3.2 Summary of Overall Risk of Bias Across Studies

Selection bias was found to be the most common potential bias in the studies of this review (see Table 11), with eight studies (Blonk 2006, Dewa 2009, Kawakami 1997, Krogh 2009, Lo Sasso 2006/Rost 2004, Schene 2007, Smith 2002, Wang 2007) demonstrating a high risk of selection bias and three a moderate risk (Knekt 2008a/2008b, Rebergen 2009a/2009b, Schoenbaum 2001). Reasons were varied, as described in the previous section. Risk of performance and attrition bias were also apparent in a number of studies, including five demonstrating a high risk of performance bias (Kawakami 1997, Knekt 2008a/2008b, Rebergen 2009a/2009b, van der Feltz-Cornelis 2010, Wang 2007), namely due to the potential for contamination and co-interventions. Four studies demonstrated a high risk of attrition bias (Kawakami 1997, Knekt 2008a/2008b, Schoenbaum 2001, Wang 2007).

Studies were found to be less likely at risk of both measurement and reporting bias. Only two studies (Knekt 2008a/2008b and Krogh 2009) demonstrated a high risk of measurement bias, owing to the lack of blinded outcome assessment in these studies. Similarly, reporting bias was high in only three studies (Lo Sasso 2006/Rost 2004, Schoenbaum 2001, Wang 2007) due to the use of multiple imputation methods for a large quantity of missing data.

Overall, each study demonstrated a high risk of at least one type of bias. As a result, all 12 studies included in this review were judged to be at an overall *high* risk of bias which reduces the certainty of the findings.

Table 11: **Overall risk of bias across studies**

Bias / Author, Year*	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk Judgement
Blonk 2006	●	◒	○	◒	◒	High
Dewa 2009	●	◒	○	◒	◒	High
Smith 2002	●	○	◒	◒	◒	High
Rebergen 2009a, 2009b	○	○	●	◒	◒	High
van der Feltz-Cornelis 2010	◒	○	●	○	◒	High
Schene 2007	●	○	●	◒	◒	High
Krogh 2009	●	◒	●	●	◒	High
Kawakami 1997	●	●	○	○	◒	High
Knekt 2008a, 2008b	○	●	○	●	◒	High
Lo Sasso 2006, Rost 2004	●	○	◒	○	●	High
Wang 2007	●	●	○	◒	●	High
Schoenbaum 2001	○	●	●	○	●	High

* Ordered according to ascending number of high risk bias categories in each study

Legend: ◒ = criteria met; ○ = unclear if criteria met due to insufficient information; ● = criteria not met

3.4 Evidence Synthesis

3.4.1 Results of the GRADE assessment

As previously described in Section 2.8.1, a number of review sub-questions were formulated based on the framework of primary outcomes (*Table 9*), in order to inform the process of answering the review's main research objective. The quality of evidence available to answer each sub-question was then evaluated on six domains. It is important to note that we only graded the quality of evidence for those interventions demonstrating positive findings for the particular sub-question involved (i.e., the intervention group was statistically significantly better than the control group). We did not grade interventions found to be equivalent to the control group for a particular sub-question, but instead, summarize these interventions in a corresponding table. Results of the GRADE assessments below are presented in the manner described in Table 12 below. In all cases, the grade of evidence was considered to be "very low". The GRADE tables are found in Tables 13a to 19b.

Table 12: **Research sub-questions for which the evidence was graded**

Primary Outcome Category	Sub-Question	GRADE Table	Corresponding Null Findings Table
Prevention of Work Disability/ Sickness Absence	Among workers currently working and not on work disability leave/ sickness absence, which interventions for depression *significantly reduce* the incidence and duration of absenteeism from work (i.e., casual sick leave days/sickness absence days)?	Table 13a	Table 13b
	Among workers currently working and not on work disability leave/ sickness absence, which interventions for depression *significantly improve* job retention?	Table 14a	Table 14b
	Among workers currently working and not on work disability leave/ sickness absence, which interventions for depression *significantly increase* the number of worked hours?	Table 15a	Table 15b
Management of Work Disability/	Among workers currently on work disability leave/sickness absence, which interventions for depression	Table 16a	Table 16b

Sickness Absence	*are significant* in returning workers to work and reducing work disability/ sickness absence duration?		
	Among workers currently on work disability leave/sickness absence, which interventions for depression *significantly prevent* the transition from short-term to long-term work disability or the transition from sickness absence to work disability?	Table 17a	Table 17b
Work Functioning	Among workers currently working and not on work disability leave/ sickness absence, which interventions for depression *significantly improve* work functioning?	Table 18a	Table 18b
Recurrence of Work Disability/ Sickness Absence	Among workers who are currently working, but have had a prior episode of work disability leave/sickness absence, which interventions for depression *significantly prevent or reduce* recurrences of work disability leave/sickness absence?	Table 19a	Table 19b

Table 13a: Prevention of work disability/sickness absence - Among workers currently working and not on work disability leave/sickness absence, which interventions for depression *significantly reduce* the incidence and duration of absenteeism from work (i.e., casual sick leave days/sickness absence days)?†

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
In the short term							
None							Not Applicable
In the long term							
Worksite-wide stress reduction program (Kawakami, 1997)	Non-Randomized Study	High	Not Applicable	Population: yes[2] Outcome: yes[3]	Not Precise[4]	Not Applicable	⊙○○○
	Initial GRADE: Low ⊙⊙○○	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: 0	
Short-term psychodynamic psychotherapy (Knekt, 2008)[5]	Randomized Controlled Trial	High	Not Applicable	Population: no[6] Outcome: yes[7]	Not Precise[8]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: 0	
Long-term psychodynamic psychotherapy (Knekt, 2008)[9]	Randomized Controlled	High	Not Applicable	Population: no[10]	Not Precise[12]	Not Applicable	⊙○○○

[2] Blue collar workers at a large electric company, consisting of machine operators and technicians
[3] Length of sick leave in the past year (days) measured at 24 months
[4] Evidence from only one study
[5] Compared to brief and resource-oriented solution-focused psychotherapy
[6] Outpatients referred from various psychiatric services, consisting of white collar workers and entrepreneurs
[7] Proportion with more than 7 sick leave days during last 3 months
[8] Evidence from only one study
[9] Compared to brief and resource-oriented solution-focused psychotherapy

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
	Trial			Outcome: yes[11]			
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: 0	
Strength training (Krogh, 2009)[13,14]	Randomized Controlled Trial	High	Not Applicable	Population: no[15] Outcome: yes[16]	Not Precise[17]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: 0	

†Compared to usual care, unless otherwise specified
*In alphabetical order by first author's last name due to equivalent GRADEs
**Only assessed in studies with economic analyses

[10] Outpatients referred from various psychiatric services, consisting of white collar workers and entrepreneurs
[12] Evidence from only one study
[11] Proportion with more than 7 sick leave days during last 3 months
[13] Compared to relaxation training (active control)
[14] Note that study sample included workers who were unemployed, employed and working, and on sick leave at baseline. It is unclear whether evidence based on this study outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, this data is also presented in Table 16a under management of work disability/sickness absence
[15] Patients referred from general practitioners, private practice psychiatrists, psychologists and psychiatric wards.
[16] Percentage of days absent from work in the last 10 days
[17] Evidence from only one study

Table 13b: Prevention of work disability/sickness absence - Among workers currently working and not on work disability/sickness absence leave, which interventions for depression *are not different from control* in reducing the incidence and duration of absenteeism from work (i.e., casual sick leave/sickness absence days)?†

Intervention*	Specific Study Outcome(s)	Comments
In the short term		
Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)	1. Number of sick leave days during last 3 months 2. Proportion with more than 7 sick leave days during last 3 months	- Compared to long-term psychotherapy (active control) and short-term psychotherapy
Short-term psychotherapy (Knekt 2008)	1. Number of sick leave days during last 3 months 2. Proportion with more than 7 sick leave days during last 3 months	- Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy
Long-term psychotherapy (Knekt, 2008)	1. Number of sick leave days during last 3 months 2. Proportion with more than 7 sick leave days during last 3 months	- Compared to short-term psychotherapy and brief and resource-oriented solution-focused psychotherapy
Strength training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 4 months 2. Percentage of days absent from work in last 10 days measured at 4 months	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Aerobic training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 4 months 2. Percentage of days absent from work in last 10 days measured at 4 months	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Relaxation training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 4 months 2. Percentage of days absent from work in last 10 days measured at 4 months	- Compared to strength training and aerobic training - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence

Intervention*	Specific Study Outcome(s)	Comments
In the long term		
Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)	1. Number of sick leave days during last 3 months	- Compared to long-term psychotherapy (active control) and short-term psychotherapy
Short-term psychodynamic psychotherapy (Knekt 2008)	1. Number of sick leave days during last 3 months 2. Proportion with more than 7 sick leave days during last 3 months	- Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy for outcome 1 - Compared to long-term psychotherapy (active control) for outcome 2
Long-term psychodynamic psychotherapy (Knekt 2008)	1. Number of sick leave days during last 3 months 2. Proportion with more than 7 sick leave days during last 3 months	- Compared to short-term psychotherapy and brief and resource-oriented solution-focused psychotherapy for outcome 1 - Compared to short-term psychotherapy for outcome 2
Strength training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 12 months	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Aerobic training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 12 months 2. Percentage of days absent from work in last 10 days measured at 12 months	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Relaxation training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 12 months 2. Percentage of days absent from work in last 10 days measured at 12 months	- Compared to strength training and aerobic training for outcome 1 - Compared to aerobic training only for outcome 2 - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Enhanced depression care delivered by primary care physicians and nurses (Lo Sasso,	1. Number of days absent over 24 months	- Not measured in the short term - Applicable to individuals who were consistently employed over the follow-up.

Intervention*	Specific Study Outcome(s)	Comments
2006)		
Quality improvement program for improved psychotherapy with primary care clinicians, combined with quality improvement program for improved access to medication in primary care (Schoenbaum, 2001)	1. Days missed from work due to illness over 24 months	- Not measured in the short term

†Compared to usual care, unless otherwise specified
*In alphabetical order by first author's last name

Table 14a: Prevention of work disability/sickness absence - Among workers currently working and not on work disability leave/sickness absence, which interventions for depression *significantly improve* job retention?†

Intervention	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit*	Final GRADE
In the short term							
Telephone screening, outreach and care management (Wang 2007)	Randomized Controlled Trial	High	Not Applicable	Population: yes[18] Outcome: yes[19]	Not Precise[20]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: 0	
In the long term							
Enhanced depression care delivered by primary care physicians and nurses (Smith, 2002)	Randomized Controlled Trial	High	Not Applicable	Population: no[21] Outcome: no[22]	Not Precise[23]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -2	GRADE Adjustment: -1	GRADE Adjustment: 0	

†Compared to usual care, unless otherwise specified
*Only assessed in studies with economic analyses

[18] Employees from one of 16 companies covered by a managed care program
[19] Job retention measured by the Health Performance Questionnaire at 6 months
[20] Evidence from only one study
[21] Patients presenting for routine visit at participating community primary care practices
[22] Subsequent employment status measured at 12 months
[23] Evidence from only one study

Table 14b: Prevention of work disability/sickness absence - Among workers currently working and not on work disability leave/sickness absence, which interventions for depression *are not different from control* in improving job retention?†

Intervention*	Specific Study Outcome(s)	Comments
In the short term		
Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)	1. Current employment status	- Compared to long-term psychotherapy (active control) and short-term psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Short-term psychotherapy (Knekt 2008)	1. Current employment status	- Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Long-term psychotherapy (Knekt, 2008)	1. Current employment status	- Compared to short-term psychotherapy and brief and resource-oriented solution-focused psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Strength training (Krogh, 2009)	1. Unemployment status	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Aerobic training (Krogh, 2009)	1. Unemployment status	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on

Intervention*	Specific Study Outcome(s)	Comments
		these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Relaxation training (Krogh, 2009)	1. Unemployment status	- Compared to strength training and aerobic training - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Enhanced depression care delivered by primary care physicians and nurses (Smith, 2002)	1. Subsequent employment at 6 months	
In the long term		
Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)	1. Current employment status	- Compared to long-term psychotherapy (active control) and short-term psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Short-term psychotherapy (Knekt 2008)	1. Current employment status	- Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Long-term psychotherapy (Knekt, 2008)	1. Current employment status	- Compared to short-term psychotherapy and brief and resource-oriented solution-focused psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence

Intervention*	Specific Study Outcome(s)	Comments
Strength training (Krogh, 2009)	1. Unemployment status	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Aerobic training (Krogh, 2009)	1. Unemployment status	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Relaxation training (Krogh, 2009)	1. Unemployment status	- Compared to strength training and aerobic training - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the prevention of work disability/sickness absence. Therefore, data are also presented in **Table 16b** under management of work disability/sickness absence
Psychiatric treatment with adjuvant occupational therapy (Schene, 2007)	1. Proportion working at least 2 days or 16 hours per week over 42 months	- Not measured in the short term - Study sample included workers who were working and not working at baseline. Unclear whether evidence based on this outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, data also presented in **Table 16b** under management of work disability/sickness absence
Telephone screening, outreach and care management (Wang 2007)	1. Job retention measured by the Health Performance Questionnaire at 12 months	

†Compared to usual care, unless otherwise specified
*In alphabetical order by first author's last name

Table 15a: Prevention of work disability/sickness absence - Among workers currently working and not on work disability leave/sickness absence, which interventions for depression *significantly increase* the number of worked hours?†

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
In the short term							
Psychiatric treatment with adjuvant occupational therapy (Schene, 2007)[24]	Randomized Controlled Trial	High	Not Applicable	Population: no[25] Outcome: no[26]	Not Precise[27]	Yes[28]	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -2	GRADE Adjustment: -1	GRADE Adjustment: +1	
In the long term							
Psychiatric treatment with adjuvant occupational therapy (Schene, 2007)[29]	Randomized Controlled Trial	High	Not Applicable	Population: no[30] Outcome: no[31]	Not Precise[32]	Yes[33]	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -2	GRADE Adjustment: -1	GRADE Adjustment: +1	

[24] Note that study sample included workers who were working and not working at baseline. It is unclear whether evidence based on this study outcome is strictly applicable to the prevention of work disability. Therefore, this data is also presented in Table 16a under management of work disability/sickness absence
[25] Patients presenting to an outpatient mood disorder clinic
[26] Total hours worked between 7 and 12 months.
[27] Evidence from only one study
[28] Compared to the workers in the control group, the additional mean net benefit per worker in the intervention group was $3,952.
[29] Note that study sample included workers who were working and not working at baseline. It is unclear whether evidence based on this study outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, this data is also presented in Table 16a under management of work disability
[30] Patients presenting to an outpatient mood disorder clinic
[31] Total hours worked between 13 and 18 months.
[32] Evidence from only one study
[33] Compared to the workers in the control group, the additional mean net benefit per worker in the intervention group was $3,952.

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
Quality improvement program for improved psychotherapy with primary care clinicians (Schoenbaum, 2001)[34]	Randomized Controlled Trial	High	Not Applicable	Population: no[35] Outcome: yes[36]	Not Precise[37]	Yes[38]	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: +1	

†Compared to usual care, unless otherwise specified
*In alphabetical order by first author's last name due to equivalent GRADEs
**Only assessed in studies with economic analyses

[34] Note that study sample included workers who were working and not working at baseline. It is unclear whether evidence based on this study outcome is strictly applicable to the prevention of work disability. Therefore, this data is also presented in Table 16a under management of work disability/sickness absence
[35] Patients presenting to primary care clinics
[36] Days of employment over 24 months, calculated as number of days worked over each 6-month follow-up
[37] Evidence from only one study
[38] The incremental costs were within the range of many accepted medical interventions for depression, and substantially below the estimated value for a year of life.

Table 15b: Prevention of work disability/sickness absence - Among workers currently working and not on work disability leave/sickness absence, which interventions for depression *are not different from control* in increasing the number of worked hours?†

Intervention*	Specific Study Outcome(s)	Comments
In the short term		
Telephone screening, outreach and care management (Wang 2007)	1. Number of weekly hours worked measured at 6 months	
In the long term		
Psychiatric treatment with adjuvant occupational therapy (Schene, 2007)	1. Total hours worked between 19 and 42 months	- Study sample included workers who were working and not working at baseline. Unclear whether evidence based on this study outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, this data is also presented in **Table 16b** under management of work disability/sickness absence
Quality improvement program for improved access to medication in primary care (Schoenbaum, 2001)	1. Days of employment over 24 months, calculated as number of days worked over each 6-month follow-up	- Not measured in the short term - Study sample included workers who were working and not working at baseline. Unclear whether evidence based on this study outcome is strictly applicable to the prevention of work disability/sickness absence. Therefore, this data is also presented in **Table 16b** under management of work disability/sickness absence
Telephone screening, outreach and care management (Wang 2007)	1. Number of weekly hours worked measured at 12 months	

†Compared to usual care, unless otherwise specified
*In alphabetical order by first author's last name

Table 16a: Management of work disability/sickness absence - Among workers currently on work disability leave/sickness absence, which interventions for depression *are significant* in returning workers to work and reducing work disability/sickness absence duration?†

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
In the short term							
Psychiatric treatment with adjuvant occupational therapy (Schene, 2007)[39]	Randomized Controlled Trial	High	Not Applicable	Population: no[40] Outcome: no[41]	Not Precise[42]	Yes[43]	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -2	GRADE Adjustment: -1	GRADE Adjustment: +1	
Enhanced role of occupational physician with psychiatric consultations (van der Feltz-Cornelis, 2010)	Randomized Controlled Trial	High	Not Applicable	Population: yes[44] Outcome: yes[45]	Not Precise[46]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: 0	
In the long term							

[39] Note that study sample included workers who were working and not working at baseline. It is unclear whether evidence based on this study outcome is strictly applicable to the management of work disability/sickness absence. Therefore, this data is also presented in Table 15a under prevention of work disability/sickness absence

[40] Patients presenting to an outpatient mood disorder clinic

[41] Total hours worked between 7 and 12 months.

[42] Evidence from only one study

[43] Compared to the workers in the control group, the additional mean net benefit per worker in the intervention group was $3,952.

[44] Sick-listed patients seen by occupational physicians at companies providing occupational health care, consisting of legislators, senior officials and managers, professionals, association professionals, craft and related trades workers, technicians, manual labourers, clerks, service workers, and shop and market sales workers

[45] Full RTW status (yes/no) measured at 3 months

[46] Evidence from only one study

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
Brief cognitive behavioural therapy-based stress management delivered by labour experts with a focus on improving workplace processes (Blonk, 2006)[47]	Randomized Controlled Trial	High	Not Applicable	Population: yes[48] Outcome: yes[49]	Not Precise[50]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: 0	
Enhanced disability management consisting of an additional collaborative mental health program delivered by psychiatrists to workers on short-term disability leave for psychiatric disorders (Dewa, 2009)[51]	Non-Randomized Study	High	Not Applicable	Population: yes[52] Outcome: yes[53]	Not Precise[54]	Yes[55]	⊙○○○
	Initial GRADE: Low ⊙⊙○○	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: +1	
Strength training (Krogh, 2009)[56,57]	Randomized Controlled Trial	High	Not Applicable	Population: no[58] Outcome: yes[59]	Not Precise[60]	Not Applicable	⊙○○○

[47] Compared to cognitive behavioural therapy (active control), as well as usual care
[48] Employed workers insured for sick leave, consisting of self-employed workers in agricultural, service, construction, health care, trade, and other industries.
[49] Time until full RTW measured at 12 months
[50] Evidence from only one study
[51] Compared to usual disability management practices consisting of treatment by primary care physician and third-part psychiatrist claim adjudication
[52] Employees of a nationwide financial insurance sector company on short-term disability leave for a psychiatric disorder
[53] Two outcomes: return to work and days on short-term disability measured at 12 months
[54] Evidence from only one study
[55] Compared to the control group, the intervention resulted in 22% more disabled workers returning to work, resulted in 15 fewer STD days lost, and cost $355 less per treated worker, thus from a cost-effectiveness perspective, the intervention dominated.
[56] Compared to relaxation training (active control)
[57] Note that study sample included workers who were unemployed, employed and working, and on sick leave at baseline. It is unclear whether evidence based on this study outcome is strictly applicable to the management of work disability/sickness absence. Therefore, this data is also presented in Table 13a under prevention of work disability/sickness absence
[58] Patients referred from general practitioners, private practice psychiatrists, psychologists and psychiatric wards.
[59] Percentage of days absent from work in the last 10 days

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: 0	
Dutch guideline-based care for employees on sick leave due to mental health problems delivered by occupational physicians (Rebergen, 2009)	Randomized Controlled Trial	High	Not Applicable	Population: yes[61] Outcome: no[62]	Not Precise[63]	Yes[64]	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: +1	
Psychiatric treatment with adjuvant occupational therapy (Schene, 2007) [65]	Randomized Controlled Trial	High	Not Applicable	Population: no[66] Outcomes: yes[67], no[68]	Not Precise[69]	Yes[70]	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -2	GRADE Adjustment: -1	GRADE Adjustment: +1	
Quality improvement program for improved psychotherapy with primary care clinicians	Randomized Controlled Trial	High	Not Applicable	Population: no[72] Outcome:	Not Precise[74]	Yes[75]	⊙○○○

[60] Evidence from only one study
[61] Employees from police departments on sick leave for mental health problems, consisting of executive and administrative police department workers
[62] Immediate full RTW versus partial RTW measured at 12 months
[63] Evidence from only one study
[64] The net mean benefit of the intervention compared to control was €3,582
[65] Note that study sample included workers who were working and not working at baseline. It is unclear whether evidence based on the study outcome 'total hours worked between 13 and 18 months' is strictly applicable to the management of work disability/sickness absence. Therefore, the data for this outcome is also presented in Table 15a under prevention of work disability/sickness absence
[66] Patients presenting to an outpatient mood disorder clinic
[67] Time until any work resumption measured over 42 months
[68] Total hours worked between 13 and 18 months
[69] Evidence from only one study
[70] Compared to the workers in the usual care group, the additional mean net benefit per worker in the intervention group was $3,952

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
(Schoenbaum, 2001)[71]				yes[73]			
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: +1	

†Compared to usual care, unless otherwise specified
*In alphabetical order by first author's last name due to equivalent GRADEs
**Only assessed in studies with economic analyses

[72] Patients presenting to primary care clinics
[74] Evidence from only one study
[75] The incremental costs were within the range of many accepted medical interventions for depression, and substantially below the estimated value for a year of life.
[71] Note that study sample included workers who were working and not working at baseline. It is unclear whether evidence based on this study outcome is strictly applicable to the management of work disability/sickness absence. Therefore, this data is also presented in Table 15a under prevention of work disability/sickness absence
[73] Days of employment over 24 months, calculated as number of days worked over each 6-month follow-up

Table 16b: Management of work disability/sickness absence - Among workers currently on work disability leave/sickness absence, which interventions for depression *are not different from control* in returning workers to work and reducing work disability/sickness absence duration?†

Intervention*	Specific Study Outcome(s)	Comments
In the short term		
Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)	1. Current employment status	- Compared to long-term psychotherapy (active control) and short-term psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in **Table 14b** under prevention of work disability/sickness absence
Short-term psychotherapy (Knekt 2008)	1. Current employment status	- Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in **Table 14b** under prevention of work disability/sickness absence
Long-term psychotherapy (Knekt, 2008)	1. Current employment status	- Compared to short-term psychotherapy and brief and resource-oriented solution-focused psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in **Table 14b** under prevention of work disability/sickness absence
Strength training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 4 months 2. Percentage of days absent from work in last 10 days measured at 4 months 3. Unemployment status	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the management of work disability/sickness absence. Therefore, data are so presented in **Tables 13b and 14b** under prevention of work disability/sickness absence
Aerobic training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 4 months	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and

Intervention*	Specific Study Outcome(s)	Comments
	2. Percentage of days absent from work in last 10 days measured at 4 months 3. Unemployment status	working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the management of work disability/sickness absence. Therefore, data are so presented in **Tables 13b and 14b** under prevention of work disability/sickness absence
Relaxation training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 4 months 2. Percentage of days absent from work in last 10 days measured at 4 months 3. Unemployment status	- Compared to strength training and aerobic training for outcomes 1 and 2 - Compared to aerobic training for outcome 3 - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the management of work disability/sickness absence. Therefore, data are so presented in Tables **13b and 14b** under prevention of work disability/sickness absence
Enhanced role of occupational physician with psychiatric consultations (van der Feltz-Cornelis, 2010)	1. Full RTW status (yes/no) measured at 6 months	- Mixed results in the short term, as the intervention was effective at 3 months for this outcome
In the long term		
Brief cognitive behavioural therapy-based stress management delivered by labour experts with a focus on improving workplace processes (Blonk, 2006)	1. Time until partial return to work	- Compared to cognitive behavioural therapy, as well as usual care - Not measured in the short term
Cognitive behavioural therapy delivered by psychologists (Blonk, 2006)	1. Time until partial return to work 2. Time until full return to work	- Compared to brief cognitive behaviour therapy-based stress management with a focus on workplace processes, as well as usual care for outcome 1 - Compared to usual care for outcome 2 - Not measured in the short term
Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)	1. Current employment status	- Compared to long-term psychotherapy (active control) and short-term psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in **Table 14b** under prevention

Intervention*	Specific Study Outcome(s)	Comments
		of work disability/sickness absence
Short-term psychotherapy (Knekt 2008)	1. Current employment status	- Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in **Table 14b** under prevention of work disability/sickness absence
Long-term psychotherapy (Knekt, 2008)	1. Current employment status	- Compared to short-term psychotherapy and brief and resource-oriented solution-focused psychotherapy - Study sample included workers who were unemployed, employed and working, or studying at baseline. Unclear whether evidence based on this outcome is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented **in Table 14b** under prevention of work disability/sickness absence
Strength training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 12 months 3. Unemployment status	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in **Tables 13b and 14b** under prevention of work disability/sickness absence
Aerobic training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 12 months 2. Percentage of days absent from work in last 10 days measured at 12 months 3. Unemployment status	- Compared to relaxation training (active control) - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in Tables **13b and 14b** under prevention of work disability/sickness absence
Relaxation training (Krogh, 2009)	1. Proportion of individuals on sick leave measured at 12 months 2. Percentage of days absent from work in last 10 days measured at 12 months 3. Unemployment status	- Compared to strength training and aerobic training for outcomes 1 and 3 - Compared to aerobic training for outcome 2 - Study sample included workers who were unemployed, employed and working, and on sick leave at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in **Tables 13b and 14b** under prevention of work disability/sickness absence
Enhanced depression	1. Number of days absent over 24	- Not measured in the short term

Intervention*	Specific Study Outcome(s)	Comments
care delivered by primary care physicians and nurses (Lo Sasso, 2006)	months	- Applicable to individuals who were inconsistently employed over the follow-up.
Dutch guideline-based care for employees on sick leave due to mental health problems delivered by occupational physicians (Rebergen, 2009)	1. Duration of sick leave days until partial return to work measured at 12 months 2. Duration of sick leave days until full return to work measured at 12 months 3. Duration of sick leave days, including recurrences, until full RTW measured at 12 months	- Not measured in the short term
Psychiatric treatment with adjuvant occupational therapy (Schene, 2007)	1. Proportion working at least 2 days or 16 hours per week over 42 months 2. Total hours worked between 19 and 42 months	- Not measured in the short term - Study sample included workers who were working and not working at baseline. Unclear whether evidence based on these outcomes is strictly applicable to the management of work disability/sickness absence. Therefore, data are also presented in Tables **14b and 15b** under prevention of work disability/sickness absence
Quality improvement program for improved access to medication in primary care (Schoenbaum, 2001)	1. Days of employment over 24 months, calculated as number of days worked over each 6-month follow-up	- Not measured in the short term - Study sample included workers who were working and not working at baseline. Unclear whether evidence based on this study outcome is strictly applicable to the management of work disability/sickness absence. Therefore, this data is also presented in **Table 15b** under prevention of work disability/sickness absence
Enhanced role of occupational physician with psychiatric consultations (van der Feltz-Cornelis, 2010)	1. Time until full RTW measured at 12 months	- Not measured in the short term

[†]Compared to usual care, unless otherwise specified
[*]In alphabetical order by first author's last name

Table 17a: Management of work disability/sickness absence - Among workers currently on work disability leave/sickness absence, which interventions for depression *significantly prevent* the transition from short-term to long-term work disability or the transition from sickness absence to work disability?†

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit*	Final GRADE
In the short term							
None							Not Applicable
In the long term							
Enhanced disability management consisting of an additional collaborative mental health program delivered by psychiatrists to workers on short-term disability leave for psychiatric disorders[76] (Dewa, 2009)	Non-Randomized Study	High	Not Applicable	Population: yes[77] Outcome: yes[78]	Not Precise[79]	Yes[80]	⊙○○○
	Initial GRADE: Low ⊙⊙○○	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: +1	

†Compared to usual care, unless otherwise specified
*Only assessed in studies with economic analyses

Table 17b: Management of work disability/sickness absence - Among workers currently on work disability leave/sickness absence, which interventions for depression *are not different from control* in preventing the transition from short-term to long-term work disability or the transition from sickness absence to work disability?

Intervention	Specific Study Outcome(s)	Comments
In the short term		
None		
In the long term		
None		

[76] Compared to usual disability management practices consisting of treatment by primary care physician and third-part psychiatrist claim adjudication
[77] Employees of a nationwide financial insurance sector company on short-term disability leave for a psychiatric disorder
[78] Transition from short-term to long-term disability (yes/no) measured at 12 months
[79] Evidence from only one study
[80] Compared to the control group, the intervention resulted in fewer workers transitioning from short- to long-term disability and cost $355 less per treated worker, thus from a cost-effectiveness perspective, the intervention dominated

Table 18a: Work functioning - Among workers currently working and not on work disability leave/sickness absence, which interventions for depression *significantly improve* work functioning?†

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
In the short term							
Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)[81]	Randomized Controlled Trial	High	Not Applicable	Population: no[82] Outcome: yes[83]	Not Precise[84]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: 0	
Short-term psychodynamic psychotherapy (Knekt 2008)[85]	Randomized Controlled Trial	High	Not Applicable	Population: no[86] Outcome: yes[87]	Not Precise[88]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: 0	
Telephone screening, outreach and care management (Wang 2007)	Randomized Controlled Trial	High	Not Applicable	Population: yes[89] Outcome: yes[90]	Not Precise[91]	Not Applicable	⊙○○○

[81] Compared to long-term psychotherapy (active control)
[82] Outpatients referred from various psychiatric services, consisting of white collar workers and entrepreneurs
[83] Adequate work ability (modified Work-Ability Index score ≥37)
[84] Evidence from only one study
[85] Compared to long-term psychotherapy (active control)
[86] Outpatients referred from various psychiatric services, white collar workers and entrepreneurs
[87] Self-estimated work ability (using the modified Work-Ability Index) measured at 7 months
[88] Evidence from only one study
[89] Employees from one of 16 companies covered by a managed care program

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: 0	
In the long term							
Long-term psychodynamic psychotherapy (Knekt 2008)[92]	Randomized Controlled Trial	High	Not Applicable	Population: no[93] Outcome: yes[94]	Not Precise[95]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: 0	
Enhanced depression care delivered by primary care physicians and nurses (Lo Sasso 2006)	Randomized Controlled Trial	High	Not Applicable	Population: no[96] Outcome: yes[97]	Not Precise[98]	Yes[99]	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: -1	GRADE Adjustment: +1	

[90] Effective weekly hours worked (hours worked weighted by job performance from the Health Performance Questionnaire) measured at 6 months
[91] Evidence from only one study
[92] For outcomes 1 and 3 listed: compared to short-term psychodynamic psychotherapy. For outcome 2 listed, compared to brief and resource-oriented solution-focused psychotherapy
[93] Outpatients referred from various psychiatric services, consisting of white collar workers and entrepreneurs
[94] Three outcomes, all measured at 36 months: 1) Self-estimated work ability (modified Work-Ability Index) 2) Adequate work ability (modified Work-Ability Index score ≥37), and 3) Work role functioning (Work subscale of SAS-SR)
[95] Evidence from only one study

[96] Patients presenting for routine visits at participating community primary care practices, consisting of professional/administrators, managers/salespeople, clerical/services
[97] Employee's rating of their productivity at work during the previous 2 weeks measured at 24 months
[98] Evidence from only one study
[99] The mean net benefit to the employer of at least $1409 per treated worker in Year One and $5136 per treated worker in Year Two. The return on investment (ROI) over 2 years was at least 302%

Intervention*	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit**	Final GRADE
Telephone screening, outreach and care management (Wang 2007)	Randomized Controlled Trial	High	Not Applicable	Population: yes[100] Outcome: yes[101]	Not Precise[102]	Not Applicable	⊙○○○
	Initial GRADE: High ⊙⊙⊙⊙	GRADE Adjustment: -2	GRADE Adjustment: 0	GRADE Adjustment: 0	GRADE Adjustment: -1	GRADE Adjustment: 0	

†Compared to usual care, unless otherwise specified
*In alphabetical order by first author's last name due to equivalent GRADEs
**Only assessed in studies with economic analyses

[100] Employees from one of 16 companies covered by a managed care program
[101] Effective weekly hours worked (hours worked weighted by job performance from the Health Performance Questionnaire) measured at 12 months
[102] Evidence from only one study

Table 18b: Work functioning - Among workers currently working and not on work disability leave/sickness absence, which interventions for depression *are not different from control* in improving work functioning?†

Intervention*	Specific Study Outcome(s)	Comments
In the short term		
Brief and resource-oriented solution-focused psychotherapy (Knekt, 2008)	1. Self-estimated work ability (using the modified Work-Ability Index) 2. Adequate work ability (modified Work-Ability Index score ≥37) 3. Work role functioning (Work subscale of SAS-SR)	- Compared to long-term psychotherapy (active control) and short-term psychotherapy for outcome 1 - Compared to short-term psychotherapy for outcome 2 - Compared to long-term psychotherapy (active control) and short-term psychotherapy for outcome 3
Short-term psychodynamic psychotherapy (Knekt, 2008)	1. Self-estimated work ability (using the modified Work-Ability Index) 2. Adequate work ability (modified Work-Ability Index score ≥37) 3. Work role functioning (Work subscale of SAS-SR)	- Compared to brief and resource-oriented solution-focused psychotherapy for outcome 1 - Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy for outcome 2 - Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy for outcome 3
Long-term psychodynamic psychotherapy (Knekt, 2008)	1. Self-estimated work ability (using the modified Work-Ability Index) 2. Adequate work ability (modified Work-Ability Index score ≥37) 3. Work role functioning (Work subscale of SAS-SR)	- Compared to brief and resource-oriented solution-focused psychotherapy for outcome 1 - Compared to short-term psychotherapy for outcome 2 - Compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy for outcome 3
Telephone screening, outreach and care management (Wang, 2007)	1. On-the-job performance (measured using the Health Performance Questionnaire)	
In the long term		
Brief and resource-oriented solution-focused psychotherapy (Knekt, 2008)	1. Self-estimated work ability (using the modified Work-Ability Index) 2. Adequate work ability (modified Work-Ability Index score ≥37) 3. Work role functioning (Work subscale of SAS-SR)	- Compared to long-term psychotherapy (active control) and short-term psychotherapy for outcome 1 - Compared to short-term psychotherapy for outcome 2 - Compared to long-term psychotherapy (active control) and short-term psychotherapy for outcome 3

Short-term psychodynamic psychotherapy (Knekt, 2008)	1. Self-estimated work ability (using the modified Work-Ability Index) 2. Adequate work ability (modified Work-Ability Index score ≥37) 3. Work role functioning (Work subscale of SAS-SR)	- Compared to brief and resource-oriented solution-focused psychotherapy for outcome 1 - Compared to long-term psychotherapy (active control) and brief and resource-oriented solution-focused psychotherapy for outcome 2 - Compared to brief and resource-oriented solution-focused psychotherapy for outcome 3
Long-term psychodynamic psychotherapy (Knekt, 2008)	1. Self-estimated work ability (using the modified Work-Ability Index) 2. Adequate work ability (modified Work-Ability Index score ≥37) 3. Work role functioning (Work subscale of SAS-SR)	- Compared to brief and resource-oriented solution-focused psychotherapy for outcome 1 - Compared to short-term psychotherapy for outcome 2 - Compared to brief and resource-oriented solution-focused psychotherapy for outcome 3
Enhanced depression care delivered by primary care physicians and nurses (Lo Sasso, 2006)	1. Employee's rating of their productivity at work during the previous 2 weeks measured at 24 months	- Specific to workers who were inconsistently employed (i.e., not employed at at least one follow-up) - Not measured in the short term
Telephone screening, outreach and care management (Wang, 2007)	1. On-the-job performance (measured using the Health Performance Questionnaire)	

[†]Compared to usual care, unless otherwise specified
[*]In alphabetical order by first author's last name

Table 19a: Recurrence of work disability/sickness absence - Among workers who are currently working, but have had a prior episode of work disability/sickness absence, which interventions for depression *significantly prevent* or reduce recurrences of work disability leave/sickness absence?†

Intervention	Study Design	Risk of Bias	Consistency of Evidence	Directness of Population and Outcome	Precision of Evidence	Economic Benefit*	Final GRADE
In the short term							
None							Not Applicable
In the long term							
None							Not Applicable

†Compared to usual care, unless otherwise specified
*Only assessed in studies with economic analyses

Table 19b: Recurrence of work disability/sickness absence - Among workers who are currently working, but have had a prior episode of work disability/sickness absence, which interventions for depression *are not different from control* to prevent or reduce recurrences of work disability leave/sickness absence?†

Intervention	Specific Study Outcome(s)	Comments
In the short term		
None		
In the long term		
Dutch guideline-based care for employees on sick leave due to mental health problems delivered by occupational physicians (Rebergen, 2009)	1. Number of recurrences of sick leave periods 2. Duration of recurrences of sick leave periods	- Not measured in the short term

†Compared to usual care, unless otherwise specified

3.4.2 Summary of Findings and Key Messages - Primary Outcomes

For each primary outcome category, we found interventions that demonstrated positive findings (i.e., the intervention group was statistically significantly better than the control group at $p<0.05$), negative findings (i.e., the control group was statistically significantly better than the intervention at $p<0.05$), and neutral findings (i.e., not statistically different from the control group, $p\geq0.05$). A detailed description of individual study findings can be found in Appendix H, Table 8. We have also summarized the findings for all interventions evaluated in the included studies in a detailed table in Appendix J.

Using the summary data in Appendix J, key messages for each intervention approach were
extracted using the framework provided in Table 7 (*see Section 2.8.2 of the methods*). We extracted key messages from ten of the 12 studies employing an inactive control group (e.g., usual care). As previously mentioned, we did not draw key messages from the two studies using active control groups (Krogh 2009; Knekt 2008a, Knekt 2008b). Instead, they were used only to inform future research.

Overall, we did not find evidence to recommend any of these intervention approaches because all studies were judged to have a high risk of bias and, therefore, all graded evidence was rated 'very low". In addition, the majority of the interventions demonstrated mixed or contradictory results for the primary outcomes. Furthermore, given the significant policy differences across jurisdictions care should be taken in generalizing from non-Canadian jurisdictions. Consequently, all of the interventions require further investigation in future studies in Ontario and other Canadian jurisdictions.

3.4.2.1 Enhanced primary care

Enhanced primary care involved physicians and nurses working in the primary care centres or managed care organizations. All studies that assessed this type of intervention approach were conducted in the USA. The predominant components of this approach were education for physicians and nurses on guideline-concordant care and reinforcement to adhere to these guidelines.

We found three large trials reported in four main publications (n total=1,944) that assessed this type of intervention approach (Lo Sasso 2006, Rost 2004, Smith 2002, Schoenbaum 2001). Outcomes assessed included measures of work disability (prevention and management) and work functioning, all measured in the long-term. All three studies demonstrated a high risk of bias and were assessed as providing a "very low" grade of evidence. The findings were also mixed (positive and neutral). Therefore, no recommendation could be made for this intervention approach.

In the trial by **Lo Sasso et al** (2006)/Rost (2004), patients presenting for routine visits at participating community primary care practices were recruited. Between 69% and 85% of participants were employed and working full-time at baseline. Depression was diagnosed by interview with administrative staff using the World Health Organization Composite International Diagnostic Interview (WHO-CIDI) and the Inventory to Diagnose Depression (IDD). Eligible participants were then randomized to enhanced care (n= 158) or usual care (n= 168). In the enhanced care group, physicians were informed when patients screened positive for depression. Physicians and nurses were trained in guideline-concordant pharmacotherapy and psychotherapy and nurses contacted patients for assessment, education, adherence to treatment, and follow-up. Physicians received status summaries and reminders to adjust treatment for symptomatic patients.

This study (Rost 2004) showed that enhanced care leads to a significant improvement ($p< 0.05$) in work functioning at work over 24 months, increasing productivity by 6.1%. There was a non-significant ($p= 0.06$) reduction in absenteeism over 24 months by 22.8% or 10.6 days in the enhanced care group. The authors of this trial analyzed subgroups of the population sample according to their employment status. Among the consistently employed individuals, enhanced care significantly improved work functioning compared to the usual care group ($p= 0.03$), leading to an 8.2% increase in productivity over 24 months and led to a non-significant 28% reduction in absenteeism ($p= 0.08$). However, there was no significant impact on work functioning ($p= 0.99$) or absenteeism ($p= 0.64$) among those who were inconsistently employed ($p= 0.99$).

The study by **Smith et al** (2002) was based on the same trial as that described above by Lo Sasso et al. (2006)/Rost (2004). However, in this particular analysis, data were only analyzed among participants who were working at baseline (n= 129 enhanced care, n= 133 usual care). In the long term (by 12 months), this study demonstrated that enhanced care significantly prevented work disability, with the enhanced care group reporting a significantly higher proportion of employed individuals (92.1%) at follow-up than participants in the usual care group (82%) ($p= 0.04$). In the short-term (at six months), however, this relationship was not apparent.

In the trial by **Schoenbaum et al** (2001) patients presenting to primary care clinics in community-based managed care organizations were recruited. The proportion of individuals at baseline that were working was not reported. Depression was identified using the WHO-CIDI. Eligible patients were randomized to a quality improvement program for improved access to medication (QI Meds, n= 424), to a quality improvement program for improved psychotherapy (QI Therapy, n= 489), or to usual care (n= 443). In

the QI Meds group, nurse specialists presented antidepressant medications and psychotherapy as equally effective treatments to the patient. The nurse contacted the patients monthly and assisted with management of antidepressant medications and support of adherence. In the QI Therapy group, the primary care clinician formulated a treatment plan, and if the clinician determined that psychotherapy was appropriate, patients were referred to CBT-trained therapists. Medication was available, but no medication management was provided.

This study showed that the QI Therapy group had an average of 20.9 more employed days than the control group over the 24 months period (p= 0.03). This may represent both prevention of work disability or management of work disability, because the groups included both employed and non-employed patients at baseline. The QI Meds group also had more employed days (17.9 days) than the control group, but this was non-significant (p= 0.07). Among individuals who were working, QI Therapy and QI Meds combined had no impact on the number of sick days over 24 months.

Key Message: No recommendation can be made for **enhanced primary care delivered by physician and nurse** because 1) the grade of the evidence to support this intervention is "very low" and 2) there were mixed findings (positive and neutral). We recommend that more research is conducted for this intervention.

Information relevant for future studies: There is suggestive evidence from one study (Lo Sasso 2006) that enhanced primary care delivered by physician and nurse is more effective among individuals who are consistently employed and when the enhanced care involves improved psychotherapy (Schoenbaum 2001). All three studies took place in the USA. There is a need to verify if this intervention is effective in other countries which have a different compensation system, specifically in Ontario.

3.4.2.2 Enhanced psychiatric care

This intervention approach involved out-patient psychiatric treatment enhanced by occupational therapy.

We found one small trial (n total= 62) conducted in the Netherlands that assessed the addition of occupational therapy to psychiatric care (with antidepressants, if indicated) (**Schene et al.**) (2007). This study demonstrated a high risk of bias and provided a "very low" grade of evidence. The findings were also mixed (positive and neutral). Therefore, no recommendation could be made for this intervention approach. Participants were recruited from an outpatient mood disorder clinic and diagnosis of depression was made by interview (by psychiatrist and trained staff) using the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) and

the Beck Depression Inventory (BDI). At baseline, between 15.6% and 23.3% of the participants were working and not on sickness absence. Eligible participants were randomized to the enhanced group of psychiatric care plus occupational therapy (n=30) or to the usual care with psychiatrist (n= 32). Occupational therapy included three phases: diagnostic, therapeutic phase and follow-up phase, including contact with the patient's employer and a plan for work reintegration.

This study found that among patients not working at baseline, time until any work resumption over the long term was significantly shorter for those in the enhanced psychiatric care group (207 days) than among those in the usual psychiatric care group (299 days) (p= 0.01). This study also demonstrated that participants in the enhanced care group worked significantly more hours over the short term, as well as over the long term, although only up to 18 months. Beyond this time, there was no relationship between the intervention and total hours worked. This finding may suggest evidence for both prevention and management of sickness absence due to the fact that the intervention groups included individuals who were working, as well as not working at baseline. On the other hand, the enhanced psychiatric care group was equivalent to usual psychiatric care when examining the outcome of proportion working at least two days or 16 hours per week over the long term.

Key Message: No recommendation can be made for **enhanced psychiatric care with occupational therapy** because 1) the grade of the evidence to support this intervention is "very low" and 2) there were mixed findings (positive and neutral). We recommend that more research is conducted for this intervention.

Information relevant for future studies: There was only one study found to examine this intervention and was conducted in The Netherlands. There is a need to verify if this intervention is effective in other countries that have a different compensation system, specifically Ontario.

3.4.2.3 Enhanced occupational physician role

The studies in this category evaluated an intervention approach that was aimed at establishing a more active role for the occupational physician (OP) in the management of work disability and in the prevention of work disability recurrences.

We found two trials (n total= 300) conducted in the occupational health setting in The Netherlands (van der Feltz-Cornelis 2010, Rebergen 2009a, 2009b). A decision was made to report these two studies together, but it is acknowledged that important differences exist in the way these interventions were conducted. In the study by van der Feltz-Cornelis, the

intervention involved OPs being trained in the diagnosis and treatment of depressive disorders with the addition of supportive psychiatric consultation. In the study by Rebergen et al., OPs were trained according to the recommendations from the Dutch guidelines for management of common mental health problems by OPs, which was introduced in The Netherlands in 2000. The guideline promotes the role of the OP as case and care manager in sickness absence management (i.e., as a RTW facilitator). The guidelines recommend that the OP and employee should meet regularly as long as the employee has not fully returned to work. After a full RTW, at least one meeting should take place to focus on relapse prevention.

Neither trial demonstrated better outcomes related to the management or recurrence of sickness absence, except for a higher proportion of workers in the intervention group who had more partial RTW before full RTW in one study (Rebergen 2009a). However, both trials demonstrated a high risk of bias and, therefore, were assessed as providing a “very low” grade of evidence. The findings were also mixed (positive and neutral). Therefore, no recommendation could be made for this intervention approach. Below we describe each trial in more detail.

In the cluster-randomized controlled trial conducted in collaboration with two occupational health services by **van der Feltz-Cornelis et al.** (2010), employees on sick leave from various companies were recruited following the allocation of the OP. Depression was diagnosed by self-administered questionnaire using the Patient Health Questionnaire (PHQ)-depression subscale. Eligible patients were randomized to a group where the OPs received training in diagnosis and treatment of depressive disorders, with the addition of supportive psychiatric consultation whereby OPs were provided advice on managing employees on sick leave for common mental health disorders (n= 29), or to a group where they received care as usual from the OP (n= 31).

Over the short term, this study demonstrated that workers in the enhanced group (58%) were significantly more likely than workers in the usual OP care group (44%) to make a full RTW by three months (p= 0.0093). By six months, however, there was no longer a significant difference in the proportion of workers who made a full RTW (85% in the enhanced group and 84% in the control group) (p= 0.0574). There was also no significant difference between groups in the time taken until a full RTW was made (122 days in the enhanced group, 190 days in the control group, p= 0.078).

The trial by **Rebergen et al.** (2009a) included employees from police departments on sick leave for mental health problems in the Netherlands. Depression was diagnosed by self-administered questionnaire using the Depression Anxiety Stress Scale (DASS) and the Hospital Anxiety

Depression Scale (HADS). Eligible participants were randomized to a group where the OPs received training on the Dutch guideline which promotes a more active role of the OP as a case and care manager facilitating RTW (n= 125) or to a usual care group where there was minimal involvement of the OP and access to treatment by a psychologist (n= 115).

This study showed that those in the intervention group were more likely to experience a gradual RTW process. In The Netherlands, a gradual RTW (i.e., partial RTW) can be part of the RTW process leading towards a full RTW. This study showed that, over the long term, significantly more people had a partial RTW (69%) before full RTW compared to the control group (54%) (p= 0.01). However, there were no significant differences between the two groups with respect to the management of work disability/sickness absence outcome of duration of sick leave until partial or full RTW or for recurrence of work disability (number and duration of recurrences of sick leave periods). Finally, the mean number of days of total productivity loss (duration of sick leave days, including recurrences, until full RTW) was not significantly different between the groups (p= 0.28).

Key Message: No recommendation can be made for **enhanced occupational physician role** because 1) the grade of the evidence is "very low" and 2) there were mixed findings (positive and neutral). We recommend that more research is conducted for this intervention.

Information relevant for future studies: One study (Rebergen 2009a) found that type of work was a significant effect modifier of total productivity loss, with workers in administrative functions benefiting more from the intervention than workers with executive functions. The severity of the disorder also showed a significant interaction effect in this study: workers with a more severe depressive or anxiety state had better outcomes in the intervention group compared to usual care by an occupational physician. The two studies examining this intervention approach took place in The Netherlands, therefore, there is a need to verify if this intervention is effective in other countries which have a different compensation system, specifically Ontario.

3.4.2.4 Psychological interventions

The interventions in this category involved psychological treatments, such as cognitive behavioural therapy (CBT) and psychotherapy, which are normally delivered by psychologists or psychotherapists. However, in one trial, one of the interventions (brief CBT-based stress management with a focus on improving workplaces processes) was delivered by "labour experts" (Blonk et al 2006).

We found two trials (n total= 448) that assessed the effectiveness of CBT therapies and psychotherapy. One trial was conducted in The Netherlands (Blonk 2006) and the other in Finland (Knekt 2008a). They measured sickness absence (prevention and management) and work functioning outcomes. One study showed that CBT combined with a workplace-focused technique promotes full RTW approximately 200 days earlier than CBT alone or an “inactive” encounter with a general practitioner (Blonk 2006). The other study (Knekt 2008a) showed that shorter psychotherapies (between five and eight months duration) improve work functioning more quickly than longer therapies (duration up to three years). But in the long-term, longer psychotherapies (duration up to three years) are more effective than shorter psychotherapies. However, both trials demonstrated a high risk of bias and, therefore, were assessed as providing a “very low” grade of evidence. The findings were also mixed (positive and neutral; negative and neutral). Therefore, no recommendation could be made for this intervention approach. Below we describe each trial in details.

The trial by **Blonk et al.** (2006) was conducted in The Netherlands with self-employed individuals on sickness absence. Depression was diagnosed via a telephone interview by a psychologist using the WHO-CIDI and a self-administered questionnaire using the Depression Anxiety Stress Scale (DASS). Eligible participants were randomized into one of two intervention groups or the control group. The first intervention consisted of brief CBT-based stress management program administered by labour experts combined with a workplace-focused technique consisting of psycho-education on work stress, registration of symptoms and situations, relaxation, self-help books, time-management, and writing and homework assignments. The labour expert also provided advice about work processes and provided suggestions on how to lower workload and job demands and to increase decision latitude. They also encouraged partial RTW (n= 40). The second intervention consisted of CBT given by psychologist who followed a highly structured protocol (n= 40). The control group consisted of two brief sessions with a general practitioner whose role was to check the validity of the work disability claim, with no actual treatment (n= 42).

This study found that the combined CBT intervention promoted full RTW a median of 207 days earlier than the CBT alone and 198 days earlier than the control group ($p < 0.01$ for both unadjusted and adjusted analyses). There was no difference between CBT alone and the control group for time until full RTW. With regards to partial RTW, the combined CBT intervention was better than CBT alone (median 17 days earlier) and control group (median 30 days earlier), but these differences were not significant in the adjusted analyses. There was no difference between CBT alone and the control group for partial RTW.

The trial by **Knekt et al.** (2008a) was conducted in Finland, where they recruited outpatients referred from various psychiatric services. The majority (between 75% and 85%) were employed or students. Depression was diagnosed by interview (not reported by whom) using the DSM-IV, 17-item Hamilton Rating Scale for Depression (HAM-D17) and BDI. Eligible participants were randomized to three active treatment groups: a) solution-focused therapy, brief and resource-oriented, conducted by a therapist for one session every second or third week to a limit of 12 sessions for up to eight months (n= 97); or b) short-term psychodynamic psychotherapy conducted by a therapist using a brief transference-based approach which helps patients by exploring and working through intra-psychic and interpersonal conflicts for 20 weekly treatment sessions for a duration of 5-6 months (n= 101); or c) long-term psychodynamic psychotherapy conducted by a therapist which offered an open-ended, intensive, transference-based therapeutic approach which helps patients by exploring and working through a broad area of intra-psychic and interpersonal conflicts for 2-3 times a week for up to three years (n= 128).

This study demonstrated that, over the long-term and among employed workers, both short-term psychotherapy and long-term psychotherapy resulted in a significantly lower proportion of workers taking more than seven sick leave days in the previous three months, than workers in solution-focused psychotherapy ($p< 0.05$). However, no differences in this outcome were demonstrated over the short term, nor were there significant group differences in the number of sick leave days taken during the previous three months over the short or long term. Finally, no group differences were seen over the short or long term in current employment status. Given the fact that participants were a mix of employed and non-employed at baseline, the latter may reflect a lack of evidence for both prevention and management of work disability/sickness absence.

With regards to work functioning, this latter study showed that the two shorter therapies (short-term psychotherapy and solution-focused therapy) improve work functioning in the short term, but not in the long-term. It also showed that long-term psychotherapy improves work functioning in the long-term, but not in the short-term. In each case, the findings varied according to which work functioning outcome was being analyzed.

Key Message: Based on the one study to use an inactive control (**Blonk 2006**), no recommendation can be made for **psychological interventions** because 1) the grade of the evidence to support this intervention is “very low” and 2) there were mixed (positive and neutral; negative and neutral) findings. We recommend that more research is conducted for this intervention.

Information relevant for future studies: In one study (Knekt 2008a), psychotherapy of shorter duration (between five and eight months) seems to have a greater impact on work functioning in the short-term, but not in the long-term, while psychotherapy of longer duration (up to three years) seems to have a longer impact on work functioning outcomes that are not visible when measured in the short-term. These studies were conducted in Europe (Finland and The Netherlands). There is a need to verify if this intervention is effective in other countries which have different compensation systems, specifically Ontario.

3.4.2.5 Exercise

We found one trial (n total= 165), with high risk of bias and very low grade of evidence conducted in Denmark, which assessed three different types of exercises (strength, aerobic and relaxation training) without an "inactive" control group (**Krogh et al.** 2009). Patients were referred from general practitioners, private practicing psychiatrists, psychologists or psychiatric wards. Depression was diagnosed by interview with a psychologist or research assistant using the Major Depression Inventory (MDI), HAM-D17, Montgomery-Asberg Depression rating Scale and the BDI. At baseline, the proportion of participants who were employed varied from 45% to 64%, and the proportion of participants not on sickness leave ranged from 47% to 58%. Eligible participants could be randomized to one of the three exercise training groups: strength training (n= 55), aerobic training (n= 55) or relaxation training (n= 55). This intervention involved participation in twice a week, group exercises, for a total of 32 sessions in a four-month period. All sessions lasted 1.5 hours and were supervised by a physiotherapist experienced in instructing psychiatric patients.

This study showed that participants in the strength training group had a significantly lower mean percentage in "days absent from work" at 12 months (-12.1 days, 95%CI -21.1 to -3.1, p=0.009) compared to participants in the relaxation training group. The sample at baseline included a mix of individuals who were employed/not employed and on sick leave/not on sick leave, therefore, this finding could apply to either prevention or management of sickness absence. No significant differences were noted at measures taken at four months or between aerobic training and relaxation training at either four or 12 months. There were also no significant differences between any of the three groups on unemployment or sick leave at four or 12 months.

Key Message: No key messages can be derived for this intervention because the only trial found did not have an inactive control group.

3.4.2.6 Worksite-based stress reduction intervention

We found one non-randomized study (n total= 285), with high risk of bias and a very low grade of evidence conducted in Japan that assessed the effectiveness of a worksite-based stress reduction program compared to no intervention (**Kawakami et al** 1997). Blue collar workers were recruited from a large electric company. At baseline, 52% of the participants in the intervention group and 33% in the control group reported 1-5 days of sick leave in the past year; and the proportion who reported zero days of sick leave was 40% and 53% respectively. Depression was diagnosed by self-administered questionnaire using the Zung Self-Rating Depression Scale. The control worksites were matched for selection on mean age, major products and occupations, worksite size, and mean depression scores. The intervention group received the worksite stress reduction program (n= 110), and the control group received no activities for reducing work stress (n= 175). The intervention consisted of a stress reduction program in which supervisors were asked to list possible work stressors in their worksites and to make plans to reduce these stressors while a working committee made the plans feasible. The supervisors started stress reduction activities and the committee monitored their activity periodically.

This study found that the worksite stress reduction program reduced sick leave measured at two years follow-up: sick leave (measured as 1-5 days of sick leave in the last year) was reduced to 34% in the intervention group and it remained stable (37%) in the control group. More participants reported zero days of sick leave in the intervention group (61%) than in the control group (58%) at two years (group x time interaction: p= 0.034).

Key Message: Despite the consistency in findings from the one study examining this intervention approach, no recommendation can be made for **worksite stress reduction program** because the grade of the evidence to support this intervention is “very low”. We recommend that more research be conducted for this intervention.

Information relevant for future studies: This non-randomized trial did not adjust for important baseline imbalances that were observed. This intervention needs to be confirmed in a randomized trial design with appropriate statistical analysis. This study was conducted in Japan, therefore, there is a need to verify if this intervention is effective in other jurisdictions, specifically Ontario.

3.4.2.7 Systems integration and care management

Systems integration and care management interventions refer to interventions conducted at the organizational or health-care system level. In

this systematic review, the interventions were aimed at appropriate diagnosis, adherence to treatment, adequate follow-up, and ensuring collaboration among all individuals involved in the care management of workers with depression.

We found one randomized controlled trial (**Wang et al.** 2007) and one non-randomized trial (**Dewa et al.** 2009) which assessed the effectiveness of some kind of integrative approach at the health-care system level to prevent and manage work disability and to improve work functioning in a population with depression (n total= 728). The trial by Wang et al. was conducted in the USA and the non-randomized trial by Dewa et al. was conducted in Ontario, Canada and both were judged to have a high risk of bias. The randomized trial found that the telephone outreach intervention was better than the control group in preventing work disability and improving work functioning in both short and long term. The non-randomized study found that the collaborative mental health program was better than the control group to manage work disability. However, both studies demonstrated a high risk of bias and, therefore, were assessed as providing a “very low” grade of evidence. The findings were also mixed (positive and neutral; negative and neutral). Therefore, no recommendation could be made for this intervention approach. Below we describe each study in detail.

The trial by **Wang et al.** (2007) recruited participants from 16 companies covered by a specific managed behavioural health care company. At baseline, 100% of the participants were working and not on work disability. Depression was diagnosed in two phases: first by a self-administered questionnaire (K-6 psychological distress screen) and second by a telephone interview using the Quick Inventory of Depression Symptoms Self-Report (QIDS-SR). Eligible participants were randomized to a “telephone screening, outreach, and care management program” which assessed the need for treatment, facilitated entry into in-person treatment and medication use, as necessary, with monitoring and support for treatment adherence. All participants received a psycho-educational workbook. For those declining in-person treatment, care managers maintained regular telephone contacts and, if experiencing significant symptoms after two months, they were provided with a structured psychotherapy intervention by telephone (n= 304). The control group received usual care normally available through the insurance benefit or service but not the additional telephone care management components (n= 300).

This study found that a telephone outreach intervention was better than the control group to prevent work disability (measured by job retention) in the short-term (six months), but not in the long-term (12 months). At six months, the intervention group had significantly (p= 0.007) higher job retention (96%)

than the control group (90%). At 12 months, the intervention group continued to demonstrate higher job retention (93%) compared to the control group (88%), but this difference was non-significant (p= 0.07). There were significant differences between groups in “effective weekly hours worked” both at the short and long-term, with individuals in the telephone outreach group demonstrating a higher number of effective weekly hours worked at both six and 12 months. However, there were no significant differences between the “actual hours worked” between groups in any of the time points. With regards to work functioning, there was no difference between groups in any time point in on-the-job performance scores measured by the World Health Organization Health and Work Performance Questionnaire (HPQ).

The study by **Dewa et al.** (2009) recruited subjects from a large insurance service company in Canada who were on work disability (short-term disability) related to a psychiatric disorder. There is no description on how depression was diagnosed. All data was drawn from company administrative data routinely collected for people who received short-term disability benefits. A quasi-experimental design was used to form two groups: participants referred to the Collaborative Mental Health Program (CMHP) or a comparison group composed of subjects who were on short-term disability leave in the year before the program implementation and who would have met the screening criteria for the program. The CMHP structure was based on a collaborative care concept, including psychiatric assessment and treatment recommendation, short-term management by a psychiatrist (if referred by a primary care physician), psychiatric support for management by the primary care physician, and the availability of psychiatric consultation for non-referred workers. The goal of this collaborative program was to return patient’s care to the primary care physician as soon as possible (n= 73). The control group consisted of usual practice where individuals were referred to a third-party psychiatrist for the purpose of adjudicating the claim only. Once the diagnosis and severity of the disorder were established, the primary care physician continued to treat the employee until they returned to work. The employee was referred to a psychiatrist at the discretion of the primary care physician (n= 51).

This non-randomized study showed that the collaborative program significantly improved management of work disability: at 12 months, RTW was achieved in 85% of program’s participants compared to 63% in the control group (p= 0.005); the average number of days on short-term disability leave was significantly shorter for the program’s participants (62 days) than the control group (76 days) (p= 0.03); and transition to long-term disability was significantly lower among the program’s participants (7%) than the control group (31%) (p< 0.001).

Key Message: No recommendation can be made for **systems integration and care management** because 1) the grade of the evidence to support this intervention is "very low" and 2) there were mixed findings (positive and equal). We recommend that more research is conducted for this intervention.

Information relevant for future studies: Screening, outreach, and a care management program can be delivered by telephone (**Wang et al. 2007**). There is a need to confirm the findings of the collaborative program study in randomized trial design. One study took place in the USA and the other in Canada, therefore, there is a need to verify if this intervention is effective in other countries which have a different compensation system.

3.4.3 Summary of Findings and Key Messages - Economic Results

Five studies in this review (Dewa 2009, Lo Sasso 2006, Schoenbaum 2001, Rebergen 2009b, Schene 2007) measured the economic outcomes of the intervention compared to the comparison intervention. A detailed description of the economic findings of these individual studies can be found in Appendix H, Tables 6 and 7.

Appraising the strength of the evidence from the economic evaluations involves many of the same considerations as appraisal of the main study outcomes – for example, the risk of bias analyses discussed elsewhere in the results are relevant when appraising economic outcomes. However, the economic outcomes of a study may diverge from clinical or other outcomes. For example in the **Schene et al.** (2007) study, the addition of occupational therapy to psychiatric care when compared to psychiatric care alone did not improve depression outcomes, but from a societal perspective did result in a net economic benefit. Economic evaluations must also be appraised in the labour and health system context in which they occurred. As these contexts vary significantly between countries, the generalizability of the results across countries may be constrained.

The perspective adopted by the study determined what costs and benefits were included, and how the benefits were valued. Two studies adopted an employer perspective (Lo Sasso 2006, Dewa 2009), two studies adopted the societal perspective (Schoenbaum 2001, Schene 2007) and one study included both employer and societal perspectives (Rebergen 2009b). The two studies that adopted an employer perspective (Lo Sasso 2006, Dewa 2009) conducted a cost-benefit analysis that included employer-borne costs and a monetized value of the benefits realized by the employer. The study that provided an economic evaluation from both the employer and societal perspective (Rebergen 2009b) also included a cost-benefit analysis from the societal perspective. The remaining two studies that adopted the

societal perspective (Schoenbaum 2001, Schene 2007) conducted a cost-effectiveness analysis, with the primary outcome expressed as the difference in cost (between the intervention and comparison interventions) divided by the difference in outcome, where the outcome was not expressed in monetary units.

The economic evaluation conducted by **Lo Sasso et al** (2006) was a cost-benefit analysis from the employer perspective. Treatment costs (borne by the employer) for the enhanced care intervention were $735 in Year 1 and $353 in Year 2, representing an additional $158 and $130 of costs compared to the comparison treatment in Year 1 and Year 2 respectively. However, the value of the benefits realized by the employer exceeded the intervention's costs and the net economic benefit realized was at least $1,409 per treated worker in Year 1 and $5,136 per treated worker in Year 2. The evaluation includes sensitivity analyses to cover situations where the cost of a worker's absence (or reduced on-the-job performance) can result in costs to the firm greater than the worker's full wage, such as team production, expensive substitute labour, and penalties for output shortfalls resulting from absences/impaired productivity. The analyses also include scenarios where the costs of the enhanced depression care may be higher (due to including spouses of employees in the benefit), or where the benefits of the intervention may be lower (due to turnover of workers that received the benefit). In each of these scenarios the net benefit was retained. The extensive sensitivity analyses included in the study widen the relevance of the findings to firms with diverse production processes, benefits arrangements, and employer turnover rates, and the study exemplifies methods of economic evaluation that increase the scope and relevance of the findings.

The economic analysis reported by **Schoenbaum et al** (2001) also examined enhanced primary care, but was a cost-effectiveness analysis conducted from the societal perspective. Total health care costs for usual care averaged US$3,835 with an additional cost for the quality improvement intervention involving medications (QI-Meds) of US$419 and an additional cost for the quality improvement intervention involving psychotherapy (QI-Therapy) of US$485. QALYs were estimated by two methods, using a health utility index derived from the Short-Form 12-item health survey, and from a measure of depression burden days assigned utility scores. The incremental cost per QALY fell between $15,331 and $36,467 for QI-Meds, while incremental cost per QALY for the QI-Therapy intervention had a more favourable range between $9,478 and $21,478. These incremental cost-effectiveness ratios fall within the range of many accepted medical interventions. However, the economic implications of greatest relevance to employers were not reported, as the employer perspective was not adopted,

and the costs of workers' absence (or reduced on-the-job performance) were not included in the cost-effectiveness calculations.

The study of enhanced depression care delivered by occupational physicians reported by **Rebergen et al.** (2009b) included two types of economic evaluation: a cost-effectiveness analysis from the societal perspective, and a cost-benefit analysis from the employer perspective. The mean cost of occupational health care was higher in the enhanced care group compared to usual care, but the costs of psychological treatment were significantly greater in the usual care group. Total health care costs (including routine health care, psychological care, and the costs of the enhanced occupational physician care) averaged 2,665 Euros per treated case in the usual care intervention, 520 Euros higher than the mean treated case cost in the enhanced care group, with most of the increase attributable to greater use of psychological services in usual care. Although worker absences and productivity loss costs did not differ between the study groups, the incremental cost-effectiveness ratio was estimated to be -736 Euros per worker absent day. In the cost-benefit analysis from the employer perspective, a mean net benefit of 3582 Euros over one year was realized for the enhanced occupational physician intervention, based upon valuing lost productivity days at the mean daily wage of 125 Euros over the entire period of absence. The magnitude of the mean net benefit was reduced in alternate analyses that assumed that productivity losses due to absence days were restored after a maximum friction cost period of 154 days. The results of the cost-benefit analysis and the cost-effectiveness analysis in this study were driven by the higher costs for psychological services in the usual care group, who were described as having "easy access" to a psychologist. If access to psychological services between groups was not equal, the cost-effectiveness and cost-benefit results of the study may be attributable to this inequity and not to the effects of the occupational physician intervention.

The small trial reported by **Schene et al** (2007) included a cost-benefit analysis from the societal perspective. The average total costs (including health care, occupational therapy, medications, parking, and travel) of $3,149 dollars in the intervention group exceeded the costs of usual care, which were US$1,891. The intervention resulted in more hours worked compared to the treatment as usual group. Valuing the additional hours worked by the average Dutch wage of US$36.88 resulted in a mean net benefit (earnings minus costs) of US$3,952. The mean net benefit rose to US$5,370 with the exclusion of an outlier in the intervention group.

Finally, the Canadian study of **Dewa et al** (2009) was a cost-benefit and cost-effectiveness analysis of a collaborative mental health program (compared to usual care and disability management) for persons receiving

short-term disability benefits related to psychiatric disorders. For the study intervention, costs from the employer's perspective were limited to those services provided by the collaborative mental health program that were not covered under the public health care system. For the comparison group, the major cost covered by the employer was the cost of independent medical examinations conducted by an independent psychiatrist. The study found that the employer-borne costs per worker in the intervention group were $355 lower than the usual care group or $503 when adjusted for age differences. As the intervention group also experienced shorter average disability durations, from a cost-benefit perspective the mean net benefit of the intervention was $355 (or $503 adjusted) even when the study authors assigned no value to the reduced number of disability days. From a cost-effectiveness perspective, the intervention group dominated (lower costs, better outcomes) the control group.

Key message: From the economic evaluations conducted, three interventions showed evidence of a net economic benefit to the employer: enhanced primary care, enhanced occupational physician role, and system integration and care management.

From the economic evaluations conducted, three interventions showed evidence from the societal perspective of either an economic net benefit or cost-effectiveness: enhanced primary care, enhanced occupational physician role, and psychiatric care enhanced by occupational therapy.

Information relevant for future studies: Studies conducting economic evaluations improve the relevance of the evaluations to employers by including cost-benefit analyses from the employer perspective. These analyses should include costs borne by the employer, quantitative measures of workplace disability, and estimates of the monetized value of reductions in workplace disability from an employer perspective.

The relevance of cost-benefit analyses from the employer perspective is enhanced by including sensitivity analyses that reflect features of employment and benefits, as demonstrated in the study of Lo Sasso (2006). Sensitivity analyses that are often relevant include modeling characteristics of production (team production, costly substitute labour, penalties for production shortfalls) and high worker turnover rates, which when present reduce the value to the employer over time of reductions in workplace disability.

In summary, the following are interventions to be considered **in future research:**

- Enhanced Primary Care
- Enhanced Psychiatric Care
- Enhanced Role for the Occupational Physician
- Psychological Interventions
- Worksite Stress Reduction
- Systems Integration and Care Management

3.5 Results for the secondary outcomes

3.5.1 Summary of Findings – Secondary Outcomes

With respect to secondary outcomes, we found interventions that were positive (i.e., the intervention group was statistically significantly better than the control group), interventions that were negative (i.e., the control group was statistically significantly better than the intervention), and interventions that were not different from the control group. A detailed description of the individual study findings pertaining to secondary outcomes can be found in Appendix H, Table 8. We have also summarized the findings for all interventions evaluated in the included studies in a detailed Appendix K for the following secondary outcome categories: improved psychosocial work outcomes, reduction in critical workplace incidents, reduction in depression symptom severity, and depression remission.

3.5.1.1 Enhanced Primary Care

Two trials (Rost 2004; Smith 2002), both based on the same intervention approach of enhanced depression care delivered by primary physicians and nurses, examined the secondary outcomes improvement in psychosocial work outcomes (Smith 2002) and reduction in depression symptom severity (Rost 2004). In the trial by Smith et al., individuals in the enhanced care group were significantly less likely than those in the usual care group over the long term to report having a workplace conflict in the past year. In the Lo Sasso study, findings regarding the effect of enhanced care on reduction in depression symptom severity were somewhat conflicting. Using the CES-D, symptom reduction was found to be significantly more likely to occur in the enhanced care group than the usual care group over the long term, but only among consistently employed individuals. Among those were inconsistently employed, there was no association demonstrated.

3.5.1.2 Enhanced Psychiatric Care

The only trial to evaluate enhanced psychiatric care (Schene 2007) examined the following key secondary outcomes: improvement in psychosocial work outcomes, reduction in depression symptom severity, and depression remission. In the both the short and long term, psychiatric treatment with adjuvant occupational therapy was no different than usual psychiatric care in improving work stress or leading to a depression remission. Similarly, in the short term, individuals receiving this intervention approach were no different than those receiving usual care in depression symptom severity and, over the long term, actually fared significantly worse than usual care in symptom severity.

3.5.1.3 Enhanced Role for the Occupational Physician

Only one of the two trials in this category (van der Feltz-Cornelis 2010) evaluated the impact of the enhanced OP role on the secondary outcome of reduction in depression symptom severity, finding that the provision of a consultation by a psychiatrist providing treatment advice to the occupational physician for employees on sick leave was no different than usual care in terms of reducing depression symptom severity over the long term.

3.5.1.4 Psychological interventions

The two trials falling into this category (Blonk 2006, Knekt 2008b) both looked at reduction in depression symptom severity as an outcome. In the **Blonk et al.** (2006) study, both methods of CBT (the brief CBT-based stress management delivered by labour experts and the extensive CBT delivered by psychotherapists) were equivalent to usual care in the reduction of depression symptom severity. In the **Knekt et al.** (2008b) study, long-term psychotherapy was significantly less likely to result in a reduction in symptom severity than short-term psychotherapy or solution-focused psychotherapy. However, over the long term, the opposite was true, with individuals receiving long-term psychotherapy more likely to experience a reduction in symptom severity than those receiving one of the short-term therapies. Similar findings were seen for depression remission in the short-term. However, over the long-term, the three groups were generally equivalent in their effects on depression remission. It should be noted that these findings varied somewhat, depending on the instrument used to assess depression severity and remission. See Appendix K for further details.

3.5.1.5 Exercise

The one trial (Krogh 2009) to evaluate three types of physical activity interventions (strength training, aerobic training, and relaxation training) also examined the effects of these activities on depression symptom severity and depression remission over both the short and long term. Compared to

relaxation training, both strength and aerobic training did not significantly reduce depression symptom severity or result in depression remission.

3.5.1.6 Stress Reduction

Worksites receiving a worksite-wide stress reduction program (Kawakami et al. 1997) were not significantly different than worksites without this program with respect to improvement in workplace stressors of overtime, overload, little chance to learn new knowledge, lack of control over workplace, problems with supervisor, or problems with co-workers over the long term. However, workers within the intervention group were significantly more likely to demonstrate a reduction in depression symptom severity over the long term.

3.5.1.7 Systems Integration and Care Management

One trial within this category (Wang 2007) found that a telephonic outreach and care managed program encouraging workers to participate in outpatient care for depression significantly reduced depression scores and was significantly more likely to lead to depression remission when compared to usual care. However, it had no significant effect on the reduction of critical workplace incidents.

4.0 Discussion

4.1 Summary of Key Findings

Our systematic review was designed to answer the question: "which intervention approaches to manage depression in the workplace have been successful and yielded value for employers in developed economies." We included ten randomized trials and two non-randomized studies from various countries and jurisdictions that evaluated a wide range of intervention practices to manage the impact of mild to moderate depression in the workplace. With feedback from our stakeholders, we developed a framework that combined the type of population and intervention to answer the overall research question.

- First, we reviewed interventions aimed at the prevention of work disability/sickness absence among workers with mild to moderate depression who are currently working and not on a work disability leave/sickness absence. The goal of these interventions are to maintain people at work, by reducing sick leave days used to cope with their depression and preventing a worker from taking a disability leave/sickness absence leave
- Second, we reviewed interventions aimed at returning people to work following a work disability leave/sickness absence. These were usually interventions to promote or hasten return to work, to prevent the transition from short-term to long-term work disability leave, or to prevent the transition from sickness absence to work disability.
- Third, for a population with mild to moderate depression still working, or who have just returned to work after a period of work disability leave/sickness absence, we reviewed the intervention approaches to improve work functioning by means of measuring performance or productivity.
- Last, we reviewed approaches to maintain workers at work after they had returned to work due to a mild to moderate depression episode, as measured by recurrences of work disability leave/sickness absence.

The evidence derived from all studies and intervention approaches for the primary outcomes of interest was graded as "very low" in all cases. A combination of factors contributed to this grade of evidence including:

1. All included studies were judged to be at a high risk of bias
2. Evidence for specific interventions was always based on data from one study. This precluded the ability to examine consistency of the

evidence from various studies and affected the precision of the evidence
3. The population included in the studies was often not considered generalizable to the population of interest for this review.
4. There was imprecise data for all primary outcomes because in all instances there was only one study that provided evidence.

Consequently, there is no one intervention that we have found that can be recommended as effective for the prevention of work disability/sickness absence, the management of work disability/sickness absence, the prevention and management of recurrent work disability/sickness absence, or the improvement of work functioning among workers with mild to moderate depression. At best, we have identified the following interventions as recommended for future research:

- Enhanced Primary Care
- Enhanced Psychiatric Care
- Enhanced Role for the Occupational Physician
- Psychological Interventions
- Worksite Stress Reduction
- Systems Integration and Care Management

The following are interventions demonstrating a net economic benefit to the employer:

- Enhanced Primary Care
- Enhanced Role for the Occupational Physician
- Systems Integration and Care Management

The following are interventions demonstrating evidence from the societal perspective of either an economic net benefit or cost-effectiveness:

- Enhanced Primary Care
- Enhanced Role for the Occupational Physician
- Enhanced Psychiatric Care

Work disability/sickness absence and work functioning were considered primary outcomes in our review, but we also examined the effects of these interventions on psychosocial work outcomes (for example: work stress), critical workplace incidents, depression symptom severity, and depression remission.

Half of the studies assessed some sort of enhanced physician role, i.e. primary care physicians, psychiatrists, and occupational physicians.

However, the type of enhancements and quality improvements were very different because these studies were conducted in different settings where the compensation systems and health care systems are very distinct. Primary care enhancements were conducted in the USA and included education on guideline-concordant interventions, screening for depression, and more frequent contacts with patients especially regarding adherence to treatments (pharmacological or psychotherapy). Psychiatrist care was enhanced by the addition of an occupational therapist that focused on contacting the worker and the employer to discuss a program for work reintegration, and it was conducted in The Netherlands. The enhanced occupational physician role occurred in The Netherlands where they attempted a more active role of the physician by guideline-based education and facilitation of RTW.

4.2 Strengths and Limitations of the Review

Our review was conducted by an international and multidisciplinary team, who received input and feedback from a Canadian stakeholder group during the execution of this review. The research team included physicians, economists, epidemiologists, kinesiologists, information scientists from Canada, USA and The Netherlands. The involvement of the stakeholders group in this review was essential to shape the research question, to suggest terms for the literature search, to prioritize outcome measures and to interpret the key findings. The stakeholders represented the Ontario Ministry of Health and Long Term Care, Ministry of Government Services, the Workplace Safety and Insurance Board (WSIB) of Ontario, association of people with mental health disorders in Canada, private insurers, employers, and union representatives.

The main limitation of our review is the paucity and low quality of the research evidence contributed by existing studies addressing this review's research question. Despite our extensive literature search of multiple databases, as well as hand searches of reference lists, we found only 12 studies that met our inclusion criteria. All of these studies demonstrated a high risk of bias and either generated "very low" grade evidence in support of a particular intervention, meaning any estimate of effect was considered "very uncertain", or null findings. At best, the key messages identified in this review suggest that some intervention approaches deserve further evaluation in future studies. However, there remain many unanswered questions for which stakeholders require answers. Namely:

- It is not yet known which intervention approaches for depression are effective in work disability/sickness absence and work functioning.

- It is also unclear when in the course of a depression episode and work disability/sickness absence an intervention should be administered.
- It is not clear from this review whether the findings from an intervention that demonstrated positive results in the context of other compensation and health care systems are generalizable and effective in Ontario.
- It is not clear why some interventions achieved positive results in the short-term, but results were not maintained in the long-term (for example: short-term psychodynamic psychotherapy) (Knekt 2008a) or vice-versa (long-term psychodynamic psychotherapy) (Knekt 2008a).
- It is not clear why the same intervention often demonstrated conflicting findings for outcomes attempting to measure similar concepts. For example: return to part-time before full-time work, duration of sick leave days until partial RTW, duration of sick leave days until full RTW, and duration of sick leave days (including recurrences) until full RTW (Rebergen 2009a).

Given the prevalence of depression in the workplace and the costs associated with work disability and productivity loss at work, even a small effect size with economic benefits may be regarded as relevant to employers and employees. Although ten of the 12 included studies used a randomized controlled design, there were many features of study design, study performance or analyses that risked the validity of these studies. For instance, due to the inherent nature of these interventions, all included studies lacked the ability to appropriately blind intervention providers and participants to the intervention, and it is acknowledged that controlled studies without blinding are prone to performance and measurement biases. In addition, participation and adherence to the proposed interventions was also not well documented, nor were the potential differences between participants and non-participants, potential differences between remaining participants and those lost to follow-up, or the methods used to randomly allocate individuals to their respective intervention groups. Contamination was also a problem in several studies, while some studies did not account for baseline differences between groups in the analysis. Even with the use of the best available statistical analyses, these inherent biases could not be accounted for, and, therefore, the totality of the evidence in this review was judged as “high risk”.

There are many potential interventions that could have been studied and included in our review, but we did not find studies that met our inclusion criteria, for example: employee assistance programs (EAPs), work re-organization, healthy workplace strategies, work accommodation practices,

self-management approaches (written, web-based or community resources), etc.

4.3 Interpretation and Implications in the Context of the Totality of Evidence

A recently published Cochrane review evaluated the effects of interventions aimed at reducing work disability in depressed workers.(Nieuwenhuijsen 2009) They included 11 studies, of which three were included in our review (Rost 2004, Schene 2006 and Schoenbaum 2001). This Cochrane review used the instrument of Downs and Black to assess the quality of the evidence and they concluded that four out of 11 studies were of high methodological quality, including the study by Schene 2007. The interventions included in the Cochrane review were: adjuvant occupational therapy, antidepressant medications, psychodynamic therapy, enhanced primary care, and psychological treatment. The authors concluded that, based on the heterogeneous sample of studies, there was no evidence of an effect of medication alone, enhanced primary care, psychological interventions or combinations on sickness absence of depressed workers. Our conclusions did not conflict with the conclusion of this Cochrane review. Given the similarities with our own review, we were conscious of the design and results of the Nieuwenhuijsen et al. (2009) review, particularly as Dr. Bültmann is a co-author on both reviews. As a result, we sought to be more inclusive with the outcomes and interventions considered in our review. Furthermore, our review also included a search of the literature published after the time period of the Nieuwenhuijsen et al. study.

In May 2010, the Occupational Health and Safety Agency for Healthcare in British Columbia (Canada) released a report on Best Practices for Return-to-Work/Stay-at-Work Interventions for Workers with Mental Health Conditions (Pomaki, 2010). They performed a systematic review to answer the following questions: 1. Are workplace-based interventions effective in improving return to work or stay at work outcomes for workers with mental health conditions? 2. What are key elements of effective interventions? and 3. Are any interventions specific to the healthcare sector? Five best practices principles were developed and classified according to the following levels of interventions: organizational-level; disability management practice-level; and individual-level:

1. Organizational-level interventions: Clear, detailed, and well-communicated organizational workplace mental health policy supports the return to work/stay at work process
2. Disability management practice-level interventions: Return to work coordination and structured, planned, close communication between workers, employers, unions, healthcare providers, and other

disability management stakeholders are required to optimize return to work and stay at work outcomes
3. Disability management practice-level interventions: Application of systematic, structured and coordinated return to work practices improves return to work outcomes
4. Disability management practice-level interventions: Work accommodations are an integral part of the return to work process and the context of their implementation determines their effectiveness
5. Individual-level interventions: Facilitation of access to evidence-based treatment reduces work absence

It is important to note that this review differs from ours as it included studies of individuals with various mental health conditions, not just depression or depressive symptoms. This review also examined the literature from a much shorter time frame than that considered in our review.

Various other systematic reviews and meta-analyses have demonstrated the efficacy of enhanced primary care programs to screen, manage, and follow populations with depression. (Gilbody 2006, Badamgarav 2003, Williams 2007, Kates 2007) There are reviews (Markowitz 2008, Myhr 1996) showing that cognitive-behavioural therapy and interpersonal therapy reduce work disability and are cost-effective. Corbiere et al, in 2009 (Corbiere 2009) conducted a systematic review of studies with preventive interventions aiming to promote mental health or well being for employees at the workplace. They included 24 studies, of which, none overlapped with the studies included in our review: eight were classified as focusing on primary prevention, 14 on secondary prevention and 2 were considered mixed (both primary and secondary). There was a predominance of studies utilizing skills training. They found positive and significant results with regards to work and mental health outcomes. A meta-analysis of exercises for adults with depression highlighted the lack of good quality research in this area, however, this review did not have a primary focus on the working population. (Lawlor 2001)

Although it is well established that depression significantly reduces work functioning and leads to significant work disability (Adler 2006, Wang 2004, Kessler 1999, Lerner 2004), many workplaces have yet to adopt strategies to deal with the issue. Putnam and McKibbin propose organizational and individual barriers that prevent companies from effectively managing employee depression. Organizational barriers include information gaps, lack of data to justify increased investment in employee mental health programs, and employers' ambiguous roles in addressing depression. Individual barriers, such as an inability to recognize signs and symptoms, stigma, confidentiality and privacy concerns, and unavailability of easily accessible,

quality resources can also keep employees who are depressed from seeking treatment (Putnam 2004).

4.4 Future Research Directions

Our review underscored the paucity and low quality of existing research on intervention approaches available to address the problem of work disability and poor work functioning in workers with mild to moderate depression. In future studies, we encourage researchers to consider addressing some of the pertinent research questions for which evidence is lacking, as outlined in Section 4.2 of this report. We commend the authors of the published studies for their efforts in conducting research in this area. However, we still need to aim for research with low risk of bias in order to make solid recommendations for practice. Future studies need to focus on controlled design, especially randomized trials. As we saw in this review, it is possible to conduct randomized trials in this field. Future randomized trials need to attempt blinding participants to the kind of interventions they are receiving, especially given the subjective nature of depression and work functioning outcomes. It is recognized that it is difficult to blind the workers in this kind of workplace interventions, but the use of cluster randomized trial designs may help facilitate this. Also, authors of published studies should adhere to the CONSORT statement (http://www.consort-statement.org/) for description and reporting. In many situations, we were not able to judge the adequacy of the methodology used, even after many attempts to contact the authors of these studies by e-mail.

Future studies should attempt to analyze the participants according to their baseline working status (working or on disability/sick leave) in order to more specifically address whether an intervention is effective to prevent work disability/sickness absence or to manage work disability/sickness absence. This delineation was not clear from the findings of many of the included studies. There is also a need for better outcome measures, and a consensus of what should be measured when approaching productivity or loss of productivity.

Studies conducting economic evaluations improve the relevance of the evaluations to employers by including cost-benefit analyses from the employer perspective. These analyses should include costs borne by the employer, quantitative measures of workplace disability, and estimates of the monetized value of reductions in workplace disability from an employer perspective. The relevance of cost-benefit analyses from the employer perspective is enhanced by including sensitivity analyses that reflect features of employment and benefits, as demonstrated in the study of Lo Sasso (2006). Sensitivity analyses that are often relevant include modeling characteristics of production (team production, costly substitute labour,

penalties for production shortfalls) and high worker turnover rates, which when present reduce the value to the employer over time of reductions in workplace disability.

The problem of depression in the workplace is complex, with consequences to the worker and their families, co-workers, supervisors and employers, disability insurers, and government. No single intervention approach was shown to be effective to tackle the issue, but perhaps the solution lies in multifaceted and layered approaches that aim to break down both individual and organizational barriers and are coordinated to achieve best outcomes for prevention, management and recurrences of work disability/sickness absence, while maximizing work functioning of those people with mild to moderate depression who choose to stay at work.

5.0 References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. [text revision]. 4th ed. Washington, DC: American Psychiatric Association; 2000.

2. DistillerSR. (Web-based systematic review software) [computer program]. Ottawa, Ontario: Evidence Partners Incorporated; 2010.

3. Statistics Canada. Canadian community health survey mental health and well-being. The Daily 2003 Sep 3.

4. The Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions version 5.0.2 [updated September 2009]. http://www cochrane-handbook org/ 2009 [cited 2011 Jan 5];

5. The Cochrane Collaboration. The Cochrane policy manual [updated 6 December 2010]. www cochrane org/policy-manual/welcome 2010 [cited 2010 Dec 10];

6. The CONSORT Group. CONSORT Statement. http://www consort-statement org/ 2010Available from: URL: http://www.consort-statement.org/

7. The GRADE working group. The GRADE working group. http://www gradeworkinggroup org/ 2010Available from: URL: http://www.gradeworkinggroup.org

8. Watson Wyatt Worldwide. Staying at work survey: preliminary results. Watson Wyatt Worldwide; 2005.

9. World Health Organization. International statistical classification of diseases and related health problems. 10th revision. Version for 2007. http://apps who int/classifications/apps/icd/icd10online/ 2006Available from: URL: http://apps.who.int/classifications/apps/icd/icd10online/

10. Adler DA, McLaughlin TJ, Rogers WH, Chang H, Lapitsky L, Lerner D. Job performance deficits due to depression. Am J Psychiatry 2006 Sep;163(9):1569-76.

11. Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil 2000 Dec;81(12 Suppl 2):S15-S20.

12. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490.

13. Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, et al. Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. BMC Health Serv Res 2005 Mar 23;5(1):25.

14. Badamgarav E, Weingarten SR, Henning JM, Knight K, Hasselblad V, Gano A, Jr., et al. Effectiveness of disease management programs in depression: a systematic review. Am J Psychiatry 2003 Dec;160(12):2080-90.

15. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. J Pers Assess 1996 Dec;67(3):588-97.

16. Birnbaum HG, Kessler RC, Kelley D, Ben Hamadi R, Joish VN, Greenberg PE. Employer burden of mild, moderate, and severe major depressive disorder: mental health services utilization and costs, and work performance. Depress Anxiety 2010;27(1):78-89.

17. Blonk RWB, Brenninkmeijer V, Lagerveld SE, Houtman ILD. Treatment of work- related psychological complaints: a randomized field experiment among self-employed. Work and Stress 2006;20:129-44.

18. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-91.

19. Brewer S, Van Eerd D, Amick BC, III, Irvin E, Daum KM, Gerr F, et al. Workplace interventions to prevent musculoskeletal and visual symptoms and disorders among computer users: a systematic review. J Occup Rehabil 2006 Sep;16(3):325-58.

20. Burton WN, Pransky G, Conti DJ, Chen CY, Edington DW. The association of medical conditions and presenteeism. J Occup Environ Med 2004 Jun;46(6 Suppl):S38-S45.

21. Cole D, Rivilis I, Van Eerd D, Cullen K, Irvin E, Kramer D. Effectiveness of participatory ergonomic interventions: A systematic review. Toronto: Institute for Work & Health; 2005.

22. Cooper H, Hedges LV. The handbook of research synthesis. 1st ed. New York: Russell Sage Foundation; 1994.

23. Corbiere M, Shen J, Rouleau M, Dewa CS. A systematic review of preventive interventions regarding mental health issues in organizations. Work 2009;33(1):81-116.

24. Crawford JR, Henry JD. The Depression Anxiety Stress Scales (DASS): normative data and latent structure in a large non-clinical sample. Br J Clin Psychol 2003 Jun;42(Pt 2):111-31.

25. Dewa CS, Hoch JS, Carmen G, Guscott R, Anderson C. Cost, effectiveness, and cost-effectiveness of a collaborative mental health care program for people receiving short-term disability benefits for psychiatric disorders. Can J Psychiatry 2009 Jun;54(6):379-88.

26. Franche RL, Cullen K, Clarke J, MacEachen E, Frank J, Sinclair S. Workplace-based return-to-work interventions: a systematic review of the quantitative and qualitative literature. Toronto: Institute for Work & Health; 2004.

27. Franche RL, Cullen K, Clarke J, Irvin E, Sinclair S, Frank J. Workplace-based return-to-work interventions: a systematic review of the quantitative literature. J Occup Rehabil 2005 Dec;15(4):607-31.

28. Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. Arch Intern Med 2006 Nov 27;166(21):2314-21.

29. Goetzel RZ, Ozminkowski RJ, Sederer LI, Mark TL. The business case for quality mental health services: why employers should care about the mental health and well-being of their employees. J Occup Environ Med 2002 Apr;44(4):320-30.

30. Goldner E, Bilsker D, Gilbert M, Myette L, Corbiere M, Dewa CS. Disability management, return to work and treatment. Healthc Pap 2004;5(2):76-90.

31. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden of depression in 1990. J Clin Psychiatry 1993 Nov;54(11):405-18.

32. Hansson M, Chotai J, Nordstom A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. Br J Gen Pract 2009 Sep;59(566):e283-e288.

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

34. Ilmarinen J, Rantanen J. Promotion of work ability during ageing. Am J Ind Med 1999 Sep;Suppl 1:21-3.

35. Jaycox LH, Miranda J, Meredith LS, Duan N, Benjamin B, Wells K. Impact of a primary care quality improvement intervention on use of psychotherapy for depression. Ment Health Serv Res 2003 Jun;5(2):109-20.

36. Kates N, Mach M. Chronic disease management for depression in primary care: a summary of the current literature and implications for practice. Can J Psychiatry 2007 Feb;52(2):77-85.

37. Kawakami N, Araki S, Kawashima M, Masumoto T, Hayashi T. Effects of work-related stress reduction on depressive symptoms among Japanese blue-collar workers. Scand J Work Environ Health 1997 Feb;23(1):54-9.

38. Kessler RC, Barber C, Birnbaum HG, Frank RG, Greenberg PE, Rose RM, et al. Depression in the workplace: effects on short-term disability. Health Aff (Millwood) 1999 Sep;18(5):163-71.

39. Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). JAMA 2003 Jun 18;289(23):3095-105.

40. Kessler RC, Barber C, Beck A, Berglund P, Cleary PD, McKenas D, et al. The World Health Organization Health and Work Performance Questionnaire (HPQ). J Occup Environ Med 2003 Feb;45(2):156-74.

41. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-27.

42. 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. Am J Psychiatry 2006 Sep;163(9):1561-8.

43. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. J Psychosom Res 2010 Apr;68(4):319-23.

44. Knekt P, Lindfors O, Harkanen T, Valikoski M, Virtala E, Laaksonen MA, et al. Randomized trial on the effectiveness of long-and short-term psychodynamic psychotherapy and solution-focused therapy on psychiatric symptoms during a 3-year follow-up. Psychol Med 2008 May;38(5):689-703.

45. Knekt P, Lindfors O, Laaksonen MA, Raitasalo R, Haaramo P, Jarvikoski A, et al. Effectiveness of short-term and long-term psychotherapy on work ability and functional capacity--a randomized clinical trial on depressive and anxiety disorders. J Affect Disord 2008 Apr;107(1-3):95-106.

46. Kristensen TS. Intervention studies in occupational epidemiology. Occup Environ Med 2005 Mar;62(3):205-10.

47. Kroenke K, Spitzer RL, Williams JB, Lowe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry 2010 Jul;32(4):345-59.

48. Krogh J, Petersen L, Timmermann M, Saltin B, Nordentoft M. Design paper: the DEMO trial: a randomized, parallel-group, observer-blinded clinical trial of aerobic versus non-aerobic versus relaxation training for patients with light to moderate depression. Contemp Clin Trials 2007 Jan;28(1):79-89.

49. 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. J Clin Psychiatry 2009;70(6):790-800.

50. Lamoureux BE, Linardatos E, Fresco DM, Bartko D, Logue E, Milo L. Using the QIDS-SR16 to identify major depressive disorder in primary care medical patients. Behav Ther 2010 Sep;41(3):423-31.

51. Lawlor DA, Hopker SW. The effectiveness of exercise as an intervention in the management of depression: systematic review and meta-regression analysis of randomised controlled trials. BMJ 2001 Mar 31;322(7289):763-7.

52. Lerner D, Adler DA, Chang H, Lapitsky L, Hood MY, Perissinotto C, et al. Unemployment, job retention, and productivity loss among employees with depression. Psychiatr Serv 2004 Dec;55(12):1371-8.

53. Lerner D, Henke RM. What does research tell us about depression, job performance, and work productivity? J Occup Environ Med 2008 Apr;50(4):401-10.

54. Lo Sasso AT, Rost K, Beck A. Modeling the impact of enhanced depression treatment on workplace functioning and costs: a cost-benefit approach. Med Care 2006 Apr;44(4):352-8.

55. Markowitz JC. Evidence-based psychotherapies for depression. J Occup Environ Med 2008 Apr;50(4):437-40.

56. Miranda J, Duan N, Sherbourne C, Schoenbaum M, Lagomasino I, Jackson-Triche M, et al. Improving care for minorities: can quality improvement interventions improve care and outcomes for depressed minorities? Results of a randomized, controlled trial. Health Serv Res 2003 Apr;38(2):613-30.

57. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 2009 Aug 22;374(9690):609-19.

58. Mitchell AJ, Rao S, Vaze A. Can general practitioners identify people with distress and mild depression? A meta-analysis of clinical accuracy. J Affect Disord 2010 Aug 11;[Epub ahead of print].

59. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. Lancet 1997 May 24;349(9064):1498-504.

60. Myhr G, Payne K. Cost-effectiveness of cognitive-behavioural therapy for mental disorders: implications for public health care funding policy in Canada. Can J Psychiatry 2006 Sep;51(10):662-70.

61. Nieuwenhuijsen K, de Boer AG, Verbeek JH, Blonk RW, van Dijk FJ. The Depression Anxiety Stress Scales (DASS): detecting anxiety disorder and depression in employees absent from work because of mental health problems. Occup Environ Med 2003 Jun;60 Suppl 1:i77-i82.

62. Owens DK, Lohr KN, Atkins D, Treadwell JR, Reston JT, Bass EB, et al. AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions--agency for healthcare research and quality and the effective health-care program. J Clin Epidemiol 2010 May;63(5):513-23.

63. Oxman AD, Schunemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. Health Res Policy Syst 2006;4:20.

64. Patten SB. Progress against major depression in Canada. Can J Psychiatry 2002 Oct;47(8):775-80.

65. Pomaki G, Franche RL, Khushrushahi N, Murray E, Lampinen T, Mah P. Best practices for return-to-work/stay-at-work interventions for workers with mental health conditions. Final report. Vancouver, BC: Occupational Health and Safety Agency for Healthcare in BC (OHSAH); 2010.

66. Putnam K, McKibbin L. Managing workplace depression: an untapped opportunity for occupational health professionals. AAOHN J 2004 Mar;52(3):122-9.

67. Rebergen DS, Bruinvels DJ, Van Der Beek AJ, van Mechelen W. Design of a randomized controlled trial on the effects of counseling of mental health problems by occupational physicians on return to work: the CO-OP-study. BMC Public Health 2007;7:183.

68. Rebergen DS, Bruinvels DJ, Van Tulder MW, Van Der Beek AJ, van Mechelen W. Cost-effectiveness of guideline-based care for workers with mental health problems. Journal of Occupational and Environmental Medicine 51(3)()(pp 313-322), 2009 Date of Publication: March 2009 2009;(3):313-22.

69. Rebergen DS, Bruinvels DJ, Bezemer PD, Van Der Beek AJ, van Mechelen W. Guideline-based care of common mental disorders by occupational physicians (CO-OP study): A randomized controlled trial. Journal of Occupational and Environmental Medicine 51(3)()(pp 305-312), 2009 Date of Publication: March 2009 2009;(3):305-12.

70. Rost K, Nutting PA, Smith J, Werner JJ. Designing and implementing a primary care intervention trial to improve the quality and outcome of care for major depression. Gen Hosp Psychiatry 2000 Mar;22(2):66-77.

71. Rost K, Nutting P, Smith J, Werner J, Duan N. Improving depression outcomes in community primary care practice: a randomized trial of the quEST intervention. Quality Enhancement by Strategic Teaming. J Gen Intern Med 2001 Mar;16(3):143-9.

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

73. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996 May;26(3):477-86.

74. Sanderson K, Tilse E, Nicholson J, Oldenburg B, Graves N. Which presenteeism measures are more sensitive to depression and anxiety? J Affect Disord 2007 Aug;101(1-3):65-74.

75. Schene AH, Koeter MWJ, Kikkert MJ, Swinkels JA, McCrone P. Adjuvant occupational therapy for work-related major depression works: Randomized trial including economic evaluation. Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences 2007;37(3):351-62.

76. 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 Sep 19;286(11):1325-30.

77. Schoenbaum M, Miranda J, Sherbourne C, Duan N, Wells K. Cost-effectiveness of interventions for depressed Latinos. The Journal of Mental Health Policy & Economics 2004 Jun;7(2):69-76.

78. Sherbourne CD, Wells KB, Duan N, Miranda J, Unutzer J, Jaycox L, et al. Long-term effectiveness of disseminating quality improvement for depression in primary care. Arch Gen Psychiatry 2001 Jul;58(7):696-703.

79. Simon GE, VonKorff M, Barlow W. Health care costs of primary care patients with recognized depression. Arch Gen Psychiatry 1995 Oct;52(10):850-6.

80. Smith JL, Rost KM, Nutting PA, Libby AM, Elliott CE, Pyne JM. Impact of primary care depression intervention on employment and workplace conflict outcomes: is value added? The Journal of Mental Health Policy & Economics 2002 Mar;5(1):43-9.

81. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-44.

82. Steer RA, Clark DA, Beck AT, Ranieri WF. Common and specific dimensions of self-reported anxiety and depression: the BDI-II versus the BDI-IA. Behav Res Ther 1999 Feb;37(2):183-90.

83. Stephens T, Joubert N. The economic burden of mental health problems in Canada. Chronic Dis Can 2001;22(1):18-23.

84. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. JAMA 2003 Jun 18;289(23):3135-44.

85. Terracciano L, Brozek J, Compalati E, Schunemann H. GRADE system: new paradigm. Current Opinion in Allergy and Clinical Immunology 2010;10(4):377-83.

86. Tompa E, Trevithick S, McLeod C. Systematic review of the prevention incentives of insurance and regulatory mechanisms for occupational health and safety. Scand J Work Environ Health 2007 Apr;33(2):85-95.

87. van der Feltz-Cornelis CM, Meeuwissen JA, de Jong FJ, Hoedeman R, Elfeddali I. Randomised controlled trial of a psychiatric consultation model for treatment of common mental disorder in the occupational health setting. BMC Health Serv Res 2007;7:29.

88. Van der Feltz-Cornelis CM, Hoedeman R, de Jong FJ, Meeuwissen JAC, Drewes HW, van der Laan NC. Faster return to work after psychiatric consultation for sicklisted employees with common mental disorders compared to care as usual. A randomized clinical trial. Neuropsychiatric Disease and Treatment 2010;6:375-85.

89. Van Eerd D, Brewer S, Amick IBC, Irvin E, Daum K, Gerr F. Workplace interventions to prevent musculoskeletal and visual symptoms and disorders among computer users: a systematic review. Toronto: Institute for Work & Health; 2006.

90. Waddell G, Burton AK. Concepts of rehabilitation for the management of common health problems. London: The Stationery Office (TSO); 2004.

91. Wang PS, Beck AL, Berglund P, McKenas DK, Pronk NP, Simon GE, et al. Effects of major depression on moment-in-time work performance. Am J Psychiatry 2004 Oct;161(10):1885-91.

92. Wang PS, Simon GE, 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: a randomized controlled trial. JAMA 2007 Sep 26;298(12):1401-11.

93. Wells KB, Schoenbaum M, Unutzer J, Lagomasino IT, Rubenstein LV. Quality of care for primary care patients with depression in managed care. Arch Fam Med 1999 Nov;8(6):529-36.

94. 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 Jan 12;283(2):212-20.

95. Williams JW, Jr., Gerrity M, Holsinger T, Dobscha S, Gaynes B, Dietrich A. Systematic review of multifaceted interventions to improve depression care. Gen Hosp Psychiatry 2007 Mar;29(2):91-116.

Appendices

Appendix A

MEDLINE Search Strategy – FINAL. June 21, 2010 (3pm)
N=1109

Database: Ovid MEDLINE(R) <1950 to June Week 2 2010>
Search Strategy:

---

1 affective disorder?.ti,ab. (11009)
2 Affective Symptoms/ (9490)
3 Depression/ (56175)
4 exp Depressive Disorder/ (65451)
5 depress$.ti,ab. (239652)
6 dysthymia.ti,ab. (1436)
7 Mood Disorders/ (8461)
8 mood symptom?.ti,ab. (598)
9 or/1-8 (287924)
10 apprentice?.ti,ab. (651)
11 (boss or bosses).ti,ab. (587)
12 (branch or branches).ti,ab. (63804)
13 (company or companies).ti,ab. (26266)
14 contractor?.ti,ab. (1048)
15 department$.ti,ab. (147222)
16 employee?.ti,ab. (25038)
17 employer?.ti,ab. (10104)
18 Employment/ (30886)
19 facilit$.ti,ab. (275724)
20 (factory or factories).ti,ab. (8286)
21 firm?.ti,ab. (15287)
22 Health Services/ (16503)
23 exp Hospitals/ (170136)
24 Industry/ (18528)
25 institution?.ti,ab. (85803)
26 isolation pay$.ti,ab. (0)
27 laborer?.ti,ab. (1018)
28 labourer?.ti,ab. (425)
29 leader?.ti,ab. (21859)
30 manager?.ti,ab. (21117)
31 office?.ti,ab. (43437)
32 operator?.ti,ab. (23502)
33 Organizations/ (5747)
34 organisation$.ti,ab. (19548)
35 organization$.ti,ab. (188453)
36 personnel.mp. (213563)
37 plant?.ti,ab. (179617)
38 retail.ti,ab. (3300)
39 skilled trade$.ti,ab. (22)
40 staff$.ti,ab. (82091)
41 supervisor?.ti,ab. (6014)
42 team?.ti,ab. (59468)

43 telecommunications/ (3161)
44 union$.ti,ab. (26751)
45 Work/ (7545)
46 work environment?.ti,ab. (4509)
47 work site?.ti,ab. (939)
48 worksite?.ti,ab. (1894)
49 worker?.ti,ab. (97285)
50 "work$ at home".ti,ab. (148)
51 "work$ from home".ti,ab. (780)
52 Workplace/ (9172)
53 workplace?.ti,ab. (16964)
54 work place?.ti,ab. (1544)
55 or/10-54 (1556578)
56 "access to care".ti,ab. (3739)
57 accommodat$.ti,ab. (24251)
58 (acute stress adj2 manag$).ti,ab. (6)
59 adjust$.ti,ab. (211827)
60 advocate?.ti,ab. (26375)
61 affinity group?.ti,ab. (57)
62 alternat$ duty.ti,ab. (5)
63 alternat$ duties.ti,ab. (1)
64 "assessment and referral".ti,ab. (187)
65 benefit?.ti,ab. (248374)
66 Insurance Benefits/ (2209)
67 Case Management/ (6842)
68 (chronic stress adj2 manag$).ti,ab. (3)
69 club member$.ti,ab. (119)
70 coaching.ti,ab. (1232)
71 Social Welfare/ (6949)
72 contracted ombud$ service?.ti,ab. (0)
73 ombud$ service?.ti,ab. (2)
74 Counseling/ (23465)
75 cultural resource?.ti,ab. (56)
76 depression screen$.ti,ab. (548)
77 disability management program$.ti,ab. (18)
78 diversity resource?.ti,ab. (2)
79 early intervention?.ti,ab. (6575)
80 Education/ (16059)
81 "education and training".ti,ab. (4326)
82 e-learning.ti,ab. (499)
83 elearning.ti,ab. (26)
84 embrace diversity.ti,ab. (4)
85 employee assistance program$.ti,ab. (266)
86 EAP program$.ti,ab. (2)
87 employee satisfaction survey?.ti,ab. (6)
88 employer resource group?.ti,ab. (0)
89 engagement.ti,ab. (13111)
90 enhanced access.ti,ab. (109)
91 fitness group?.ti,ab. (76)
92 flexible work.ti,ab. (106)
93 functional capacity assessment?.ti,ab. (12)
94 functionality.ti,ab. (14501)

95 Gardening/ (231)
96 grassroot?.ti,ab. (818)
97 gym member$.ti,ab. (2)
98 "health and wellness".ti,ab. (383)
99 health risk management.ti,ab. (59)
100 healthy workplace strateg$.ti,ab. (1)
101 horticulture.ti,ab. (261)
102 independent medical evaluation?.ti,ab. (15)
103 IMEs.ti,ab. (64)
104 "inviting an organi#ation in".ti,ab. (0)
105 job control.ti,ab. (409)
106 (job adj2 modifi$).ti,ab. (49)
107 joint labor management initiative?.ti,ab. (0)
108 joint labour management initiative?.ti,ab. (0)
109 long-term disabilit$.ti,ab. (986)
110 LTD benefit?.ti,ab. (1)
111 LTD depression screen$.ti,ab. (0)
112 management of individual?.ti,ab. (621)
113 medical surveillance.ti,ab. (974)
114 mental health promotion.ti,ab. (222)
115 mental job analys#s.ti,ab. (0)
116 mentoring.ti,ab. (2104)
117 (modifi$ adj2 duties).ti,ab. (15)
118 (modifi$ adj2 duty).ti,ab. (20)
119 (modifi$ adj2 work).ti,ab. (322)
120 nature.ti,ab. (213002)
121 Occupational Health Services/ (8868)
122 organizational culture/ (9821)
123 organizational policy/ (10644)
124 "organizational polic$ and practice?".ti,ab. (14)
125 "organisational polic$ and practice?".ti,ab. (1)
126 "Organization and Administration"/ (14049)
127 OPPs.ti,ab. (120)
128 Pastoral Care/ (2869)
129 peer support.ti,ab. (854)
130 performance management.ti,ab. (220)
131 Animal Assisted Therapy/ (13)
132 positive psychology.ti,ab. (105)
133 practice guideline/ (14360)
134 Practice Guidelines as Topic/ (56609)
135 prayer room?.ti,ab. (2)
136 preferred provider network?.ti,ab. (5)
137 prevention.ti,ab. (256219)
138 (promot$ adj2 recovery).ti,ab. (1858)
139 psychological safety.ti,ab. (33)
140 psychological risk factor?.ti,ab. (194)
141 quiet room?.ti,ab. (116)
142 quiet space?.ti,ab. (2)
143 reflection room?.ti,ab. (0)
144 Rehabilitation/ (14884)
145 reintegrat$.ti,ab. (1657)
146 (relapse adj2 prevent$).ti,ab. (3552)

147 (resilienc? adj2 train$).ti,ab. (14)
148 "return$ to work".ti,ab. (5495)
149 RTW.ti,ab. (195)
150 Reward/ (8841)
151 "Referral and Consultation"/ (43536)
152 self help.ti,ab. (3577)
153 Self-Help Groups/ (6800)
154 Self Care/ (18032)
155 self-care program$.ti,ab. (80)
156 shared-care.ti,ab. (617)
157 short term disabilit$.ti,ab. (116)
158 STD benefit?.ti,ab. (0)
159 spiritual care.ti,ab. (605)
160 spirituality/ (3280)
161 "stay$ at work".ti,ab. (25)
162 (stress$ adj2 manag$).ti,ab. (2890)
163 support group?.ti,ab. (3447)
164 (support$ adj2 "small business$").ti,ab. (3)
165 (support$ adj3 leader$).ti,ab. (536)
166 (support$ adj3 manage$).ti,ab. (3816)
167 (support$ adj3 supervis$).ti,ab. (781)
168 (task? adj2 modifi$).ti,ab. (376)
169 Time Management/ (2112)
170 training.ti,ab. (168694)
171 treatment support?.ti,ab. (250)
172 universal access.ti,ab. (679)
173 Rehabilitation, Vocational/ (7788)
174 wellness strateg$.ti,ab. (11)
175 (work$ adj2 intervention?).ti,ab. (1477)
176 (work$ adj2 reorgani$).ti,ab. (78)
177 work$ adjustment?.ti,ab. (129)
178 or/56-177 (1350837)
179 Absenteeism/ (6235)
180 accommodat$.ti,ab. (24251)
181 benefit duration.ti,ab. (7)
182 Cost-Benefit Analysis/ (48453)
183 (co-worker? adj2 conflict?).ti,ab. (4)
184 (coworker? adj2 conflict?).ti,ab. (6)
185 cultural shift?.ti,ab. (116)
186 disability pension?.ti,ab. (731)
187 employee satisfaction.ti,ab. (200)
188 engagement.ti,ab. (13111)
189 job match.ti,ab. (5)
190 job turnover.ti,ab. (79)
191 labo?r force participation.ti,ab. (737)
192 long-term disabilit$.ti,ab. (986)
193 lost time.ti,ab. (348)
194 lost workday?.ti,ab. (142)
195 new employer?.ti,ab. (21)
196 new job?.ti,ab. (321)
197 presenteeism.ti,ab. (152)
198 Efficiency/ (10018)

199 productivity ratio.ti,ab. (10)
200 reassign$.ti,ab. (1415)
201 re-assign$.ti,ab. (72)
202 recovery.ti,ab. (225784)
203 reduced cost?.ti,ab. (1222)
204 (reduc$ adj2 complaint?).ti,ab. (233)
205 (reduc$ adj2 harassment).ti,ab. (14)
206 reemploy$.ti,ab. (103)
207 re-employ$.ti,ab. (132)
208 remission.mp. (90165)
209 Resilience, Psychological/ (279)
210 resilienc$.ti,ab. (3847)
211 "return on investment".ti,ab. (542)
212 "return$ to work".ti,ab. (5495)
213 RTW.ti,ab. (195)
214 "short-term disabilit$".ti,ab. (116)
215 Sick Leave/ (2546)
216 sick$ absence?.ti,ab. (971)
217 "stay$ at work".ti,ab. (25)
218 stigma.ti,ab. (5370)
219 (support$ adj2 "work$ solution?").ti,ab. (1)
220 Aptitude/ (2494)
221 talent?.ti,ab. (1641)
222 "time on benefit?".ti,ab. (3)
223 Unemployment/ (4337)
224 vocational assessment.ti,ab. (44)
225 wage replacement.ti,ab. (30)
226 wellness strateg$.ti,ab. (11)
227 work abilit$.ti,ab. (418)
228 work absence?.ti,ab. (199)
229 (work$ adj1 adapt$).ti,ab. (144)
230 (work$ adj1 adjust$).ti,ab. (290)
231 work$ capacity.ti,ab. (5680)
232 work$ disabilit$.ti,ab. (1145)
233 work$ functioning.ti,ab. (113)
234 work$ impairment.ti,ab. (110)
235 work$ limit$.ti,ab. (264)
236 work$ loss$.ti,ab. (399)
237 work$ performance.ti,ab. (1378)
238 (work$ adj2 re-entry).ti,ab. (26)
239 (work$ adj2 reentry).ti,ab. (47)
240 (work$ adj2 reintegrat$).ti,ab. (50)
241 (work$ adj2 resumption).ti,ab. (201)
242 (work$ adj2 retention).ti,ab. (153)
243 Workers' Compensation/ (6098)
244 work$ compensation.ti,ab. (2805)
245 work-life balance.ti,ab. (122)
246 or/179-245 (455310)
247 9 and 55 and 178 and 246 (1109)

EMBASE Search Strategy – FINAL. June 21, 2010
N=1431

Database: EMBASE <1980 to 2010 Week 24>
Search Strategy:
---

1 emotional disorder/ (4128)
2 exp depression/ (177158)
3 depress$.ti,ab. (214031)
4 dysthymia/ (3614)
5 mood disorder/ (13114)
6 mood symptom?.ti,ab. (643)
7 or/1-6 (288337)
8 apprentice?.ti,ab. (425)
9 (boss or bosses).ti,ab. (344)
10 (branch or branches).ti,ab. (50542)
11 (company or companies).ti,ab. (30215)
12 contractor?.ti,ab. (852)
13 department$.ti,ab. (119423)
14 employee/ (6114)
15 employee?.ti,ab. (15761)
16 employer/ (2924)
17 employer?.ti,ab. (5924)
18 employment/ (14270)
19 facilit$.ti,ab. (232087)
20 (factory or factories).ti,ab. (7128)
21 firm?.ti,ab. (12522)
22 health service/ (46656)
23 exp hospital/ (174101)
24 industry/ (8452)
25 institution$.ti,ab. (93206)
26 isolation pay$.ti,ab. (1)
27 laborer?.ti,ab. (570)
28 labourer?.ti,ab. (337)
29 leadership/ (9798)
30 manager?.ti,ab. (11181)
31 office?.ti,ab. (28901)
32 office worker/ (1369)
33 operator/ (1415)
34 operator?.ti,ab. (19288)
35 organization/ (19083)
36 organi#ation$.ti,ab. (148165)
37 personnel.ti,ab. (25624)
38 plant?.ti,ab. (138392)
39 retail.ti,ab. (2347)
40 skilled trade$.ti,ab. (14)
41 staff$.ti,ab. (52868)
42 supervisor?.ti,ab. (3460)
43 team?.ti,ab. (43039)
44 telecommunication/ (6912)
45 trade union/ (918)

46 union$.ti,ab. (21373)
47 work/ (6652)
48 work environment/ (12127)
49 work site?.ti,ab. (714)
50 worksite?.ti,ab. (1538)
51 worker/ (5056)
52 worker?.mp. (81768)
53 "work$ at home".ti,ab. (107)
54 "work$ from home".ti,ab. (585)
55 workplace/ (11865)
56 workplace?.ti,ab. (14558)
57 work place?.ti,ab. (1282)
58 or/8-57 (1181845)
59 "access to care".ti,ab. (2657)
60 accommodat$.ti,ab. (19432)
61 job accommodation/ (121)
62 (acute stress adj2 manag$).ti,ab. (3)
63 adjustment/ (3385)
64 job adaptation/ (305)
65 advocate?.ti,ab. (20986)
66 affinity group?.ti,ab. (43)
67 alternat$ duty.ti,ab. (4)
68 alternat$ duties.ti,ab. (4)
69 "assessment and referral".ti,ab. (149)
70 benefit?.ti,ab. (226050)
71 case management/ (914)
72 (chronic stress adj2 manag$).ti,ab. (2)
73 club member$.ti,ab. (86)
74 coaching.ti,ab. (900)
75 community service?.ti,ab. (1447)
76 contracted ombud$ service?.ti,ab. (0)
77 ombud$ service?.ti,ab. (1)
78 counseling/ (10362)
79 cultural resources.ti,ab. (44)
80 depression screen$.ti,ab. (477)
81 disability management program$.ti,ab. (16)
82 diversity resource?.ti,ab. (0)
83 early intervention?.ti,ab. (6317)
84 education/ (71062)
85 "education and training".ti,ab. (2934)
86 e-learning.ti,ab. (376)
87 elearning.ti,ab. (26)
88 embrace diversity.ti,ab. (6)
89 personnel management/ (6109)
90 employee assistance program$.ti,ab. (292)
91 EAP program$.ti,ab. (8)
92 employee satisfaction survey?.ti,ab. (2)
93 employee resource group?.ti,ab. (0)
94 engagement.ti,ab. (11580)
95 enhanced access.ti,ab. (90)
96 fitness group?.ti,ab. (68)
97 flexible work.ti,ab. (69)

98 functional capacity assessment?.ti,ab. (15)
99 functionality.ti,ab. (14502)
100 gardening/ (155)
101 grassroot?.ti,ab. (287)
102 gym member$.ti,ab. (6)
103 "health and wellness".ti,ab. (252)
104 health risk management.ti,ab. (40)
105 healthy workplace strateg$.ti,ab. (1)
106 horticulture/ (267)
107 independent medical evaluation?.ti,ab. (13)
108 IMEs.ti,ab. (67)
109 "invit$ an organi#ation in".ti,ab. (0)
110 job control.ti,ab. (387)
111 (job adj2 modifi$).ti,ab. (43)
112 joint labor management initiative?.ti,ab. (0)
113 joint labour management initiative?.ti,ab. (0)
114 long term disability benefit?.ti,ab. (5)
115 long-term disabilit$.ti,ab. (915)
116 LTD benefit?.ti,ab. (1)
117 "management of individual?".ti,ab. (580)
118 periodic medical examination/ (1574)
119 mental health promotion.ti,ab. (163)
120 mental job analys#s.ti,ab. (0)
121 mentoring.ti,ab. (1106)
122 (modifi$ adj2 duties).ti,ab. (16)
123 (modifi$ adj2 duty).ti,ab. (14)
124 (modifi$ adj2 work).ti,ab. (328)
125 nature.ti,ab. (204728)
126 occupational health service/ (1981)
127 organi#ational culture.ti,ab. (384)
128 "organisational polic$ and practice?".ti,ab. (2)
129 "organizational polic$ and practice?".ti,ab. (13)
130 OPPs.ti,ab. (99)
131 pastoral care.ti,ab. (100)
132 peer group/ (2601)
133 performance management.ti,ab. (140)
134 pet therapy/ (54)
135 positive psychology.ti,ab. (91)
136 practice guideline/ (120194)
137 prayer room?.ti,ab. (0)
138 preferred provider network?.ti,ab. (2)
139 prevention/ (88155)
140 (promot$ adj2 recovery).ti,ab. (1752)
141 psychological safety.ti,ab. (22)
142 psychological risk factor?.ti,ab. (190)
143 quiet room?.ti,ab. (88)
144 quiet space?.ti,ab. (3)
145 reflection room?.ti,ab. (0)
146 rehab$.ti,ab. (65667)
147 reintegrat$.ti,ab. (1497)
148 (relapse adj2 prevent$).ti,ab. (3708)
149 (resilienc$ adj2 train$).ti,ab. (9)

150 "return$ to work".ti,ab. (4753)
151 RTW.ti,ab. (183)
152 reward?.ti,ab. (12426)
153 second opinion.ti,ab. (630)
154 self help/ (3493)
155 self-help program$.ti,ab. (139)
156 shared care.ti,ab. (499)
157 short-term disabilit$.ti,ab. (118)
158 STD benefit?.ti,ab. (0)
159 spiritual care/ (467)
160 religion/ (16953)
161 "stay$ at work".ti,ab. (23)
162 stress management/ (830)
163 support group/ (4433)
164 (support$ adj2 "small business$").ti,ab. (1)
165 (support$ adj3 leader$).ti,ab. (286)
166 (support$ adj3 manage$).ti,ab. (3071)
167 (support$ adj3 supervis$).ti,ab. (550)
168 (task? adj2 modifi$).ti,ab. (327)
169 time management/ (1364)
170 training.ti,ab. (130536)
171 treatment support?.ti,ab. (240)
172 universal access.ti,ab. (460)
173 vocational rehabilitation/ (2957)
174 wellness strateg$.ti,ab. (7)
175 (work$ adj2 intervention?).ti,ab. (1225)
176 (work$ adj2 reorgani$).ti,ab. (57)
177 work$ adjustment?.ti,ab. (106)
178 or/59-177 (977637)
179 absenteeism/ (6440)
180 accommodat$.ti,ab. (19432)
181 benefit duration.ti,ab. (8)
182 "cost effectiveness analysis"/ (64622)
183 (co-worker? adj2 conflict?).ti,ab. (3)
184 (coworker? adj2 conflict?).ti,ab. (6)
185 cultural shift?.ti,ab. (80)
186 disability pension?.ti,ab. (621)
187 employee satisfaction.ti,ab. (79)
188 engagement.ti,ab. (11580)
189 job match.ti,ab. (8)
190 job turnover.ti,ab. (52)
191 labo?r force participation.ti,ab. (227)
192 long-term disabilit$.ti,ab. (915)
193 lost time.ti,ab. (314)
194 lost workday?.ti,ab. (127)
195 new employer?.ti,ab. (7)
196 new job?.ti,ab. (176)
197 presenteeism.ti,ab. (146)
198 productivity/ (10333)
199 productivity ratio.ti,ab. (8)
200 reassign$.ti,ab. (1260)
201 re-assign$.ti,ab. (68)

202 recovery.ti,ab. (201642)
203 reduced cost?.ti,ab. (1047)
204 (reduc$ adj2 complaint?).ti,ab. (225)
205 (reduc$ adj2 harassment).ti,ab. (9)
206 work resumption/ (2704)
207 reemploy$.ti,ab. (60)
208 re-employ$.ti,ab. (104)
209 remission/ (31527)
210 resilienc$.ti,ab. (3045)
211 "return on investment".ti,ab. (425)
212 "return$ to work".ti,ab. (4753)
213 RTW.ti,ab. (183)
214 short-term disabilit$.ti,ab. (118)
215 medical leave/ (863)
216 sick$ absence?.ti,ab. (890)
217 "stay$ at work".ti,ab. (23)
218 stigma/ (1711)
219 (support$ adj2 "work$ solution?").ti,ab. (1)
220 talent?.ti,ab. (1073)
221 "time on benefit?".ti,ab. (2)
222 unemploy$.ti,ab. (5955)
223 vocational assessment.ti,ab. (58)
224 wage replacement.ti,ab. (22)
225 wellness strateg$.ti,ab. (7)
226 work abilit$.ti,ab. (345)
227 work absence?.ti,ab. (192)
228 (work$ adj1 adapt$).ti,ab. (102)
229 (work$ adj1 adjust$).ti,ab. (248)
230 work capacity/ (4966)
231 work disability/ (2565)
232 work$ functioning.ti,ab. (106)
233 work$ impairment.ti,ab. (110)
234 work$ limit$.ti,ab. (228)
235 work$ loss$.ti,ab. (370)
236 work$ performance.ti,ab. (1105)
237 (work$ adj2 re-entry).ti,ab. (27)
238 (work$ adj2 reentry).ti,ab. (33)
239 (work$ adj2 reintegrat$).ti,ab. (55)
240 work resumption/ (2704)
241 (work$ adj2 retention).ti,ab. (112)
242 workman compensation/ (3165)
243 work$ compensation.ti,ab. (1957)
244 work-life balance.ti,ab. (84)
245 or/179-244 (370709)
246 7 and 58 and 178 and 245 (1431)

Appendix B

Level 1: Title and Abstract Screening - Reviewer Guide

A screening tool has been developed to help reviewers evaluate an article's relevance to our research question based on a series of relevancy questions. You will be asked to answer three questions, all of which are designed to exclude articles that are not relevant to the research question. When reviewing titles and abstracts to determine their relevance, please continue to bear in mind our primary research question:

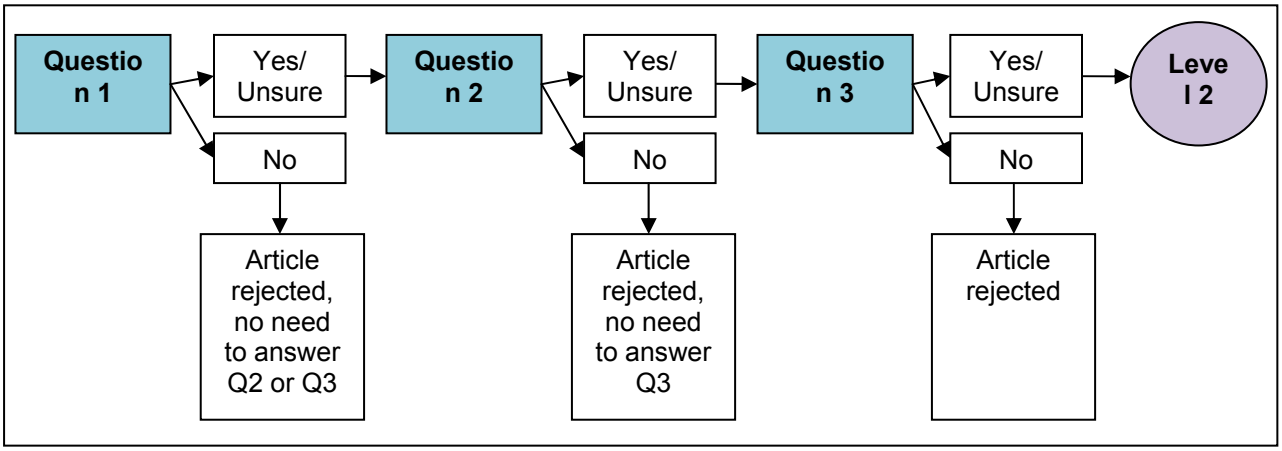
Which intervention approaches to manage depression in the workplace have been successful and yielded value for employers in developed economies?

This guide is intended to be a resource to help reviewers determine whether the articles yielded from the literature search are relevant to the research question. As you review the titles and abstracts you will need to consider various terms and concepts that could be interpreted in different ways. The guide provides descriptions to help with this.

Each reviewer should become thoroughly familiar with the guide prior to conducting the title and abstract screening. Inter-rater variability will be minimized by each rater's familiarity with the guide. Please print (or view onscreen) a copy of the guide and refer to it as you review the titles and abstracts.

How to answer questions at Level 1 (Title and Abstract screening):
For each question there are three predetermined answers. Please select the most appropriate answer to each question. For instance, question one (Q1), asks you to choose whether the article includes people of working age with depression by selecting "Yes", "No" or "Uncertain". Please answer "Yes" if you are sure that the title/abstract meets the relevance criteria, "No" if you are certain that the article is not relevant and if you are uncertain or have any doubt, please select "Uncertain". At this stage of the review we will be erring on the side of inclusiveness and responses marked "Uncertain" will continue to the next Level (see Figure 1).

Figure 1: Flow diagram demonstrating the path for questions at Level 1



Reviewer Guide: Inclusion and Exclusion Criteria

	Inclusion criteria	Exclusion criteria
Population		
Q1	Men and/or women of working age (approximately 18-65 years old) with current or remitted depression (mild, moderate or severe*). Persons with any co-morbidity as long as the study population is defined as having mild, moderate or severe* depression. Any business size (small, medium, or large), as well as any sector (health care, services, industry, mining, forestry, etc.)	Studies that focus on persons with a serious mental disorder (i.e. bipolar disorder or schizophrenia), (these studies are only eligible *if* co-morbid with depression *and* depression is the primary focus of the article) Studies where the primary focus is on persons with alcohol or other substance abuse or dependence (these studies are only eligible *if* co-morbid with depression *and* depression is the primary focus of the article) Studies that focus on military personnel and veterans. Studies that focus on seniors, the elderly population and children. Studies that focus on depression related to pregnancy unless the article is about working mothers and coming back to work after a maternity leave, and they had some kind of intervention to screen/manage post-partum depression. Studies that focus on bereavement or burnout.
Intervention		
Q2	Interventions or programs that are workplace-based or may be explicitly implemented and/or facilitated by the workplace. Workplace or work setting is defined as any location where a worker is performing his or her assigned work.	An in-patient intervention or drug trial.
Comparison		
Q3	Studies with a comparison group: before-and-after comparison within the same population, comparison between different populations, or randomized trials.	Studies with only a post-intervention measure.

*Note: at Level 1 we are intentionally aiming to be more inclusive by including people with severe depression. At Level 2 (the full article screen) we will be more specific about whether studies with this population will be included or excluded from the review.

Definitions of key terms

The review team defined the key terms as follows:

Co-morbidity

Any co-occurring mental disorder or chronic health condition that is not a serious mental disorder, as defined below.

Current or remitted depression

Depression that is determined by any of the following methods: screening questionnaires or instruments, clinician-derived diagnosis, or diagnoses verified according to formal standardized diagnostic criteria (DSM, ICD).

In-patient intervention

Any health or psychosocial intervention that occurs when a client is admitted to a hospital or psychiatric facility.

Return-to-work (RTW)

The return of employees to the workplace following a period of depression-related work absence.

Serious mental disorder

For example, bipolar disorder or schizophrenia.

Stay-at-work (SAW)

The prevention of depression-related work absence and/or improvements in work productivity and performance.

Workplace or work setting

Any location where a worker is performing his or her assigned work.

Specific questions about articles:

- Regarding “burnout”, only include clinically diagnosed depression. Because "burnout" is a lay term, we will exclude these UNLESS the article suggests that the burnout really was depression. If this is the case, please mark it as "Unsure"
- If ages (or age ranges) are not specified and terms such as "elderly", "geriatric" or “older adults” are used, and if there is no potential of a stratified sample of a working population in the article, then please exclude.
- If the age range provided is “60+”, please exclude.
- If the study population has an 'adjustment disorder', please exclude.
- If the depression is related to bereavement, please exclude.
- If the depression is related to pregnancy (i.e. antenatal and postpartum), then please exclude. UNLESS the article is about working mothers and coming back to work after a maternity leave, and they had some kind of intervention to screen/manage post-partum depression.
- All articles relating to veterans (even if the population is recruited from veterans health administration hospitals), please exclude.
- For studies relating to traumatic brain injury populations who have developed depression the team decided to exclude these studies as depression is not their primary focus.

- For articles that are systematic reviews, meta analyses or reviews that meet the inclusion criteria, when these articles come up, the team is to “flag” them to me by sending me an email with the refID (and please review as usual).

Appendix C

Level 2: Full Article Screening - Reviewer Guide

A screening tool has been developed to help reviewers evaluate an article's relevance to our research question based on a series of relevancy questions. You will be asked to answer four relevancy questions, all of which are designed to exclude articles that are not relevant to the research question. In addition, there are three other questions aimed to help us capture other potentially relevant information from the articles. When reviewing titles and full articles to determine their relevance, please continue to bear in mind our primary research question:

Which intervention approaches to manage depression in the workplace have been successful and yielded value for employers in developed economies?

This guide is intended to be a resource to help reviewers determine whether the articles yielded from the literature search are relevant to the research question. As you review the titles and full articles you will need to consider various terms and concepts that could be interpreted in different ways. The guide provides descriptions to help with this.

Each reviewer should become thoroughly familiar with the guide prior to conducting the title and full article screening. Inter-rater variability will be minimized by each rater's familiarity with the guide. Please print (or view onscreen) a copy of the guide and refer to it as you review the titles and full articles.

How to answer questions at Level 2 (Full Article screening):
For each question there are predetermined answers. Please select the most appropriate answer to each question. For instance, question one (Q1), asks you to choose whether the article includes people of working age with depression by selecting "Yes" or "No". Please answer "Yes" if you are sure that the article meets the relevance criteria, "No" if you are certain that the article is not relevant. If after reading the full article you are still uncertain, please select "Unsure" and use the text box provided to make any notes to yourself that will be helpful during the conflict resolution stage. Please note, reviewers are to use "Unsure" as a last resort only.

Reviewer Guide: Inclusion and Exclusion Criteria

	Inclusion criteria	Exclusion criteria
Population		
Q1	Men and/or women of working age (approximately 18-65 years old) with current or remitted depression (mild or moderate). Studies in which 50% or more of the population has depression Persons with any co-morbidity as long as the study population is defined as having mild or moderate depression. Any business size (small, medium, or large), as well as any sector (health care, services,	Studies that focus on chronic severe depression Studies that focus on persons with a serious mental disorder or mental health issues (i.e. bipolar disorder or schizophrenia), (these studies are only eligible *if* co-morbid with depression *and* depression is the primary focus of the article, with 50% or more of the population having depression) Studies where the primary focus is on persons with alcohol or other substance abuse or dependence (these studies are only eligible *if* co-

	industry, mining, forestry, etc.)	morbid with depression *and* depression is the primary focus of the article, with 50% or more of the population having depression) Studies in which less than 50% of the population have depression Studies that focus on military personnel and veterans. Studies that focus on seniors, the elderly population and children. If studies do not provide ages, but use terms such as "elderly" or "geriatric", and if there is no potential of a stratified sample of a working population in the article, then please exclude. Studies that focus on depression related to pregnancy unless the article is about working mothers and coming back to work after a maternity leave, and they had some kind of intervention to screen/manage post-partum depression. Studies that focus on bereavement or burnout.
Intervention		
Q2	Interventions or programs that are workplace-based or may be explicitly implemented and/or facilitated by the workplace.	An in-patient intervention or drug trial.
Comparison		
Q3	Studies with a comparison group: before-and-after comparison within the same population, comparison between different populations, or randomized trials.	Studies with only a post-intervention measure.
Outcome		
Q4	Primary outcomes that are relevant to employers may include: -Changes in productivity - including changes in productivity while depressed workers are still at work -Changes in sickness absence, absenteeism, worker turnover, and long-term disability -Changes in on-the-job performance and health-related performance -Changes in rates of job-related accidents, or -Economic outcomes	Articles that assess only secondary outcomes, i.e. clinical improvement of depression, general well-being (for example, SF-36), patient satisfaction, or quality of life.
Other		
Q5	Please state any reasons why this article should be included.	These articles will not proceed to QA, but will be deemed useful as background articles that may be helpful in context setting or as part of the Introduction to the report.
Q6	Please select whether the article is a review or not	Please select whether the article is review or not
Q7	Please state (by copying and pasting citation information) whether there are any relevant references	Please state (by copying and pasting citation information) whether there are any relevant references

Definitions of key terms

The review team defined the key terms as follows:

Co-morbidity

Any co-occurring mental disorder or chronic health condition that is not a serious mental disorder, as defined below.

Current or remitted depression

Depression that is determined by any of the following methods: screening questionnaires or instruments, clinician-derived diagnosis, or diagnoses verified according to formal standardized diagnostic criteria (fulfilling criteria of the DSM IV or other classifications) or validated self-report instruments.

In-patient intervention

Any health or psychosocial intervention that occurs when a client is admitted to a hospital or psychiatric facility.

Return-to-work (RTW)

The return of employees to the workplace following a period of depression-related work absence.

Serious mental disorder

For example, bipolar disorder or schizophrenia.

Chronic severe depression

Chronic depression with onset in adolescence or early adulthood that has precluded meaningful labour market participation.

Stay-at-work (SAW)

The prevention of depression-related work absence and/or improvements in work productivity and performance. This may include any measure of labour force attachment, including presenteeism, unemployed but actively seeking work, work accommodation, modified hours, etc.

Workplace or work setting

Any location where a worker is performing his or her assigned work.

Appendix D

Level 3: Quality Appraisal – Reviewer Guide

Quality assessment will be conducted on the articles that remain following the exclusion stages (Level 1 and 2). The quality assessment process involves a review of the full article to evaluate the overall quality of the article and provide a quality ranking. The ranking determines if the article should continue to Level 4: Data Extraction. When assessing articles, please continue to bear in mind our primary research question:

Which intervention approaches to manage depression in the workplace have been successful and yielded value for employers in developed economies?

Each reviewer should become thoroughly familiar with the guide prior to conducting the quality assessment review as inter-rater variability will be minimized by each reviewer's familiarity with the guide. Please print (or view onscreen) a copy of the guide and refer to it as you assess the articles.

Please note that the textboxes provided with each answer for the quality assessment questions gives you the ability to make brief comments that would be helpful in resolving potential conflicts with other reviewers and to help remind you of your thoughts and justifications for selecting a particular response. It is not a requirement for these text boxes to be filled out for each response.

Screening question

1. **Should this article continue to QA?**
   This is to provide an additional layer of quality control to ensure that articles have been appropriately assigned. Does the article meet our inclusion criteria for the population, intervention, comparison and outcomes? (Please see appendix A for inclusion/exclusion criteria)
   a) Yes _____
   b) No (please indicate why) _____

Design and objectives

2. **Is the research question clearly stated?**
   If the aim of the study is not clearly stated, then the results are likely of limited value. A clear, explicit statement of objectives/purposes should be included in the study. Consider if the question is focused in terms of the population studied, the intervention and the outcomes considered.
   a) Yes _____
   b) No _____

3. **Were comparison group(s) used?**
   In this review, only articles that had a comparison group passed to this level. A comparison group is important to document and account for the potential effects of unexpected changes. Having a closely analogous comparison group, with similar exposure to causal risk factors as the intervention subjects is a major strength of a study. A comparison can receive a placebo or usual care and thus be considered a comparison (i.e. a control) or it can be presented by use of a before and after comparison (i.e. a pre-post design). Please note, comparison groups are actual groups of individuals, *statistically generated references created for comparison do not constitute a control.*
   a) Yes, a distinct comparison group was used _____
   b) Yes, a pre-post comparison group was used _____
   c) No, this article should not be here _____

4. **Was an intervention allocation method performed adequately?**
   Allocation is a process of determining which participants will receive the treatment or intervention and which will not. An inadequate description of the exposure/intervention allocation strategy makes it impossible to know how likely it is that the intervention group will differ from the comparison group in important baseline characteristics that could influence the outcome. To score yes, the method of randomization AND allocation concealment have to be adequate.
   Adequate methods of randomization include: referring to a random number table; using a computer random number generator; coin tossing; shuffling cards or envelopes; throwing dice; drawing of lots; minimization (minimization may be implemented without a random element, and this is considered to be equivalent to being random). Examples of inadequate methods are: alternation, birth date, social insurance/security number, date in which they are invited to participate in the study, or hospital registration number.
   Adequate concealment of allocation occurs if assignment was generated by an independent person not responsible for determining the eligibility of the patients. This person has no information about the persons included in the trial and has no influence on the assignment sequence or on the decision about the eligibility of the patient, and could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: central allocation (including telephone, web-based, and pharmacy-controlled, randomization); sequentially numbered drug containers of identical appearance; sequentially numbered, opaque, sealed envelopes. Examples of inadequate methods are: using an open random allocation schedule (e.g. a list of random numbers); assignment envelopes were used without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered); alternation or rotation; date of birth; case record number; other explicitly unconcealed procedure.
   a) Yes, adequately _____
   b) No, not adequately _____
   c) No, not described _____

Level of recruitment

5. **Was recruitment (or participation) rate reported and adequate?**
   Recruitment (or participation) rate is the ratio of those who agreed to participate in the study over those who were approached and eligible to participate in the study. (The denominator includes those who participated but also those who were eligible but chose not to participate). Sometimes the information to calculate a recruitment (or participation rate) must

be abstracted from information reported in tables. If the participation rate is greater than 35%, please select option 'a'. If the participation rate is less than 35%, please select 'b'.
a) Yes, and rates were greater than 35% _____
b) Yes, and rates were less than 35% _____
c) No, not described _____

6. **Did the author(s) examine whether important differences existed between those who participated and those who did not?**
Comparisons should be made between those who agreed to participate in the study and those who were approached to participate in the study but chose not to participate to determine to what extent the results are generalizable. An example of "Not applicable" would be if there was a 100% participation rate for the study.
a) Yes, described and no major differences _____
b) Yes, described, and there were major differences _____
c) Not described _____
d) Not applicable _____

7. **Were pre-intervention (baseline) characteristics described and appropriately balanced?**
These may include job related factors, individual characteristics, and factors related to exposures and outcomes. A description of pre-intervention characteristics allows us to identify any important pre-intervention characteristics that could potentially confound the relationship between the intervention and the outcome. It is important to measure potential confounders/effect modifiers as they could mask any true associations that may be present and therefore threaten internal validity of a given study.
a) Yes, described and no major differences between comparison groups _____
b) Yes, described, and there were major differences between comparison groups _____
c) Not applicable, differences not described _____
d) Not applicable, was a pre-post comparison _____

8. **Was loss to follow up (attrition) less than 35%?**
The percentage lost to follow up introduces the potential for exclusion bias, reduces the available sample size and reduces the confidence in the results obtained. The 35% cut-off is to be used as an arbitrary guide. The most important point is to determine if the reasons for drop out are likely related to the intervention or not. For example, if drop outs were due to participants being laid off and losing their health benefits (an event unrelated to the intervention), one could accept a drop-out rate higher than 35% and use the text box provided to explain this. However, if drop outs were due to poorly managed depression, then the reviewer should consider lowering the threshold of an acceptable drop-out rate
a) Yes, less than 35% _____
b) No, more than 35% _____
c) Not described _____

9. **Did the author(s) examine whether important differences existed between the remaining and drop-out participants after the intervention?**
Differential attrition of subjects poses a major threat to internal validity. Exclusion bias can result if certain subjects are systematically more likely to be lost to follow-up than others.

Comparisons should be made between drop-out and remaining participants on pre-intervention characteristics or other demographic variables, as available. When there are no statistical differences between these groups, one can be more confident that attrition bias did not occur.

a) Yes, described and no major differences between groups _____
b) Yes, described, and there were major differences between groups _____
c) Not described _____
d) Not applicable _____

Intervention characteristics

10. Was the intervention process adequately described to allow for replication?

Inadequate description of the intervention strategy makes it impossible to reproduce the intervention in another population. The setting of the intervention, (i.e., where it was carried out) what was changed and how, are important aspects to document. To be considered an adequate description, it should be clearly stated to be reproducible by others.

a) Yes, described adequately _____
b) No, not described adequately _____
c) No, not described _____

11. Was there any potential for contamination and/or co-intervention?

Is it likely that participants received unintended intervention (contamination or co-intervention) that may influence the results? Please use the text box to describe.

a) Yes (describe) _____
b) No _____

Intervention intensity

12. Was compliance with the intervention in all groups described and adequate?

Examining the compliance with the intervention is important in order to understand the intensity of the intervention, how adequately the intervention was received, and the likelihood that the intervention actually could have resulted in the described outcome. One aspect of evaluating compliance and intensity is examining, when applicable, how well the intervention was implemented within the workplace. This might be assessed is by looking at the extent to which the workplace actually participated in the intervention. Other sources of information (e.g. description of treatment logs, documentation on compliance with exercise) may also be described in the paper.

a) Yes, described and adequate _____
b) Yes, described but not adequate _____
c) No, not described _____

Outcomes

13. Were the instruments used to assess the outcomes valid and reliable?

If the outcomes were not collected using a systematic method and did not use an established measurement instrument, the validity of the data may be poor. If the outcomes were systematically collected with instruments with known validity and reliability properties, confidence in the outcome should be regarded as higher. Please use the text box provided

to describe any answers that are only partially a "Yes" (e.g. the instruments used were valid yet not reliable and vice versa). Please also use the text box provided to name the instrument used.

a) Yes _____
b) No _____
c) Not described _____

14. Were the outcomes described at baseline and follow-up?

Baseline is defined as "at the time of the intervention" or "information retrieved from/for years prior to the intervention" (i.e. the intervention started in 2000 and records from 1997 were reported as pre-intervention data). Follow-up is defined as "the period of time that the individual, group or initially defined population is observed following the completion of the intervention"

a) Yes, described at baseline and follow-up _____
b) No, only described at baseline _____
c) No, only described at follow-up _____

15. Was the length of follow-up three months or greater?

Length of follow-up refers to the period of time that the individual, group or initially defined population is observed following the completion of the intervention. A minimum of three months of follow-up provides a realistic window in which the intended effects of the intervention could occur, and also provides a more realistic assessment of the long-term effects of the intervention.

a) Yes _____
b) No _____

Analysis

16. Was there adjustment for pre-intervention differences (if necessary)?

Statistical adjustment allows the researchers to control for factors that may potentially confound the relationship between the intervention and outcome. Possible adjustment methods include stratifying based on the difference (i.e. separate analyses for males and females), or including the variable in the statistical model, which controls for its effect on the association of interest.

a) Yes, baseline differences were observed and adjusted for _____
b) Yes, baseline differences were observed but not adjusted for _____
c) Not applicable, differences not described _____
d) Not applicable, was a pre-post comparison _____

17. Were the statistical analyses appropriate?

Did the investigators use all the information that was available/collected when they did their analysis? Did they reasonably collect all that they could have in order to perform an optimal analysis?

a) Yes _____
b) No _____

18. Were all participants' outcomes analyzed by the groups to which they were originally allocated (intention-to-treat analysis)?

An estimated treatment effect may be biased if some participants are analyzed according to the intervention they received, rather than the intervention to which they were allocated. Intention-to-treat analysis aims to include all participants recruited into a trial and analyzes individuals according to the intervention groups to which they were originally allocated (minus missing values) irrespective of non-compliance and co-interventions. An example of "Not applicable" would be if the study design did not use random or quasi-random allocation to the intervention(s).

a) Yes _____
b) No _____
c) Not applicable _____
d) Not described _____

19. Was there a direct between group comparison?

The direct between group comparison could be a statistical test, an estimate of effect size or expressed as a magnitude of effect. There must be a clear direct comparison between the intervention group and the control group to determine the extent to which the intervention produces an effect.

a) Yes _____
b) No _____
c) Not applicable, only one group _____

20. Should this article proceed to data extraction?

Is there any fatal flaw in the article that has reduced your confidence in the results and provides a reason why this article should not proceed to data extraction? A fatal flaw is something that would invalidate the study completely, for example, a drop-out rate of 80%. Please summarize in the text box provided.

a) Yes _____
b) No (please specify why) _____

~~~~~~~~~~~~

Reminder: Inclusion/Exclusion criteria

Inclusion criteria	Exclusion criteria
Men and/or women of working age (approximately 18-65 years old) with current or remitted depression (mild or moderate). Studies in which 50% or more of the population has depression Persons with any co-morbidity as long as the study population is defined as having mild or moderate depression. Any business size (small, medium, or large), as well as any sector (health care, services, industry, mining, forestry, etc.)	Studies that focus on chronic severe depression Studies that focus on persons with a serious mental disorder or mental health issues (i.e. bipolar disorder or schizophrenia), (these studies are only eligible *if* co-morbid with depression *and* depression is the primary focus of the article, with 50% or more of the population having depression) Studies where the primary focus is on persons with alcohol or other substance abuse or dependence (these studies are only eligible *if* co-morbid with depression *and* depression is

	the primary focus of the article, with 50% or more of the population having depression) Studies in which less than 50% of the population have depression Studies that focus on military personnel and veterans. Studies that focus on seniors, the elderly population and children. If studies do not provide ages, but use terms such as "elderly" or "geriatric", and if there is no potential of a stratified sample of a working population in the article, then please exclude. Studies that focus on depression related to pregnancy unless the article is about working mothers and coming back to work after a maternity leave, and they had some kind of intervention to screen/manage post-partum depression. Studies that focus on bereavement or burnout.
Interventions or programs that are workplace-based or may be explicitly implemented and/or facilitated by the workplace.	An in-patient intervention or drug trial.
Studies with a comparison group: before-and-after comparison within the same population, comparison between different populations, or randomized trials.	Studies with only a post-intervention measure.
Primary outcomes that are relevant to employers may include: -Changes in productivity - including changes in productivity while depressed workers are still at work -Changes in sickness absence, absenteeism, worker turnover, and long-term disability -Changes in on-the-job performance and health-related performance -Changes in rates of job-related accidents, or -Economic outcomes	Articles that assess only secondary outcomes, i.e. clinical improvement of depression, general well-being (for example, SF-36), patient satisfaction, or quality of life.
Please state any reasons why this article should be included.	These articles will not proceed to QA, but will be deemed useful as background articles that may be helpful in context setting or as part of the Introduction to the report.
Please select whether the article is a review or not	Please select whether the article is review or not
Please state (by copying and pasting citation information) whether there are any relevant references	Please state (by copying and pasting citation information) whether there are any relevant references

Appendix E

Level 4: Data Extraction (DE) Questions and Reviewer Guide

Guide to the Data Extraction form for Depression in the Workplace systematic review:

Please read this guide before beginning data extraction. It may be helpful to print this guide and have it available to refer to while doing data extraction. Please extract the data from the articles you review by completing the form on Distiller and entering text in the provided areas. Please read the questions carefully, especially the instructions which provide detail on how to enter the data.

All of the questions in the Distiller form should have an answer when you are complete. If an article does not have the information necessary to answer a particular question, please enter "**NR**" (for "not reported") in the text box for that question. It is very important that all questions have answers because if we allow blank responses we will not know if an article did not have the information or a reviewer forgot to enter it. **Remember**, try not to interpret or extrapolate - just provide the data that is presented in the article. It may be useful to cut and paste the information directly from the article. This can be done for all articles except those that have been scanned. To do this, just select and copy the text you want from the pdf and paste it into the appropriate text box in Distiller.

SCREENING QUESTION

1. Should this article be excluded from DE because it does not meet our inclusion criteria for the population, intervention, comparison and outcomes?
The goal of this question is to provide an additional layer of quality control to ensure that articles have been appropriately assigned. Please see appendix A for inclusion/exclusion criteria.

a) Yes (please indicate why) _____
b) No

STUDY DESIGN AND SETTING

2. State the research question/objective(s)
Please use the exact wording from the article. If more than one objective is stated, please list all objectives. Be sure to only include the *objectives tested* and not the broader objectives described.

Open text box response _____

3. Write the last name of the first author and the year of publication
Write only the first author's last name and the year (using 4 digits) that the article was published.

a) First author's last name _____
b) Year of publication _____

4. State the jurisdiction where the study was completed
Provide all information regarding the country, province, region, state and city where the study was carried out as described in the article. Remember to enter "NR" where information is not available.

a) Country _____
b) Province or State _____
c) Region _____
d) City _____

5. Describe the source population from which the participants were recruited
Please provide a brief description of the population where participants were recruited from.

Open text box response _____

6. Describe the type of setting/workplace/work setting the study was conducted in
Please use the language from the article to describe succinctly. Describe the organization and the unit, as applicable. For example, the organization may be a hospital but the units are only surgical units in the hospital.

Open text box response _____

7. List the job titles/classification of the participants that participated in the study
Provide the level of detail given in the study for the *actual* participants (not for the potential participants from the source population) or enter “NR” if the information is not available.

Open text box response _____

8. Describe how the presence of depression among potential participants was determined
State how depression was determined, for example, a self-report screener, a physician diagnosis, etc.

Open text box response _____

9-11. Please clearly list the inclusion criteria described in the study
Describe how the study selected their worksite and/or participants for inclusion. This may be found in the setting description or in the outcome description. Please use the appropriate box to summarize the level for inclusion criteria or enter “NR” where information is not available.

9. Worksite characteristics _____
10. Individual characteristics _____
11. Other _____

12-14. Please clearly list the exclusion criteria described in the study
Describe the criteria selected for worksite and/or participant exclusion. This may be found in the setting description or in the outcome description. Please use the appropriate box to summarize the level for exclusion criteria or enter “NR” where information is not available. For example, use “individual characteristics” to describe criteria such as people on sick leave for more than one year, people with psychological co-morbidities, etc.

12. Worksite characteristics _____
13. Individual characteristics _____

14. Other _____

15. What is the study design?
Please describe any unique characteristics about the study design in the comment boxes beside the choice you make.

a) Randomized Trial _____
b) Non-randomized study (with a separate control group) _____
c) Non-randomized study (pre-post) _____
d) Other (please specify) _____

16. Was the study protocol reviewed and approved by a Research Ethics Board (REB)?
Please use the option "Unsure" if this was not reported in the paper

a) Yes
b) No
c) Unsure (not reported) _____

INTERVENTION CHARACTERISTICS

17-18. What was the intervention?
Describe the nature of the interventions provided with a specific focus on the components addressing depression. Please make sure that you describe the interventions provided to all intervention and control groups (if applicable) using the language of I (intervention) and C (control). For multiple intervention and/or control groups, please label as I_1, I_2, etc. and C_1, C_2, etc. as applicable. Please use the same group names throughout the data extraction form.

17. Intervention group(s) _____
18. Control group(s) _____

19-20. How often was the intervention applied?
Please state how often the intervention was applied. Please make sure that you describe all intervention and control groups. For multiple intervention and/or control groups, please label as I_1, I_2, etc. and C_1, C_2, etc. as applicable. Please use the same group names throughout the data extraction form.
For example: I_1 – daily; I_2 - every week, etc.

19. Intervention group(s) _____
20. Control group(s) _____

21-22. What was the duration of the intervention? (note: this is not the follow-up time but the actual duration of the intervention implementation).
Please indicate in months if possible, if not, indicate in weeks or days, or enter "NR". Please make sure that you describe all intervention and control groups. For multiple intervention and/or control groups, please label as I_1, I_2, etc. and C_1, C_2, etc. as applicable. Please use the same group names throughout the data extraction form.
For example: I_1 – computer therapy for 1 hour per week for 4 weeks; I_2 – computer therapy for 1 hour per week for 4 weeks and therapy sessions for 50 minutes weekly; C_1 – no intervention

21. Intervention group(s) _____
22. Control group(s) _____

23. Indicate the time period between the baseline measurement and all subsequent follow-up measurements

Indicate time in months if possible, if not, indicate in weeks or days, or enter "NR". For example, questionnaires were administered at 6, 12, and 18 months. Please make sure that you describe all intervention and control groups. For multiple intervention and/or control groups, please label as I_1, I_2, etc. and C_1, C_2, etc. as applicable. Please use the same group names throughout the data extraction forms.

For example: Baseline data collected on May 1st 2000. Intervention implemented June 1st 2000 and continues until June 1st 2001. Follow-up data collected on May 1st 2002. Note: this information may be presented in a number of ways (tables, figures, timelines, etc). In this example the length of follow-up is I_1 =24 months.

Open text box response _____

SAMPLE CHARACTERISTICS AT BASELINE

24. Describe the intervention group at baseline

Provide answers for each category. Enter "NR" in all text boxes where information is not available. If there is more than one intervention group, please use I_1 and I_2, to describe the responses below. When describing "g) loss to follow up", please do so for each level of follow-up: **immediate follow-up** (immediately after the intervention), **short-term follow-up** (from immediately after the intervention up to 3 months after), **intermediate follow-up** (from 3 months to 6 months after) and **long term follow-up** (from 6 months to 12 months, or longer, after).

a) Sample Size _____
b) Age (mean, SD, range, median) _____
c) % female _____
d) Education _____
e) % with depression _____
f) Any depression scale(s) used (mean, SD, range, median) _____
g) % employed/working _____
h) Loss to follow up (% or N) _____
i) Other (please describe in question 25 below)

25. Describe the intervention group, "other": _____

26. Describe the control group at baseline

Provide answers for each category. Enter "NR" in all text boxes where information is not available. If there is more than one control group, please use C_1 and C_2 to describe the responses below. When describing "g) loss to follow up", please do so for each level of follow-up: **immediate follow-up** (immediately after the intervention), **short-term follow-up** (from immediately after the intervention up to 3 months after), **intermediate follow-up** (from 3 months to 6 months after) and **long term follow-up** (from 6 months to 12 months, or longer, after).

a) Sample Size _____
b) Age (mean, SD, range, median) _____
c) % female _____
d) Education _____
e) % with depression _____
f) Any depression scale(s) used (mean, SD, range, median) _____
g) % employed/working _____
h) Loss to follow up (% or N) _____

i) Other (please describe in question 27 below)

27. Describe the control group, "other": _____

28. Describe the overall (study) group at baseline (please answer this question only if it is reported in the article)
This question is here in case the study did not separate characteristics by the intervention and control groups. Please answer this question only if the data is presented in the article, and do not calculate the data independently. If there is more than one control group, please use C_1 and C_2 to describe the responses below. When describing "g) loss to follow up", please do so for each level of follow-up: **immediate follow-up** (immediately after the intervention), **short-term follow-up** (from immediately after the intervention up to 3 months after), **intermediate follow-up** (from 3 months to 6 months after) and **long term follow-up** (from 6 months to 12 months, or longer, after).

a) Sample Size _____
b) Age (mean, SD, range, median) _____
c) % female _____
d) Education _____
e) % with depression _____
f) Any depression scale(s) used (mean, SD, range, median) _____
g) % employed/working _____
h) Loss to follow up (% or N) _____
i) Other (please describe in question 29 below)

29. Describe the overall (study) group, "other": _____

COVARIATE QUESTIONS

30. When were potential covariates/confounders measured? (please select all that apply and also list in the text box)
If covariates were measured any time prior to the intervention, this will be counted as baseline. Please check and list the covariates/confounders measured. If unsure then please describe.
For example: a) Baseline: age, b) Follow-up: age and gender.

a) Baseline (i.e. pre-intervention), please list: _____
c) Follow-up (i.e. post-intervention), please list: _____
d) Unsure (please describe) _____
e) Not Applicable (not measured) _____

31. Were covariates/confounders ultimately controlled for in the final analysis?
Covariates/confounders, for example, can include gender, age, non-work activities, education, etc. Psychosocial and organizational risk factors can include: social support, job satisfaction, control over one's job, etc. Temporal confounding factors can include the season of year (e.g. with agricultural workers).

a) Yes (please list) _____
b) No (please describe why not) _____
c) Not applicable

OUTCOMES

32. Provide a list of outcome variables used to evaluate intervention effectiveness, that are relevant to our review project

Please only list those outcomes that are relevant to our review project – see Appendix A for a list of relevant outcomes.

When listing the outcome variables, please identify the following three issues for each: **a)** what the outcome(s) are, **b)** how the outcome(s) were assessed and **c)** when the outcome(s) were measured. Please also be sure to indicate if any measures of presenteeism or absenteeism were used in the study. Presenteeism refers to how health impacts on-the-job productivity. Absenteeism can mean any measure of work absence from sick leave, short-term disability and long term disability.

Open text box response _____

33. Were direct and indirect costs associated with the intervention measured?

If direct costs were measured, please list those cost items and dollar value. Direct costs refer to observable expenditures. For example, any equipment that was purchased, the costs of professional services, etc. Indirect (opportunity) costs refer to costs that are implied by an action but not directly observed. For instance, someone could attend college for a year and the direct costs would be tuition and books. The indirect costs would be the foregone income from the job that you would have worked had you not gone to college. In the context of these interventions the likely relevant indirect costs would be the value of the employees' time if they are required to participate in training.

a) Yes, direct costs were measured (please list the cost items and dollar value) _____

b) Yes, indirect (opportunity) costs were measured (please list the cost items and dollar value) _____

c) No, direct and indirect costs were not measured

34. Were any outcome measures monetized (converted into a dollar figure)?

Please list whether any outcome measures were created into a monetary figure.

a) Yes, (please list the measures) _____

b) No _____

35. Was a cost-effectiveness analysis (CEA) conducted?

Cost-effectiveness studies have a non-monetary measure in the numerator and a dollar cost in the denominator. For instance, "in a clinical trial the antidepressant drug Nortriptyline created two additional depression-free days/dollar cost."

a) Yes _____

b) No _____

36. Was a cost-benefit analysis (CBA) used?

Cost-benefit analyses use monetary measures both for the outcome and for the costs. Usually this takes the form of a benefit to cost ratio. For example, "for every dollar spent on Nortriptyline there is $2 in savings from lowered health expenditures."

a) Yes _____

b) No _____

37. If the answer to question 35 or 36 was "Yes", was the CEA or CBA done from the point of view of society or of the employer?

There are benefits from interventions that may not be captured by the employer. For instance, consider how one would value the benefit of a reduction in suicide risk. One could consider this only as reducing the probability of incurring costs with having to hire a new employee. That would be the employer perspective. Or, one could value all of the benefits of the intervention irrespective of which party experiences the benefits. This would be the societal perspective. In the example above, the societal perspective would include employer benefits and the statistical value of the lives saved.

a) Point of view of society _____
b) Point of view of the employer _____
c) Not applicable, no CEA or CBA was used _____

38. What time frame was used for the CEA or CBA?

Please indicate the time frame used for CEA or CBA, as stated in the article. Or write "N/A" for not applicable.

Open text box response _____

39. Are the results sensitive to the time frame used?

The benefit to cost ratios can often be very sensitive to the time frame used to measure benefits. For instance suppose the intervention costs $200 as a one time expenditure but results in a monthly benefit of $10. A one year time frame would result in a benefit to cost ratio of (benefits = 12 x $10 = $120)/(costs = $200) = 0.6 and thus one would conclude that the intervention was not recommended. A two year time frame would result in a benefit to cost ratio of 1.1 ($220/$200) implying that the intervention was a success.

a) Yes _____
b) No _____
c) Not reported _____

40. If the answer to question 36 was "Yes", given the flow of benefits and the costs, did the authors calculate how long it would take to recoup the costs?

a) Yes _____
b) No _____
c) Not applicable (answer to question 36 was "No")

41. Was there an inflation adjustment?

a) Yes _____
b) No _____

42. Did the CEA or CBA perform discounting?

Discounting is the process of converting future dollars and future health outcomes to their present value. Discounting is typically used in economic evaluations with follow-up periods of greater than one year. The discount rate is the interest used to compute the present value, or (in other words) the interest rate used to discount future sums.

a) Yes (please enter the discount rate as a percentage) _____
b) No _____

c) Not applicable _____

STATISTICAL ANALYSIS AND RESULTS

43. Please check the types of final analyses done for testing the observed effects of the intervention from the list below and provide details for which outcome in the text box
You should select the option that represents the final test, not the preliminary analyses. Provide details of which outcome in the text box. Please provide details if you select "other".
For example: a) ANOVA: for days of work, b) Percentage of change: for presenteeism, etc.

a) ANOVA (ANCOVA) _____
b) MANOVA _____
c) Linear Regression _____
d) Logistic Regression _____
e) Survival Regression _____
f) Poisson Regression _____
g) Percentage of change _____
h) Nonparametric tests _____
i) Nonparametric matched test _____
j) Nonparametric unmatched test _____
k) Other parametric matched test _____
l) Other parametric unmatched test _____
m) No statistical test _____
n) Other (please describe) _____

44. Describe for each outcome of interest, the observed intervention effects
Be brief and concise, for instance, enter "effect size", "risk ratio", "rate differences", "mean differences", etc, the actual number and associated outcome. If there is more than one outcome of interest please number them as done in the form previously.
For example.: I_1 – LWD Rate 13% change pre vs post, I_1 = left arm RR 1.3

Open text box response _____

45. Were additional statistical analyses conducted to increase your confidence in the observed effects?
For example, if there was a significant loss to follow-up and/or movement between study arms then an intention-to-treat analysis may be appropriate.

a) Yes (please describe) _____
b) No

46. Remark on the findings or enter information that is unique about the study that may not be adequately captured in the other DE questions.
Be clear and concise. Please note that this is your last opportunity to provide overall comments on the study.

Open text box response _____

ARTICLE CONSENSUS

47. Is this the consensus (final) version of the DE form?
Please select "no" until absolutely all conflicts have been resolved and consensus has been completed.

a) Yes
b) No

~~~~~~~~~~

Reminder: Inclusion/Exclusion criteria

Inclusion criteria	Exclusion criteria
Men and/or women of working age (approximately 18-65 years old) with current or remitted depression (mild or moderate). Studies in which 50% or more of the population has depression Persons with any co-morbidity as long as the study population is defined as having mild or moderate depression. Any business size (small, medium, or large), as well as any sector (health care, services, industry, mining, forestry, etc.)	Studies that focus on chronic severe depression Studies that focus on persons with a serious mental disorder or mental health issues (i.e. bipolar disorder or schizophrenia), (these studies are only eligible *if* co-morbid with depression *and* depression is the primary focus of the article, with 50% or more of the population having depression) Studies where the primary focus is on persons with alcohol or other substance abuse or dependence (these studies are only eligible *if* co-morbid with depression *and* depression is the primary focus of the article, with 50% or more of the population having depression) Studies in which less than 50% of the population have depression Studies that focus on military personnel and veterans. Studies that focus on seniors, the elderly population and children. If studies do not provide ages, but use terms such as "elderly" or "geriatric", and if there is no potential of a stratified sample of a working population in the article, then please exclude. Studies that focus on depression related to pregnancy unless the article is about working mothers and coming back to work after a maternity leave, and they had some kind of intervention to screen/manage post-partum depression. Studies that focus on bereavement or burnout.
Interventions or programs that are workplace-based or may be explicitly implemented and/or facilitated by the workplace.	An in-patient intervention or drug trial.
Studies with a comparison group: before-and-after comparison within the same population,	Studies with only a post-intervention measure.

comparison between different populations, or randomized trials.	
Primary outcomes that are relevant to employers may include: **-Changes in productivity - including changes in productivity while depressed workers are still at work** **-Changes in sickness absence, absenteeism, worker turnover, and long-term disability** **-Changes in on-the-job performance and health-related performance** **-Changes in rates of job-related accidents, or** **-Economic outcomes**	Articles that assess only secondary outcomes, i.e. clinical improvement of depression, general well-being (for example, SF-36), patient satisfaction, or quality of life.
Please state any reasons why this article should be included.	These articles will not proceed to QA, but will be deemed useful as background articles that may be helpful in context setting or as part of the Introduction to the report.
Please select whether the article is a review or not	Please select whether the article is review or not
Please state (by copying and pasting citation information) whether there are any relevant references	Please state (by copying and pasting citation information) whether there are any relevant references

Appendix F

Evaluation criteria for methods of depression identification

Category	Definition	Criteria for Grades
Reliability	Does the tool give a consistent answer?	Internal Consistency Coefficient Alpha *** ≥ 0.80 ** < 0.80, > 0.70 * < 0.70
Validity	Does the tool measure what it purports to measure, in this study of the diagnosis of depression? Either comparison to a gold standard (psychiatrist/psychologist structured interview), or another diagnostic test	ROC Analysis, AUC *** ≥ 0.90 ** < 0.90, > 0.80 * < 0.80 OR if ROC analysis not available Correlation, r *** ≥ 0.60 ** < 0.60, > 0.30 * < 0.30
Respondent Burden	Time taken to complete the instrument.	*** <5 minutes ** 5 minutes to 15 minutes * ≥ 15 minutes
Cost and Availability		*** Freely available for download ** Free but have to ask permission from publishers * Any cost
Languages		*** English, French, and others (≥ 5 total) ** English, French only, or ≥ 5 total but excluding French * English only

Criteria for rating the instruments were modified from Andresen, 2000
For each measure: excellent= ***, adequate= **, poor= *

Appendix G

Social security systems – Europe and Ontario

The Netherlands

Type of program: Social insurance system.

Coverage:
Cash sickness and maternity benefits: Coverage is mostly through private providers. (Under the Civil Code, employers must pay 70% of wages during sick leave periods for up to 104 weeks.)
Social insurance covers workers who have no employer **(self-employed)** or no longer have an employer (and, in a few special circumstances, wage earners and salaried employees), including employees who have lost their jobs in the first 2 years of incapacity, incapacitated unemployed persons, temporary workers on sick leave without a permanent contract, the voluntarily insured, apprentices, organ donors, vocationally rehabilitated persons, and women incapacitated due to pregnancy or childbirth. (Entrepreneurs and directors with a major shareholding in a company are excluded.)

Source of Funds
Insured person: A flat-rate contribution set by the private insurer on annual earnings up to €33,189; 12.15% of annual earnings up to €32,738 for exceptional medical expenses insurance.
For sickness and maternity benefits, see source of funds under Unemployment Benefits.
Self-employed person: A flat-rate contribution set by the private insurer for medical benefits; 4.95% of taxable income up to €33,189 a year for medical benefits; 12.15% of income up to €32,738 a year for exceptional medical expenses insurance.
Employer: 7.05% of covered payroll for medical benefits.
The maximum annual earnings used to calculate contributions are €33,189
For sickness and maternity benefits, see source of funds under Unemployment Benefits.
Government: An annually determined contribution for medical benefits.

Sickness benefit: The benefit is 70% of earnings, up to €186.65 a day, and is paid for up to 104 weeks; may be extended for an additional 52 weeks.

(all data from: http://www.ssa.gov/policy/docs/progdesc/ssptw/2010-2011/europe/index.html)

Denmark

Type of program: Universal (medical benefits) and employment-related (cash benefits) system.

Coverage
Cash sickness and maternity benefits: All employed and self-employed persons.
Medical benefits: All persons residing in Denmark.

Source of Funds
Insured person: None.
Self-employed person: Voluntary contributions to finance cash benefits during the first 2 weeks of incapacity.

Employer: The total cost of cash benefits for the first 3 weeks of incapacity if the employee worked for the same employer for at least 8 weeks before the incapacity began. No contribution is made for medical benefits.
The employer's contributions also finance temporary disability benefits under Work Injury.
Government: Local (municipal) government meets the total cost of cash benefits from the third week (from day 1 if the insured is ineligible for the 3-week benefit from the employer). Local (municipal) government is reimbursed fully by central government up to the end of the fourth week; thereafter, the cost is split equally between local and central governments. Local government (county level) finances the total cost of medical benefits.
Government contributions also finance temporary disability benefits under Work Injury.

Sickness benefit: Up to 3,760 kroner a week is paid, based on the insured's hourly wage; for employees, the benefit is paid from the first day of incapacity; for self-employed persons, the benefit is paid from the third week of incapacity (may insure voluntarily for the first 3 weeks). The employer is reimbursed by local government for the cost of sickness benefits paid directly to employees (the benefit paid for the first day of incapacity is not reimbursed.)
The weekly benefits provided under the national cash benefit program are paid for 52 weeks within any 18-month period; may be extended under specified circumstances.
Local government assesses the incapacity every 8 weeks.
Partial benefit: A reduced benefit is paid for a partial incapacity to work.

(all data from: http://www.ssa.gov/policy/docs/progdesc/ssptw/2010-2011/europe/index.html)

Finland

Type of program: Social insurance system.
Note: Health care is provided by both a private-sector sickness insurance program and a public-sector (municipal) health services program financed primarily by local and national taxes.

Coverage
Cash sickness and maternity benefits: All persons residing in Finland.
Medical benefits: All persons residing in Finland.

Source of Funds
Insured person
Cash sickness and maternity benefits: 0.93% of gross monthly earnings.
Medical benefits: 1.47% of gross monthly earnings; 1.64% of gross monthly earnings for pensioners and other beneficiaries.
Contributions are calculated on all earnings.
Self-employed person
Cash sickness and maternity benefits: 0.93% or 1.05% of gross monthly earnings.
Medical benefits: 1.47% of net monthly earnings.
Contributions are calculated on all earnings.
Employer
Cash sickness and maternity benefits: 2.23% of monthly payroll (private employers and local and central government).
Medical benefits: 2.23% of monthly payroll.
Government
Cash sickness and maternity benefits: Subsidies as required; 100% of the cost of minimum daily allowances.
Medical benefits: 50% of the cost of medical benefits.

Sickness benefit: The benefit is 70% of daily earnings for annual earnings up to €32,892, 40% of daily earnings for annual earnings of €32,893 to €50,606, and 25% of daily earnings for annual earnings of €50,607 or more.
The benefit is paid after a 10-day waiting period for up to 300 days (excluding Sundays). (The employer pays 100% of earnings for the first 9 days for employees who have worked for at least a month; otherwise, 50%.) Insured persons who have been unemployed during the last 4 months receive at least 86% of the unemployment benefit.
The minimum daily benefit is €22.04.
Sickness allowance (means-tested): The allowance is paid after 55 days of incapacity provided that annual earnings are less than €1,264. The daily benefit is €22.04.
Rehabilitation benefit: The benefit is 70% of daily earnings for annual earnings up to €32,832, 40% of daily earnings for annual earnings of €32,893 to €50,606, and 25% of daily earnings for annual earnings of €50,607 or more.
Special sickness benefit: The allowance is paid for up to 60 days in hospital and 60 days at home (90 days if the treatment is ongoing). The benefit is 70% of daily earnings for annual earnings up to €32,892, 40% of daily earnings for annual earnings of €32,893 to €50,606, and 25% of daily earnings for annual earnings of €50,607 or more.
The minimum daily benefit is €22.04.

(all data from: http://www.ssa.gov/policy/docs/progdesc/ssptw/2010-2011/europe/index.html)

The Ontario system of work disability

The Employment Standards Act (ESA) 2000 sets out the rights and responsibilities of both employees and employers in Ontario workplaces. Employees who work for employers that regularly employ at least 50 employees are entitled to a **leave of absence** called “personal emergency leave” in certain situations. Personal emergency leave is unpaid, job-protected leave of up to 10 days each year. It may be taken in the case of a personal illness, injury or medical emergency, or a death, illness, injury, medical emergency of, or urgent matter relating to, certain relatives. Some employers have paid benefit plans for sickness, bereavement and other leaves of absence. These plans aren't required by the ESA.

When the illness is prolonged and the worker needs additional time off, he/she may apply to a **private disability benefit** that is usually arranged between the employer and a private insurance company, therefore it is highly variable in terms of waiting period, benefits, and duration. The worker will need a physician’s letter indicating the nature of the illness and the expected prognosis. Generally, the file is reviewed after a certain period (variable between 30 days and 2 years) when the disability benefit is divided into short-term (STD) or long-term disability (LTD). In general, LTD policies any type of injury or illness that prevents the person from being able to work, but this coverage varies from policy to policy. Some policies exclude illnesses that are work-related injuries (which are covered by the Ontario Workplace Safety and Insurance Board – WSIB) Most plans provide a two-year period of benefits for persons unable to perform their pre-disability occupation and who have medical evidence of a permanent or long-duration impairment; thereafter beneftis are only provided if the beneficiary is unable to perform any occupation for which the person is reasonably trained or educated.

Usually, if the worker is still on private disability benefits after a period of two years, he/she is encouraged to apply to the Canadian Pension Plan (**CPP) disability benefits**. The Canada Pension Plan is a contributory, earnings-related social insurance program. It ensures a measure

of protection to a contributor and his or her family against the loss of income due to retirement, disability and death.

The **Ontario Disability Support Program** (ODSP) Income Support helps people with disabilities who are in financial need pay for living expenses, like food and housing. To be eligible the individual must be 18 years of age or older, live in Ontario, be in financial need, and have a substantial physical or mental disability that is expected to last a year or more, which makes it hard for to care for oneself, take part in community life or work. To determine eligibility to receive Income Support, ODSP will look at financial situation, and disability status.

Service Canada on behalf of Human Resources and Skills Development Canada (HRSDC) offers the **Employment Insurance (EI) Sickness Benefit** provides benefits for a maximum of 15 weeks for periods of temporary disability. EI Sickness Benefits are administered as a "last-payer" benefit program, i.e. they are reduced where beneficiaries receive contributions from workers' compensation, group insurance income, accident compensation for loss of wages, CPP-D and provincial social assistance programs. In order to qualify, the worker must show that regular weekly earnings have decreased more than 40 per cent due to disability, and that 600 insured hours have accumulated over the last 52 weeks, or since the worker's last claim.

There are other systems available to Canadians with disability. For more information see "A patchwork quilt: Income security for Canadians with disabilities" at http://www.iwh.on.ca/briefings/a-patchwork-quilt

Appendix H

Table 1: Study Characteristics

First Author Year Country	Research Question	Study Design	Setting/ Workplace Setting	Inclusion Criteria	Exclusion Criteria
1. Blonk 2006 The Netherlands*	To investigate the effectiveness of two interventions (extensive CBT by psychotherapists and a brief CBT-derived intervention by labour experts with a focus on the workplace) in a sample of self-employed individuals reporting sick to their insurance company owing to work-related psychological complaints	Randomized trial	At home or at the workplace of self-employed participants	1. Met criteria for presence of an adjustment disorder, such as burnout and job stress, on the shortened version of the (WHO-CIDI)	1. Suffering from a serious psychiatric disorder (major depression, addictive disorders, posttraumatic stress disorder, or other anxiety disorders) 2. Individuals who did not want to postpone their current psychiatric treatment
2. Dewa 2009 Canada*	To examine the cost, effectiveness and cost-effectiveness of a collaborative mental health care pilot program for people on short-term disability leave for psychiatric disorders	Non-randomized study (with a separate control group)	A large financial insurance services company	1. A disability leave related to a psychiatric disorder 2. Prior history of psychiatric illness but did not have disability leave in the year before the demonstration project's implementation 3. Not under the care of a psychiatrist 4. No terminal illness 5. Would have been referred for an independent medical evaluation under usual disability management practices	NR

First Author Year Country	Research Question	Study Design	Setting/ Workplace Setting	Inclusion Criteria	Exclusion Criteria
3. Kawakami 1997 Japan*	To determine the effects of a stress reduction program on any decrease in depressive symptoms and to examine changes in systolic and diastolic blood pressures and sick leave during the stress reduction program among blue-collar workers	Non-randomized study (with a separate control group)	Five worksites from a large electric company	Individual: 1. All workers in the identified worksites were included Worksite: 1. Worksites with mean depression scores from a community-wide stress survey higher than the mean plus 1 standard deviation 2. Among the worksites identified, worksites that first appeared in the worksite directory 3. Control worksites were selected for matching mean age, major products and occupations, worksite size and mean depression scores	NR
4. Knekt 2008a Finland* Knekt, 2008b	To study improvement in work ability and functional capacity due to solution-focused therapy and short-term and long-term psycho-dynamic psychotherapy among patients suffering from depressive and anxiety disorders To compare the effectiveness of long- and short-term psychodynamic psychotherapy as well as solution-focused therapy in the treatment of depressive and anxiety disorders	Randomized trial	Psychiatric service centres	1. 20-45 years of age 2. Suffering from a long-standing (>1 year) disorder causing dysfunction in work ability 3. Met DSM-IV criteria for anxiety or mood disorders 4. Psychodynamic assessment of suffering from neurosis to higher level borderline disorder, according to Kernberg's classification of personality organization	1. Psychotic disorder or severe personality disorder, adjustment disorder, or substance-related disorder (according to DSM-IV criteria) 2. Organic brain disease or other diagnosed severe organic disease 3. Mental retardation 4. Treated with psychotherapy within the previous 2 years 5. Psychiatric health employees

First Author Year Country	Research Question	Study Design	Setting/ Workplace Setting	Inclusion Criteria	Exclusion Criteria
5. Krogh 2009 Denmark* Supplemental: Krogh, 2007	To assess the benefits and harms of strength versus aerobic versus relaxation training in patients with depression	Randomized trial	Copenhagen University Hospital	1. 18-55 years of age 2. Referred by a medical doctor or psychologist 3. Fulfilling the ICD-10 criteria for unipolar depression (F32.0, F32.1, F33.0, ICD-10 verified F33.1) using the Major Depression Inventory 4. Living in Greater Copenhagen 5. Able to read and understand the consent statement	1. Engaging in regular sports activity for more than 1 hour per week 2. Ongoing alcohol or substance abuse 3. At risk of suicide (score >2 on item 3 of the HAM-D17) 4. Poor Danish language skills 5. A medical condition that contraindicated physical exercise 6. On sickness leave for more than 24 consecutive months 7. Psychotic symptoms (as measured using the SCL-92)

First Author Year Country	Research Question	Study Design	Setting/ Workplace Setting	Inclusion Criteria	Exclusion Criteria
6. Lo Sasso 2006 USA*, ** Rost 2004 USA*, ** Supplemental: Rost, 2000 Rost, 2001	To test whether an intervention to improve primary care depression management significantly improves productivity at work and absenteeism To construct a cost-benefit analysis of this treatment under different workplace assumptions better reflecting the nature of employment	Randomized trial	Community primary care practices	Individual: First stage: 1. 18 years of age or more 2. Not pregnant, breastfeeding or less than 3 months post-partum 3. Sufficient literacy in English and cognitive function to complete surveys requiring 6 month recall 4. No acute life threatening physical conditions 5. Access to a telephone 6. Experienced ≥2 weeks during last year of feeling sad, empty, depressed, or loss of interest in things they normally enjoyed (WHO-CIDI) 7. Reported ≥1 week of the above symptoms during the past month (WHO-CIDI) Second stage: 1. Reported 5 or more of the 9 DSM-III Revised criteria for major depression in the past two weeks (IDD) 2. Could be currently taking anti-depressants and/or seeing a specialist 3. Could report suicidal ideation Other Criteria: 1. Practices must employ two primary care physicians willing to participate in the study, a nurse willing to deliver the nursing intervention, and administrative staff willing to screen patients	First stage: 1. Depressive symptoms that began after the loss of a loved one within the last 2 months (WHO-CIDI) 2. No intention of receiving ongoing care in the clinic during the next year Second stage: 1. Screening positive by self-report for lifetime mania, use of lithium, or current alcohol dependence Other criteria: 1. Practices where primary care physicians routinely could refer depressed patients to mental health specialists employed by the practice

First Author Year Country	Research Question	Study Design	Setting/ Workplace Setting	Inclusion Criteria	Exclusion Criteria
7. Rebergen 2009a The Netherlands* Rebergen 2009b The Netherlands* Supplemental: Rebergen, 2007	To evaluate the effectiveness of guideline-based care (GBC) of workers with mental health problems, which promotes counselling by Occupational Physicians (OPs) to facilitate RTW To conduct an economic evaluation of GBC in reducing productivity loss costs, from both a societal and a company perspective	Randomized trial	Occupational health service	1. Mental health problems as per OP diagnosis 2. Was on sick leave due to mental health problems at the time of inclusion 3. Sick leave did not start before 2002	1. Mental health symptoms caused by somatic illness 2. Disagreement between OP and employee about the diagnosis 3. Lack of confidence in the relation between OP and employee
8. Schene 2007 The Netherlands*	To conduct an RCT comparing treatment as usual versus treatment as usual plus occupational therapy (OT), to determine the impact of the addition of OT to treatment as usual on recovery from depression, work reintegration, and work stress	Randomized trial	An outpatient mood disorder clinic	1. >18 years of age 2. Met DSM-IV criteria for major depressive disorder, single episode or recurrent, without psychotic features 3. A BDI score >15 4. Work reduction of at least 50% of regular hours worked per week because of depression, for a minimum of 10 weeks and a maximum of 2 years	1. History of psychosis 2. Manic, hypomanic, or cyclothymic features 3. History of active drug or alcohol abuse or dependence

First Author Year Country	Research Question	Study Design	Setting/ Workplace Setting	Inclusion Criteria	Exclusion Criteria
9. Schoenbaum 2001 USA* Supplemental: Schoenbaum, 2004 Jaycox, 2003 Miranda, 2003 Sherbourne, 2001 Wells, 1999 Wells, 2000	To determine the cost-effectiveness from a societal perspective of two quality improvement interventions to improve treatment of depression in primary care and their effects on patient employment relative to usual care	Randomized trial	Primary care clinics	Individual: 1. 17 years of age or more 2. Intention to use the clinic as a source of care for the next year 3. Fluent in English or Spanish 4. Had insurance or a public-pay arrangement that covered the intervention care 5. Reported at least 1 week of depressed mood in the last 30 days, plus 2 or more weeks of depressed mood or loss of interest in pleasurable activities in the last year or persistent depression over the year (WHO-CIDI) Other Criteria: 1. Primary care practices in one of the 6 managed care organizations 2. Practices with at least 2 clinicians	1. An acute medical emergency 2. Did not have an eligible insurance plan

First Author Year Country	Research Question	Study Design	Setting/ Workplace Setting	Inclusion Criteria	Exclusion Criteria
10. Smith 2002 USA* Supplemental: Rost, 2000 Rost, 2001	To assess the impact of the Quality Enhancement by Strategic Teaming (QuEST) intervention on subsequent employment and workplace conflict among depressed primary care patients	Randomized trial	Community primary care practices	Individual: First stage: 1. 18 years of age or more 2. Not pregnant, breastfeeding or less than 3 months post-partum 3. Sufficient literacy in English and cognitive function to complete surveys requiring 6 month recall 4. No acute life threatening physical conditions 5. Access to a telephone 6. Experienced ≥2 weeks during last year of feeling sad, empty, depressed, or loss of interest in things they normally enjoyed (WHO-CIDI) 7. Reported ≥1 week of the above symptoms during the past month (WHO-CIDI) Second stage: 1. Reported 5 or more of the 9 DSM-III Revised criteria for major depression in the past two weeks (IDD) 2. Could be currently taking anti-depressants and/or seeing a specialist 3. Could report suicidal ideation 4. Need to be employed at baseline 5. <64 years of age at baseline Other Criteria: 1. Practices must employ two primary care physicians willing to participate in the study, a nurse willing to deliver the nursing intervention, and administrative staff willing to screen patients	First stage: 1. Depressive symptoms that began after the loss of a loved one within the last 2 months (WHO-CIDI) 2. No intention of receiving ongoing care in the clinic during the next year Second stage: 1. Screening positive by self-report for lifetime mania, use of lithium, or current alcohol dependence Other criteria: 1. Practices where primary care physicians routinely could refer depressed patients to mental health specialists employed by the practice

First Author Year Country	Research Question	Study Design	Setting/ Workplace Setting	Inclusion Criteria	Exclusion Criteria
11. van der Feltz-Cornelis 2010 The Netherlands* Supplemental: van der Feltz-Cornelis, 2007	To test the effectiveness of psychiatric consultation aimed at diagnosis and treatment of common mental disorders in employees on sick leave with a focus on RTW, as compared to care as usual	Randomized trial	Two occupational health services, related to various companies	1. If after at least 6 weeks absenteeism, no plan for RTW within another 6 weeks 2. PHQ score >8 on the depression subscale, ≥8 on the panic disorder subscale, or >3 on the generalized anxiety disorder subscale OR WI score of >3 for somatoform disorders	1. Suicidal 2. Addicted to drugs or alcohol 3. Psychotic 4. Suffering from dementia 5. Insufficient knowledge of the Dutch language to complete the questionnaires 6. Involved in a legislative procedure for unemployment compensation 7. On sick leave for longer than 52 weeks
12. Wang 2007 USA*	To evaluate the effects of a depression outreach-treatment and care management program on workplace outcomes (depression symptom relief, job retention, decreased sickness absence and work productivity)	Randomized trial	Sixteen companies covered by a specific managed behavioural health care company	First stage: 1. 18 years of age or more 2. K-6 psychological distress score of ≥9 Second stage: 2. A QIDS-SR score of >8 (at least moderate depression severity)	Second stage: 1. Positive responses to a Composite International Diagnostic Interview (CIDI) short-form screening for a history of mania or substance dependence 2. Suicidal ideation or attempts in the prior week 3. Treatment by a mental health specialist in the prior year

* Denotes primary articles from which the data was extracted. All references, primary and supplemental, may be found in the bibliography.

** Denotes articles based on the same overall study, but using different sub-samples of the original recruited sample.

NR: Not reported, **RCT**: Randomized Controlled Trial, **DSM-IV**: Diagnostic and Statistical Manual of Mental Disorders-IV, **RTW**: Return to Work, **BDI**: Beck Depression Inventory, **ICD**: International Classification of Diseases, **CBT**: Cognitive Behavioural Therapy, **WHO-CIDI**: World Health Organization's Composite International Diagnostic Interview, **HAM-D17:** 17-item Hamilton Rating Scale for Depression, **SCL-92**: Symptom Checklist, **IDD**: Inventory to Diagnose Depression, **QIDS-SR**: Quick Inventory of Depression Symptoms Self-Report, **PHQ**: Patient Health Questionnaire, **WI**: Whitely Index

Table 2: Intervention Characteristics

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
1. Blonk, 2006	I1 (n=40): Combined intervention consisting of brief CBT-based stress management administered by labour experts. Consisted of psycho-education on work stress, registration of symptoms and situations, relaxation, self-help books, time-management, and writing and homework assignments. Also provided advice about work processes and provided suggestions on how to lower workload and job demands and to increase decision latitude. Also encouraged partial return to work I2 (n=40): The CBT intervention was given by psychologists who followed a highly structured protocol. Six initial sessions focused on cognitive restructuring and registration of symptoms and situations, followed by five sessions focusing further on cognitive restructuring, as well as on work resumption, time-management, workplace interventions, conflict management, and fatigue	C (n=42): Two brief sessions with a GP whose role was to check the validity of the work disability claim, no actual treatment	I1: Twice-weekly sessions, 5 -6 sessions in total. I2: Twice-weekly sessions, 11 sessions in total	2 visits	I1: 1 hour sessions, over 3 weeks I2: 45 minute sessions, over 5 -6 weeks	"Brief" GP sessions after the claim was initiated, and 4 months later	4, 10, and 12 months	4 months: I1: 7 (18%) I2: 8 (20%) 10 months: I1: 10 (25%) I2: 10 (25%)	4 months: 9 (21%) 10 months: 14 (33%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
2. Dewa, 2009	I (n=73): Enhanced disability management process created by adding a Collaborative Mental Health Care (CMHC) component delivered by psychiatrists. Individuals who met criteria for an independent medical exam were referred to a CMHC psychiatrist. The program structure was based on collaborative care concepts, including psychiatric assessment and treatment recommendation, short-term management by the psychiatrist (if referred by primary care physician), psychiatric support of management by the primary care physician, and the availability of psychiatric consultation for non-referred workers. The goal of was to return patient's care to primary care physician as soon as possible.	C (n=51): Usual practice. Individuals who met criteria for an independent medical exam were referred to a third-party psychiatrist for the purpose of adjudicating the claim only. Once the diagnosis and severity of the disorder were established, the primary care physician continued to treat the employee until they returned to work. The employee was referred to a psychiatrist at the discretion of the primary care physician	For workers referred to CMHC psychiatrists by their primary care physicians, psychiatrists provided 2-4 sessions Visits to primary care physician, NR	Visits to primary care physician, NR	NR	NR	12 months	0 (0%)	0 (0%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
3. Kawakami, 1997	I (n=110): A worksite stress reduction program in which supervisors were asked to list possible work stressors in their worksites and to make plans to reduce these stressors while a working committee made the plans feasible. The supervisors started stress reduction activities and the committee monitored their activity periodically	C (n=175): No intervention. No activities for reducing work stress were conducted	NR	NR	12 months	12 months	12 and 24 months	12 months: NR 24 months: 31 (28%)	12 months: NR 24 months: 67 (38%)
4. Knekt, 2008a and 2008b	I1 (n=97): Solution-focused therapy. Brief and resource-oriented. Conducted by a therapist, a goal-focused therapeutic approach which helps clients change by constructing solutions I2 (n=101): Short-term psychodynamic psychotherapy. Conducted by a therapist, a brief transference-based approach which helps patients by exploring and working through intra-psychic and interpersonal conflicts	C (n=128): Long-term psychodynamic psychotherapy. Conducted by a therapist, an open-ended, intensive, transference-based therapeutic approach which helps patients by exploring and working through a broad area of intrapsychic and interpersonal conflicts	I1: 1 session every 2nd or 3rd week to a limit of 12 sessions I2: 20 weekly treatment sessions	2-3 times a week	I1: Up to 8 months I2: 5-6 months	Up to 3 years	3, 7, 9, 12, 18, 24, and 36 months	12 months: I1: 8 (8%) I2: 10 (10%) 24 months: I1: 26 (27%) I2: 18 (18%) 36 months: I1: 21 (22%) I2: 18 (18%)	12 months: 13 (10%) 24 months: 26 (20%) 36 months: 21 (16%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
5. Krogh, 2009	I1 (n=55): Strength training group. Designed to increase muscular strength with initial repetitions of 50% of repetition maximum (RM) and participants progress to 75% of RM. It is a circle-training program involving large muscle groups including machines, free weights, and sand bags. I2 (n=55): Aerobic training group. Designed to increase fitness as measured by maximal oxygen uptake (VO2max). The exercise program involves 10 different aerobic exercises using large muscle groups and using a variety of equipment, including machines, trampolines, and jump rope. Initial sessions involve exercises done at 70% of maximal heart rate, gradually increasing to those performed at 89% of maximal heart rate.	C (n=55): Relaxation training group. The goal is to avoid muscular contractions or stimulation of the cardiovascular system and patients do not engage in activity perceived higher than 12 on the Borg scale. Includes exercises on mattresses, light balance exercises, and relaxation exercises	For I1 and I2: 2 times per week for a total of 32 sessions	2 times per week for a total of 32 sessions	For I1 and I2: Sessions 1.5 hours each for 4 months	Sessions 1.5 hours each for 4 months	4 and 12 months	4 months: I1: 8 (15%) I2: 7 (13%) 12 months I1: 9 (16%) I2: 9 (16%)	4 months: 13 (24%) 12 months: 18 (33%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
6. Lo Sasso, 2006 Rost, 2004	I1 (n=96) and I2 (n=62): Enhanced care among consistently employed* (I1) and among inconsistently employed** people (I2). Physicians were informed when patients screened positive for depression. Nurses/physicians were trained to provide high quality depression care during the acute and continuation phase of treatment. Training emphasized the need to encourage patients to initiate guideline-concordant pharmacotherapy or psychotherapy during the acute phase of treatment. Nurses contacted patients to assess symptoms, educated them about depression and treatment, and gave homework assignments to increase their readiness to engage in active treatments. Training also emphasized the need to encourage continued treatment adherence if symptoms were resolving, to adjust treatment if symptoms were not resolving, or terminate treatment when remitted patients did not require maintenance therapy following the continuation phase. Physicians received symptom/treatment status summaries, along with reminders to adjust treatment for symptomatic patients.	C1 (n=102) consistently employed: Usual care C2 (n=66) inconsistently employed: Usual care Physicians were not informed when patients screened positive for depression. No regular nurse contacts during initial or continuation phases of treatment.	Patients returned after the index visit and then for weekly sessions with the nurse Sessions followed by regular telephone contact (those with ≥3 out of 9 depressive symptoms were contacted monthly by nurses, those with ≤2 out of 9 symptoms were contacted every 3 months) Physicians monitored patients monthly.	NR	10-15 minute sessions for 5-7 weeks Telephone contacts over 1 year Duration of physician's contacts, NR	NR	6, 12, 18 and 24 months	I1: 6 months: 15 (16%) 12 months: 25 (26%) 18 months: 33 (34%) 24 months: 38 (40%) I2: 6 months: 4 (6%) 12 months: 5 (8%) 18 months: 12 (19%) 24 months: 13 (21%)	C1: 6 months: 7 (7%) 12 months: 14 (14%) 18 months: 23 (23%) 24 months: 24 (24%) C2: 6 months: 2 (3%) 12 months: 1 (2%) 18 months: 8 (12%) 24 months: 13 (20%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
7. Rebergen, 2009a Rebergen, 2009b	I (n= 125): Guideline-based care (GBC) by Occupational Physicians (OPs). Involved treatment by OPs according to the Dutch guideline of employees on sick leave due to mental health problems. The guide promotes a more active role of the OP as case and care manager facilitating RTW of the employee. OPs received training in the guideline. The course reflected the guideline by teaching multiple cognitive-behavioural prescriptive interventions to stimulate the patients' acquisition of problem solving skills, and to structure the patients' daily activities. Information was given and OPs were trained to differentiate between disorders. In addition, a graded activity treatment approach was introduced, which was based on a three stages model and resembled stress inoculation training	C (n=115): Usual care. Minimal involvement of the OP and access to treatment by a psychologist. Optimal usual care was provided in this study. This consisted of the advice to OPs to refer to a psychologist, whose treatment was fully funded by the insurance company	Unclear, though, there were 2 follow-up consultations with the OP	Unclear, though, there were 2 follow-up consultations with the OP	NR	NR	Actual times not specified, although total is 12 months. Baseline= Time 1, follow-up= Time 2, Time 3, Time 4	Time 1: 13 (10%) Time 3: 24 (19%)	Time 1: 14 (12%) Time 3: 26 (22%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
8. Schene, 2007	I (n=30): Treatment as usual (including outpatient psychiatric clinical management plus anti-depressants, if indicated) as well as occupational therapy which included three phases: 1) Diagnostic phase: including a detailed occupational history, role-playing work situations, contact with an Occupational Physician (OP) from the patient's employer, and a plan for work reintegration. 2) Therapeutic phase: Group and individual sessions including preparation of work reintegration, contacting the place of work, and starting to work, if possible. Individual sessions involved further analyses of the relationship between work and depression, exploration of work problems, and support and evaluation of work resumption. 3) Follow-up phase	C (n=32): Treatment as usual, including outpatient psychiatric clinical management and anti-depressants if indicated	Treatment as usual visits every 2-3 weeks plus 1. Diagnostic phase, 5 contacts 2. Therapeutic phase, 24 weekly group sessions and 12 individual sessions 3. Follow-up phase, 3 visits	Every 2-3 weeks	Treatment as usual visits lasted 30 minutes Diagnostic phase, 4 weeks Therapeutic phase, group sessions 2 hours each, individuals sessions NR, all over 24 weeks Follow-up phase, 20 weeks	Visits lasted 30 minutes	3, 6, 12, and 42 months	3 months: NR 6 months: 3 (9%) 12 months: 4 (13%) 42 months: 8 (25%)	3 months: NR 6 months: 0 (0%) 12 months: 1 (3%) 42 months: 6 (20%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
9. Schoenbaum, 2001	I1 (n= 424): Quality improvement (QI) QI for improving access to medication, "QI Meds". Nurse specialists presented antidepressant medications and psychotherapy as equally effective treatments to the patient. The primary care clinician formulated a treatment plan with the patient. The nurse contacted patients monthly (calls or visits) and helped with management of antidepressant medications and support of adherence. I2 (n= 489): Quality Improvement (QI) for improving psychotherapy, "QI Therapy". Primary care clinicians formulated a treatment plan with the patient. If the clinician determined that psychotherapy was appropriate, patients were referred to CBT-trained therapists. Local psychotherapists provided individual and group CBT sessions. Medication was available, but no medication management was provided	C (n=443): Usual care. No study resources were available, although clinic medical directors were mailed written copies of national depression practice guidelines	I1: Nurse contacted patients monthly I2: NR	NR	I1: 6 or 12 months I2: 12 to 16 sessions	NR	6, 12, 18 and 24 months	6 months: I1: 56 (13%) I2: 87 (18%) 12 months: I1: 65 (15%) I2: 96 (20%) 18 months: I1: 70 (17%) I2: 106 (22%) 24 months: I1: 53 (13%) I2: 88 (18%) 24 months, economic survey: I1: 55 (13%) I2: 83 (17%)	6 months: 57 (13%) 12 months: 69 (16%) 18 months: 76 (17%) 24 months: 57 (13%) 24 months, economic survey: 62 (14%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
10. Smith, 2002	I (n=129): Enhanced care. Physicians were informed when patients screened positive for depression. Nurses/physicians were trained to provide high quality depression care during the acute and continuation phase of treatment. Training emphasized the need to encourage patients to initiate guideline-concordant pharmacotherapy or psychotherapy during the acute phase of treatment. Nurses contacted patients to assess symptoms, educated them about depression and treatment, and gave homework assignments to increase their readiness to engage in active treatments. Training also emphasized the need to encourage continued treatment adherence if symptoms were resolving, to adjust treatment if symptoms were not resolving, or terminate treatment when remitted patients did not require maintenance therapy following the continuation phase. Physicians received symptom/treatment status summaries, along with reminders to adjust treatment for symptomatic patients.	C (n=133): Usual care Physicians were not informed when patients screened positive for depression. No regular nurse contacts during initial or continuation phases of treatment	Patients returned after the index visit and then for weekly sessions with the nurse Sessions followed by regular telephone contact (those with ≥3 out of 9 depressive symptoms were contacted monthly by nurses, those with ≤2 out of 9 symptoms were contacted every 3 months) Physicians monitored patients monthly	NR	10-15 minute sessions for 5-7 weeks Telephone contacts over 1 year Duration of physicians' contacts, NR	NR	6 and 12 months	6 months: NR 12 months: 27 (21%)	6 months: NR 12 months: 16 (12%)

Author, Year	Nature of Intervention		Frequency of Intervention		Duration of Intervention		Follow-Up Period(s)	Loss to Follow-Up n (%)	
	Intervention Sample Size (n)	Control Sample Size (n)	Intervention	Control	Intervention	Control		I	C
11. van der Feltz-Cornelis, 2010	I (n=29): OPs were trained in diagnosis and treatment of employees with depressive disorders, anxiety disorders or somatoform disorders. Supportive psychiatric consultations by two psychiatrists aimed at delivering a diagnosis and treatment plan, including suggestions for RTW adapted to the specific needs of the patients due to their specific disorder. Psychiatrists spoke to patients once and reported to OP by consult letter	C (n=31): Care as usual from the OP	NR	NR	NR	NR	3 and 6 months	3 months: 0 (0%) 6 months: 4 (14%)	3 months: 0 (0%) 6 months: 7 (23%)
12. Wang, 2007	I (n=304): Telephone intervention program assessing need for treatment, facilitating entry into in-person treatment and medication, as necessary, with monitoring and support for treatment adherence. All participants received a psycho-educational workbook. For those declining in-person treatment, care managers maintained regular telephone contacts and, if experiencing significant symptoms after 2 months, they were provided with a structured psychotherapy intervention by telephone	C (n=300): Usual care. Any normally available insurance benefit or service, but not the additional telephone care management components provided to those in the I group	For those consenting to in-person treatment, NR For those declining in-person treatment, intervals, care manager contacts ranged from weekly to bimonthly. The telephone intervention was provided in 8 weekly sessions	NR	For those consenting to in-person treatment, NR For those declining in-person treatment, care manager contact duration was NR. Telephone intervention sessions were 30-40 minutes. Overall duration was NR	NR	6 and 12 months	6 months: 35 (12%) 12 months: 44 (14%)	6 months: 22 (7%) 12 months: 30 (10%)

* Consistently employed: reported full-time or part-time employment at each follow-up over 2 years
** Inconsistently employed: reported full-time or part-time employment at one ore or more follow-ups and no employment at one or more follow-ups over 2 years
NR: Not Reported, **QI**: Quality Improvement, **CBT**: Cognitive Behavioural Therapy, **OPs**: Occupational Physicians, **AHRQ**: Agency for Healthcare Research and Quality, **RTW**: Return to Work, **GP**: General Practitioner

Table 3: Depression Characteristics

Author, Year	Instrument(s) Used to Determine the Presence and Severity of Depression	Method of Instrument Administration	Depression Scores at Baseline Mean (SD)		% of Participants with Depression at Baseline	
			Intervention	Control	Intervention	Control
1. Blonk, 2006	1. Shortened version of the Composite International Diagnostic Interview (WHO-CIDI) 2. Depression Anxiety Stress Scales (DASS) depression subscale	WHO-CIDI: Telephone interview (by psychologist) DASS: Self-administered questionnaire	DASS depression subscale: I1: 15.3 (9.8) I2: 20.0 (8.9)	DASS depression subscale: 20.0 (10.4)	DASS: I1: 54.8 I2: 77.1	DASS: 69.7
2. Dewa, 2009	NR	NR	NR	NR	67	71
3. Kawakami, 1997	1. Zung Self-Rating Depression Scale	Self-administered questionnaire	41.6 (NR)	40.6 (NR)	NR	NR
4. Knekt, 2008a and 2008b	1. Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) 2. 17-item Hamilton Rating Scale for Depression (HAM-D17) 3. BDI	Interview (NR by whom)	HAM-D17: I1: 15.8 (0.49) I2: 15.4 (0.48) BDI: I1: 18.2 (0.81) I2: 17.9 (0.79)	HAM-D17: 15.8 (0.43) BDI: 18.7 (0.70)	Mood disorder (DSM-IV): I1: 86.6, I2: 78.2 Anxiety disorder: I1: 46.4, I2: 49.5	Mood disorder (DSM-IV): 88.3 Anxiety disorder: 36.7
5. Krogh, 2009	1. Major Depression Inventory (MDI) (used to establish ICD-10 and DSM-IV criteria) 2. HAM-D17 3. Montgomery-Asberg Depression Rating Scale 4. BDI	Interview (by psychologist & research assistant)	HAM-D17: I1: 18.2 (3.6) I2: 18.2 (3.8) Montgomery-Asberg Depression Rating Scale: I1: 22.0 (5.6) I2: 22.9 (5.5) BDI: I1: 30.6 (8.8) I2: 30.5 (6.9)	HAM-D17: 16.7 (3.8) Montgomery-Asberg Depression Rating Scale: 21.6 (4.7) BDI: 31.8 (8.3)	Based on MDI: I1: 100 I2: 100 Based on DSM-IV criteria for major depressive disorder: I1: 70.9% I2: 69.1%	Based on MDI: 100 Based on DSM-IV criteria for major depressive disorder: 63.6%

Author, Year	Instrument(s) Used to Determine the Presence and Severity of Depression	Method of Instrument Administration	Depression Scores at Baseline Mean (SD)		% of Participants with Depression at Baseline	
			Intervention	Control	Intervention	Control
6. Lo Sasso, 2006 Rost, 2004	First and second stage screening: 1. WHO-CIDI 2. Inventory to Diagnose Depression (IDD) 3. Modified Center for Epidemiologic Studies Depression Scale (m-CESD)	Interviews (by administrative staff)	IDD: I1: 6.7 (1.4)* I2: 6.8 (1.4)	IDD: C1: 6.6 (1.4)* C2: 6.6 (1.5)	I1 and I2: 100	C1 and C2: 100
7. Rebergen, 2009a Rebergen, 2009b	1. DASS depression subscale 2. Hospital Anxiety Depression Scale (HADS) depression subscale	Self-administered questionnaire	DASS-depression: 6.9 (7.4). HADS-depression: 11.5 (4.4)	DASS-depression: 6.6 (7.4). HADS-depression: 11.8 (4.5)	DASS: 34.8	DASS: 29.0
8. Schene, 2007	1. DSM-IV 2. BDI	Interview (by psychiatrist and trained staff)	BDI: 27.1 (SD 9.4)	BDI: 23.6 (SD 9.1)	100	100
9. Schoenbaum, 2001**	1. WHO-CIDI 2. 12-Item Short Form Health Survey (SF-12) - Mental Health Composite Score (MCS-12)	Self-administered questionnaire for screening Following screening, self-administered questionnaire or telephone interviews	MCS-12: I1: 36.0 (10.8) I2: 34.9 (10.4)	MCS-12: 36.4 (10.9)	WHO-CIDI: I1 and I2: 100	WHO-CIDI: 100
10. Smith, 2002	First and second stage screening: 1. WHO-CIDI 2. IDD 3. m-CESD	Interview (by administrative staff)	m-CESD: 56.5 (20.2)	m-CESD: 50.1 (20.2)	100	100
11. van der Feltz-Cornelis, 2010	1. Patient Health Questionnaire (PHQ) depression subscale	Self-administered questionnaire	NR	NR	Major depressive disorder: 37.0 Other depressive disorder: 17.0	Major depressive disorder: 35.0 Other depressive disorder: 13.0

Author, Year	Instrument(s) Used to Determine the Presence and Severity of Depression	Method of Instrument Administration	Depression Scores at Baseline Mean (SD)		% of Participants with Depression at Baseline	
			Intervention	Control	Intervention	Control
12. Wang, 2007	First stage: 1. K-6 psychological distress screen Second stage: 1. Quick Inventory of Depression Symptoms Self-Report (QIDS-SR)	Phase 1: Self-administered questionnaire Phase 2: Telephone interview (by survey interviewers)	QIDS-SR: 13.3 (3.3)	QIDS-SR: 13.8 (3.6)	100	100

* For Lo Sasso, 2006 the economic results are presented for a subgroup analysis only. The subgroup includes I1 and C1 (consistently employed persons)
** For Schoenbaum, 2001 baseline values were not provided. Values are instead provided by the supplemental paper by Sherbourne, 2001. The sample is restricted to individuals who completed at least 1 follow-up (n=1299) and the values are weighted for the probability of non-enrollment and non-response to the eligible sample.

I1: Intervention 1, I2: Intervention 2, etc.; C1: Control group 1, C2: Control group 2, etc.

NR: Not Reported, **MCS-12**: Mental Health Composite Score, **SF-12**: 12 item Short Form Health Survey. The SF-12 is a multipurpose short form survey with 12 questions, all selected from the 36-item Short-Form Health Survey score and standardized to a general population mean (SD) of 50(10), **BDI**: Beck Depression Inventory, **m-CESD**: Modified Center for Epidemiologic Studies Depression scale

Table 4: Participant Characteristics

Author, Year	Source Population	Job Title(s)	Sample Size n		Age Mean (SD)		% Female		Education Status (%)		% Working at Baseline	
			I	C	I	C	I	C	I	C	I	C
1. Blonk, 2006	Self-employed workers insured for work disability through a private insurer and reporting to their insurance company for disability benefits owing to psychological complaints	Self-employment in: agricultural, service, construction, health care, trade, and other industries	I1: 40 I2: 40	42	NR	NR	NR	NR	NR	NR	I1: 0 I2: 0	0
2. Dewa, 2009	Employees of a nationwide financial-insurance sector company on short-term disability leave	NR	73	51	44 (8.7)	49 (8.2)	90	82	NR	NR	0	0
3. Kawakami, 1997	Blue collar employees of a large electric company	Machine operators and technicians (main occupation)	110	175	33 (12)	35 (12)	24	44	NR	NR	100	100
4. Knekt, 2008a and 2008b	Outpatients referred from various psychiatric services	White collar workers and entrepreneurs (majority of participants)	I1: 97 I2: 101	128	I1: 33.6 (7.2) I2: 32.1 (7.0)	31.6 (6.6)	I1: 74.2 I2: 74.3	78.9	I1: 28.9* I2: 19.8*	75.4*	Employed or student: I1: 83.2 I2: 85.1	Employed or student: 75.4
5. Krogh, 2009	Patients referred from GPs, private practicing psychiatrists, psychologist and psychiatric wards	NR	I1: 55 I2: 55	55	I1: 41.9 (8.7) I2: 38.1 (9.0)	36.7 (8.7)	I1: 81.8 I2: 78.2	61.8	NR	NR	Not on sick leave: I1: 47.3 I2: 58.2 Employed: I1: 58.2 I2: 45.5	Not on sick leave: 56.4 Employed: 63.6

Author, Year	Source Population	Job Title(s)	Sample Size n		Age Mean (SD)		% Female		Education Status (%)		% Working at Baseline	
			I	C	I	C	I	C	I	C	I	C
6. Lo Sasso, 2006 Rost, 2004	Patients presenting for routine visits at participating community primary care practices	Professional/ administrators, managers/ salespeople, clerical/services	I1: 96 I2: 62	C1: 102 C2: 66	I1: 37.6 (9.5) I2: 38.3(12.8)	C1: 40.1 (10.3) C2: 40.4 (10.5)	I1: 84.4 I2: 83.8	C1: 82.4 C2: 90.9	I1: HS: 88.5 I2: HS: 80.7	C1: HS: 93.1 C2: HS: 75.8	Employed full-time: I1: 85.4 I2: 68.8	Employed full-time: C1: 74.5 C2: 72.2
7. Rebergen, 2009a Rebergen, 2009b	Employees from police departments on sick leave for mental health problems	Executive and administrative police department workers	125	115	38.8 (8.4)	40.0 (9.5)	48.8	39.5	NR	NR	0	0
8. Schene, 2007	Patients presenting to an outpatient mood disorder clinic	NR	30	32	46.6 (7.4)	45.2 (7.5)	50	53	<HS 63 ≥HS: 37	<HS: 59 ≥HS: 41	23.3	15.6
9. Schoenbaum, 2001**	Patients presenting to primary care clinics in community-based managed care organizations	NR	I1: 424 I2: 489	443	I1: 44.0 (14.7) I2: 44.9 (16.0)	42.2 (13.9)	I1: 66.7 I2: 75.8	69.1	I1 <HS: 16.2 HS: 29.3 >HS: 54.5 I2 <HS: 19.2 HS: 26.5 >HS: 54.2	<HS: 20.2 HS: 33.6 >HS: 46.2	NR	NR

Author, Year	Source Population	Job Title(s)	Sample Size n		Age Mean (SD)		% Female		Education Status (%)		% Working at Baseline	
			I	C	I	C	I	C	I	C	I	C
10. Smith, 2002	Patients from community primary care practices	Professionals, managers and administrators, craftsmen, clerical and sales workers, labourers and operatives	129	133	37.9 (9.5)	40.4 (10.1)	81.2	79.1	HS: 86.4	HS: 92.6	100	100
11. van der Feltz-Cornelis, 2010	Sick-listed patients seen by occupational physicians at companies providing occupational health care	Legislators, senior officials and managers, professionals, associate professionals, craft and related trades workers, technicians, manual labourers, clerks, service workers, shop and market sales workers	29	31	42.0 (NR)	42.0 (NR)	52.0	64	Low: 7* Middle: 50* High: 43*	Low: 17* Middle: 47* High: 37*	0	0
12. Wang, 2007	Employees from one of 16 companies covered by a managed behavioural health plan	NR	304	300	40.7 (10.5)	42.4 (10.8)	70.7	77	College graduate: 38	College graduate: 43.8	100	100

* Refers to receiving an academic education (level of education, e.g. "HS", is not specified)
** For Schoenbaum, 2001 baseline values were not provided. Values are instead provided by the supplemental paper by Sherbourne, 2001. The sample is restricted to individuals who completed at least 1 follow-up (n=1299) and the values are weighted for the probability of non-enrollment and non-response to the eligible sample.

Note: For Lo Sasso, 2006 the economic results are presented for a subgroup analysis only. The subgroup includes I1 and C1 (consistently employed persons) I1: Intervention 1, I2: Intervention 2, etc.; C1: Control group 1, C2: Control group 2, etc.

NR: Not Reported, **HS**: High School education

Table 5: Categories of Primary and Secondary Outcomes of Interest to this Systematic Review

Author, Year	Primary Outcomes			Secondary Outcomes		Timing of Outcome Measurement
	Work Functioning	Work Disability & Recurrences of Work Disability	Economic Outcomes	Depression Outcomes	Other Outcomes	
1. Blonk, 2006	None	1. Time until partial RTW (part-time) 2. Time until full RTW (full-time)	None	1. Depression symptoms (DASS depression subscale)	1. Psychological complaints (DASS anxiety, and stress subscales, MBI-NL emotional exhaustion, depersonalization, and professional efficacy subscales)	Baseline, 4, 10, & 12 months
2. Dewa, 2009	None	1. RTW (yes/no) 2. Transition to long-term disability (yes/no) 3. Days on short-term disability (in days)	1. Healthcare costs 2. Intervention costs	None	None	12 months
3. Kawakami, 1997	None	1. Length of sick leave in past year (days)	None	1. Depression symptoms (Zung SDS)	1. Perceived workplace stressors (6 stressors, yes/no to each) 2. Blood pressure (systolic & diastolic)	Baseline, 12, & 24 months
4. Knekt, 2008a and 2008b	1. Self-estimated work ability (modified Work-Ability Index, WAI) 2. Adequate work ability (WAI score ≥37) 3. Work role functioning (Work subscale of SAS-SR)	1. Current employment status 2. Number of sick leave days in past 3 months 3. More than 7 sick-leave days during last 3 months	None	1. Depression symptoms (HAM-D, BDI) 2. Remission from depressive symptoms (BDI score <10) 3. Recovery from psychiatric diagnosis of major depressive disorder (DSM-IV)	1. Anxiety symptoms (SCL-90-Anx, HAMA) 2. General psychiatric symptoms (SCL-90-GSI) 3. Perceived psychological functioning (Perceived Psychological Functioning scale) 4. Recovery from psychiatric diagnosis of mood disorder, and anxiety disorder (DSM-IV)	Baseline, 3, 7, 9, 12, 18, 24, & 36 months

Author, Year	Primary Outcomes			Secondary Outcomes		Timing of Outcome Measurement
	Work Functioning	Work Disability & Recurrences of Work Disability	Economic Outcomes	Depression Outcomes	Other Outcomes	
5. Krogh, 2009	None	1. Unemployment (yes/no) 2. Sick leave (yes/no) 3. % of days absent from work in last 10 days 4. Job status (full-time, half=time, <20 hr/week)	None	1. Depression symptoms (HAM-D17, Montgomery-Asberg depression rating scale, BDI) 2. Remission of depression symptoms (defined as HAM-D17 <8 and not fulfilling ICD-10 criteria for depression)	1. Quality of life (WHO-5)	Baseline, 4 & 12 months
6. Lo Sasso, 2006 Rost, 2004	1. Productivity (employee's rating of their productivity at work during the previous 2 weeks)	1. Absenteeism (calculated from employee reports of how many full and partial workdays missed due to illness or doctor visits in the past 4 weeks)	1. Annual earnings, estimated by employee estimates of the income they received the previous calendar year 2. Intervention costs	1. Depression severity (CES-D)	1. Specialty care counselling during past 6 months (yes/no) 2. Emotional role functioning (SF-36) 3. Antidepressant medication use, calculated as number of months in past 6 months employee took antidepressants at guideline-concordant doses	Baseline, 6, 12, 18 & 24 months

Author, Year	Primary Outcomes			Secondary Outcomes		Timing of Outcome Measurement
	Work Functioning	Work Disability & Recurrences of Work Disability	Economic Outcomes	Depression Outcomes	Other Outcomes	
7. Rebergen, 2009a Rebergen, 2009b	None	1. RTW process (immediate full RTW versus partial RTW) 2. Number of recurrences of sick leave periods 3. Duration of recurrences of sick leave periods 4. Duration of sick leave days until partial RTW 5. Duration of sick leave days until full RTW 6. Total productivity loss, defined as the duration of sick leave days until full RTW added with number of days of recurrences on sick leave	1. Healthcare costs 2. Intervention costs	None	None	Baseline, during treatment (timeline not reported) & 12 months
8. Schene, 2007	None	1. Time until any work resumption 2. Total hours worked within each 6-month period up to 42 months 3. Proportion of patients working at least 2 days or 16 hours per week	1. Healthcare costs	1. Presence of major depression (DSM-IV) 2. Depression severity (BDI)	1. Work stress (Psychic Strains Section of the QOS)	Baseline, 3, 6, 12, & 42 months
9. Schoenbaum, 2001	None	1. Days of employment over 24 months, calculated as number of days worked over each 6-month follow-up 2. Days missed from work due to illness in the previous 4 weeks	1. Healthcare costs 2. QALY (estimated costs)	1. Days of depression burden	1. Quality-Adjusted Life-Years (QALYs)	Baseline, 6, 12, 18 & 24 months

Author, Year	Primary Outcomes			Secondary Outcomes		Timing of Outcome Measurement
	Work Functioning	Work Disability & Recurrences of Work Disability	Economic Outcomes	Depression Outcomes	Other Outcomes	
10. Smith, 2002	None	1. Subsequent employment, defined as persons working full-time at baseline reporting full-time work at follow-up; persons working part-time at baseline reporting part-/full-time work at follow-up	None	None	1. Workplace conflict in previous 12 months (arguments/difficulties with co-workers, yes/no)	Baseline, 6 & 12 months
11. van der Feltz-Cornelis, 2010	None	1. Full RTW status, with full RTW defined as RTW for at least 4 weeks without relapse 2. Time to full RTW (MOS-SF20, OHS database)	None	1. Depression severity (SCL-90, PHQ-9, PHQ-15)	1. Quality-adjusted life years (QALY from EQ-5D)	Baseline, 3, 6 & 12 months
12. Wang, 2007	1. Effective weekly hours worked (hours worked weighted by job performance from the HPQ) 2. On-the-job performance (HPQ)	1. Actual weekly hours worked 2. Job retention (HPQ)	None	1. Depression severity (QIDS-SR) 2. Substantial improvement in depression symptoms (≥50% reduction in QIDS-SR score) 3. Complete remission of depression (QIDS-SR score ≤5)	1. Critical workplace incidents (HPQ)	Baseline, 6 & 12 months

RTW: Return to Work, **DASS** (Depression Anxiety Stress Scales), **MBI-NL**: Maslach Burnout Inventory, **CES-D**: Center for Epidemiologic Studies Depression Scale, **SF-36**: Short-Form 36, **QOS**: Questionnaire Organization Stress, **DSM**: Diagnostic and Statistical Manual of Mental Disorders, **BDI**: Beck Depression Inventory, **WAI**: Work Ability Index, **SAS-SR**: Social Adjustment Scale, **HPQ**: World Health Organization Health and Productivity Questionnaire, **WHO-5**: World Health Organization-5 Well-Being Index, **ICD-10**: International Classification of Diseases-10, **MOS-SF20**: Medical Outcomes Study Short-Form 20, **OHS**: occupational health service, **EQ-5D**: Euroqol, **SCL-90**: Symptom Checklist, SCL-90-Anx: Symptom Checklist Anxiety Scale, **SCL-90-GSI**: Symptom Checklist Global Severity Index, **PHQ**: Patient Health Questionnaire, **QOS**: Questionnaire Organization Stress, **HAM-D**: Hamilton Depression Rating Scale, **HAM-D17**: 17 item Hamilton Depression Rating Scale, **HAMA**: Hamilton Anxiety Rating Scale

Table 6: Studies with Economic Analyses - Main Results

Author, Year	CEA or CBA	Perspective	Time-Frame	Results Sensitive to Time-Frame	Main Economic Findings
2. Dewa, 2009	CEA and CBA	Employer	1 year	No	**Summary of results:** The intervention cost per treated worker was $355 less than the control. The intervention also resulted in fewer average days lost, higher rates of return to work, and lower transitions to long-term disability than the control. Therefore, from a cost-effectiveness perspective, the intervention dominated (lower costs, better outcomes) the control The intervention resulted in 15 fewer short-term disability days lost per treated worker, in 22% more disabled workers returning to work, and in 25% fewer disabled workers transitioning to long-term disability, compared to the control. The expected net disability savings per worker was $503 **Further details:** The cost and effect data were analyzed simultaneously while adjusting for age using net benefit regression. When the willingness to pay was zero, the intervention was $503 less costly than the control (95% CI=996 less to 11 less). When the willingness-to-pay values were varied at $10, $50, and $100 the CMHC were of greater value than the extra costs.
6. Lo Sasso, 2006 Rost, 2004	CBA	Employer	3 years	Yes	**Summary of results:** The intervention resulted in a mean net benefit to the employer of at least $1409 per treated worker in Year 1 and $5136 per treated worker in Year 2. The return on investment (ROI) over 2 years was at least 302% The intervention reduced absenteeism by 22.8% or 10.6 days over 2 years at an estimated annual value of $539 per depressed Full Time Equivalent (FTE) The intervention significantly improved productivity by 6.1% over 2 years at an estimated annual value of $1491 per depressed FTE **Further details:** The incremental costs of the intervention per treated worker in Years 1 and 2 were $158 and $130 (respectively). The intervention's ROI increases in firms that rely on team production, have higher substitute labour costs, or realize penalties for output shortfalls. There was a mean net benefit to the employer even under extreme assumptions (for example, valuing workers' subjectively reported decreases in productivity at only 50% of the reported value) that would bias the analysis against finding a net mean benefit.

Author, Year	CEA or CBA	Perspective	Time-Frame	Results Sensitive to Time-Frame	Main Economic Findings
7. Rebergen, 2009a Rebergen, 2009b	CEA and CBA	CEA: Society CBA: Employer	1 year	No	**Summary of results:** CEA: mean total health care costs were 520 Euros less in the intervention group compared with the control group. There were no significant differences in productivity losses (sick leave days) between the two groups. The incremental cost-effectiveness ratio (incremental cost per sick-leave day) was -736. CBA: the net mean benefit of the intervention compared to control was 3,582 Euros. No significant differences were found in mean sick leave days between the intervention and control group. **Further details:** Despite the intervention group receiving guideline-based care from an OP, the mean cost of OP services per treated patient in the intervention group (310 Euros) was only 7 Euros greater than the mean cost of OP services in the control group (307 Euros). Mean psychological treatment costs were significantly higher in the control group compared to the intervention group (1233 and 534 Euros, respectively). The mean total health care costs of the intervention group were 2145 Euros, lower than the mean of the control group (2664 Euros).

Author, Year	CEA or CBA	Perspective	Time-Frame	Results Sensitive to Time-Frame	Main Economic Findings
8. Schene, 2007	CBA	Society	1 year	NR	**Summary of results:** Compared to workers in the control group, the additional mean net benefit per worker in the intervention group was $3,952 The number of days to any return to work was significantly lower in the intervention (mean=207) than the control group (mean=299) Over the first 18 months, the control group worked fewer total hours than the intervention group (medians, months 0-6: 0 v 20.25; months 7-12: 0.85 v 261.75; months 12-18: 156.42 v 456.25) **Further details:** Total service costs were 67% higher for the intervention patients but this was not statistically significant. One patient with psychotic depression had substantially higher costs. With this outlier removed, service costs in the intervention group were 27% higher than the control. The intervention resulted in more hours worked. Multiplying hours by US$36.88 (average hourly Dutch wage) showed a mean net benefit (earnings minus costs) of US$14, 850 in the intervention and US$10,898 in the control group. The difference in net benefit increased from US$3,952 to US$5,370 with the exclusion of the outlier. Fig. 2 of the paper shows that the likelihood of intervention being the most cost-effective option rises at a decreasing rate as the value of an hour's work rises. At the value of US$36.88 per hour the probability is 75.5%, which falls to 68.4% for a value of US$20 per hour and to 52.5% for a value of US$10 per hour. With the removal of the outlier from the analyses, the probability that the intervention is the most cost-effective option increases to 82.5% at value of US$36.88.
9. Schoenbaum, 2001	CEA	Society	2 years	Yes	**Summary of results:** Compared to usual care, the incremental cost per QALY for I1 (QI-meds) group was $36,467, and for the I2 (QI-therapy) group the incremental cost per QALY was $21,478. Participants in the I1 group (QI-meds) had a mean of 17.9 more employed days over two years compared to the usual care group. Participants in the I2 group (QI-therapy) had 20.9 more employed days. **Further details:** 1. Relative to usual care, average health care costs increased $419 (11%) in QI-meds (I1; P=.35) and $485 (13%) in QI-therapy (I2; P=.28). 2. In the usual care group, the mean number of employment days (over two years) was 279.2.

Note: For Lo Sasso, 2006 the economic results are presented for a subgroup analysis only. The subgroup includes I1 and C1 (consistently employed persons)
CEA: Cost-effectiveness analysis, **CBA**: Cost-benefit analysis, **I**: Intervention, **C**: Control, **QALY**: Quality-Adjusted Life Year

Table 7: Studies with Economic Analyses - Additional Details

Author, Year	Costs Associated with Intervention Measured		Outcome Measures Monetized into a Dollar Figure	Inflation Adjustment	Did the Authors Calculate How Long it Would Take to Recoup Costs?	Discounting
	Direct Costs	Indirect Costs				
2. Dewa, 2009	Health care costs covered by employer: costs of the Collaborative Mental Health Care program, an costs of independent medical evaluations	None	Days lost, Return to work, Long-term disability	No	No	No
6. Lo Sasso, 2006 Rost, 2004	Intervention costs: Training costs (physicians, nurses, and administrative staff), screening costs	Employee earnings	Productivity, absenteeism	Yes	No	No
7. Rebergen, 2009a Rebergen, 2009b	Intervention costs: Trainer costs, room/equipment costs, study material for OPs	Administration costs	Lost work days (absenteeism)	No	No	No
8. Schene, 2007	Healthcare delivery costs: Outpatient psychiatric care, GP care, hospitalization costs, travel costs	Earnings	Value of additional work hours	No	No	No
9. Schoenbaum, 2001	Intervention costs: Screening costs, intervention materials, health care delivery costs (nurse intervention) Excluded inpatient costs. Average cost in 1998 dollars using a national database of Ingenix, a benefits consulting firm	Patient time cost for obtaining health care - priced patient's time using reported hourly wage at baseline and sex-specific mean wage for those not working at baseline	Mean cost per outpatient medical visit, mental health visits and emergency department visits. Also psychotropic medications.	No	Yes	No

Note: For Lo Sasso, 2006 the economic results are presented for a subgroup analysis only. The subgroup includes I1 and C1 (consistently employed persons)

Table 8: Main Findings

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
1. Blonk, 2006	Cox proportional hazards regression Repeated measures MANOVA ANOVA (post-hoc analyses)	Gender, education, age, and number of employees	**PRIMARY OUTCOMES** **Work Disability and Recurrences of Work Disability Outcomes:** 1. Time until partial RTW (part-time): In unadjusted analyses, there was a significant main effect of intervention group on time until partial RTW (χ^2=6.28, p<0.05). Based on median days, partial RTW occurred 17 and 30 days earlier for I1 (brief CBT-based stress management with workplace focus) than for I2 (CBT) and C (control) groups, respectively, but there was no difference between I2 and C groups. In adjusted analyses, the effect of intervention group on time until partial RTW was no longer significant (χ^2=2.17, p>0.05). 2. Time until full RTW (full time): In unadjusted analyses, there was a significant main effect of intervention group on time until full RTW (χ^2=14.95, p<0.01). Based on median days, full RTW occurred 207 and 198 days earlier for I1 than for I2 and C groups, respectively, but there was no difference between I2 and C groups. In adjusted analyses, the effect of intervention group on time until full RTW remained significant (χ^2=9.64, p<0.01). **SECONDARY OUTCOMES** **Depression Outcomes:** 1. Psychological complaints (DASS depression, anxiety, and stress subscales): All subscale scores significantly decreased over 10 months (F=8.90, p<0.01). However, there was no significant difference in scores between intervention groups (F=1.33, p<0.05). There was also no significant group x time interaction (F=0.99, p<0.05), such that all DASS subscores decreased over time, irrespective of the intervention received. **Other Secondary Outcomes:** 1. Psychological complaints (MBI-NL emotional exhaustion, depersonalization, and professional efficacy subscales): Burnout scores decreased significantly over 10 months (F=4.47, p<0.01). However, there was no significant difference in MBI-NL scores between intervention groups, nor was there a significant group x time interaction (data not shown). That is, MBI-NL scores decreased over time, irrespective of the intervention received.
2. Dewa, 2009	t test Chi-square test Linear regression	Age (economic outcomes only)	**PRIMARY OUTCOMES** **Work Disability and Recurrences of Work Disability Outcomes:** 1. RTW (%): A significantly higher proportion of participants in the I (enhanced disability management with additional collaborative mental health care) group (85%) compared to the C (usual care) group (63%) returned to work by 12 months (χ^2=8.06, p=0.005). 2. Transition to long-term disability (%): A significantly lower proportion of participants in the I group (7%) compared to the C group (31%) transitioned to long-term disability by 12 months (χ^2=12.84, p<0.001). 3. Days on short-term disability: The average number of days on short-term disability leave was significantly shorter for the I group (62 days) than the C group (76 days) (t=2.17, p=0.03). **Economic Outcomes:** See Table 6 **SECONDARY OUTCOMES** None

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
3. Kawakami, 1997	ANCOVA Generalized logit analysis with repeated measurements	Age, gender	**PRIMARY OUTCOMES** **Work Disability and Recurrences of Work Disability Outcomes:** 1. Length of sick leave in the past year: The proportion of participants reporting 1-5 days of sick leave decreased in the I (stress reduction) group (52% at baseline, 34% at 2 year follow-up), but remained relatively stable in the C (control) group (33% at baseline, 37% at 2 years). On the other hand, more participants reported 0 days of sick leave over time in the I group (40% at baseline, 61% at 2 years), but not in the C group (53% at baseline, 58% at 2 years). No differences over time were seen in either group in the proportion reporting ≥6 days of sick leave. Overall, both the group x time interaction (χ^2=10.4, p=0.034) and the main group effect (χ^2=18.4, p<0.001) were significant, suggesting stress reduction was effective in decreasing length of sick leave. No difference was observed between men and women (data not shown). **SECONDARY OUTCOMES** **Depression Outcomes:** 1. Depression symptoms (Zung SDS scale): Mean scores on the Zung SDS scale decreased from 41.4 (SD 7.7) at baseline to 38.6 (SD 6.4) at 2 year follow-up for the I group, while they remained relatively stable in the C group (41.2, SD 7.1 at baseline, 41.0, SD 7.3 at 2 years), resulting in a statistically significant intervention (group x time) effect (F=3.41, p=0.035). For males, a statistically significant intervention effect was observed for the depression score (F=4.96, p= 0.025), with mean depression scores decreasing in the I group, but remaining stable in the C group. The intervention effect was not statistically significant for females (F=0.34, p= 0.716). **Other Secondary Outcomes:** 1. Workplace stressors: The proportion of participants working overtime (>30h per month) increased in the I group from baseline (44%) to 2 year follow-up (78%), but not in the C group (57% at baseline and 2 year follow-up). However, this group x time interaction was not significant (χ^2=3.0, p=0.229). The proportion of participants reporting work overload increased in the I group (26% at baseline, 43% at 2 years), but not in the C group (28% at baseline, 26% at 2 years) (group x time interaction χ^2=5.9, p=0.054). The group x time interaction effect was non-significant for the remaining self-reported stressors: little chance to learn new knowledge (χ^2=2.1, p=0.355), lack of control over workplace (χ^2=2.3, p=0.311), problems with supervisor (χ^2=3.2, p=0.204), and problems with co-workers (χ^2=0.9, p=0.633). No difference in the intervention effects was observed between men and women (data not shown). 2. Blood pressure: Mean systolic and diastolic blood pressure readings were similar across groups at both baseline and 2 year follow-up and remained relatively stable over time. The group x time intervention effect was non-significant (F=0.01, p=0.906 for systolic blood pressure; F=0.00, p=0.966 for diastolic blood pressure). No differences were observed between men and women (data not shown).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
4. Knekt, 2008a and 2008b	Linear mixed modeling Logistic regression Generalized estimating equations	Age, sex, marital status, education, age at onset, separation experiences, axis I and II diagnoses and baseline values of outcome measures	**PRIMARY OUTCOMES** **Work Functioning Outcomes:** 1. Self-estimated work ability (WAI): Mean scores significantly improved in all three groups from baseline (33.5, SE 0.70 in I1 [solution-focused therapy], 34.1, SE 0.68 in I2 [short-term psychotherapy], and 33.4, SE 0.61 in C [long-term psychotherapy) to 36 months (37.9, SE 0.87 in I1, 37.8, SE 0.80 in I2, 39.9, SE 0.76 in I3) ($p<0.001$). No significant mean score differences were found at any time point between I1 and I2 or I1 and C. At 7 months, I2 more effectively improved work ability than C (mean score difference 1.68, 95%CI 0.0 to 3.36), but by 36 months, C was statistically significantly more effective than I2 (mean score difference -2.48, 95%CI -4.56 to -0.39, $p<0.05$. 2. Adequate work ability (WAI scores ≥37): Among individuals demonstrating an impaired work ability at baseline (i.e. with WAI values ≤36), 55.1% (SE 6.8) of I1, 51.5% (SE 6.8) of I2, and 34.5% (SE 6.3) of C demonstrated an adequate work ability at 7 month follow-up. By 36 months, there was little change in the proportion demonstrating adequate ability in I1 (53%, SE 7.3). I2 demonstrated an increase between 7 and 36 months to 61.8% (SE 6.7), while C demonstrated the largest increase, with 74% (SE 6.2) of individuals demonstrating adequate work ability at 36 months. Both group ($p=0.012$) and time ($p<0.001$) effects were significant. No significant differences were demonstrated at any time point between I1 and I2. At 7 months, a significantly higher proportion of I1 had adequate work ability than C (mean difference 20.6, 95%CI 2.4 to 38.7, $p<0.05$). By 36 months, the opposite was true, with fewer individuals in I1 compared to C demonstrating an adequate work ability (mean difference -21.6, 95%CI -40.4 to -2.8, $p<0.05$). Similar findings were seen for I2 compared to C, but were not statistically significant (at 7 months mean difference 17.0, 95%CI -1.2 to 35.1, $p>0.05$; at 36 months, mean difference -12.8, 95%CI -30.6 to 5.0, $p>0.05$). 3. Work-role functioning (SAS-SR Work subscale): A statistically significant improvement in scores over time was evident ($p<0.001$), with the largest improvement in the C group. No significant mean score differences were found at any time point between I1 and I2. At 7 months, both I1 (mean score difference -0.13, 95%CI -0.28 to 0.02) and I2 (mean score difference -0.10, 95%CI -0.25 to 0.04) were more effective than C at improving SAS-SR scores, though not significantly ($p>0.05$). However, by 36 months, C was statistically significantly more effective than the two interventions (mean difference comparing I1 to C 0.16, 95%CI -0.01 to 0.33; mean score difference comparing I2 to C 0.21, 95%CI 0.05 to 0.37, $p<0.05$). **Work Disability and Recurrences of Work Disability Outcomes:** 1. Current employment status: The proportion of individuals employed or studying at baseline was 83.1% (SE 4.0) in I1, 85.1% (SE 3.9) in I2, and 75.5% (SE 3.5) in C, changing very little by 36 months (76.6%, SE 4.9 in I1, 80.4%, SE 4.5 in I2, and 76.3%, SE 4.0 in C) ($p=0.41$). There were also no significant differences found between the groups at any point in the follow-up. 2. Number of sick leave days during last 3 months: Among the employed, the mean number of sick leave days at baseline was 6.10 (SE 1.36) in I1, 4.60 (SE 1.30) in I2, and 5.33 (SE 1.20) in C. By 36 months, only the C group demonstrated a significant reduction in number of sick leave days (mean 2.44, SE 1.37 at 36 months compared to 5.42, SE 1.57 in I1 and 4.81, SE 1.41 in I2). At 36 months, the mean differences in number of sick leave days between I1 and C (3.28, 95%CI -0.83 to 7.38) and I2 and C (2.45, 95%CI -1.39 to 6.30) were not statistically significant. 3. More than 7 sick leave days during last 3 months: Among the employed, the proportion of individuals with more than 7 sick leave days at baseline was 22.7% (SE 5.3) in I1, 19.9% (SE 5.1) in I2, and 16.5% (SE 4.7) in C. I1 did not experience a reduction by 36 months (20.7%, SE 4.8), while both I2 (10.9%, SE 4.3) and C (9.4%, SE 4.2) demonstrated lower proportions at 36 months. At 36 months, there was a significant mean difference between I1 and C (14.1, 95%CI 0.0 to 28.1, $p<0.05$) and I2 and I1 (-14.9, 95%CI -29.1 to -0,7, $p<0.05$). There was no significant difference between I2 and C at 36 months.

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
4. Knekt, 2008a and 2008b (continued)			**SECONDARY OUTCOMES** **Depression Outcomes:** 1. Depression symptoms (HAM-D): A statistically significant reduction in symptoms over time was evident in all three treatment groups (p<0.001). No significant mean score differences were found at any time point between I1 and I2. Over the first 12 months, I2 was significantly more effective in reducing symptoms of depression than C. At 7 months, the mean score difference between I2 and C was -3.4 (95%CI -5.6 to -1.3, p<0.05), while at 12 months, the mean score difference was -2.6 (95%CI -5.0 to -0.3, p<0.05). I1 was also significantly more effective than C in reducing depression symptoms, but only over the first 7 months. At 7 months, the mean score difference between I1 and C was -3.7 (95%CI -5.8 to -1.5, p<0.05). By 36 months, the opposite was true, with C being statistically significantly more effective than the two interventions in reducing depression symptoms (mean difference comparing I1 to C 2.9, 95%CI 0.4 to 5.5, p<0.05; mean difference comparing I2 to C 3.8, 95%CI 1.4 to 6.2, p<0.05). 2. Depression symptoms (BDI): A statistically significant reduction in symptoms over time was evident in all three treatment groups (p<0.001). No significant mean score differences were found at any time point between I1 and I2. Over the first 12 months, I2 was significantly more effective in reducing symptoms of depression than C. At 7 months, the mean score difference between I2 and C was -1.8 (95%CI -3.3 to -0.3, p<0.05), while at 12 months, the mean score difference was -1.9 (95%CI -3.6 to -0.3, p<0.05). I1 was also more effective than C in reducing depression symptoms, but only over the first 7 months, though not significantly. At 7 months, the mean score difference between I1 and C was -1.4 (95%CI -2.9 to 0.1, p>0.05). By 36 months, the opposite was true, with C being statistically significantly more effective than the two interventions in reducing depression symptoms (mean score difference comparing I1 to C 1.8, 95%CI 0.1 to 3.5, p<0.05; mean difference comparing I2 to C 1.9, 95%CI 0.3 to 3.5, p<0.05). 3. Remission from depressive symptoms (BDI score<10): Remission from depressive symptoms was significantly more likely to occur in I1 than C at both 3 months (OR=3.07, 95%CI NR, p<0.05) and 7 months (OR=3.21, 95%CI 1.65 to 6.27, p<0.05). There were no significant differences between I1 and C at subsequent follow-ups, until 36 months when the direction of the relationship reversed and remission was more likely to occur in C than in I1 (OR=0.51, 95%CI 0.25 to 1.03, p>0.05), though non-significantly. Similar findings were seen for I2 compared to C. Over months 7 to 12, I2 was more likely to experience a remission from symptoms than C, with odds ratios at 7 months being 2.10 (95%CI NR, p<0.05) and at 12 months 2.21 (95%CI 1.20 to 4.07, p<0.05). By 36 months, C was non-significantly more likely to experience a remission than I2 (OR=0.57, 95%CI 0.30 to 1.08, p>0.05). Comparisons between I1 and I2 demonstrated that I1 was more likely to result in a more rapid remission than I2, with an OR of 0.52 (95%CI 0.27 to 1.02, p>0.05) after 3 months of follow-up. Thereafter, there were no significant differences between I1 and I2. 4. Remission from major depressive disorder (DSM-IV): No significant differences were found between I1 and I2 at any of the follow-up periods. Compared to C, I1 was significantly more likely to demonstrate a remission from a diagnosis of major depressive disorder at 7 months (OR=3.31, 95%CI 1.40 to 7.84, p<0.05) and 12 months (OR=2.35, 95%CI 1.08 to 5.11, p<0.05). For I2, a similar relationship with C was evident, but only at 7 months (OR=2.94, 95%CI 1.21 to 7.12, p<0.05). There was no significant difference between groups at 36 months. **Other Secondary Outcomes:** 1. Anxiety symptoms (SCL-90-Anx): A statistically significant reduction in symptoms over time was evident in all three treatment groups (p<0.001). No significant mean score differences were found at any time point between I1 and I2. There was no significant difference between I1 and C at any of the follow-up periods until 36 months, at which point C was significantly more effective at reducing symptoms than I1 (mean score difference 0.19, 95%CI -0.00 to 0.38, p<0.05). On the other hand, I2 was more effective at reducing anxiety symptoms at 7 months than C (mean score difference -0.19, 95%CI -0.37 to -0.01, p<0.05), with a non-significantly similar, but attenuated difference at 9 months (mean score difference -0.15, 95%CI -0.31 to 0.01, p>0.05). By 36 months, C was significantly more effective than I2 (mean score difference 0.20, 95%CI 0.02 to 0.38, p<0.05).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
4. Knekt, 2008a and 2008b (continued)			2. Anxiety symptoms (HAMA): A statistically significant reduction in symptoms over time was evident in all three treatment groups (p<0.001). No significant mean score differences were found at any time point between I1 and I2. There was no significant difference between I1 and C at any of the follow-up periods until 36 months, at which point C was more effective at reducing anxiety symptoms than I1 (mean score difference 2.0, 95%CI 0.5 to 3.5, p<0.05). On the other hand, I2 was more effective than C at reducing anxiety symptoms at 7 months (mean score difference -1.6, 95%CI -3.0 to -0.2, p<0.05) and 12 months (mean score difference -1.5, 95%CI -3.0 to 0.0, p<0.05). By 36 months, there was a non-statistically significant difference in HAMA scores between I2 and C, with C demonstrating more improved scores than I2 (mean score difference 1.3, 95%CI -0.1 to 2.8, p>0.05). 3. General psychiatric symptoms (SCL-90-GSI): A statistically significant reduction in symptoms over time was evident in all three treatment groups (p<0.001). No significant mean score differences were found at any time point between I1 and I2. Over the first 12 months, I2 was significantly more effective in reducing general psychiatric symptoms than C. At 7 months, the mean score difference between I2 and C was -0.14 (95%CI -0.28 to -0.00, p<0.05), while at 12 months, the mean score difference was -0.15 (95%CI -0.28 to -0.01, p<0.05). I1 was also more effective than C in reducing depression symptoms, but only over the first 9 months. At 7 months, the mean score difference between I1 and C was -0.16 (95%CI -0.30 to -0.01, p<0.05) and 9 months it was -0.15 (95%CI -0.29 to -0.01, p<0.05). By 36 months, the opposite was true, with C being more effective than the two interventions in reducing general psychiatric symptoms (mean score difference comparing I1 to C 0.15, 95%CI -0.01 to 0.31, p>0.05; mean score difference comparing I2 to C 0.16, 95%CI 0.01 to 0.32, p<0.05). 4. Remission from mood disorder (DSM-IV): No significant differences were found between I1 and I2 at any of the follow-up periods. Compared to C, at 7 and 12 months, both I1 and I2 were significantly more likely to demonstrate a remission from a diagnosis of mood disorder (I1 7 months OR=2.86, 95%CI 1.27 to 6.44; I1 12 months OR=2.66, 95%CI 1.25 to 5.65; I2 7 months OR=3.04, 95%CI 1.35 to 6.84; I2 12 months OR=2.41, 95%CI 1.13 to 5.16, all p<0.05). There was no significant difference between groups at 36 months. 5. Remission from anxiety disorder (DSM-IV): No significant differences were found between I1 and I2 at any of the follow-up periods. There was no significant difference found between I1 and C at 7 and 12 months. I2 was significantly more likely to result in a remission from anxiety disorder at 7 months only (OR=3.39, 95%CI 1.24 to 9.28, p<0.05). At 36 months, C was significantly more likely to result in a remission from anxiety disorder when compared to both I1 (OR=0.21, 95%CI 0.05 to 0.88, p<0.05) and I2 (OR=0.23, 95%CI 0.06 to 0.96, p<0.05). 6. Perceived psychological functioning scale: A statistically significant improvement in scores over time was evident (p<0.001). No significant mean score differences were found at any time point between I1 and I2. At 7 months, both I1 (mean score difference -1.90, 95%CI -3.49 to -0.30, p<0.05) and I2 (mean score difference -2.40, 95%CI -3.97 to -0.84, p<0.05) were more effective than C at improving scores. However, by 36 months, C was statistically significantly more effective than the two interventions (mean score difference comparing I1 to C 1.67, 95%CI 0.08 to 3.26, p<0.05; mean score difference comparing I2 to C 2.14, 95%CI 0.64 to 3.65).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
5. Krogh, 2009	ANOVA Chi-Square Test Repeated-measurement likelihood-based mixed model analysis	None	**PRIMARY OUTCOMES** **Work Disability and Recurrences from Work Disability Outcomes:** 1. Unemployment: By 4 months, the proportion of individuals who were unemployed in I1 (strength training group) was 42.6% (compared to 41.8% at baseline), in I2 (aerobic training) 37.5% (down from 54.5% at baseline), and in C (relaxation training) 40.5% (compared to 36.4% at baseline). However, there was no significant difference in the proportion unemployed at 4 months between I1 and C (OR=1.1, 95%CI 0.5 to 2.6, p=0.8) or between I2 and C (OR=0.9, 95%CI 0.4 to 2.1, p=0.8). By 12 months, all 3 groups experienced a further reduction in the proportion unemployed (32.6% unemployed in I1 and I2, 18.9% in C), but there remained no significant difference at 12 months between I1 and C (OR=1.7, 95%CI 0.6 to 5.0, p=0.3) or between I2 and C (OR=1.8, 95%CI 0.6 to 5.0, p=0.3). 2. Sick leave: Between baseline and 4 months, all groups experienced a reduction in the proportion of individuals on sick leave, decreasing from 52.7% to 34.0% in I1, 41.8% to 27.2% in I2, and 43.6% to 31.0% in C. When comparing groups, there was no significant difference in the proportion on sick leave at 4 months between I1 and C (OR=1.1, 95%CI 0.5 to 2.8, p=0.8) or between I2 and C (OR=1.0, 95%CI 0.4 to 2.7, p=0.9). By 12 months, I2 did not experience a reduction (28.3% still on sick leave), while I1 and C demonstrated further reductions (19.6% and 24.3%, respectively on sick leave). However, there was no significant difference in the proportion on sick leave at 12 months between I1 and C (OR=0.8, 95%CI 0.3 to 2.4, p=0.7) or between I2 and C (OR=1.2, 95%CI 0.5 to 3.4, p=0.7). 3. Percentage of days absent from work in last 10 days: The mean percentage of days absent from work at baseline decreased by 4 months in all groups - in I1 from 17.8 (SD 31.5) to 4.3 (SD 9.0), in I2 from 30.0 (SD 34.7) to 10.0 (SD 17.6), and in C from 26.6 (SD 35.3) to 16.9 (SD 33.3). There was, however, no statistically significant difference in mean percentage of days absent between I1 and C (-12.5, 95%CI -28.9 to 4.0, p=0.1) or between I2 and C (-7.9, 95%CI -24.1 to 8.3, p=0.3). By 12 months, the mean percentage remained relatively stable for I2 (11.2, SD 19.2) and C (14.5, SD 20.7) and there was no significant difference between means for these two groups (-2.7, 95%CI -11.7 to 6.2, p=0.5). I1 experienced a slight reduction in mean percentage to 1.4 (SD 4.4) and there was a significant difference in mean percentage for I1 compared to C (-12.1, 95%CI -21.1 to -3.1, p=0.009), such that individuals in I1 experienced a lower mean percentage of days absent from work than those in C. **SECONDARY OUTCOMES** **Depression Outcomes:** 1. Depression symptoms - 17-item Hamilton rating scale for depression (HAM-D17): All 3 groups experienced a large reduction in mean scores between baseline and 4 months, with the I1, I2, and C groups decreasing from a mean of 18.2 (SD 3.6), 18.2 (SD 3.8), and 16.7 (SD 3.8) at baseline, respectively, to 10.0 (SD 6.4), 12.1 (SD 6.4), and 10.6 (SD 5.6) at 4 months, respectively. At 4 months, HAM-D17 scores did not differ significantly between the 3 groups. The mean difference in scores between the I1 and C groups was -1.3 (95%CI -3.7 to 1.2, p=0.3), while the mean difference between I2 and C was 0.4 (95%CI -2.0 to 2.9, p=0.3). Similarly, the mean difference in scores between I1 and I2 was non-significant (-1.7, 95% CI -4.1 to 0.6, p=0.15). Mean scores at 12 months were similar to those at 4 months and mean differences in scores between I1 and C (-0.2, 95%CI -2.7 to 2.3, p=0.8), I2 and C (0.6 (95%CI -1.9 to 3.1, p=0.6), and I1 and I2 (-0.8, 95%CI -3.2 to 1.6, p=0.5) were all non-significant. 2. Depression symptoms - Montgomery-Aasberg depression rating scale: At 4 months, scores did not differ significantly, with the mean difference in scores between the I1 and C groups being -1.5 (95%CI -4.9 to 1.9, p=0.4) and between I2 and C 0.2 (95%CI -3.2 to 3.6) (p=0.9). Mean differences in scores at 12 months between I1 and C (0.9, 95%CI -2.7 to 4.4, p=0.6) and I2 and C (1.3, 95%CI -2.2 to 4.8, p=0.5) were also non-significant. 3. Depression symptoms - Beck depression inventory (BDI): At 4 months, scores did not differ significantly between I2 and C, with the mean difference in scores being 0.4 (95%CI -4.4 to 5.2, p=0.9). While not statistically significant, the mean difference between I1 and C was larger, with I1 demonstrating a lower BDI score than C (-3.2, 95%CI -8.1 to 1.6, p=0.2).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
5. Krogh, 2009 (continued)			Mean differences in scores at 12 months between I1 and C (-0.3, 95%CI -5.1 to 4.3, p=0.9) and I2 and C (0.1, 95%CI -4.7 to 4.9, p=0.9) were all non-significant. 4. Depression remission (HAM-D17 <8 and not fulfilling ICD-10 criteria for depression): At 4 months, a higher percentage of individuals in I1 met the criteria for remission (40.4%) compared to 29.2% in I2 and 31.7% in C, though the result was not statistically significant (χ^2=1.462, p=0.48). Similar results were seen at 12 months: 40.4% in I1, 32.6% in I2, and 37.8% in C (χ^2=0.628, p=0.73). **Other Secondary Outcomes:** 1. Quality of Life - WHO-5: At 4 months, scores did not differ significantly between I2 and C, with the mean difference in scores being -1.0 (95%CI –10.0 to 8.0, p=0.8). While not statistically significant, the mean difference between I1 and C was larger, with I1 demonstrating a higher WHO-5 score than C (8.3, 95%CI -0.7 to 17.3, p=0.07). Mean differences in scores at 12 months between I1 and C (1.2, 95%CI -8.7 to 11.0, p=0.8) and I2 and C (1.7, 95%CI -8.3 to 11.7, p=0.7) were all non-significant.

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
6. Lo Sasso, 2006 Rost, 2004	Time trend model with unstructured variance/ covariance matrix	Age, minority status, education, co-morbid dysthymia, treatment resistant depression risk, emotional functioning, physical co-morbidity, full/part time employment, paid sick leave benefits, and time co-varying health insurance status	**PRIMARY OUTCOMES** **Work Functioning Outcomes:** 1. Productivity (self-reported): When I1 (enhanced depression primary care among consistently employed) and I2 (enhanced depression primary care among inconsistently employed) were combined to create an overall enhanced care group and compared to C1 (usual care among consistently employed) and C2 (usual care among inconsistently employed) combined (usual care group), enhanced care led to a significant improvement in productivity over 24 months, increasing productivity by 6.1% (-2LL χ^2=6.0, p<0.05). Among consistently employed individuals only, I1 (enhanced care) significantly improved productivity when compared to C1 (usual care) (-2LL χ^2=7.8, p=0.03), leading to an 8.2% increase in productivity over 24 months. Among inconsistently employed individuals only, I2 (enhanced care) had no significant impact on productivity when compared to C2 (usual care) (-2LLχ^2=0.0, p=0.99). **Work Disability and Recurrences of Work Disability Outcomes:** 1. Absenteeism: When I1 and I2 were combined and compared to C1 and C2 combined, enhanced care led to a non-significant reduction in absenteeism over 24 months, reducing absenteeism by 22.8% or 10.6 days over 24 months (-2LL χ^2=5.6, p=0.06). In the analysis of consistently employed individuals, I1 tended to lead to improve absenteeism when compared to C1 (-2LL χ^2=5.1, p=0.08), reducing absenteeism non-significantly by 28.4% or 12.3 days over 24 months. Among inconsistently employed individuals, I2 had no significant impact on absenteeism when compared to C2 (-2LL χ^2=0.0, p=0.64). **Economic Outcomes:** See Table 6 **SECONDARY OUTCOMES** **Depression Outcomes:** 1. Depression Severity (CES-D): When I1 and I2 were combined and compared to C1 and C2 combined, enhanced care had minimal impact on depression severity over 24 months (F=2.72, p=0.09). Among consistently employed individuals, I1 significantly decreased depression severity when compared to C1 (F=5.15, p=0.02), while it had no impact among inconsistently employed individuals (F=0.24, p=0.62). **Other Secondary Outcomes:** 1. Specialty Care Counselling: When I1 and I2 were combined and compared to C1 and C2 combined, enhanced care led to a significant increase in the proportion of individuals reporting specialty care counselling over 24 months (40.9% versus 25.8%, p<0.05). Among consistently employed individuals, I1 significantly increased specialty care counselling use when compared to C1 (41.3% versus 23.2%, p<0.05), while among inconsistently employed individuals, I2 had no significant impact on specialty care counselling use when compared to C2 (36.7% versus 35.1%, p>0.20). 2. Emotional Role Functioning (SF-36): When I1 and I2 were combined and compared to C1 and C2 combined, enhanced care had no significant impact on emotional role functioning (-2LL χ^2=1.8, p=0.41). Among consistently employed individuals, I1 significantly increased emotional role functioning when compared to C1 (-2LL χ^2=8.1, p<0.03), while it had no impact among inconsistently employed individuals (-2LL χ^2=3.6, p=0.16). 3. Anti-Depressant Use: When I1 and I2 were combined and compared to C1 and C2 combined group, enhanced care led to a non-significant increase in the number of months of anti-depressant use over 24 months (8.9 versus 8.0, p=0.10). Among consistently employed individuals, I1 significantly increased the number of anti-depressant months when compared to C1 (9.1 versus 7.8, p<0.05), while in the analysis of inconsistently employed individuals, I2 had no significant impact on anti-depressant months when compared to C2 (8.3 versus 8.7, p>0.20).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
7. Rebergen, 2009a Rebergen, 2009b	Kaplan-Meier curves Cox proportional hazards analysis t test Chi square test	For RTW process, number of recurrences, and duration of recurrences: none For duration of sick leave until partial or full RTW: treating OP, HADS total score, children, and number of sick leave periods in the previous year For total productivity loss: DASS depression/anxiety, work relatedness of the disorder, and number of sick leave periods in the previous year	**PRIMARY OUTCOMES** **Work Disability and Recurrences of Work Disability Outcomes:** 1. RTW process: Individuals in the I group (guideline-based care for mental health disorders by OPs) were significantly more likely to have a partial RTW (69%) before full RTW, compared to the C group (usual care) (54%) (χ^2 NR, p=0.01) 2. Number of recurrences of sick leave periods: The mean number of recurrences was not significantly different between I (1.7, SD 1.9) and C (1.4, SD 1.5) (t-test NR, p=0.08). 3. Duration of recurrences of sick leave periods: The mean duration of recurrences was not significantly different between I (19.4, SD 39.0) and C (18.6, SD 39.1) (t-test NR, p=0.95). 4. Duration of sick leave until partial RTW: Mean duration until partial RTW was not significantly different between I (53.1, SD 56.3) and C (50.6, SD 78.4) (t-test value NR, p=0.28) in unadjusted analyses. In adjusted analyses, the median number of sick leave days until partial RTW was also not significantly different between I (50 days, 95%CI 34 to 66) and C (47 days, 95%CI 31 to 63) (HR 0.99, 95%CI 0.75 to 1.31, p=0.94). 5. Duration of sick leave until full RTW: In adjusted analyses, the median number of sick leave days until full RTW was not significantly different between I (105 days, 95%CI 84 to 126) and C (104 days, 95%CI 81 to 127) (HR 0.96, 95%CI 0.73 to 1.27, p=0.78). In ancillary analyses, type of work function was found to be a significant modifier, with workers in administrative functions benefiting more from the intervention than workers with executive functions (p=0.03, statistics NR). 6. Total productivity loss (duration of sick leave days, including recurrences, until full RTW): In adjusted analyses, the mean number of days of total productivity loss was not significantly different between I (151 days, SD 97) and C (147 days, SD 102) (HR 1.21, 95%CI 0.86 to 1.71, p=0.28). In ancillary analyses, a significant interaction effect (p=0.02) was found for the severity of the disorder (DASS-depression/anxiety) with the intervention - for workers with a depressive or anxiety state, C seemed to be more effective than I in reducing total productivity loss. For this subgroup, the hazard ratio, however, was not significant (HR=0.67, 95%CI 0.36 to 1.26, p=0.21, adjusted for diagnosis of the OP, the treating OP, and gender). **Economic Outcomes:** See Table 6

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
8. Schene, 2007	Cox regression analyses Generalized estimating equations Generalized linear model	For work stress: baseline score on the QOS For presence of major depression, depression severity, and working at least 2 days or 16 hours per week: baseline BDI score For time until any work resumption and total hours worked within each 6-month period: NR	**PRIMARY OUTCOMES** **Work Disability and Recurrences of Work Disability Outcomes:** 1. Time until any work resumption: Among patients not working at baseline, the mean number of days from baseline to any work resumption was significantly lower in the I (psychiatric treatment, plus OT) group (207 days) compared to the C (treatment as usual) group (299 days) (HR=2.71, 95%CI 1.16 to 6.29, p=0.01). 2. Total hours worked within each 6-month period up to 42 months: Over the first 18 months, individuals in I worked significantly more hours than those in C. Namely, between 7 and 12 months since baseline, the median number of hours worked was 261.75 for I and 0.85 for C (χ^2=4.13, p=0.042), while between 13 and 18 months, the median number of hours was 456.25 for I and 156.42 for C (χ^2=4.46, p=0.035). This trend continued from months 19 to 42, but for each 6-month period, the differences in median hours worked between the intervention groups was non-significant (months 19 to 24, 456.25 for I, 91.25 for C, χ^2=1.42, p=0.234; months 25 to 30, 397.58 for I, 0.0 for C, χ^2=0.44, p=0.509; months 31 to 36, 391.07 for I, 130.35 for C, χ^2=1.11, p=0.293; months 37 to 42, 404.10 for I, 0.0 for C, χ^2=0.62, p=0.431). 3. The proportion of working at least 2 days or 16 hours per week: From months 0 to 18, the proportion of patients working at least 2 days or 16 hours per week significantly increased in both groups (χ^2=15.81, p=0.001), from 9% in I and 11% in C in months 0 to 6, to 52% in I and 22% in C in months 13 to 18. From months 19 to 42 (p=0.387), further increases were small in both groups and by 42 months, 57% of the I group, 42% of the C group were working at least 2 days/16 hours per week. There was no significant difference between I and C in both months 0 to 18 (χ^2=6.27, p=0.099) and 19 to 42 (χ^2=3.12, p=0.374) **Economic Outcomes:** See Table 6 **SECONDARY OUTCOMES** **Depression Outcomes:** 1. Presence of major depression (DSM-IV): The proportion of patients continuing to meet criteria for major depression fell to 44% and 29% at 12 months for I and C, respectively. Recovery was significant over the first 6 months (χ^2=7.28, p=0.007), but not over the second 6 months (χ^2=1.41, p=0.234). When I and C were compared, the difference in the proportion continuing to meet criteria for depression between 0 and 6 months (χ^2=2.59, p=0.107) and 7 and 12 months (χ^2=0.19, p=0.667) was not statistically significant. 2. Depression severity (BDI): The mean BDI total score significantly decreased between baseline and 12 months from 27.1 to 17.1 for the I group and from 23.6 to 13.8 for the C group (F=60.83, p=0.000). From 13 to 42 months, there was a further decrease in the I group to 12.3, but the mean BDI score remained stable at 14.0 (F=3.19, p=0.80). When I and C were compared, the difference in depression severity over time was not significant in the first 12 months (F=0.00, p=0.950), but became statistically significant for BDI scores between months 13 and 42 (F=4.82, p=0.032). **Other Secondary Outcomes:** 1. Work stress (QOS): Examined in two specific subgroups: patients working a mean of at least half a day (4 hours) per week over a 6-month period and patients working at least 2 days (16 hours) per week over a 6-month period. In both subgroups, there was no increase in work stress between baseline and 6 months and baseline and 12 months (data not shown) and there was no significant difference between I and C (treatment as usual) (data not shown).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
9. Schoenbaum 2001	Mixed effect linear regression	Age, sex, marital status, education, rank in the distribution of household wealth, employment status, medical co-morbidity, depressive disorder status, the SF-12, aggregate HRQOL measures, presence of co-morbid anxiety disorder, and practice randomization block.	**PRIMARY OUTCOMES** **Work Disability and Recurrences of Work Disability Outcomes:** 1. Days of employment over 24 months: Compared to the C group (usual care), which had a mean of 279.2 days of employment over 24 months (95%CI 270.2 to 288.1), I1 (quality improvement to medication access in primary care) had, on average, 17.9 more employed days (95%CI −1.6 to 37.4, t=1.87, p=0.07), though non-significant and I2 (quality improvement in psychotherapy in primary care) had 20.9 more employed days (95%CI 2.4 to 39.3, t=2.30, p=0.03). 2. Days missed from work due to illness over 24 months: For the subgroup of participants who were working, I and C groups did not differ substantially or statistically with respect to sick days at any follow-up period. For instance, at the 12-month follow-up, the number of reported sick days in the previous 4 weeks was similar between groups (1.2 days in I1 and I2 combined and 1.1 days in C, 95%CI for the difference -0.5 to 0.6, p=0.81). Data for other follow-up periods not shown, but authors remark that results for other periods are similar. **Economic Outcomes:** See Table 6 **SECONDARY OUTCOMES** **Depression Outcomes:** 1. Days of depression burden: Compared to the C group, which had a mean of 419.9 days of depression burden over 24 months (95%CI 398.9 to 441.0), I1 had 25 (95%CI -63.1 to 13.2) fewer depression burden days, on average, though the result was not statistically significant (t=-1.33, p=0.19). On the other hand, I2 had an average of 46.7 (95%CI -83.1 to -10.3) fewer depression burden days, a finding that was statistically significant (t=-2.61, p=0.01). **Other Secondary Outcomes:** 1. Quality-adjusted life-years (QALYs): Mean QALYs over 24 months for the C group was 1.6624 (95%CI 1.628 to 1.697). The incremental increase in QALYs over 24 months for I1 (QI-Meds) over C was not significantly different (0.0115, 95%CI -0.004 to 0.027, t=1.49, p=0.15), while the increase in QALYs for I2 (QI-therapy) over C was significant (0.0226, 95%CI 0.008 to 0.038, t=2.94, p=0.006).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
10. Smith, 2002	Logistic regression	Age, gender, minority status, marital status, educational attainment, household income adjusted by family size, health insurance status, baseline depression symptom severity, depression diagnosis, physical co-morbidity, recent depression treatment, patient receptivity to antidepressant treatment, occupation, paid time off for doctor visits, unemployment rate in patients county of residence (subsequent employment outcome only), and baseline workplace conflict (workplace conflict outcome only)	**PRIMARY OUTCOMES** **Work Disability and Recurrences of Work Disability Outcomes:** 1. Subsequent employment: At 6 months, there was no significant difference in the proportion of individuals in the I (enhanced depression primary care) and C (usual care) groups reporting subsequent employment (data no shown). However, by 12 months, the proportion in the I group reporting subsequent employment (92.1%) was significantly higher than in the C group (82.0%), a difference of 10.1% (χ^2=4.42, 90%CI 2.8 to 17.4, p=0.04). **SECONDARY OUTCOMES** **Other Secondary Outcomes:** 1. Workplace conflict in previous 12 months (arguments/difficulties with co-workers): Among those employed at 12 months, individuals in the I group were significantly less likely than those in the C group to report having a workplace conflict in the past year (8.1% versus 18.9%, respectively) (χ^2=4.11, p=0.04).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
11. van der Feltz-Cornelis, 2010	Logistic regression (propensity score calculation) Cox proportional hazards regression Chi-square test	Propensity score of income and functioning	**PRIMARY OUTCOMES** **Work Disability and Recurrences of Work Disability Outcomes:** 1. Full RTW status: At 3 month follow-up, a significantly higher proportion of individuals in the I group (enhanced OP role with psychiatric consultation and advice) (58%) had made a full RTW compared to 44% in the C group (usual care) (χ^2=NR, p=0.0093). At 6 month follow-up, there was no longer a difference in the proportion making a full RTW (85% in I, 84% in C, x2=NR, p=0.0574). 2. Time to full RTW: Time until full RTW from baseline was faster for the I group (122 days, 95%CI 77 to 166) than for the C group (190 days, 95%CI 134 to 246) for a mean difference of 68 days (χ^2=3.101, p=0.078). **SECONDARY OUTCOMES** **Depression Outcomes:** 1. Depression severity (PHQ-9 and PHQ-15): Both I and C demonstrated decreases between baseline and 6 month follow-up in PHQ-9 scores (mean difference -4.00, SD 6.94 in I; -4.708, SD 4.53 in C) and PHQ-15 scores (mean difference -4.05, SD 5.08 in I; -3.750, SD 4.17 in C), but the difference in degree of decrease between I and C was not significant for either PHQ-9 15 (β=0.913, 95%CI -2.62 to 4.45, p=0.605) or PHQ-15 (β=-0.178, 95%CI -3.01 to 2.66, p=0.900). 2. Depression severity (SCL-90): Both I and C demonstrated decreases between baseline and 6 month follow-up in SCL-90 scores (mean difference -2.475, SD 0.73 in I; -3.05, SD 0.50 in C), but the difference in degree of decrease between I and C was not significant (β=0.03, 95%CI -0.35 to 0.41, p=0.872). **Other Secondary Outcomes:** 1. Quality-adjusted life years (EQ-5D): Both I and C demonstrated increases between baseline and 6 month follow-up in quality-adjusted life years (mean difference 0.378, SD 0.06 in I; 0.374, SD 0.11 in C), but the difference in degree of increase between I and C was not significant (β=0.005, 95%CI -0.05 to 0.06, p=0.869).

Author, Year	Final Statistical Analyses	Covariates/ Confounders Controlled for in Analyses	Main Findings
12. Wang, 2007	Linear regression Logistic regression	Outcomes at the prior assessment, expected hours of work at the prior assessment, age, sex, and education	**PRIMARY OUTCOMES** **Work Functioning Outcomes:** 1. Effective weekly hours worked (HPQ), hours (SD): Mean effective weekly hours were significantly higher at 6 month follow-up in the I group (telephone outreach and care management group) (30.1 hours, SD 14.5) compared to the C group (usual care) (27.1 hours, SD 15.5) (β=3.0, 95%CI 0.4 to 5.6, p=0.03). Similarly, mean effective weekly hours were significantly higher at 12 month follow-up in the I group (29.5 hours, SD 14.5) compared to the C group (usual care) (26.0 hours, SD 15.8) (β=3.3, 95%CI 0.9 to 5.8, p=0.008). 2. On-the-job performance (HPQ): At 6 months, there was no significant difference in on-the-job performance scores between the I group (0.8, SD 0.2) compared to the C group (0.7, SD 0.2) (β=0.2, 95%CI -0.2 to 0.5, p=0.35) and at 12 months, there continued to be no difference between I (0.8, SD 0.2) and C (0.7, SD 0.2) (β=0.2, 95%CI -0.2 to 0.6, p=0.40). **Work Disability and Recurrences from Work Disability Outcomes:** 1. Actual weekly hours worked among the employed: At 6 months, there was no significant difference in the mean number of actual weekly hours worked between the I group (42.0 hours, SD 15.4) and C group (40.1 hours, SD 15.6) (β=1.8, 95%CI -0.8 to 4.4, p=0.18). At 12 months, there continued to be no difference between I (42.3 hours, SD 13.4) and C (39.5 hours, SD 13.7) (β=2.1, 95%CI -0.4 to 4.5, p=0.09). 2. Job retention: (HPQ): At 6 months, the I group had significantly higher job retention (96.1%, SD 3.7) than the C group (90.1%, SD 8.9) (OR=2.5, 95%CI 1.2 to 5.0, p=0.007). At 12 months, the I group continued to demonstrate higher job retention (92.6%, SD 6.8) compared to the C group (88.0%, SD 10.6) (OR=1.7, 95%CI 1.0 to 3.3, p=0.07), though not significantly. **SECONDARY OUTCOMES** **Depression Outcomes:** 1. Depression severity (QIDS-SR score): By 6 month follow-up, mean QIDS-SR scores were significantly lower in the I group (10.2, SD 4.8) than the C group (11.2, SD 4.9) (B=-1.0, 95%CI -1.8 to 0.2, p=0.01). Similarly, at 12 months, mean scores were also significantly lower in I (8.9, SD 4.8) than C (10.0, SD 4.7) (B=-1.1, 95%CI -1.8 to 0.3, p=0.005). 2. Substantial improvement in depression symptoms (≥50% reduction in QIDS-SR score): The proportion of participants with substantial improvement was similar for the I (21.7%, SD 17.0) and C (17.4%, SD 14.4) groups at 6 months (OR=1.2, 95%CI 0.8 to 2.0, p=0.20). However, by 12 months, a significantly higher proportion of individuals in the I group (30.9%, SD 21.4) compared to the C group (21.6%, SD 16.9) had experienced a substantial symptom improvement (OR=1.7, 95%CI 1.1 to 2.5, p=0.01). 3. Complete remission (QIDS-SR score ≤5): By 6 month follow-up, the proportion of participants in complete remission was significantly higher in the I group (18.2%, SD 14.9) than the C group (12.6%, SD 11.0) (OR=1.7, 95%CI 1.0 to 2.5, p=0.05). Similarly, at 12 month follow-up, the proportion of participants in complete remission was significantly higher in the I group (26.2%, SD 19.3) than the C group (17.7%, SD 14.6) (OR=1.7, 95%CI 1.1 to 2.4, p=0.01). **Other Secondary Outcomes:** 1. Critical workplace incidents (HPQ): At 6 months, there was no significant difference in the mean number of critical workplace incidents between the I group (-0.2, SD 0.6) compared to the C group (-0.2, SD 0.6) (β=0.00, 95%CI -0.1 to 0.1, p=0.93) and at 12 months, there continued to be no difference between I (-0.2, SD 0.6) and C (-0.2, SD 0.6) (β=0.05, 95%CI -0.0 to 0.2, p=0.29).

NA: Not Applicable, **NR**: Not Reported, **FTE**: Full Time Equivalent, **HRQOL:** Health-related Quality of Life, **QOS**: Questionnaire Organization Stress, **TAU**: Treatment as Usual, **OT**: Occupational Therapy, **RTW**: Return to Work, **LTD**: Long Term Disability, **STD**: Short Term Disability, **HR**: Hazard Ratio, HADS: Hospital Anxiety Depression Scale, **OP**: Occupational Physician, **QIDS-SR**: Quick Inventory of Depression Symptoms Self-Report, **MBI-NL**: Maslach Burnout Inventory, **DASS**: Depression Anxiety Stress Scale, **ANOVA**: Analysis of Variance, **MANOVA**: Multivariate Analysis of Variance, **CI**: confidence interval, **CBT**: Cognitive Behavioural Therapy, **OR**: Odds Ratio
+ I1, I2, etc., -2LL: minus 2 log-likehood

Appendix I

Detailed assessment of risk of bias

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
1. Blonk 2006 The Netherlands*	*Criteria Met:* 1. Participation rate >65% *Criteria Not Met:* 2. Intention-to-treat analyses not completed. After randomization, investigators removed participants who did not receive the intervention because of a misunderstanding (8 in total), breaking the randomization. *Unclear if Criteria Met:* 3. Described as randomized, but no description of randomization procedure 4. No description of potential differences between participants and non-participants 5. No description of baseline characteristics between groups at	*Criteria Met:* 1. Losses to follow-up <35% of sample 2. No major differences between remaining participants and those lost to follow-up *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for co-interventions appears minimal *Criteria Not Met:* None *Unclear if Criteria Met:* 3. Potential for contamination. Participants receiving psychotherapy pre-study asked to postpone therapy until study completion. Unclear whether other participants were also asked to post-pone therapy 4. No description of compliance with intervention 5. Intervention providers not blinded to actual intervention. However, unclear whether providers	*Criteria Met:* 1. Valid and reliable instruments 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) 4. Blinding of outcome assessors not applicable due to self-administered questionnaires *Criteria Not Met:* None *Unclear if Criteria Met:* 5. No description of blinding of participants (see previous column) **Risk Judgement:** Low Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

	baseline 6. No adjustment for baseline differences, but unclear whether required (see item 5) **Risk Judgement:** High Risk		were blinded to the study hypothesis and/or presence of a comparison treatment. 6. Participants not blinded to actual intervention. However, unclear whether participants were blinded to the study hypothesis and/or presence of a comparison treatment. **Risk Judgement:** Moderate Risk			

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
2. Dewa 2009 Canada*	*Criteria Met:* 1. Active participation not applicable – intervention and control treatments were considered standard at the time each was provided and administrative data for outcomes were used. Data from 100% of controls were used, while 97.3% of intervention group were included (2.7% had missing age data and were excluded). Omission of these 2 individuals led to more conservative results, but likely had a minimal effect 2. Major differences in baseline age found. However, adjustment was made in the analyses to account for these differences 3. Intention-to-treat analyses not applicable *Criteria Not Met:* 4. Non-randomized	*Criteria Met:* 1. Losses to follow-up not applicable – all data drawn from an administrative database *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for co-interventions appear minimal 3. Potential for contamination appears minimal *Criteria Not Met:* 4. Providers not blinded to intervention or its purpose 5. Participants not blinded to intervention *Unclear if Criteria Met:* 6. No description of compliance with intervention **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Valid and reliable outcome assessment 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) 4. Participants not blinded to actual intervention, but should have no effect on outcome measurement as outcomes are objective measures obtained from administrative data 5. Blinding of outcome assessors not applicable as data extracted solely from an administrative database *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

	study, with the intervention occurring between June 2006 and May 2007 and historical controls selected from individuals receiving formerly standard treatment in the year prior to intervention implementation *Unclear if Criteria Met:* None **Risk Judgement:** High Risk					

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
3. Kawakami 1997 Japan*	*Criteria Met:* 1. Intention-to-treat analyses completed *Criteria Not Met:* 2. Major differences in baseline characteristics (sex, sick leave) found between groups. While adjustment was made for sex in the analyses, there was no adjustment to account for baseline differences in sick leave 3. Intervention allocation method not performed adequately. Allocation occurred at the worksite level. First two worksites appearing within the company directory meeting eligibility criteria were allocated to intervention. Control worksites matched to intervention worksites on a number of factors. *Unclear if Criteria Met:* 4. Rate of participation	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* 2. Major differences in age and sex between remaining participants and those lost to follow-up *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	*Criteria Met:* 1. Potential for co-interventions appears minimal 2. Potential for contamination appears minimal *Criteria Not Met:* 3. Intervention process not adequately described *Unclear if Criteria Met:* 4. No description of compliance with intervention 5. Intervention providers not blinded to actual intervention. However, unclear whether providers were blinded to the study hypothesis and/or presence of a comparison treatment. 6. Participants not blinded to actual intervention. However, unclear whether participants were blinded to the study hypothesis and/or presence of a comparison	*Criteria Met:* 1. Valid and reliable instruments 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) *Criteria Not Met:* None *Unclear if Criteria Met:* 4. No description of blinding of participants (see previous column) 5. No description of blinding of outcome assessors **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

	unclear. All individuals in the respective worksites participated, but recruitment was at the level of the worksite. It's unclear how many worksites were approached and if any declined participation 5. No description of potential differences between participants and non-participants, but unclear if required (see item 4 above) **Risk Judgement:** High Risk		treatment. **Risk Judgement:** Moderate Risk			

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
4. Knekt 2008a Finland* Knekt, 2008b	*Criteria Met:* 1. Allocation method adequate 2. Participation rate >65% 3. Major differences in baseline characteristics (education, occupation, psychiatric diagnosis) found between groups. However, adjustment was made in the analyses to account for these differences 4. Intention-to-treat analyses completed *Criteria Not Met:* None *Unclear if Criteria Met:* 5. No description of potential differences between participants and non-participants **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* 2. Some information provided on differences between remaining participants and those lost to follow-up – namely, higher symptoms and perceived need for psychiatric treatment were more common in the solution-focused therapy group *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for contamination appears minimal 3. Compliance with intervention adequate. *Criteria Not Met:* None *Unclear if Criteria Met:* 4. Potential for co-interventions. Auxiliary use of psychotherapy more common in solution-focused and short-term therapy groups than long-term therapy. Overall, 40% used psychotrophic medications (not specified by group). Also, 3.1% treated in psychiatric hospital (none in solution-focused). 5. Intervention providers not blinded to actual intervention. However, unclear whether providers	*Criteria Met:* 1. Valid and reliable instruments 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) *Criteria Not Met:* 4. Outcome assessors not blinded *Unclear if Criteria Met:* 5. No description of blinding of participants (see previous column) **Risk Judgement:** High Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

			were blinded to the study hypothesis and/or presence of a comparison treatment. 6. Participants not blinded to actual intervention. However, unclear whether participants were blinded to the study hypothesis and/or presence of a comparison treatment. **Risk Judgement:** Moderate Risk			

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
5. Krogh 2009 Denmark* Supplemental: Krough, 2007	*Criteria Met:* 1. Allocation method adequate 2. Participation rate >65% 3. Performed intention-to-treat analyses *Criteria Not Met:* 4. Major differences in baseline characteristics (sex, HAM-D depression score, time since dx, use of anti-depressants, employment status, % of days absent, % on sick leave) found between groups. No adjustment was made in the analyses to account for these differences *Unclear if Criteria Met:* 5. No description of potential differences between participants and non-participants **Risk Judgement:** High Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample 2. No major differences between remaining participants and those lost to follow-up *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Intervention process adequately described *Criteria Not Met:* 2. Potential for contamination and co-interventions as both providers and participants were not blinded 3. Compliance with the intervention inadequate. Mean participation was 18.0 (56.2%), 16.2 (50.6%), and 10.5 (32.8%) sessions of the 32 sessions in the strength, aerobic, and relaxation groups, respectively 4. Intervention providers not blinded 5. Participants not blinded *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	*Criteria Met:* 1. Valid and reliable instruments 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) 4. Outcome assessors blinded to depression-related outcomes *Criteria Not Met:* 5. Outcome assessors not blinded to work-related outcomes 6. Participants not blinded *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
6. Lo Sasso 2006 USA* Rost 2004 USA* Supplemental: Rost, 2000 Rost, 2001	*Criteria Met:* 1. Participation rate >65% 2. Major differences in baseline characteristics (marital status, occupation, full-time or part-time employment) found between groups. However, adjustment was made in the analyses to account for these differences 3. Performed intention-to-treat analyses *Criteria Not Met:* 4. Major differences between participants and non-participants with respect to sex, age, minority status, and depression symptoms. *Unclear if Criteria Met:* 5. Described as randomized, but no description of randomization procedure **Risk Judgement:** High Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* None *Unclear if Criteria Met:* 2. No description of potential differences between remaining participants and those lost to follow-up **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for co-interventions appears minimal 3. Potential for contamination appears minimal 4. Compliance with intervention adequate *Criteria Not Met:* 5. Intervention providers not blinded 6. Participants not blinded to actual intervention, but appear to have been blinded to the study's research question *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Valid and reliable instruments 2. Follow-up of sufficient duration (>3 months) 3. Outcome assessors blinded *Criteria Not Met:* 4. Outcomes described at follow-up only 5. Participants not blinded to actual intervention *Unclear if Criteria Met:* None **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison *Criteria Not Met:* 3. Statistical analyses inappropriate. No baseline data existed for the productivity outcome, so regression-based statistical forecasting methods were used to estimate baseline values for all participants *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	**High Risk of Bias**

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
7. Rebergen 2009a The Netherlands* Rebergen 2009b The Netherlands* Supplemental: Rebergen, 2007	*Criteria Met:* 1. Allocation method adequate 2. Baseline characteristics described and balanced 3. Performed intention-to-treat analyses *Criteria Not Met:* None *Unclear if Criteria Met:* 4. No description of participation rate 5. No description of potential differences between participants and non-participants **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* None *Unclear if Criteria Met:* 2. No description of potential differences between remaining participants and those lost to follow-up **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for co-interventions appears minimal *Criteria Not Met:* 3. Potential contamination. All occupational physicians (OPs) were trained in the guideline intervention, but also treated individuals in the control group, risking that OP may have provided guideline-based care to the control group. Also, it was demonstrated that 38% of participants in the guideline-based intervention group received a referral for a psychologist. 4. Intervention provider not blinded 5. Participants not blinded *Unclear if Criteria Met:* 6. No description of compliance with the intervention. **Risk Judgement:** High Risk	*Criteria Met:* 1. Valid and reliable instruments 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) 4. Participants not blinded to actual intervention, but should have no effect on outcome measurement as outcomes are objective measures obtained from administrative data 5. Blinding of outcome assessors not applicable as data extracted solely from an administrative database *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
8. Schene 2007 The Netherlands*	*Criteria Met:* 1. Allocation method adequate 2. Participation rate >65% 3. Performed intention-to-treat analyses *Criteria Not Met:* 4. Major differences in baseline characteristics (marital status, living arrangement, nature of depression episode, depression severity) found between groups. No adjustment was made in the analyses for these differences *Unclear if Criteria Met:* 5. No description of potential differences between participants and non-participants **Risk Judgement:** High Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* None *Unclear if Criteria Met:* 2. No description of potential differences between remaining participants and those lost to follow-up **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for co-interventions appears minimal 3. Compliance with intervention adequate *Criteria Not Met:* 4. Some degree of contamination. 13% in the intervention group stopped following the 4-week diagnostic phase 5. Intervention providers not blinded *Unclear if Criteria Met:* 6. Participants not blinded to actual intervention. However, no information on whether participants were blinded to the study hypothesis and/or presence of a comparison treatment. **Risk Judgement:** High Risk	*Criteria Met:* 1. Valid and reliable instruments used 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) 4. Outcome assessors blinded *Criteria Not Met:* None *Unclear if Criteria Met:* 6. No description of blinding of participants (see previous column) **Risk Judgement:** Low Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
9. Schoenbaum 2001 USA* Supplemental: Schoenbaum, 2004 Jaycox, 2003 Miranda, 2003 Sherbourne, 2001 Wells, 1999 Wells, 2001	*Criteria Met:* 1. Participation rate >65% 2. Major differences in baseline characteristics (age, education, history of lifetime disorder) found between groups. However, adjustment was made in the analyses for these differences 3. Performed intention-to-treat analyses *Criteria Not Met:* None *Unclear if Criteria Met:* 4. Described as randomized, but no indication of whether allocation was concealed 5. No description of potential differences between participants and non-participants **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* 2. Major differences in age, sex, study site, marital status, ethnicity, and education between remaining participants and those lost to follow-up *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for co-interventions appears minimal *Criteria Not Met:* 3. Compliance with the intervention appears inadequate. 30% of QI-Meds patients were not followed for the full duration, while only 40% of patients in the QI-Therapy group received CBT 4. Intervention providers were not blinded 5. Participants were not blinded *Unclear if Criteria Met:* 6. Potential for contamination. Since providers were not blinded, it is possible that usual care clinicians, while not receiving extra materials, may have been more diligent with their depression care **Risk Judgement:** High Risk	*Criteria Met:* 1. Valid and reliable instruments used 2. Follow-up of sufficient duration (>3 months) 3. No description of blinding of outcome assessors, but should have minimal effect on outcome measurement as data for the majority of outcomes were collected by self-administered questionnaires and administrative practice data. Only economic outcomes were assessed by telephone interview. *Criteria Not Met:* 4. Outcomes described at follow-up only 5. Participants were not blinded *Unclear if Criteria Met:* **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison *Criteria Not Met:* 3. Inappropriate statistical analyses. Multiple imputation used for missing items at each follow-up (anywhere from 13-22% of data missing, depending on the follow-up) *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	**High Risk of Bias**

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
10. Smith 2002 USA* Supplemental: Rost, 2001	*Criteria Met:* 1. Participation rate >65% 2. Major differences in baseline characteristics (age, symptom severity, education) found between groups. However, adjustment was made in the analyses for these differences 3. Performed intention-to-treat analyses *Criteria Not Met:* 4. Major differences between participants and non-participants with respect to sex, age, minority status, and depression symptoms. *Unclear if Criteria Met:* 5. Described as randomized, but no description of randomization procedure **Risk Judgement:** High Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* None *Unclear if Criteria Met:* 2. No description of potential differences between remaining participants and those lost to follow-up **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for co-interventions appears minimal 3. Potential for contamination appears minimal 4. Compliance with intervention adequate *Criteria Not Met:* None *Unclear if Criteria Met:* 5. Intervention providers not blinded 6. Participants not blinded to actual intervention, but appear to have been blinded to the study's research question **Risk Judgement:** Low Risk	*Criteria Met:* 1. Valid and reliable instruments used 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) 4. Outcome assessors blinded *Criteria Not Met:* 5. Participants not blinded to actual intervention *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
11. van der Feltz-Cornelis 2010 The Netherlands* Supplemental: van der Feltz-Cornelis, 2007	*Criteria Met:* 1. Allocation method performed adequately 2. Participation rate 100% 2. Major differences in baseline characteristics (sex, marital status) found between groups. However, adjustment was made in the analyses to account for these differences 3. Performed intention-to-treat analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* None *Unclear if Criteria Met:* 2. No description of potential differences between remaining participants and those lost to follow-up **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for co-interventions appears minimal *Criteria Not Met:* 3. Potential for contamination. All Occupational Physicians (OP) were trained in the intervention and only then randomized, therefore, they may have treated the controls with the intervention protocol. They "often" referred patients in the care as usual group for psychiatric treatment. 4. Provider not blinded 5. Participants not blinded *Unclear if Criteria Met:* 6. No description of compliance with the intervention **Risk Judgement:** High Risk	*Criteria Met:* 1. Valid and reliable instruments 2. Follow-up of sufficient duration (>3 months) 3. Outcome assessors blinded *Criteria Not Met:* 4. Outcomes described at follow-up only 5. Participants not blinded *Unclear if Criteria Met:* None **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison 3. Appropriate statistical analyses *Criteria Not Met:* None *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	**High Risk of Bias**

First Author Year Country	Selection Bias	Attrition Bias	Performance Bias	Measurement Bias	Reporting Bias	Overall Risk of Bias Judgement
12. Wang 2007 USA*	*Criteria Met:* 1. Baseline characteristics described and balanced 2. Performed intention-to-treat analyses *Criteria Not Met:* 3. Allocation method random, but does not appear to have been concealed *Unclear if Criteria Met:* 4. Participation rate unclear. No information on the number who declined from those eligible to participate 5. No description of potential differences between participants and non-participants, if required (see item 4) **Risk Judgement:** High Risk	*Criteria Met:* 1. Losses to follow-up <35% of sample *Criteria Not Met:* 2. Major differences between remaining participants those lost to follow-up with respect to age, education, sex, depression severity, and work performance *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	*Criteria Met:* 1. Intervention process adequately described 2. Potential for contamination appears minimal 3. Potential for co-interventions appears minimal 4. Compliance with the intervention adequate *Criteria Not Met:* 5. Participants not blinded *Unclear if Criteria Met:* 6. Intervention providers not blinded to actual intervention. However, no information on whether providers were blinded to the study hypothesis and/or presence of a comparison treatment. **Risk Judgement:** Moderate Risk	*Criteria Met:* 1. Valid and reliable instruments 2. Outcomes described at baseline and follow-up 3. Follow-up of sufficient duration (>3 months) 4. Outcome assessors blinded *Criteria Not Met:* 5. Participants not blinded *Unclear if Criteria Met:* None **Risk Judgement:** Low Risk	*Criteria Met:* 1. Research question clearly stated 2. Direct between-group comparison *Criteria Not Met:* 3. Statistical analyses inappropriate. Multiple imputation was used to impute outcome data despite that 10-15% were lost to follow-up and major differences existed between remaining participants and those lost to follow-up *Unclear if Criteria Met:* None **Risk Judgement:** High Risk	**High Risk of Bias**

Appendix J

Summary of findings for the primary outcomes†

Interventions		*Primary Outcomes*								
		Prevention of Work Disability		*Management of Work Disability*		*Work Functioning*		*Recurrence of Work Disability*		*Economic Benefit*
		Short Term	*Long Term*	*Short Term*	*Long Term*	*Short Term*	*Long Term*	*Short Term*	*Long Term*	
Psychological Interventions	Cognitive behavioural therapy delivered by psychologists (Blonk, 2006)				-[103] ⊙○○○ =[104], =[105]					
	Brief cognitive behavioural therapy-based stress management delivered by labour experts with a focus on improving workplace processes (Blonk, 2006)				+[106] ⊙○○○ =[107]					
	Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)	=[108], =[109]	-[110] ⊙○○○ =[111], =[112]	=[113]	=[114]	+[115] ⊙○○○ =[116], =[117]	-[118] ⊙○○○ =[119], =[120]			

[103] Time until full RTW, compared to brief cognitive behavioural therapy-based stress management delivered by labour experts
[104] Time until partial RTW, compared to compared to brief cognitive behavioural therapy-based stress management delivered by labour experts, as well as usual care
[105] Time until full RTW, compared to usual care
[106] Time until full RTW, compared to cognitive behavioural therapy delivered by psychologists, as well as usual care
[107] Time until partial RTW, compared to cognitive behavioural therapy delivered by psychologists, as well as usual care
[108] Two outcomes: 1) Number of sick leave days during last 3 months and 2) Proportion with more than 7 sick leave days during last 3 months, both compared to short-term and long-term psychotherapy
[109] Current employment status, compared to short-term and long-term psychotherapy. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[110] Proportion with more than 7 sick leave days during last 3 months, compared to short-term and long-term psychotherapy
[111] Current employment status, compared to short-term and long-term psychotherapy. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[112] Number of sick leave days during last 3 months; compared to short-term and long-term psychotherapy
[113] Current employment status, compared to short-term and long-term psychotherapy. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[114] Current employment status, compared to short-term and long-term psychotherapy. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

Interventions		Primary Outcomes								
		Prevention of Work Disability		Management of Work Disability		Work Functioning		Recurrence of Work Disability		Economic Benefit
		Short Term	Long Term	Short Term	Long Term	Short Term	Long Term	Short Term	Long Term	
	Short-term psychodynamic psychotherapy (Knekt 2008)	=[121], =[122]	+[123] ⊙○○○ =[124], =[125], =[126]	=[127]	=[128]	+[129] ⊙○○○ =[130], =[131]	-[132] ⊙○○○ =[133], =[134]			
	Long-term psychodynamic psychotherapy	=[135], =[136]	+[137] ⊙○○○	=[141]	=[142]	-[143] ⊙○○○	+[148] ⊙○○○			

[115] Adequate work ability (Work Ability Index score ≥37), compared to long-term psychotherapy
[116] Two outcomes: 1) Self-estimated work ability (using the Work Ability Index) and 2) Work role functioning (SAS-SR Work subscale), both compared to short-term and long-term psychotherapy
[117] Adequate work ability (Work Ability Index score ≥37), compared to short-term psychotherapy
[118] Adequate work ability (Work Ability Index score ≥37), compared to long-term psychotherapy
[119] Two outcomes: 1) Self-estimated work ability (using the Work Ability Index) and 2) Work role functioning (SAS-SR Work subscale), both compared to short-term and long-term psychotherapy
[120] Adequate work ability (Work Ability Index score ≥37), compared to short-term psychotherapy
[121] Two outcomes: 1) Number of sick leave days during last 3 months and 2) Proportion with more than 7 sick leave days during last 3 months, both compared to brief and resource-oriented solution-focused psychotherapy and long-term psychotherapy
[122] Current employment status, compared to brief and resource-oriented solution-focused psychotherapy and long-term psychotherapy. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[123] Proportion with more than 7 sick leave days during last 3 months, compared to brief and resource-oriented solution-focused psychotherapy
[124] Number of sick leave days during last 3 months, compared to brief and resource-oriented solution-focused psychotherapy and long-term psychotherapy
[125] Proportion with more than 7 sick leave days during last 3 months, compared to long-term psychotherapy
[126] Current employment status, compared to brief and resource-oriented solution-focused psychotherapy and long-term psychotherapy. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[127] Current employment status, compared to brief and resource-oriented solution-focused psychotherapy and long-term psychotherapy. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[128] Current employment status, compared to brief and resource-oriented solution-focused psychotherapy and long-term psychotherapy. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[129] Self-estimated work ability (using the Work Ability Index), compared to long-term psychotherapy
[130] Self-estimated work ability (using the Work Ability Index), compared to brief and resource-oriented solution-focused psychotherapy
[131] Two outcomes: 1) Adequate work ability (Work Ability Index score ≥37) and 2) Work role functioning (SAS-SR Work subscale), both compared to brief and resource-oriented solution-focused psychotherapy and long-term psychotherapy
[132] Two outcomes: 1) Self-estimated work ability (using the Work Ability Index) and 2) Work role functioning (SAS-SR Work subscale), both compared to long-term psychotherapy
[133] Two outcomes: 1) Self-estimated work ability (using the Work Ability Index) and 2) Work role functioning (SAS-SR Work subscale), both compared to brief and resource-oriented solution-focused psychotherapy
[134] Adequate work ability (Work Ability Index score ≥37), compared to brief and resource-oriented solution-focused psychotherapy and long-term psychotherapy
[135] Two outcomes: 1) Number of sick leave days during last 3 months and 2) Proportion with more than 7 sick leave days during last 3 months, both compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy
[136] Current employment status, compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[137] Proportion with more than 7 sick leave days during last 3 months, compared to brief and resource-oriented solution-focused psychotherapy

Interventions		Primary Outcomes								
		Prevention of Work Disability		Management of Work Disability		Work Functioning		Recurrence of Work Disability		Economic Benefit
		Short Term	Long Term	Short Term	Long Term	Short Term	Long Term	Short Term	Long Term	
	(Knekt 2008)		=[138], =[139], =[140]			-[144] ⊙○○○ =[145], =[146], =[147]	+[149] ⊙○○○ =[150], =[151]			
Enhanced Primary Care	Enhanced depression care delivered by primary care physicians and nurses (Lo Sasso 2006) (Rost, 2004) Smith, 2002)	=[152]	+[153] ⊙○○○ =[154]		=[155]		+[156] ⊙○○○ =[157]			+
	Quality improvement program for improved access to		=[158], =[159]		=[160]					

[141] Current employment status, compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[142] Current employment status, compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[143] Adequate work ability (Work Ability Index score ≥37), compared to brief and resource-oriented solution-focused psychotherapy
[148] Two outcomes: 1) Self-estimated work ability (using the Work Ability Index) and 2) Work role functioning (SAS-SR Work subscale), both compared to short-term psychotherapy
[138] Number of sick leave days during last 3 months, compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy
[139] Proportion with more than 7 sick leave days during last 3 months, compared to short-term psychotherapy
[140] Current employment status, compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[144] Self-estimated work ability (using the Work Ability Index), compared to short-term psychotherapy
[145] Self-estimated work ability (using the Work Ability Index); compared to brief and resource-oriented solution-focused psychotherapy
[146] Adequate work ability (Work Ability Index score ≥37), compared to short-term psychotherapy
[147] Work role functioning (SAS-SR Work subscale), compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy
[149] Adequate work ability (Work Ability Index score ≥37), compared to brief and resource-oriented solution-focused psychotherapy
[150] Two outcomes: 1) Self-estimated work ability (using the Work Ability Index) and 2) Work role functioning (SAS-SR Work subscale), both compared to brief and resource-oriented solution-focused psychotherapy
[151] Adequate work ability (Work Ability Index score ≥37), compared to short-term psychotherapy
[152] Subsequent employment status
[153] Subsequent employment status
[154] Number of days absent over 24 months among total sample and consistently employed
[155] Number of days absent over 24 months among inconsistently employed
[156] Employee's rating of their productivity at work during previous 2 weeks
[157] Employee's rating of their productivity at work during previous 2 weeks among inconsistently employed
[158] Days missed from work due to illness over 24 months
[159] Days of employment over 24 months. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[160] Days of employment over 24 months. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline

Interventions		Primary Outcomes								
		Prevention of Work Disability		Management of Work Disability		Work Functioning		Recurrence of Work Disability		Economic Benefit
		Short Term	Long Term	Short Term	Long Term	Short Term	Long Term	Short Term	Long Term	
	medications with primary care clinicians (Schoenbaum, 2001)									
	Quality improvement program for improved psychotherapy with primary care clinicians (Schoenbaum, 2001)		=[161] +[162] ⊙○○○		+[163] ⊙○○○					+
Enhanced Psychiatric Care	Psychiatric treatment with adjunct occupational therapy (Schene, 2007)	+[164] ⊙○○○	+[165] ⊙○○○ =[166]	+[167] ⊙○○○	+[168] ⊙○○○ +[169] ⊙○○○ =[170]					+
Enhanced Role for the Occupational Physician	Dutch guideline-based care for employees on sick leave due to mental health problems delivered by occupational physicians (Rebergen, 2009)				+[171] ⊙○○○ =[172]				=[173]	+

[161] Days missed from work due to illness over 24 months
[162] Days of employment over 24 months. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[163] Days of employment over 24 months. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[164] Total hours worked between 7 and 12 months. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[165] Total hours worked between 13 and 18 months. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[166] Two outcomes: 1) Total hours worked between 19 and 42 months and 2) Proportion working at least 2 days or 16 hours per week over 42 months. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[167] Total hours worked between 7 and 12 months. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[168] Time until any work resumption measured over 42 months
[169] Total hours worked between 13 and 18 months. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[170] Two outcomes: 1) Total hours worked between 19 and 42 months and 2) Proportion working at least 2 days or 16 hours per week over 42 months. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline
[171] Immediate full RTW versus partial RTW
[172] Three outcomes: 1) Duration of sick leave days until partial RTW, 2) Duration of sick leave days until full RTW, and 3) Duration of sick leave days (including recurrences) until full RTW
[173] Two outcomes: 1) Number of recurrences of sick leave periods and 2) Duration of recurrences of sick leave periods

Interventions		Primary Outcomes								
		Prevention of Work Disability		Management of Work Disability		Work Functioning		Recurrence of Work Disability		Economic Benefit
		Short Term	Long Term	Short Term	Long Term	Short Term	Long Term	Short Term	Long Term	
	Enhanced role of occupational physician with psychiatric consultations (van der Feltz-Cornelis, 2010)			+[174] ⊙○○○ =[175]	=[176]					
Systems Integration and Care Management	Enhanced disability management consisting of additional collaborative mental health program delivered by psychiatrists to workers on short-term disability leave for psychiatric disorders (Dewa, 2009)				+[177] ⊙○○○					+
	Telephone screening, outreach, and care management (Wang, 2007)	+[178] ⊙○○○ =[179]	=[180]			+[181] ⊙○○○ =[182]	+[183] ⊙○○○ =[184]			
Physical Activity	Strength training (Krogh, 2009)	=[185]	+[186] ⊙○○○ =[187]	=[188]	+[189] ⊙○○○ =[190]					

[174] Full RTW status at 3 months
[175] Full RTW status at 6 months
[176] Time until full RTW
[177] Three outcomes: 1) Return to work (yes/no), 2) Days on short-term disability, and 3) Transition from short-term to long-term disability
[178] Job retention (measured by the Health Performance Questionnaire)
[179] Number of weekly hours worked
[180] Two outcomes: 1) Job retention (measured by the Health Performance Questionnaire) and 2) Number of weekly hours worked
[181] Effective weekly hours worked
[182] On-the-job performance (measured by the Health Performance Questionnaire)
[183] Effective weekly hours worked
[184] On-the-job performance (measured by the Health Performance Questionnaire)
[185] Three outcomes: 1) Proportion of individuals on sick leave, 2) Percentage of days absent from work in last 10 days, and 3) Unemployment status, all compared to relaxation training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[186] Percentage of days absent from work in the last 10 days, compared to relaxation training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[187] Two outcomes: 1) Proportion of individuals on sick leave and 2) Unemployment status, both compared to relaxation training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline
[188] Three outcomes: 1) Proportion of individuals on sick leave, 2) Percentage of days absent from work in last 10 days, and 3) Unemployment status, all compared to relaxation training. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline

Interventions		*Primary Outcomes*								
		Prevention of Work Disability		*Management of Work Disability*		*Work Functioning*		*Recurrence of Work Disability*		*Economic Benefit*
		Short Term	*Long Term*	*Short Term*	*Long Term*	*Short Term*	*Long Term*	*Short Term*	*Long Term*	
	Aerobic training (Krogh, 2009)	=[191]	=[192]	=[193]	=[194]					
	Relaxation training (Krogh, 2009)	=[195]	-[196] ⊙○○○ =[197], =[198]	=[199]	-[200] ⊙○○○ =[201], =[202]					
Stress Reduction	Worksite-wide stress reduction program (Kawakami, 1997)		+[203] ⊙○○○							

†Compared to usual care, unless otherwise specified

Legend: +: intervention was statistically significantly better than the control group; -: control group was statistically significantly better than the intervention group; =no statistically significant difference between intervention and control group; ≈ mixed findings; shaded cell=no evidence

---

[189] Percentage of days absent from work in the last 10 days, compared to relaxation training. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline

[190] Two outcomes: 1) Proportion of individuals on sick leave and 2) Unemployment status, both compared to relaxation training. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline

[191] Three outcomes: 1) Proportion of individuals on sick leave, 2) Percentage of days absent from work in last 10 days, and 3) Unemployment status, all compared to relaxation training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

[192] Three outcomes: 1) Proportion of individuals on sick leave, 2) Percentage of days absent from work in last 10 days, and 3) Unemployment status, all compared to relaxation training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

[193] Three outcomes: 1) Proportion of individuals on sick leave, 2) Percentage of days absent from work in last 10 days, and 3) Unemployment status, all compared to relaxation training. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline

[194] Three outcomes: 1) Proportion of individuals on sick leave, 2) Percentage of days absent from work in last 10 days, and 3) Unemployment status, all compared to relaxation training. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline

[195] Three outcomes: 1) Proportion of individuals on sick leave, 2) Percentage of days absent from work in last 10 days, and 3) Unemployment status, all compared to strength training and aerobic training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

[196] Percentage of days absent from work in the last 10 days, compared to strength training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

[197] Two outcomes: 1) Proportion of individuals on sick leave and 2) Unemployment status, compared to strength training and aerobic training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

[198] Percentage of days absent from work in last 10 days, compared to aerobic training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

[199] Three outcomes: 1) Proportion of individuals on sick leave, 2) Percentage of days absent from work in last 10 days, and 3) Unemployment status, all compared to strength training and aerobic training. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline

[200] Percentage of days absent from work in the last 10 days, compared to strength training. Note: may also reflect evidence for prevention of work disability due to variability in employment status of study sample at baseline

[201] Two outcomes: 1) Proportion of individuals on sick leave and 2) Unemployment status, compared to strength training and aerobic training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

[202] Percentage of days absent from work in last 10 days, compared to aerobic training. Note: may also reflect evidence for management of work disability due to variability in employment status of study sample at baseline

[203] Length of sick leave in the past year

Appendix K

Summary of findings for key secondary outcomes†

Interventions		*Outcomes*							
		Improved Psychosocial Work Outcomes		*Reduction in Critical Workplace Incidents*		*Reduction in Depression Symptom Severity*		*Depression Remission*	
		Short Term	*Long Term*	*Short Term*	*Long Term*	*Short Term*	*Long Term*	*Short Term*	*Long Term*
Psychological Interventions	Cognitive behavioural therapy delivered by psychologists (Blonk, 2006)					=[204]			
	Brief cognitive behavioural therapy-based stress management delivered by labour experts with a focus on improving workplace processes (Blonk, 2006)					=[205]			
	Brief and resource-oriented solution-focused psychotherapy (Knekt 2008)					=[206] +[207] =[208],	=[209] -[210]	≈[211] =[212], +[213]	=[214] ≈[215] =[216],
	Short-term psychodynamic psychotherapy (Knekt 2008)					=[217] +[218]	=[219] ≈[220]	≈[221] =[222], +[223]	=[224] ≈[225] =[226],

[204] Using the Depression Anxiety Stress Scales depression subscale, compared to brief cognitive behavioural therapy-based stress management, as well as usual care
[205] Using the Depression Anxiety Stress Scales depression subscale, compared to cognitive behavioural therapy, as well as usual care
[206] Using the Hamilton Depression Scale and the Beck Depression Inventory, compared to short-term psychotherapy
[207] Using the Hamilton Depression Scale, compared to long-term psychotherapy
[208] Using the Beck Depression Inventory, compared to long-term psychotherapy
[209] Using the Hamilton Depression Scale and the Beck Depression Inventory, compared to short-term psychotherapy
[210] Using the Hamilton Depression Scale and the Beck Depression Inventory, compared to long-term psychotherapy
[211] Using the Beck Depression Inventory at 3 months, remission more likely compared to short-term psychotherapy; thereafter, no difference between groups
[212] Using the Diagnostic and Statistical Manual of Mental Disorders, compared to short-term psychotherapy
[213] Using both the Beck Depression Inventory and the Diagnostic and Statistical Manual of Mental Disorders, compared to long-term psychotherapy
[214] Using both the Beck Depression Inventory and the Diagnostic and Statistical Manual of Mental Disorders, compared to short-term psychotherapy
[215] Using the Diagnostic Structured Interview at 12 months, remission more likely compared to long-term psychotherapy; thereafter, no difference between groups
[216] Using the Beck Depression Inventory, compared to long-term psychotherapy
[217] Using both the Beck Depression Inventory and the Hamilton Depression Scale, compared to brief and resource-oriented solution-focused psychotherapy
[218] Using both the Beck Depression Inventory and the Hamilton Depression Scale, compared to long-term psychotherapy
[219] Using both the Beck Depression Inventory and the Hamilton Depression Scale, compared to brief and resource-oriented solution-focused psychotherapy
[220] Using both the Beck Depression Inventory and the Hamilton Depression Scale at 12 months, symptom reduction more likely, but less likely to reduce symptoms at 36 months, compared to long-term psychotherapy
[221] Using the Beck Depression Inventory remission less likely at 3 months, compared to brief and resource-oriented solution-focused psychotherapy; thereafter, no difference between groups
[222] Using the Diagnostic and Statistical Manual of Mental Disorders, compared to brief and resource-oriented solution-focused psychotherapy
[223] Using both the Beck Depression Inventory and the Diagnostic and Statistical Manual of Mental Disorders, compared to long-term psychotherapy

	Long-term psychodynamic psychotherapy (Knekt 2008)					-[227]	+[228] ≈[229]	-[230] +[231]	≈[232], =[233], ≈[234], =[235],
Enhanced Primary Care	Enhanced care delivered by primary care physician and nurse (Lo Sasso, 2006; Rost 2004; Smith, 2002)		+[236]				≈[237]		
	Quality improvement program for improved access to medications with primary care clinicians (Schoenbaum, 2001)								
	Quality improvement program for improved psychotherapy with primary care clinicians (Schoenbaum, 2001)								
Enhanced Psychiatric Care	Psychiatric treatment with adjuvant occupational therapy (Schene 2007)	=[238]	=[239]			=[240]	-[241]	=[242]	=[243]
Enhanced Role for the Occupational Physician	Guideline-based care by occupational physician (Rebergen, 2009)								
	Occupational physicians with specialized training (van der Feltz-Cornelis, 2010)					=[244]			
Systems Integration and	Collaborative mental health program (enhanced disability management) (Dewa, 2009)								

[224] Using both the Beck Depression Inventory and the Diagnostic and Statistical Manual of Mental Disorders, compared to brief and resource-oriented solution-focused psychotherapy
[225] Using the Beck Depression Inventory, remission more likely at 12 months, compared to long-term psychotherapy; thereafter no difference between groups
[226] Using the Diagnostic and Statistical Manual of Mental Disorders, compared to long-term psychotherapy
[227] Using both the Beck Depression Inventory and the Hamilton Depression Scale, compared to brief and resource-oriented solution-focused psychotherapy and short-term psychotherapy
[228] Using on both the Beck Depression Inventory and the Hamilton Depression Scale, compared to brief and resource-oriented solution-focused psychotherapy
[229] Using both the Beck Depression Inventory and the Hamilton Depression Scale, symptom reduction less likely at 12 months compared to short-term psychotherapy, but more likely to lead to improved scores on both scales by 36 months
[230] Using both the Beck Depression Inventory and the Diagnostic and Statistical Manual of Mental Disorders, compared to brief and resource-oriented solution-focused psychotherapy
[231] Using both the Beck Depression Inventory and the Diagnostic and Statistical Manual of Mental Disorders, compared to short-term psychotherapy
[232] Using the Diagnostic and Statistical Manual of Mental Disorders, remission less likely at 12 months compared to brief and resource-oriented solution-focused psychotherapy; thereafter, no difference between groups
[233] Using the Beck Depression Inventory, compared to brief and resource-oriented solution-focused psychotherapy
[234] Using the Beck Depression Inventory, remission less likely at 12 months compared to short-term psychotherapy; thereafter, no difference between groups
[235] Using the Diagnostic and Statistical Manual of Mental Disorders, compared to brief and resource-oriented solution-focused psychotherapy
[236] Workplace conflict in previous 12 months (arguments/difficulties with coworkers)
[237] Using the Center for Epidemiologic Studies Depression Scale, symptom reduction more likely than usual care among consistently employed individuals; no difference between groups among inconsistently employed individuals or in overall sample
[238] Work Stress (using the Questionnaire Organization Stress scale)
[239] Work Stress (using the Questionnaire Organization Stress scale)
[240] Using the Beck Depression Inventory
[241] Using the Beck Depression Inventory
[242] Using the Diagnostic and Statistical Manual of Mental Disorders
[243] Using the Diagnostic and Statistical Manual of Mental Disorders
[244] Using three scales: 1) Patient Health Questionnaire-9, 2) Patient Health Questionnaire-15, and 3) Symptom Checklist-90 Revised

Care Management	Telephone screening, outreach, and care management (Wang, 2007)			=[245]	=[246]	+[247] =[248],	+[249]	+[250]	+[251]
Physical Activity	Strength training (Krogh, 2009)					=[252]	=[253]	=[254]	=[255]
	Aerobic training (Krogh, 2009)					=[256]	=[257]	=[258]	=[259]
	Relaxation training (Krogh, 2009)					=[260]	=[261]	=[262]	=[263]
Stress Reduction	Worksite-wide stress reduction program (Kawakami, 1997)		=[264]				+[265]		

[†]Compared to usual care, unless otherwise specified

Legend: +: intervention was statistically significantly better than the control group; -: control group was statistically significantly better than the intervention group; =no statistically significant difference between intervention and control group; ≈ mixed findings; shaded cell=no evidence

[245] Using the Health Performance Questionnaire
[246] Using the Health Performance Questionnaire
[247] Using the continuous Quick Inventory of Depressive Symptomatology-Self-Report score
[248] Using the criteria of ≥50% reduction in Quick Inventory of Depressive Symptomatology-Self-Report score
[249] Using both the continuous Quick Inventory of Depressive Symptomatology-Self-Report score and the criteria of ≥50% reduction in Quick Inventory of Depressive Symptomatology-Self-Report score
[250] Using a Quick Inventory of Depressive Symptomatology-Self-Report score ≤5
[251] Using a Quick Inventory of Depressive Symptomatology-Self-Report score ≤5
[252] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to aerobic training and relaxation training
[253] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to aerobic training and relaxation training
[254] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to aerobic training and relaxation training
[255] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to aerobic training and relaxation training
[256] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to strength training and relaxation training
[257] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to strength training and relaxation training
[258] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to strength training and relaxation training
[259] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to strength training and relaxation training
[260] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to strength training and aerobic training
[261] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to strength training and aerobic training
[262] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to strength training and aerobic training
[263] Using three scales: 1) Hamilton Depression Scale-17, 2) Montgomery-Aasberg scale, and 3) Beck Depression Inventory, all compared to strength training and aerobic training
[264] Workplace stressors (i.e., overtime, overload, little chance to learn new knowledge, lack of control over workplace, problems with supervisor, problems with co-workers)
[265] Using the Zung SDS scale