Supplement A. *Measures of sick leave in the included studies*

|  |  |
| --- | --- |
| Measure of sick leave | Type of assessment |
| Able to work | Self-report |
| Absenteeism days | Self-report |
| Days missed at work | Assessor, Register |
| Days Missed Per Two Weeks | Self-report |
| Days of reduced absenteeism | Self-report |
| Days until Full RTW | Employer record, Register, Self-report |
| Days until Partial RTW | Employer record, Self-report |
| Days worked last month | Self-report |
| Employed days | Self-report |
| Full RTW | Employer record, Register, Self-report |
| Increased or maintained work | Register |
| Mean days on sick leave | Employer record, Register, Self-report |
| Mean weeks on sick leave | Register |
| Median days until full RTW | Register |
| Median days until partial RTW | Register |
| No Pension disability or sick leave | Register |
| No recurrence of sick leave | Register |
| No Sick leave | Self-report |
| Number of sick leave periods | Self-report |
| Part RTW | Employer record, Register, Self-report |
| Weekly hours worked | Self-report |

Supplement B. *Risk of bias: review authors’ judgements about each methodological quality item for each included study.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Author, year** | **Random sequence generation**  **(Selection bias)** | **Allocation concealment**  **(Selection bias)** | **Blinding of outcome assessment**  **(Detection bias)** | **Incomplete outcome data (Attrition bias)** | **Selective reporting**  **(Reporting bias)** |
| Adler, 2015 | ? | ? | - | + | + |
| Arends, 2014 | + | - | + | + | + |
| Bakker, 2007 | + | + | - | ? | - |
| Blonk, 2006 | ? | ? | + | + | + |
| Brouwers, 2006 | + | + | - | ? | + |
| Dahl, 2004 | ? | ? | + | + | ? |
| de Wente, 2008 | + | ? | - | - | + |
| deBoer, 2004 | ? | ? | + | - | ? |
| Duijts, 2008 | + | - | + | + | + |
| Ebert, 2014 | + | ? | - | + | + |
| Ebert, 2016 | + | + | - | + | + |
| Ejeby, 2013 | + | + | + | + | + |
| Feltz-Corneli2, 2010 | + | - | - | ? | - |
| Folke, 2012 | ? | ? | + | + | + |
| Geraedts, 2014 | + | + | - | ? | + |
| Hees, 2013 | + | + | + | + | + |
| Hollinghurst, 2013 | + | + | + | + | + |
| Kant, 2008 | + | ? | + | + | + |
| Kendrick, 2000 | + | ? | + | + | + |
| King, 2006 | + | ? | + | + | + |
| Lerner, 2013 | ? | ? | - | + | + |
| Lexis, 2011 | ? | ? | + | + | ? |
| Noordik, 2013 | ? | - | - | + | + |
| Nystuden, 2006 | + | + | + | + | + |
| Oostrom, 2010 | + | + | + | + | - |
| Patel, 2017 | + | + | + | + | + |
| Pedersen, 2015 | ? | ? | + | + | + |
| Rebergen, 2009 | + | + | + | + | + |
| Reme, 2015 | + | + | + | + | ? |
| Rhenen, 2007 | ? | ? | + | + | ? |
| Rosen, 2014 | ? | ? | - | + | + |
| Rost, 2004 | ? | - | - | ? | ? |
| Schoenbaum, 2001 | ? | - | - | + | ? |
| Schene, 2007 | + | + | - | - | ? |
| Simon, 1998 | ? | ? | - | ? | ? |
| Stenlund, 2009 | ? | ? | + | + | ? |
| Svensson, 2009 | ? | ? | - | - | ? |
| Taimela, 2008 | + | + | + | + | + |
| Thiart, 2016 | + | + | + | + | + |
| van der Klink, 2002 | + | - | + | + | ? |
| Wang, 2007 | + | + | - | + | ? |
| Warmerdam, 2010 | + | + | + | + | + |
| Willert, 2011 | + | + | + | + | + |
| Vlasveld, 2013 | + | + | + | + | + |
| Volker, 2015 | + | - | + | + | + |

Supplement C. *Forest plot of effect sizes (Hedge’s g) of interventions on sick-leave.*



Supplement D. *Forest plot of effect sizes (Hedge’s g) of interventions on psychological symptoms.*



Supplement E. *Effect sizes (Hedges’ g) on psychological symptoms at follow-up for all studies divided on comparison conditions.*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Comparison** | **k** | ***g*-value** | **95% CI** | ***z*-value** | **Q-value** | ***I*2** |
| All studies | 35 | 0.21 | 0.13-0.30 | 4.77d | 87.7b | 61 |
| All interventions vs CAU | 29 | 0.19 | 0.09-0.28 | 3.70d | 78.2d | 64 |
| All interventions vs WLC | 2 | 0.47 | 0.26-0.68 | 4.41d | 1.1 | 12 |
| All interventions vs active control | 4 | 0.27 | 0.07-0.47 | 2.66b | 0.3 | 0 |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

k = number of comparisons. A positive g-value means that the first treatment in the comparison is better and a negative that the second is better.  a *p*<0.05, b *p*<0.01, c *p*<0.001, d *p*<0.0001

CAU = Care as usual, WLC = Waitlist control