

APPENDIX

for the article: The effect of exposure to long working hours on depression: A systematic review and meta-analysis from the WHO/ILO Joint Estimates of the Work-Related Burden of Disease and Injury

Appendix 1 Search strategies for seven electronic academic databases and two grey literature databases

WHO International Clinical Trials Register Platform

Date searched: 18 July 2018

Records retrieved: 144

Strategy:

affective AND work*

adjustment AND work*

mood AND work*

depress* AND work*

dysthymi* AND work*

antidepress* AND work*

affective AND job*

adjustment AND job*

mood AND job*

depress* AND job*

dysthymi* AND job*

antidepress* AND job*

affective AND shift*

adjustment AND shift*

mood AND shift*

depress* AND shift*

dysthymi* AND shift*

antidepress* AND shift*

overwork*

overtime

condition field: depress* OR affective OR adjustment OR mood OR dysthymi* OR antidepress* AND

intervention field: work* OR job* OR shift* OR overwork* OR overtime*

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

MEDLINE (Ovid)

Date searched: 11 July 2018; Date of top-up search: 27 November 2019

Records retrieved: 1,397 (main search); 36 (top-up search)

Strategy:

1. exp “personnel staffing and scheduling”/
2. “personnel staffing and scheduling”.ti,ab,kw.
3. shift work schedule.ti,ab,kw.
4. work schedule tolerance.ti,ab,kw.
5. workload.kw.
6. workday shifts.ti,ab,kw.
7. overwork*.ti,ab,kw.
8. overtime.ti,ab,kw.
9. workweek*.ti,ab,kw.
10. (work* adj3 hour*).ti,ab,kw.
11. (work* adj3 schedul*).ti,ab,kw.
12. (work* adj3 roster).ti,ab,kw.
13. (work* adj3 organi#ation).ti,ab,kw.
14. (work* adj3 time*).ti,ab,kw.
15. (work* adj3 overload*).ti,ab,kw.
16. (work* adj3 extend*).ti,ab,kw.
17. (work* adj3 compress*).ti,ab,kw.
18. (work* adj3 week*).ti,ab,kw.
19. (work* adj3 day?).ti,ab,kw.
20. (job? adj3 hour*).ti,ab,kw.
21. (job? adj3 schedul*).ti,ab,kw.
22. (job? adj3 roster).ti,ab,kw.
23. (job? adj3 organi#ation).ti,ab,kw.
24. (job? adj3 time*).ti,ab,kw.
25. (job? adj3 overload*).ti,ab,kw.
26. (job? adj3 extend*).ti,ab,kw.
27. (job? adj3 compress*).ti,ab,kw.
28. (job? adj3 week*).ti,ab,kw.
29. (job? adj3 day?).ti,ab,kw.
30. (shift? adj3 hour*).ti,ab,kw.
31. (shift? adj3 schedul*).ti,ab,kw.
32. (shift? adj3 roster).ti,ab,kw.
33. (shift? adj3 organi#ation).ti,ab,kw.
34. (shift? adj3 time*).ti,ab,kw.
35. (shift? adj3 overload*).ti,ab,kw.
36. (shift? adj3 extend*).ti,ab,kw.
37. (shift? adj3 compress*).ti,ab,kw.
38. (shift? adj3 week*).ti,ab,kw.

1 39. (shift? adj3 day?).ti,ab,kw.
2 40. (work* and (life* or live*) and balances* or imbalances* or unbalances* or
3 interference*).ti,ab,kw.
4 41. (work* and famil* and conflict*).ti,ab,kw.
5 42. or/1-41
6 43. Behavioral Symptoms/
7 44. Affective Symptoms/
8 45. affective disorder\$.ti,ab,kw.
9 46. Mood Disorders/
10 47. (mood\$ adj3 disorder\$).ti,ab,kw.
11 48. (low\$ adj3 mood\$).ti,ab,kw.
12 49. Depression/
13 50. exp Depressive Disorder/
14 51. depress\$.ti,ab,kw.
15 52. dysthymi\$.ti,ab,kw.
16 53. Adjustment Disorder/
17 54. (adjustment\$ adj3 disorder\$).ti,ab,kw.
18 55. exp Antidepressive Agents/
19 56. antidepress\$.ti,ab,kw.
20 57. or/43-56
21 58. exp Clinical Trial/
22 59. trial\$.tw.
23 60. experiment\$.tw.
24 61. (intervention adj3 (study or studies or analys\$)).tw.
25 62. Epidemiologic Studies/
26 63. Observational Study/
27 64. ((observational or epidemiologic\$) adj (study or studies or analys\$)).tw.
28 65. exp Cohort Studies/
29 66. cohort\$.tw.
30 67. (panel\$ adj3 (study or studies or analys\$ or data)).tw.
31 68. (follow up adj (study or studies or analys\$)).tw.
32 69. (repeat\$ adj measure\$).tw.
33 70. longitudinal\$.tw.
34 71. retrospective\$.tw.
35 72. prospective\$.tw.
36 73. exp Case Control Studies/
37 74. (case\$ adj3 control\$).tw.
38 75. (exposure\$ adj4 (study or studies or analys\$)).tw.
39 76. or/58-75
40 77. 42 and 57 and 76
41
42
43
44

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

PubMed

Date searched: 11 July 2018

Records retrieved: 9,209

Strategy:

1. (“Personnel Staffing and Scheduling”[Mesh])
2. “shift work schedule” [Text Word]
3. “work schedule tolerance” [Text Word]
4. workload[Text Word]
5. “workday shifts” [Text Word]
6. overwork*[Text Word]
7. overtime[Text Word]
8. workweek*[Text Word]
9. ((work*[Text Word]) AND hour*[Text Word])
10. ((work*[Text Word]) AND schedule*[Text Word])
11. ((work*[Text Word]) AND roster[Text Word])
12. ((work*[Text Word]) AND (organization[Text Word] OR organisation[Text Word]))
13. ((work*[Text Word]) AND time*[Text Word])
14. ((work*[Text Word]) AND overload*[Text Word])
15. ((work*[Text Word]) AND extend*[Text Word])
16. ((work*[Text Word]) AND compress*[Text Word])
17. ((work*[Text Word]) AND week*[Text Word])
18. ((work*[Text Word]) AND (day[Text Word] OR days[Text Word]))
19. (((job[Text Word] OR jobs[Text Word])) AND hour*[Text Word])
20. (((job[Text Word] OR jobs[Text Word])) AND schedul*[Text Word])
21. (((job[Text Word] OR jobs[Text Word])) AND roster[Text Word])
22. (((job[Text Word] OR jobs[Text Word])) AND (organization[Text Word] OR organisation[Text Word]))
23. (((job[Text Word] OR jobs[Text Word])) AND time*[Text Word])
24. (((job[Text Word] OR jobs[Text Word])) AND overload*[Text Word])
25. (((job[Text Word] OR jobs[Text Word])) AND extend*[Text Word])
26. (((job[Text Word] OR jobs[Text Word])) AND compress*[Text Word])
27. (((job[Text Word] OR jobs[Text Word])) AND week*[Text Word])
28. (((job[Text Word] OR jobs[Text Word])) AND (day[Text Word] OR days[Text Word]))
29. (((shift[Text Word] OR shifts[Text Word])) AND hour*[Text Word])
30. (((shift[Text Word] OR shifts[Text Word])) AND schedule*[Text Word])
31. (((shift[Text Word] OR shifts[Text Word])) AND roster[Text Word])
32. (((shift[Text Word] OR shifts[Text Word])) AND (organization[Text Word] OR organisation[Text Word]))
33. (((shift[Text Word] OR shifts[Text Word])) AND time*[Text Word])
34. (((shift[Text Word] OR shifts[Text Word])) AND overload*[Text Word])
35. (((shift[Text Word] OR shifts[Text Word])) AND extend*[Text Word])
36. (((shift[Text Word] OR shifts[Text Word])) AND compress*[Text Word])
37. (((shift[Text Word] OR shifts[Text Word])) AND week*[Text Word])
38. (((shift[Text Word] OR shifts[Text Word])) AND (day[Text Word] OR days[Text Word]))

1 39. (((work*[Text Word]) AND (life*[Text Word] OR live*[Text Word])) AND (balances*[Text
 2 Word] OR imbalances*[Text Word] OR unbalances*[Text Word] OR interference*[Text Word]))
 3 40. ((work*[Text Word] AND famil*[Text Word] AND conflict*[Text Word]))
 4 41. (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13
 5 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25
 6 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37
 7 OR #38 OR #39 OR #40)
 8 42. "Behavioral Symptoms"[Mesh:NoExp]
 9 43. "Affective Symptoms"[Mesh:NoExp]
 10 44. (("affective disorder"[Text Word]) OR "affective disorders"[Text Word])
 11 45. "Mood Disorders"[Mesh:NoExp]
 12 46. (((mood*[Title]) AND disorder*[Title]) OR (("mood disorder"[Text Word]) OR "mood
 13 disorders"[Text Word]))
 14 47. ((low[Text Word]) AND mood[Text Word])
 15 48. "Depression"[Mesh:NoExp]
 16 49. "Depressive Disorder"[Mesh]
 17 50. (depress*[Title] OR (((depression[Text Word]) OR depressive[Text Word]) OR depressed[Text
 18 Word]) OR depressions[Text Word])
 19 51. Dysthymia*[Text Word]
 20 52. "Adjustment Disorders"[Mesh:NoExp]
 21 53. (("adjustment disorder"[Text Word]) OR "adjustment disorders"[Text Word] OR
 22 (((adjustment*[Title]) AND disorder*[Title])))
 23 54. "Antidepressive Agents"[Mesh]
 24 55. antidepress*[Text Word]
 25 56. (#42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53
 26 OR #54 OR #55)
 27 57. "Clinical Trial" [Publication Type]
 28 58. trial*[Text Word]
 29 59. experiment*[Text Word]
 30 60. ((Intervention[Text Word]) AND (Study[Text Word] OR Studies[Text Word] OR analysis[Text
 31 Word] OR analyses[Text Word]))
 32 61. "Epidemiologic Studies"[Mesh:NoExp]
 33 62. "Observational Study" [Publication Type]
 34 63. (((observational[Text Word] OR epidemiologic[Text Word] OR epidemiological[Text Word]))
 35 AND (Study[Text Word] OR studies[Text Word] OR analysis[Text Word] OR analyses[Text Word]))
 36 64. "Cohort Studies"[Mesh]
 37 65. cohort*[Text Word]
 38 66. ((panel*[Text Word]) AND (study[Text Word] OR studies[Text Word] OR analyses[Text Word]
 39 OR analysis[Text Word] OR data[Text Word]))
 40 67. (((follow[Text Word] AND up[Text Word])) AND (study[Text Word] OR studies[Text Word] OR
 41 analysis[Text Word] OR analyses[Text Word]))
 42 68. (((repeated[Text Word] OR repeat[Text Word])) AND (measure[Text Word] OR measures[Text
 43 Word]))
 44 69. longitudinal*[Text Word]
 45 70. retrospective*[Text Word]
 46 71. prospective*[Text Word]
 47 72. "Case-Control Studies"[Mesh]
 48 73. ((case*[Text Word]) AND control*[Text Word])

- 1 74. (((exposure[Text Word] OR exposures[Text Word])) AND (study[Text Word] OR studies[Text
- 2 Word] OR analyses[Text Word] OR analysis[Text Word]))
- 3 75. (#57 OR #58 OR #59 OR #60 OR #61 OR #62 OR #63 OR #64 OR #65 OR #66 OR #67 OR #68
- 4 OR #69 OR #70 OR #71 OR #72 OR #73 OR #74)
- 5 76. (#41 AND #56 AND #75)
- 6
- 7

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

EMBASE (Ovid)

Date searched: 11 July 2018

Records retrieved: 2,855

Strategy:

1. exp “personnel staffing and scheduling”/
2. “personnel staffing and scheduling”.ti,ab,kw.
3. shift work schedule.ti,ab,kw.
4. work schedule tolerance.ti,ab,kw.
5. workload.kw.
6. workday shifts.ti,ab,kw.
7. overwork*.ti,ab,kw.
8. overtime.ti,ab,kw.
9. workweek*.ti,ab,kw.
10. (work* adj3 hour*).ti,ab,kw.
11. (work* adj3 schedul*).ti,ab,kw.
12. (work* adj3 roster).ti,ab,kw.
13. (work* adj3 organi#ation).ti,ab,kw.
14. (work* adj3 time*).ti,ab,kw.
15. (work* adj3 overload*).ti,ab,kw.
16. (work* adj3 extend*).ti,ab,kw.
17. (work* adj3 compress*).ti,ab,kw.
18. (work* adj3 week*).ti,ab,kw.
19. (work* adj3 day?).ti,ab,kw.
20. (job? adj3 hour*).ti,ab,kw.
21. (job? adj3 schedul*).ti,ab,kw.
22. (job? adj3 roster).ti,ab,kw.
23. (job? adj3 organi#ation).ti,ab,kw.
24. (job? adj3 time*).ti,ab,kw.
25. (job? adj3 overload*).ti,ab,kw.
26. (job? adj3 extend*).ti,ab,kw.
27. (job? adj3 compress*).ti,ab,kw.
28. (job? adj3 week*).ti,ab,kw.
29. (job? adj3 day?).ti,ab,kw.
30. (shift? adj3 hour*).ti,ab,kw.
31. (shift? adj3 schedul*).ti,ab,kw.
32. (shift? adj3 roster).ti,ab,kw.
33. (shift? adj3 organi#ation).ti,ab,kw.
34. (shift? adj3 time*).ti,ab,kw.
35. (shift? adj3 overload*).ti,ab,kw.
36. (shift? adj3 extend*).ti,ab,kw.
37. (shift? adj3 compress*).ti,ab,kw.
38. (shift? adj3 week*).ti,ab,kw.
39. (shift? adj3 day?).ti,ab,kw.

1 40. (work* and (life* or live*) and balances* or imbalances* or unbalances* or
2 interference*).ti,ab,kw.
3 41. (work* and famil* and conflict*).ti,ab,kw.
4 42. or/1-41
5 43. Behavioral Symptoms/
6 44. Affective Symptoms/
7 45. affective disorder\$.ti,ab,kw.
8 46. Mood Disorders/
9 47. (mood\$ adj3 disorder\$.ti,ab,kw.
10 48. (low\$ adj3 mood\$.ti,ab,kw.
11 49. Depression/
12 50. exp Depressive Disorder/
13 51. depress\$.ti,ab,kw.
14 52. dysthymi\$.ti,ab,kw.
15 53. Adjustment Disorder/
16 54. (adjustment\$ adj3 disorder\$.ti,ab,kw.
17 55. exp Antidepressive Agents/
18 56. antidepress\$.ti,ab,kw.
19 57. or/43-56
20 58. exp Clinical Trial/
21 59. trial\$.tw.
22 60. experiment\$.tw.
23 61. (intervention adj3 (study or studies or analys\$)).tw.
24 62. Epidemiologic Studies/
25 63. Observational Study/
26 64. ((observational or epidemiologic\$) adj (study or studies or analys\$)).tw.
27 65. exp Cohort Studies/
28 66. cohort\$.tw.
29 67. (panel\$ adj3 (study or studies or analys\$ or data)).tw.
30 68. (follow up adj (study or studies or analys\$)).tw.
31 69. (repeat\$ adj measure\$).tw.
32 70. longitudinal\$.tw.
33 71. retrospective\$.tw.
34 72. prospective\$.tw.
35 73. exp Case Control Studies/
36 74. (case\$ adj3 control\$).tw.
37 75. (exposure\$ adj4 (study or studies or analys\$)).tw.
38 76. or/58-75
39 77. 42 and 57 and 76
40
41

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

Web of Science

Date searched: 11 July 2019

Records retrieved: 1,937

Databases used: Science Citation Index Expanded (1945-present), Social Sciences Citation Index (1956-present), Arts & Humanities Citation Index (1975-present), Emerging Sources Citation Index (2015-present)

Strategy:

#1: TS=("personnel staffing and scheduling") OR TS=("shift work schedule") OR TS=("work schedule tolerance") OR TS=(workload) OR TS=("workday shifts") OR TS=(overwork*) OR TS=(overtime) OR TS=(workweek*) OR TS=((work* OR job OR shift) NEAR/3 (hour* OR schedul* OR roster OR organization OR organisation OR time* OR overload* OR extend* OR compress* OR week* OR day OR days)) OR TS=(work* AND (life* OR live*) AND (balances* OR imbalances* OR unbalances* OR interference*)) OR TS=(work* AND famil* AND conflict*)

#2: TS=(affective disorder*) OR TS=(mood* NEAR/3 disorder*) OR TS=(depress*) OR TS=(dysthymia*) OR TS=(adjustment* NEAR/3 disorder*) OR TS=(antidepress*)

#3: TS=(trial*) OR TS=(experiment*) OR TS=(intervention NEAR/3 (study OR studies OR analys*)) OR TS=((observational OR epidemiologic*) NEAR/1 (study OR studies OR analys*)) OR TS=(cohort*) OR TS=(panel* NEAR/3 (study OR studies OR analys* OR data)) OR TS=(follow up NEAR/3 (study OR studies OR analys*)) OR TS=(repeat* NEAR/3 measure*) OR TS=(longitudinal*) OR TS=(retrospective*) OR TS=(prospective*) OR TS=(case* NEAR/3 control*) OR TS=(exposure* NEAR/4 (study OR studies OR analys*))

#4: #1 AND #2 AND #3

1 **CISDOC (OSH update)**

2 Date searched: 17 July 2019

3 Records retrieved: 417

4 Strategy:

5 #1 GW{affective disorder*}

6 #2 GW{mood disorder*}

7 #3 GW{low mood}

8 #4 GW{depress*}

9 #5 GW{dysthymi*}

10 #6 GW{adjustment disorder*}

11 #7 GW{antidepress*}

12 #8 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7

13 #9 GW{work*}

14 #10 GW{job*}

15 #11 GW{shift*}

16 #12 GW{overwork*}

17 #13 GW{overtime}

18 #14 #9 OR #10 OR #11 OR #12 OR #13

19 #15 DC{OUCISD}

20 #16 #8 AND #14 AND #15

21

22

1 **PsycINFO (Ebsco)**

2

3 Date searched: 12 July 2018

4 Records retrieved: 1487

5 Strategy:

6 S1: (MM "Work Scheduling") OR (MM "Work-Life Balance") OR TI ("personnel staffing and
7 scheduling") OR AB ("personnel staffing and scheduling") OR KW ("personnel staffing and
8 scheduling") OR TI ("shift work schedule") OR AB ("shift work schedule") OR KW ("shift work
9 schedule") OR TI ("work schedule tolerance") OR AB ("work schedule tolerance") OR KW
10 "work schedule tolerance") OR KW workload OR TI ("workday shifts") OR AB ("workday shifts")
11 OR KW ("workday shifts") OR TI overwork* OR AB overwork* OR KW overwork* OR TI
12 overtime OR AB overtime OR KW overtime OR TI workweek* OR AB workweek* OR KW
13 workweek* OR TI ((work* OR job# OR shift#) N3 (hour* OR schedul* OR roster OR organi?ation
14 OR time* OR overload* OR extend* OR compress* OR week* OR day#)) OR AB ((work* OR job#
15 OR shift#) N3 (hour* OR schedul* OR roster OR organi?ation OR time* OR overload* OR extend*
16 OR compress* OR week* OR day#)) OR KW ((work* OR job# OR shift#) N3 (hour* OR schedul*
17 OR roster OR organi?ation OR time* OR overload* OR extend* OR compress* OR week* OR day#)
18) OR TI ((work* AND (life* or live*) AND (balances* OR imbalances* OR unbalances* OR
19 interference*))) OR AB ((work* AND (life* or live*) AND (balances* OR imbalances* OR
20 unbalances* OR interference*))) OR KW ((work* AND (life* or live*) AND (balances* OR
21 imbalances* OR unbalances* OR interference*))) OR TI ((work* AND famil* AND conflict*)) OR
22 AB ((work* AND famil* AND conflict*)) OR KW ((work* AND famil* AND conflict*))

23

24 S2: MM "Mental Disorders" OR MM "Affective Disorders" OR TX affective N3 disorder# OR MM
25 "Affective Disorders" OR TX mood# N3 disorder# OR MM "Depression (Emotion)" OR (DE "Major
26 Depression" OR DE "Anaclitic Depression" OR DE "Dysthymic Disorder" OR DE "Endogenous
27 Depression" OR DE "Late Life Depression" OR DE "Postpartum Depression" OR DE "Reactive
28 Depression" OR DE "Recurrent Depression" OR DE "Treatment Resistant Depression") OR TX
29 depress* OR TX dysthymia# OR MM "Adjustment Disorders" OR TX adjustment# N3 disorder# OR
30 (DE "Antidepressant Drugs" OR DE "Bupropion" OR DE "Citalopram" OR DE "Fluoxetine" OR DE
31 "Fluvoxamine" OR DE "Iproniazid" OR DE "Isocarboxazid" OR DE "Lithium Carbonate" OR DE
32 "Methylphenidate" OR DE "Mianserin" OR DE "Moclobemide" OR DE "Molindone" OR DE
33 "Nefazodone" OR DE "Nialamide" OR DE "Nomifensine" OR DE "Paroxetine" OR DE "Phenelzine"
34 OR DE "Pheniprazine" OR DE "Pipradrol" OR DE "Serotonin Norepinephrine Reuptake Inhibitors"
35 OR DE "Sertraline" OR DE "Sulpiride" OR DE "Tranlycypromine" OR DE "Trazodone" OR DE
36 "Tricyclic Antidepressant Drugs" OR DE "Venlafaxine" OR DE "Zimeldine") OR TX antidepress*

37

38 S3: MM "Clinical Trials" OR TX trial* OR TX experiment* OR TX intervention N3 (study or studies
39 or analys*) OR MM "Epidemiology" OR TX (observational OR epidemiologic*) N1 (study or studies
40 or analys*) OR (MM "Cohort Analysis") OR (DE "Followup Studies") OR (MM "Repeated
41 Measures") OR (DE "Longitudinal Studies" OR DE "Prospective Studies") OR TX cohort* OR TX
42 (panel*) N3 (study OR studies OR analys* OR data) OR TX (follow up) N1 (study OR studies OR
43 analys*) OR TX repeat* N1 measure* OR TX longitudinal* OR TX retrospective* OR TX
44 prospective OR TX case* N3 control* OR TX (exposure*) N4 (study OR studies OR analys*)

45

46 S4: S1 AND S2 AND S3

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

OpenGrey (<http://www.opengrey.eu/>)

Date searched: 25 July 2018

Records retrieved: 6,450

Strategy:

("Personnel staffing and scheduling" OR "shift work schedule" OR "work life balance" OR "work schedule tolerance" OR workload OR "Workday Shifts" OR overwork* OR overtime OR workweek* OR ((work* OR job OR shift) NEAR/3 (hour* OR schedul* OR roster OR organization OR organisation OR time* OR overload* OR extend* OR compresse* OR week* OR day OR days)) OR (work* AND (life* OR live*) AND (balances* OR imbalances* OR unbalances* OR interference*)) OR (work* AND famil* AND conflict*)) AND (prevalence OR incidence OR epidemiol* OR survey OR surveillance OR "rapid assessment" OR "situation assessment" OR "situational assessment" OR rar OR cohort OR seroprevalence OR seroincidence OR seroepidemiol* OR screening OR "epidemiologic methods" OR "epidemiologic studies" OR "sentinel surveillance" OR "seroepidemiologic studies" OR "cohort studies" OR "cross sectional studies" OR "longitudinal studies" OR "follow-up studies" OR "prospective studies") AND "Behavioral Symptoms" OR "Affective Symptoms" OR "affective disorder" OR "Mood Disorders" OR "mood" OR "Depression" OR "depressive Disorder" OR "dysthymi" OR "Adjustment Disorders" OR "Antidepressive Agents" OR "Antidepress")

1
2
3
4
5
6
7
8
9
10
11

Grey Literature Report (<http://greylit.org/>)

Date searched: 5 August 2018

Records retrieved: 233

Strategy:

("work hour" OR "job hour" OR "job schedule" OR "job time" OR "shift hour" OR "shift schedule"
OR "long working hour") AND ("behavioral symptoms" OR "affective symptoms" OR "affective
disorder" OR "mood disorders" OR "mood" OR "depression" OR "depressive disorder" OR
"dysthymi" OR "adjustment disorders" OR "antidepressive agents" OR "antidepress").

1

2 *Appendix 2 Description of missing data requested and received*

3

Study ID (decision)	Description of requested missing data	Person(s) from whom missing data were requested	Date of request(s) (YYYY-MM-DD)	Data received
Ahn 2014 (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Seoyeon Ahn	2019-03-10 2019-03-30	The principal study author provided the LWH-re-categorization and some the stratified analyses.
Appelhans 2014 (excluded)	Inquiry if depression can be analysed as an outcome (was a covariate in the article).	Dr. Bradley M. Appelhans	2019-03-10	The principal study author reported that it was not possible to analyse depression as an outcome.
Berthelsen 2015 (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Mona Berthelsen	2019-03-10 2019-03-25 2019-04-19	The principal study author did not respond.
Dembe 2016 (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Allard E. Dembe	2019-03-10 2019-03-25 2019-04-19	The principal study author did not respond.
d'Errico 2011 (excluded)	Inquiring if there was a diagnostic code attached to prescription of antidepressants.	Dr. Angelo d'Errico	2019-03-10	The principal study author stated that diagnostic codes were not available.
Hannerz 2016 (excluded)	Inquiring if there was a diagnostic code attached to prescription of antidepressants.	Dr. Harald Hannerz	2019-03-10	The principal study author stated that diagnostic codes were not available.
Kato 2014 (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Rika Kato	2019-03-10	The principal study author stated that re-categorization and stratified analyses were not possible.

Study ID (decision)	Description of requested missing data	Person(s) from whom missing data were requested	Date of request(s) (YYYY-MM-DD)	Data received
Kawakami 1990 (excluded)	Inquiring if number of hours worked overtime could be re-calculated as total working hours and if adjusted estimates could be calculated.	Dr. Norito Kawakami	2019-03-10	The principal study author stated that the data was discarded and no longer available.
Kim 2013 (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Jaeyoung Kim	2019-03-10 2019-04-19	The principal study author made the original data available to us.
Kim 2016 (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Woorim Kim	2019-03-10 2019-03-25 2019-04-19	The principal study author did not respond.
Laaksonen 2012 (awaiting classification)	Inquiring if there was a diagnostic code attached to prescription of antidepressants.	Dr. Mikko Laaksonen	2019-03-10 2019-03-25 2019-04-19	The principal study author did not respond.
Ogasawara 2011 (awaiting classification)	Inquiring if number of hours worked overtime could be re-calculated as total working hours.	Dr. Kazuyoshi Ogasawara	2019-03-10 2019-03-25 2019-04-19	The principal study author did not respond.
Razavi 2015 (excluded)	Inquiring if diagnosis of depression could be extracted from diagnosis of common mental disorders.	Dr. Tahera Razawi; Dr. Stephen Stansfeld	2019-03-10 2019-03-25 2019-03-28	The principal study author stated that diagnosis of depression could not be extracted from diagnosis of common mental disorders.
Shields 2017 (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Margot Shields	2019-03-10 2019-03-25 2019-04-19	The principal study author could not be contacted (email was returned as undeliverable).
Tarumi 2003 (excluded)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age,		2019-03-10	The principal study author made the original data available to us. When analysing the data, we found out that diagnosis of

Study ID (decision)	Description of requested missing data	Person(s) from whom missing data were requested	Date of request(s) (YYYY-MM-DD)	Data received
	sex, and SES, if possible.			depression could not be extracted from diagnosis of mental disorders.
Tokuyama 2003 (awaiting classification)	Inquiring if number of hours worked overtime could be re-calculated as total working hours.		2019-03-10 2019-03-25 2019-04-19	The principal study author could not be contacted (email was returned as undeliverable).
Varma 2012 (excluded)	Inquiring if there was a diagnostic code attached to prescription of antidepressants,	Dr. Jens Peter Bonde	2019-03-10	The principal study author stated that diagnostic codes were not available.
Virtanen 2012 (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Marianna Virtanen	2019-04-23	The principal study stated that the data was no longer available to her.
Virtanen 2018 (included)	Providing estimates for the LWH categories: 41-48, 49-54 hours/week. Stratification by age, sex, and SES, if possible.	Dr. Marianna Virtanen; Dr. Mika Kivimäki	2019-03-30	The principal study author stated that providing the requested estimates and stratified analyses were not possible.
Wang 2012a (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr JianLi Wang	2019-03-10 2019-10-26	The principal study author provided the LWH-re-categorization and the stratified analyses.
Wang 2012b (included)	LWH re-categorization: 35-40, 41-48, 49-54, ≥ 55 hours/week. Stratification by age, sex, and SES, if possible.	Dr JianLi Wang	2019-03-10 2019-11-26	The principal study author provided the LWH-re-categorization but could not provide the stratified analyses.
Zadow 2019 (included)	Analyses without adjustment for psychological safety climate. Stratification by age, sex, and SES, if possible	Dr Amy Zadow	2019-10-27	The principal study author provided the requested analyses.

- 1 LWH: long working hours
- 2 The full references of the studies can be found in the reference list of the article (for included studies)
- 3 or in Appendix 3 (for excluded studies) and Appendix 4 (for studies awaiting classification)
- 4
- 5

1
2
3
4
5
6
7
8
9
10
11
12
13

Appendix 3 Selected excluded studies and reason for their exclusion

We excluded 397 studies after full text reading and rated another 3 studies as awaiting classification (see Appendix 4). Below, we state the reference and the reason for exclusion for 40 of the 397 excluded studies. This includes the 36 studies where there was disagreement among the reviewers whether these studies should be included or excluded after full text reading. These disagreements were either solved by discussions among the reviewers with involving of a third reviewer or by gathering additional information from the study authors. In addition, the list includes 4 studies (Hannerz 2016; Kawakami 1990; Tarumi 2003; Varma 2012) that were first considered for inclusion and then were excluded after we got additional information from the study authors.

Study ID (Full reference)	Reason for exclusion
Albrecht 2017 Albrecht SC, Kecklund G, Rajaleid K, Leineweber C. The longitudinal relationship between control over working hours and depressive symptoms: Results from SLOSH, a population-based cohort study. <i>Journal of Affective Disorders</i> . 2017;215:143-151.	Ineligible outcome
Appelhans 2014 Appelhans BM, Segawa E, Janssen I, Kazlauskaitė R, Thurston RC, Lewis TT, et al. Employment status, depressive symptoms, and waist circumference change in midlife women: the Study of Women's Health Across the Nation (SWAN). <i>Annals of Epidemiology</i> . 2014;24(3):187-192.	Ineligible outcome
Bannai 2014 Bannai A, Tamakoshi A. The association between long working hours and health: a systematic review of epidemiological evidence. <i>Scandinavian Journal of Work, Environment & Health</i> . 2014;40(1):5-18.	Ineligible study design
Beseler 2010 Beseler CL, Stallones L. Safety knowledge, safety behaviors, depression, and injuries in Colorado farm residents. <i>American Journal of Industrial Medicine</i> . 2010;53(1):47-54.	Ineligible outcome
Bingol 2016 Bingol F, Buzlu S. Effect of the Cognitive-Behavioral Prevention Program on Levels of Depression Symptoms Among Working Adolescents in Turkey. <i>Journal of Psychosocial Nursing and Mental Health Services</i> . 2016;54(7):43-51.	Ineligible exposure
Boini 2016 Boini S, Kolopp M, Grzebyk M, Hedelin G, Chouaniere D. Effect of change in psychosocial exposure on incidence of depressive and anxiety disorders. <i>Occupational and Environmental Medicine</i> . 2016;73(Supplement 1):A2.	Ineligible exposure

<p>Brenninkmeijer 2008 Brenninkmeijer V, Houtman I, Blonk R. Depressed and absent from work: predicting prolonged depressive symptomatology among employees. Occupational medicine (Oxford, England). 2008;58(4):295-301.</p>	<p>Ineligible outcome</p>
<p>Chevalier 1996 Chevalier A, Bonenfant S, Picot MC, Chastang JF, Luce D. Occupational factors of anxiety and depressive disorders in the French national electricity and gas company. Journal of Occupational and Environmental Medicine. 1996;38(11):1098-1107.</p>	<p>Ineligible study design</p>
<p>Cinamon 2016 Cinamon RG. Integrating work and study among young adults: Testing an empirical model. Journal of Career Assessment. 2016;24(3):527-542.</p>	<p>Ineligible study design</p>
<p>d'Errico 2011 d'Errico A, Cardano M, Landriscina T, Marinacci C, Pasian S, Petrelli A, et al. Workplace stress and prescription of antidepressant medications: a prospective study on a sample of Italian workers. International Archives of Occupational and Environmental Health. 2011;84(4):413-424.</p>	<p>Ineligible outcome</p>
<p>Fernández 2004 Fernández RJL, Álvarez dCGI, Doménech MG. An occupational health problem: Occupational stress and burnout. Medicina y Seguridad del Trabajo. 2004;1(197):65-78</p>	<p>Ineligible study design</p>
<p>Firth-Cozens 1998 Firth-Cozens J. Individual and organizational predictors of depression in general practitioners. The British Journal of General Practice: The Journal of the Royal College of General Practitioners. 1998;48(435):1647-1651.</p>	<p>Ineligible study design</p>
<p>Fried 2014 Fried EI, Nesse RM, Zivin K, Guille C, Sen S. Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. Psychological medicine. 2014;44(10):2067-2076.</p>	<p>Ineligible outcome</p>
<p>Fushimi 2013 Fushimi M, Saito S, Shimizu T. Prevalence of depressive symptoms and related factors in Japanese employees as measured by the Center for Epidemiologic Studies Depression Scale (CES-D). Community Mental Health Journal. 2013;49(2):236-242.</p>	<p>Ineligible study design</p>
<p>Geiger-Brown 2004 Geiger-Brown J, Muntaner C, Lipscomb J, Trinkoff A. Demanding work schedules and mental health in nursing assistants working in nursing homes. Work & Stress. 2004;18(4):292-304.</p>	<p>Ineligible study design</p>
<p>Geurtsen 2011 Geurtsen GJ, van Heugten CM, Martina JD, Rietveld AC, Meijer R, Geurts AC. A prospective study to evaluate a residential community reintegration program for patients with chronic acquired brain injury. Archives of Physical Medicine and Rehabilitation. 2011;92(5):696-704.</p>	<p>Ineligible exposure</p>

<p>Gjerdingen 2003 Gjerdingen DK, Center BA. First-time parents' prenatal to postpartum changes in health, and the relation of postpartum health to work and partner characteristics. <i>Journal of the American Board of Family Practice</i>. 2003;16(4):304-311.</p>	<p>Ineligible outcome</p>
<p>Hannerz 2016 Hannerz H, Albertsen K. Long working hours and use of psychotropic medicine: A follow-up study with register linkage. <i>Scandinavian Journal of Work, Environment and Health</i>. 2016;42(2):153-161.</p>	<p>Ineligible outcome</p>
<p>Hillhouse 2000 Hillhouse JJ, Adler CM, Walters DN. A simple model of stress, burnout and symptomatology in medical residents: A longitudinal study. <i>Psychology, Health and Medicine</i>. 2000;5(1):63-73.</p>	<p>Ineligible outcome</p>
<p>Houdmont 2016 Houdmont J, Randall R. Working hours and common mental disorders in English police officers. <i>Occupational Medicine</i>. 2016;66(9):713-718.</p>	<p>Ineligible study design</p>
<p>Jennings 2016 Jennings KS, Sinclair RR, Mohr CD. Who benefits from family support? Work schedule and family differences. <i>Journal of Occupational Health Psychology</i>. 2016;21(1):51-64.</p>	<p>Ineligible outcome</p>
<p>Jongen 2014 Jongen PJ, Wesnes K, van Geel B, Pop P, Sanders E, Schrijver H, et al. Relationship between working hours and power of attention, memory, fatigue, depression and self-efficacy one year after diagnosis of clinically isolated syndrome and relapsing remitting multiple sclerosis. <i>PLoS One</i>. 2014;9(5):e96444.</p>	<p>Ineligible outcome</p>
<p>Kachi 2014 Kachi Y, Otsuka T, Kawada T. Precarious employment and the risk of serious psychological distress: a population-based cohort study in Japan. <i>Scandinavian Journal of Work, Environment & Health</i>. 2014;40(5):465-472.</p>	<p>Ineligible outcome</p>
<p>Kalmbach 2015 Kalmbach DA, Pillai V, Cheng P, Arnedt JT, Drake CL. Shift work disorder, depression, and anxiety in the transition to rotating shifts: the role of sleep reactivity. <i>Sleep Medicine</i>. 2015;16(12):1532-1538.</p>	<p>Ineligible exposure</p>
<p>Kato 2010 Kato K, Inui F, Hayakawa K. The twin nurses project: A web-based cohort study of work-style and lifestyle of Japanese adults. <i>Twin Research and Human Genetics</i>. 2010;13(3):268.</p>	<p>Ineligible study design</p>
<p>Kawakami 1990 Kawakami N, Araki S, Kawashima M. Effects of job stress on occurrence of major depression in Japanese industry: a case-control study nested in a cohort study. <i>Journal of Occupational Medicine</i>. 1990;32(8):722-5.</p>	<p>Ineligible exposure</p>
<p>Levin 2012 Levin A, Besser A, Albert L, Smith D, Neria Y. The effect of attorneys' work with trauma-exposed clients on PTSD symptoms, depression, and functional</p>	<p>Ineligible outcome</p>

impairment: a cross-lagged longitudinal study. <i>Law and Human Behavior</i> . 2012;36(6):538-547.	
Marchand 2010 Marchand A, Blanc ME. The contribution of work and non-work factors to the onset of psychological distress: An eight-year prospective study of a representative sample of employees in Canada. <i>Journal of Occupational Health</i> . 2010;52(3):30-39.	Ineligible outcome
Peristera 2018 Peristera P, Westerlund H, Magnusson Hanson LL. Paid and unpaid working hours among Swedish men and women in relation to depressive symptom trajectories: results from four waves of the Swedish Longitudinal Occupational Survey of Health. <i>BMJ Open</i> . 2018;8(6):e017525.	Ineligible outcome
Perry- Jenkins 2011 Perry- Jenkins M, Smith JZ, Goldberg AE, Logan J. Working- class jobs and new parents' mental health. <i>Journal of Marriage and Family</i> . 2011;73(5):1117-1132.	Ineligible outcome
Razavi 2015 Razavi T, Clark C, Stansfeld SA. Work-family conflict as a predictor of common mental disorders in the 1958 British birth cohort. <i>Longitudinal and Life Course Studies</i> . 2015;6(3):264-278.	Ineligible outcome
Ryan 2015 Ryan KA, Eisenberg D, Kim HM, Lai Z, McInnis M, Kilbourne AM. Longitudinal impact of a collaborative care model on employment outcomes in bipolar disorder. <i>Journal of Affective Disorders</i> . 2015;188:239-242.	Ineligible outcome
Shields 2006 Shields M. Stress and depression in the employed population. <i>Health Reports</i> . 2006;17(4):11-29.	Ineligible exposure
Stavem 2003 Stavem K, Hofoss D, Aasland OG. Work characteristics and morbidity as predictors of self-perceived health status in Norwegian physicians. <i>Scandinavian Journal of Public Health</i> . 2003;31(5):375-381.	Ineligible study design
Sugihara 2008 Sugihara Y, Sugisawa H, Shibata H, Harada K. Productive roles, gender, and depressive symptoms: evidence from a national longitudinal study of late-middle-aged Japanese. <i>The journals of gerontology Series B, Psychological Sciences and Social Sciences</i> . 2008;63(4):227-234.	Ineligible outcome
Sundstrup 2017 Sundstrup E, Andersen LL. Hard Physical Work Intensifies the Occupational Consequence of Physician-Diagnosed Back Disorder: Prospective Cohort Study with Register Follow-Up among 10,000 Workers. <i>International journal of Rheumatology</i> . 2017;2017:1037051	Ineligible exposure
Tarumi 2003 Tarumi K, Hagihara A, Morimoto K. A prospective observation of onsets of health defects associated with working hours. <i>Ind Health</i> . 2003;41(2):101-108.	Ineligible outcome

<p>Varma 2012 Varma A, Marott JL, Stoltenberg CDG, Wieclaw J, Kolstad HA, Bonde JPE. With long hours of work, might depression then lurk? A nationwide prospective follow-up study among Danish senior medical consultants. <i>Scandinavian Journal of Work, Environment & Health</i>. 2012;38(5):418-426.</p>	<p>Ineligible outcome</p>
<p>Virtanen 2011 Virtanen M, Ferrie JE, Singh-Manoux A, Shipley MJ, Stansfeld SA, Marmot MG, et al. Long working hours and symptoms of anxiety and depression: a 5-year follow-up of the Whitehall II study. <i>Psychological Medicine</i>. 2011;41(12):2485-2494.</p>	<p>Ineligible outcome</p>
<p>Wang 2013 Wang J, Schmitz N, Patten S, Currie S, Sareen J. A population-based longitudinal study on work environment and major depressive disorder. <i>European Psychiatry</i>. 2013;28(Suppl. 1).</p>	<p>Duplicate (conference abstract of results that were also published in a journal article that was included in the review)</p>

1
2

1
2
3
4
5
6
7
8

Appendix 4 Studies awaiting classification

For three studies we lacked information to determine eligibility. We contacted the lead study authors but were not able to obtain the necessary information. Therefore, these studies are categorized as “awaiting classification”

Study ID (Full reference)	Rational
Laaksonen 2012 Laaksonen M, Lallukka T, Lahelma E, Partonen T. Working conditions and psychotropic medication: a prospective cohort study. <i>Social Psychiatry and Psychiatric Epidemiology</i> . 2012;47(4):663-670.	Study would be eligible if a diagnostic code was attached to prescription of antidepressants. This information could not be obtained yet.
Ogasawara 2011 Ogasawara K, Nakamura Y, Aleksic B, Yoshida K, Ando K, Iwata N, et al. Depression associated with alcohol intake and younger age in Japanese office workers: a case-control and a cohort study. <i>Journal of Affective Disorders</i> . 2011;128(1-2):33-40.	Study would be eligible if hours worked overtime could be re-calculated to total working hours. This information could not be obtained yet.
Tokuyama 2003 Tokuyama M, Nakao K, Seto M, Watanabe A, Takeda M. Predictors of first-onset major depressive episodes among white-collar workers. <i>Psychiatry and Clinical Neurosciences</i> . 2003;57(5):523-531.	Study would be eligible if hours worked overtime could be re-calculated to total working hours. This information could not be obtained yet.

9
10

1

2 *Appendix 5 Rational for selecting specific estimates from included studies*

3

Study ID	Adjustments	Why was this model chosen?
Ahn 2018	Sex, age, SES (income), marital status, industry, occupation, country (all Republic of Korea).	Re-analysis of available data in accordance with review study protocol.
Berthelsen 2015	Sex, age, SES (occupational grade, all were nurses) marital status/cohabitation, baseline level of symptoms of depression and anxiety, country (all Norway).	The only model presented in the article.
Dembe 2016	Sex, age, SES (income), race, number of years worked, smoking status, country (all USA)	The only model presented in the article
Kato 2014	Sex (men only), age, SES (job grade), years of experience, shift work, site of work, country (Japan)	Model 3 out of six models. Model 1 (age) was underadjusted; Model 2 (age and working hours at follow-up) was underadjusted; Model 4 (age, education, family income) included the key covariates but less covariates than model 3; Model 5 (age, smoking, alcohol) and model 6 (age, working hours at follow-up, education, years of experience, job grade, family income, shift work, site of work, smoking, alcohol) adjusted for potential mediators (smoking, alcohol).
Kim 2013	Sex, age, SES (education), country (all (USA)	Re-analysis of available data in accordance with review study protocol.
Kim 2016	Sex, age, SES (education, income), marital status, country (all Republic of Korea)	Model 1 out of three models. Model 2 (sex, age, education, income, marital status, job satisfaction level, chronic disease status, year) and model 3 (sex, age, education, income, marital status, job satisfaction, chronic disease status, year, permanent versus precarious employments) adjusted for potential mediators (job satisfaction, chronic disease status)

NLSY OA Cohort 2019	Sex, age, SES (by income), country (all USA)	Re-analysis of available data in accordance with review study protocol.
Shields 1999	Sex (stratified), age, SES (education, income, white collar/blue collar), self-employment, multiple jobs holdings, marital status, children under age 12 living in household, work stress (job strain, job insecurity, supervisor support), country (all Canada)	The only model presented in the article. The analysis was adjusted for a potential mediator (work stress)
Virtanen 2012	Sex, age, SES (occupational grade) marital status, country (all United Kingdom)	Model 3 out of five models. Model 1 (crude analysis) and model 2 (sex, age) were underadjusted; Model 4 (sex, age, occupational grade, marital status, physical disease, smoking, alcohol use) and model 5 (sex, age, occupational grade, marital status, physical disease smoking status, alcohol use, job strain, social support at work) adjusted for potential mediators (physical disease, smoking status, alcohol use, job strain, social support at work)
Virtanen 2018 – ACL	Sex, age, SES (type not reported), marital status, country (all USA)	The only model presented in the article.
Virtanen 2018 – DWECS	Sex, age, SES (type not reported), marital status, country (all Denmark)	The only model presented in the article.
Virtanen 2018 – ELSA	Sex, age, SES (type not reported), marital status, country (all United Kingdom)	The only model presented in the article.
Virtanen 2018 – HESSUP	Sex, age, SES (type not reported), marital status, country (all Finland)	The only model presented in the article.
Virtanen 2018 – HILDA	Sex, age, SES (type not reported), marital status, country (all Australia)	The only model presented in the article.
Virtanen 2018 – HRS	Sex, age, SES (type not reported), marital status, country (all USA)	The only model presented in the article.
Virtanen 2018 – NLSY	Sex, age, SES (type not reported), marital status, country (all USA)	The only model presented in the article.
Virtanen 2018 – PUMA	Sex, age, SES (type not reported), marital status, country (all Denmark)	The only model presented in the article.
Virtanen 2018 – SHARE	Sex, age, SES (type not	The only model presented in

	reported), marital status	the article.
Virtanen 2018 – SLOSH	Sex, age, SES (type not reported), marital status, country (all Sweden)	The only model presented in the article.
Wang 2012a	Sex, age, SES (education, income, occupational gradient), marital status, country (all Canada)	Re-analysis of available data in accordance with review study protocol.
Wang 2012b	Sex, age, SES (education, income, occupational gradient), marital status, country (all Canada)	Re-analysis of available data in accordance with review study protocol.
Zadow 2019	Sex, age, SES (income), country (all Australia)	Re-analysis of available data in accordance with review study protocol.

1

1
2
3
4
5

Appendix 6 Justification for risk of bias ratings for included studies

Ahn 2018

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from various job groups from a large national representative sample from the Republic of Korea (Korean Welfare Panel Study (KOWEPS)). Response and attrition rate not reported. However, no indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up, 6 years later. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with CES-D, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Low	Re-analyzed their data for this review. Adjusted for the three essential confounders (sex, age, socioeconomic position (education, income) plus job group, industry and country (all from Republic of Korea).
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with CES-D (covering the last 2 weeks) at baseline and after 6 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported. The author re-analyzed the data after our instructions for this review
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression.
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

6
7

1
2

Berthelsen 2015

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably high	Nurses recruited from the Norwegian Nurses Organization. Baseline response rate was low (38.1%), attrition during follow-up was 23.2%. However, no indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations. Authors did not exclude participants with baseline depression but adjusted for depression in the analyses. It is possible that participants with baseline depression were selected into specific working time arrangements because of their depression and therefore we rate here probably high risk of bias
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up, 1 year later. Investigators did not exclude individuals with depression at baseline, but instead adjusted for depression status at baseline. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably high	Self-reported working hours. We assume that individuals free of depressive symptoms are able to estimate their own working hours accurately. However, individuals with depression at baseline were not excluded, but instead the analyses adjusted for depression at baseline. Cannot be ruled out that depression at baseline had affected reporting of working hours.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with HADS which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Low	Adjusted for the three essential confounders (sex, age, socioeconomic position (occupational status as all were nurses) plus job group and industry (all were nurses) and country (all from Norway). Additional adjustment for marital status/cohabitation and children living at home. Also adjusted for baseline anxiety and depression score.
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with HADS (covering the last week) at baseline and after 1 year of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Dembe 2016

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably high	Employees from a large national survey in the USA that started in 1979, National Longitudinal Survey of Youth (NLSY) including various job groups. Participants followed up until 2010 (cohort members 46-53 years old). Working hours averaged over the 32 study period. Depression measured at 40+ and 50+ module. Therefore, onset of depression might have occurred before some measures of the working hours and individuals with depression might have been selected into specific jobs and working hours because of their depression.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up during 32 year follow-up. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably high	Self-reported working hours. We assume that individuals free of depressive symptoms are able to estimate their own working hours accurately. However, individuals might had depression when working hours were measured, therefore probably high risk of bias.
4. Were outcome assessment methods lacking accuracy?	Probably high	Depression measured with a question if a doctor ever had diagnosed a depression. As participants had to remember this over a long time period, accuracy is questionable.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex (stratified), age, socioeconomic position (education, income, occupation) plus country (all from USA)
6. Were incomplete outcome data inadequately addressed?	Probably high	One simple question, covering the whole life, likely that some cases of depression were missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because lifetime depression prior baseline was assessed with one simple question, it is possible that cases with lifetime depression prior baseline were missed. Therefore, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Kato 2014

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from two factories in Japan. Response rate 85%, attrition rate 64%. However, no indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up, 1 year later. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with CES-D, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex (all men), age, socioeconomic position (education, income) plus country (all from Japan)
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with CES-D (covering the last 2 weeks) at baseline and after 1 year of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Kim, J 2013

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Nationally representative household survey of the US population, Medical Expenditure Panel Survey (MEPS). Response rate across panels was 65% to 71%, follow-up response rate was over 90%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up, between 6 to 18 months later. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured in the interviews by asking participants about doctor-diagnosed depression and treatment. Answers were then coded into ICD-9 codes. As these were annual interviews, participants were probably able to remember episodes of depression.
5. Was potential confounding inadequately incorporated?	Probably low	Authors provided data for us, so we could re-analyze the data. Adjusted for the three essential confounders (sex, age, socioeconomic position (education) and country (all from USA)).
6. Were incomplete outcome data inadequately addressed?	Probably low	Depression measured in the interviews by asking participants about doctor-diagnosed depression and treatment. Answers were then coded into ICD-9 codes. As these were annual interviews, participants were probably able to remember episodes of depression and there is no major concern that depressive episodes between the surveys were missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported. The author re-analyzed the data after our instructions for this review
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably low	No other problems were identified. The study assessed lifetime depression.

3
4

1
2

Kim, W 2016

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from various job groups from a large national representative sample from the Republic of Korea (Korean Welfare Panel Study (KOWEPS)). Response and attrition rate not reported. However, no indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up, at 1, 2, and 3 years later. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with CES-D, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (education, income) plus marital status and country (all from Republic of Korea)
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with CES-D (covering the last 2 weeks) at baseline and after 1, 2, and 3 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

NLSY OA Cohort 2019

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from a large national survey in the USA that started in 1979, National Longitudinal Survey of Youth (NLSY) including various job groups. For the analyses, participants were followed from time of first depression measurement (age 40) to time of second depression measurement (age 50). No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline (age 40), incident depression measured at follow-up (age 50). Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably high	Depression measured with a question if a doctor ever had diagnosed a depression. As participants had to remember this over a long time period, accuracy is questionable.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex age, socioeconomic position (income)) plus country (all from USA)
6. Were incomplete outcome data inadequately addressed?	Probably high	One simple question, covering the whole life, likely that some cases of depression were missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because lifetime depression prior baseline was assessed with one simple question, it is possible that cases with lifetime depression prior baseline were missed. Therefore, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Shields 1999

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from a national household survey in Canada. Response (96%) was high and attrition (6%) rate was low. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up, 2 years later. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Low	Depression measured with clinical interview, which is the gold standard method.
5. Was potential confounding inadequately incorporated?	Probably high	Adjusted for the three essential confounders (sex (stratified), age, socioeconomic position (education, income, social class) plus country (all from Canada). However, the study also adjusted for a potential mediator (work stress) therefore probably high risk of bias.
6. Were incomplete outcome data inadequately addressed?	Probably low	Clinical interview that covers depression the last 12 months and 2 year between baseline and follow-up, i.e. it can be assumed that the vast majority of cases were identified.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2012

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	British Civil Servants from the Whitehall II study. Participants from wave 3 (1991-1993, baseline) and wave 5 (1997-1999, follow-up). Response rate was 73% (wave 1, 1985-1988), attrition from wave 3 to wave 5 was 14%. Individuals no longer employed at wave 5 were excluded. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up, 5 years later. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Low	At baseline, participants with psychiatric morbidity were identified with the GHQ and excluded. At follow-up, depression was measured with a clinical interview, which is the gold standard method.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (occupational grade) plus industry (all were civil servants) and country (all from UK).
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with GHQ at baseline and Clinical interview (covering the last 1 year) at follow-up during a 5.8 year follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported. The author re-analyzed the data after our instructions for this review
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – ACL

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from general population in the USA. Response rate at baseline 68%, loss to follow-up 17%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 3 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with CES-D, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from USA))
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with CES-D after 3 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – DWECS

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from the general population in Denmark. Response rate at baseline 75%, loss to follow-up 32%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 5 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with MHI-5, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from Denmark)
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with the MHI-5 after 4 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – ELSA

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Older employees from the general population in the United Kingdom. Response rate at baseline 67%, loss to follow-up 19%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 2 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with CES-D, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from the United Kingdom))
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with the CES-D after 2 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4
5

1
2

Virtanen 2018 – HESSUP

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from general population in Finland. Response rate at baseline 40%, loss to follow-up 21%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 5 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with BDI, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from Finland)
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with BDI after 5 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – HILDA

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from the general population in Australia. Response rate at baseline 66%, loss to follow-up 17%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 2 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with the MHI-5, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from Australia)
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with the MHI-5 after 2 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – HRS

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Older employees from general population in the USA. Response rate at baseline 82%, loss to follow-up 14%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 2 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably high	In the article it says under measure “depressive symptoms” without specifying the instrument. Based on our knowledge of the HRS study, we assume that the CES-D was used, but we cannot be sure. Therefore, we rate this as probably high risk of bias.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from USA))
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed (probably) with CES-D after 2 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – NLSY

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from general population in the USA. Response rate at baseline 87%, loss to follow-up 4%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 2 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with CES-D, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from USA)
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with CES-D after 2 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – PUMA

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees in the human service sector in Denmark. Response rate at baseline 80%, loss to follow-up 42%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 5 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with the MHI-5, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from Denmark)
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with the MHI-5 after 5 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – SHARE

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Older employees from the general population in Europe and in Israel. Response rate at baseline 85%, loss to follow-up 31%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 2 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with the EURO-D, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status.
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with the EURO-D after 2 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Virtanen 2018 – SLOSH

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from the general population in Sweden. Response rate at baseline 61%, loss to follow-up 25%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 4 years. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably high	In the article it says under measure "Symptom Check List, SCL". We assume that this was the Symptom Check List Core Depression scale (SCL-CD6), which is a validated self-administered rating scale for measuring depression. However, since there are some uncertainties, if this was indeed the SCL-CD6, we assessed risk of bias as probably high here.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (type of socioeconomic position measure not specified) plus marital status and country (all from Sweden)
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression (probably) assessed with the SCL-CD6 after 4 years of follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

3
4

1
2

Wang 2012a

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Random sample of Canadian employee recruited by random digit dialing. Response not reported, and attrition rate was 23%. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up, 1 year later. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Low	Depression measured with clinical interview, which is the gold standard method.
5. Was potential confounding inadequately incorporated?	Probably low	Re-analyzed their data for this review. Adjusted for the three essential confounders (sex, age, socioeconomic position (education, income, gradient) plus job group (partly through gradient), marital status and country (all from Canada)
6. Were incomplete outcome data inadequately addressed?	Probably low	Depression assessed with clinical interview covering the last 12 months and follow-up after 12 months, thus cases between baseline and follow-up were identified.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported. The author re-analyzed the data after our instructions for this review
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably low	No other problems were identified. The study assessed lifetime depression.

3
4

1
2

Wang 2012b

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably high	Random sample of Canadian employee recruited by random digit dialing. Only those with a history of depression (either current, last 12 months or longer than 12 months ago) were included. We were interested only in those who did not had current depression, as we wanted to examine for the review recurrence. No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations. Attrition rate was 19%. Although all individuals in the data set that we used for the review were free of depression at baseline, it is likely that individuals differed with regard to the frequency and severity of their previous depression and with regard to whether they had recovered completely or incompletely from this depression. It is possible that these differences have influenced both selection into specific work time arrangements and risk of recurrent depression during follow-up, therefore we rate this as probably high risk of bias.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, recurrent depression measured at follow-up, 1 year later. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably high	Self-reported working hours. We cannot be ruled out that incomplete recovery from a previous depression had affected reporting of working hours.
4. Were outcome assessment methods lacking accuracy?	Low	Depression measured with clinical interview, which is the gold standard method.
5. Was potential confounding inadequately incorporated?	Probably low	Re-analyzed their data for this review. Adjusted for the three essential confounders (sex, age, socioeconomic position (education, income, gradient) plus job group (partly through gradient), marital status and country (all from Canada)
6. Were incomplete outcome data inadequately addressed?	Probably low	Depression assessed with clinical interview covering the last 12 months and follow-up after 12 months, thus cases between baseline and follow-up were identified. Only cases due to non-response were not identified
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported. The author re-analyzed the data after our instructions for this review
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably low	No other problems were identified. The study assessed lifetime depression.

3
4

1
2

Zadow 2019

Bias	Authors' assessment of risk of bias	Support for assessment
1. Are the study groups at risk of not representing their source populations in the manner that might introduce selection bias?	Probably low	Employees from various job groups from three Australian regions. Baseline response rate 37%, follow-up response rate between 83% to 90% (depending on region). No indication of differential selection with regard to baseline participation or attrition during follow-up that may have biased exposure-outcome associations.
2. Was knowledge of the group assignments inadequately prevented (i.e. blinded or masked) during the study, potentially leading to subjective measurement of either exposure or outcome?	Probably low	Observational prospective cohort study, exposure measured at baseline, incident depression measured at follow-up after 1 year. Unknown if investigators (e.g., statisticians) were blinded when analyzing data, but even if not blinded, it is unlikely that this had affected analyses.
3. Were exposure assessment methods lacking accuracy?	Probably low	Self-reported working hours. We assume that individuals free of depressive symptoms (as they were at baseline) are able to estimate their own working hours accurately.
4. Were outcome assessment methods lacking accuracy?	Probably low	Depression measured with PHQ-9, which is a validated self-administered rating scale for measuring depression, although not the gold standard, which would be a clinical diagnostic interview.
5. Was potential confounding inadequately incorporated?	Probably low	Adjusted for the three essential confounders (sex, age, socioeconomic position (education, income)).
6. Were incomplete outcome data inadequately addressed?	Probably high	Depression assessed with PHQ-9 (covering the last 4 weeks) at baseline and at follow-up. Cases that occurred between baseline and follow-up and were in remission at time of follow-up were probably missed.
7. Does the study appear to have selective outcome reporting?	Low	No indication from article that results were selectively reported.
8. Did the study receive any support from a company, study author, or other entity having a financial interest in any of the exposures studied?	Low	No indication that there was support from any entity that had a financial interest related to the association between long working hours and risk of depression
9. Did the study appear to have other problems that could put it at a risk of bias?	Probably high	Because information on lifetime depression prior baseline was missing, it is uncertain how many cases during follow-up were first time incidence of depression and how many were recurrent depression.

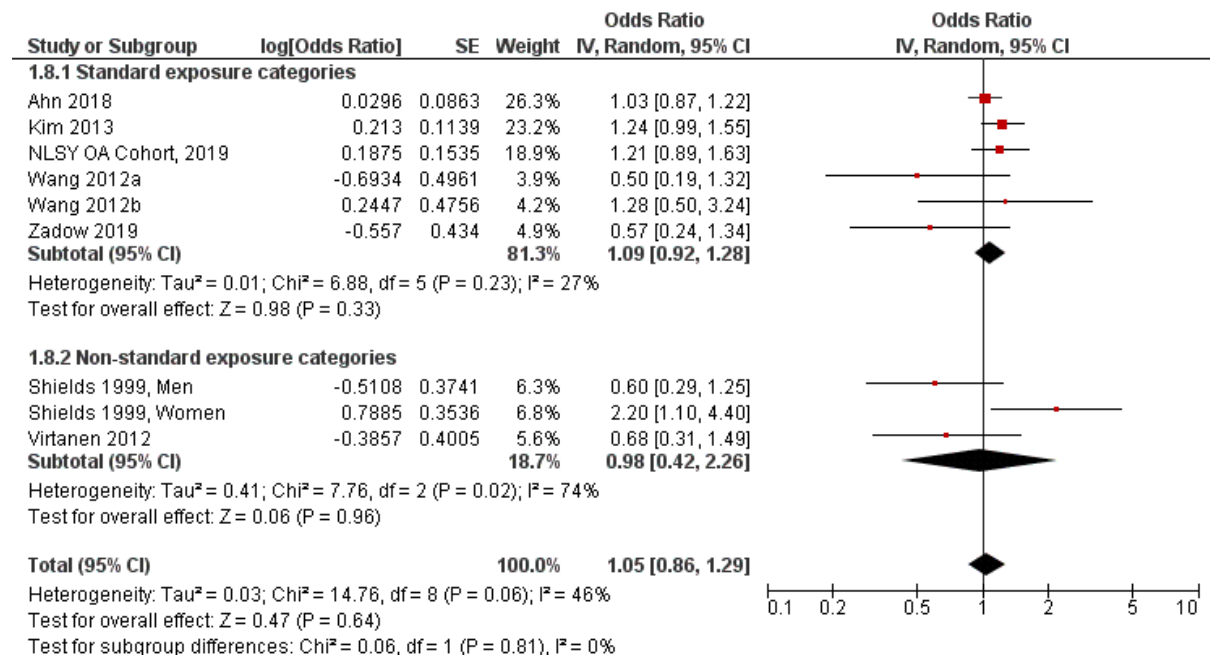
3
4
5

1
2 **Appendix 7 Explorative sensitivity analyses, comparing studies using the**
3 **standard exposure categories versus studies using approximated exposure**
4 **categories**

5
6 As explorative sensitivity analyses, we compared estimates for studies using the standard exposure
7 categories versus studies using approximated exposure categories.

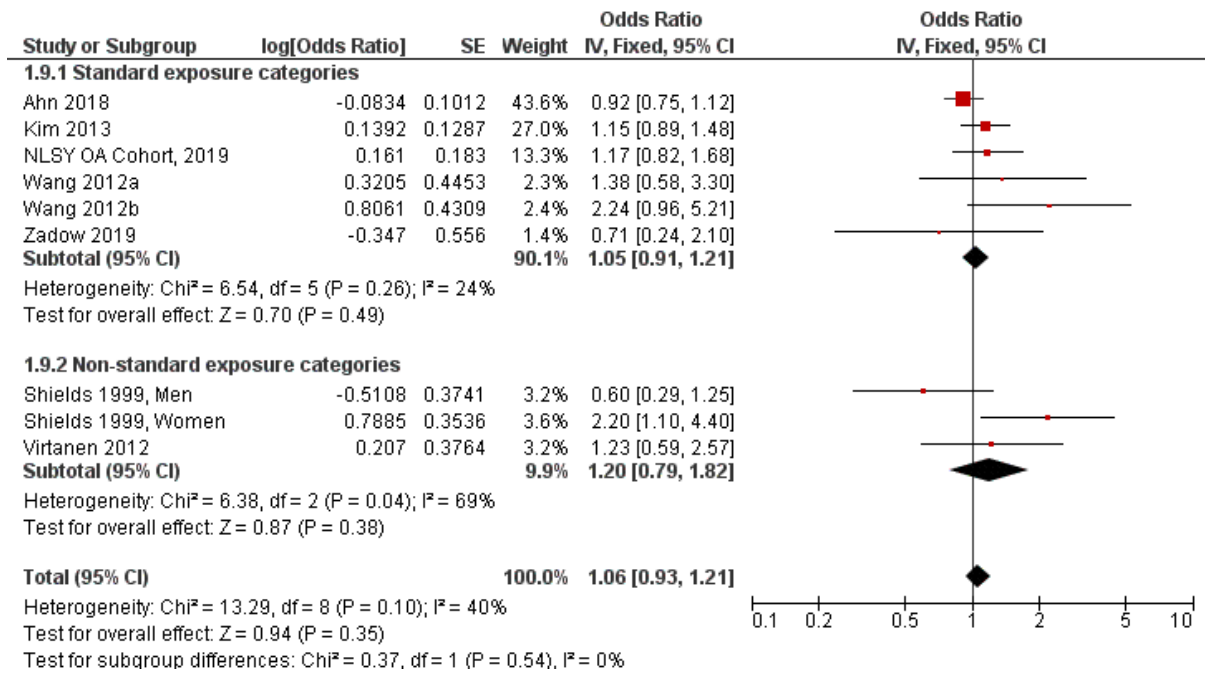
8
9 Estimates for studies using the standard exposure categories were similar to studies using
10 approximated exposures categories, both for working hours 41-48 (Fig. A7.1) (test for subgroup
11 differences p=0.81), 49-54 (Fig. A7.2) (test for subgroup differences p=0.54) and ≥55 hours/week
12 (A7.3) compared to 35-40 hours/week (test for subgroup differences p=0.48).

13
14 *Fig A7.1 Explorative sensitivity analysis, Acquired depression (standard exposure categories versus*
15 *approximated exposure categories), worked 41-48 hours/week compared with worked 35-40*
16 *hours/week*
17



18 Standard exposure categories: 41-48 hours/week compared to 35-40 hours/week.
19 None-standard exposure categories: >40 hours/week compared to 35-40 hours/week (Shields 1999,
20 Men; Shields 1999, Women) and 45 hours/week (9 hours/day) compared to 35-40 hours/week (7-8
21 hours/day) (Virtanen 2012)
22
23
24
25

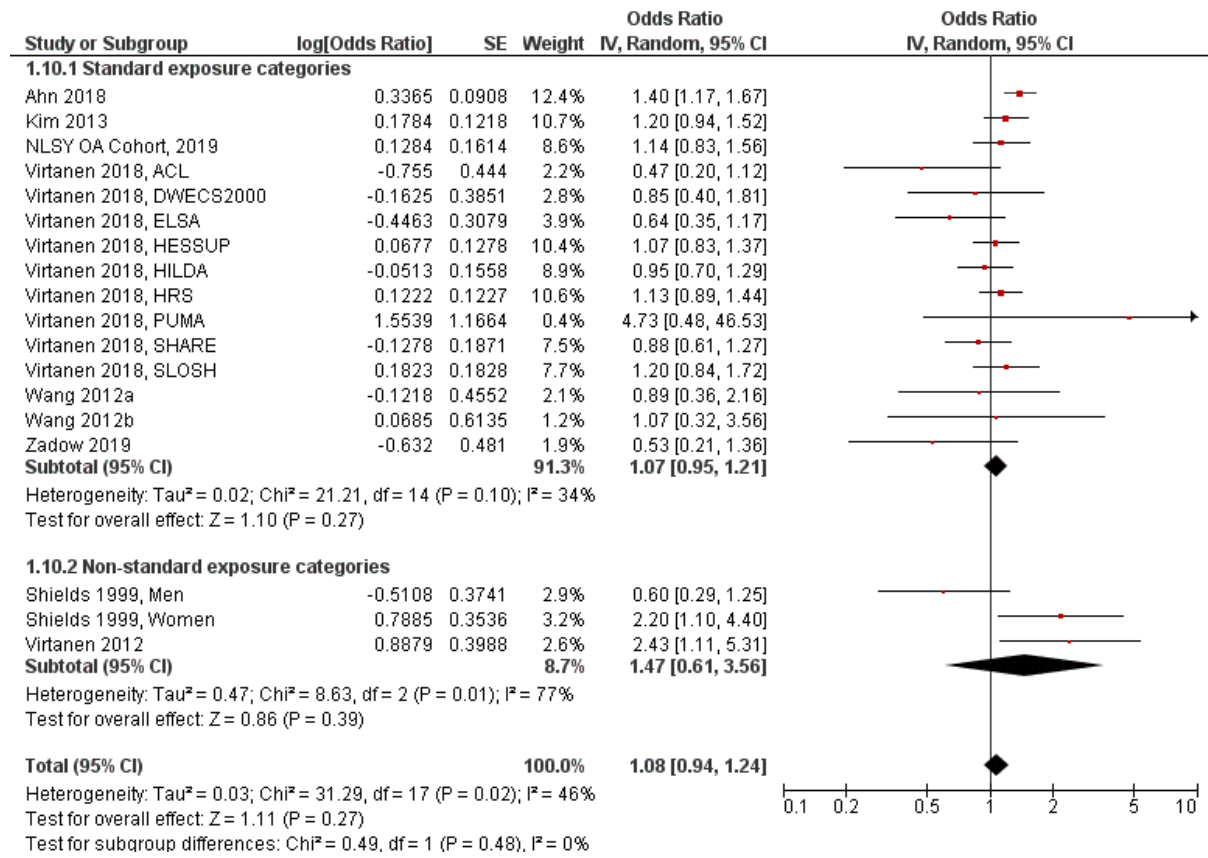
1 *Fig A7.2 Explorative sensitivity analysis, Acquired depression (standard exposure categories versus*
 2 *approximated exposure categories), worked 49-54 hours/week compared with worked 35-40*
 3 *hours/week*
 4



5
 6
 7 Standard exposure categories: 49-54 hours/week compared to 35-40 hours/week.
 8 None-standard exposure categories: >40 hours/week compared to 35-40 hours/week (Shields 1999,
 9 Men; Shields 1999, Women) and 50 hours/week (10 hours/day) compared to 35-40 hours/week (7-8
 10 hours/day) (Virtanen 2012)
 11
 12

1
2
3
4

Fig A7.3 Explorative sensitivity analysis, Acquired depression (standard exposure categories versus approximated exposure categories), worked ≥ 55 hours/week compared with worked 35-40 hours/week



5
6
7
8
9
10
11

Standard exposure categories: ≥ 55 hours/week compared to 35-40 hours/week.
 None-standard exposure categories: >40 hours/week compared to 35-40 hours/week (Shields 1999, Men; Shields 1999, Women) and 55-60 hours/week (11-12 hours/day) compared to 35-40 hours/week (7-8 hours/day) (Virtanen 2012)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Appendix 8 Additional subgroup analyses

Additional subgroup analyses that were outlined in the protocol pertained to calculate estimates by

- WHO region
- Sex
- Age groups
- SES
- Industry
- Occupational Group
- Formality of economy

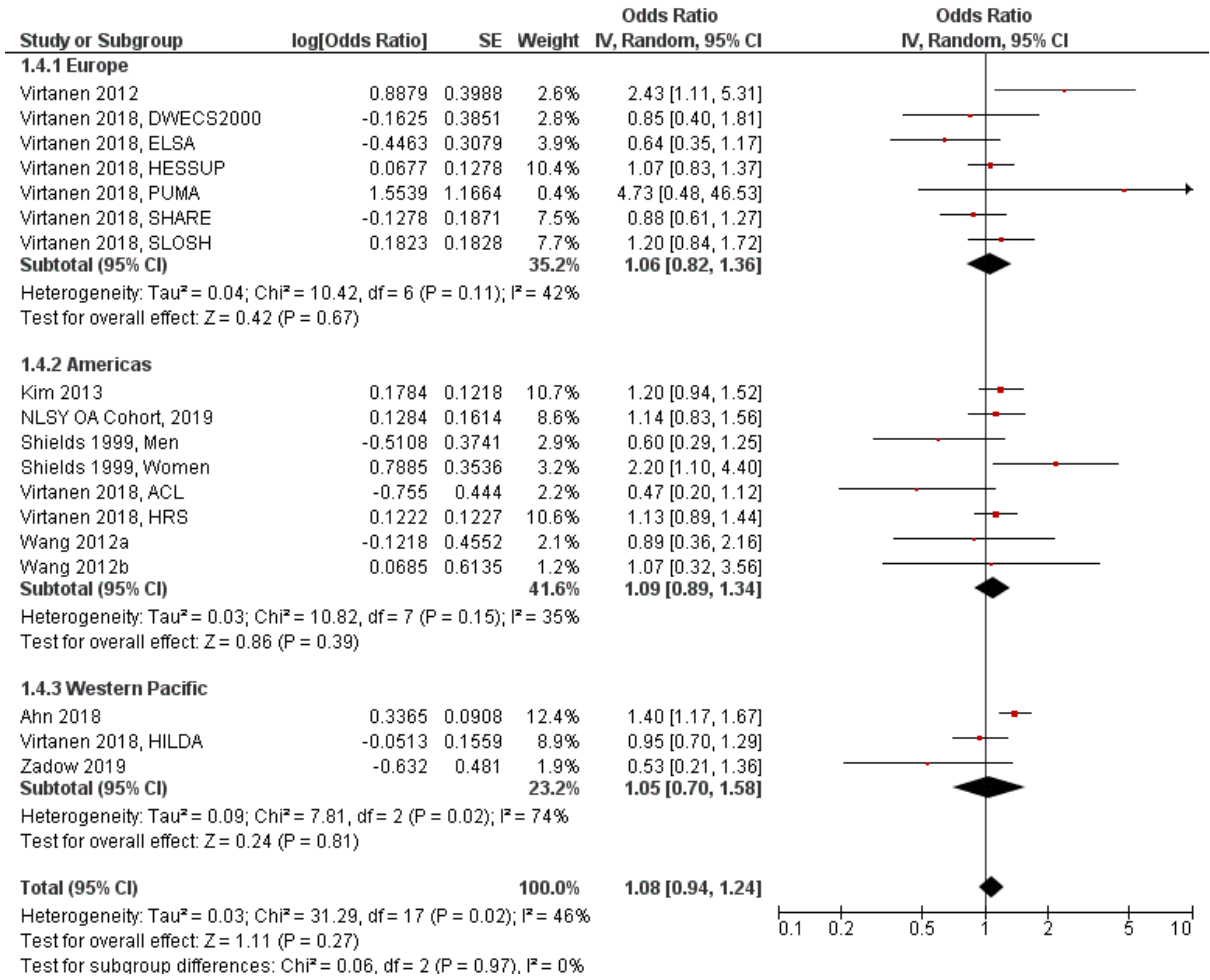
We conducted the subgroup analyses for the comparison of ≥ 55 hours/week versus 35-40 hours per week.

1
2
3
4
5
6
7
8
9
10
11
12
13
14

A8.1 By WHO region

We did not find a statistical significant difference between the three WHO regions under study (test for subgroup differences $p=0.97$) (Fig. A8.1).

Fig. A8.1 Subgroup analysis by WHO region, Acquired depression, worked ≥ 55 hours/week compared with worked 35-40 hours/week

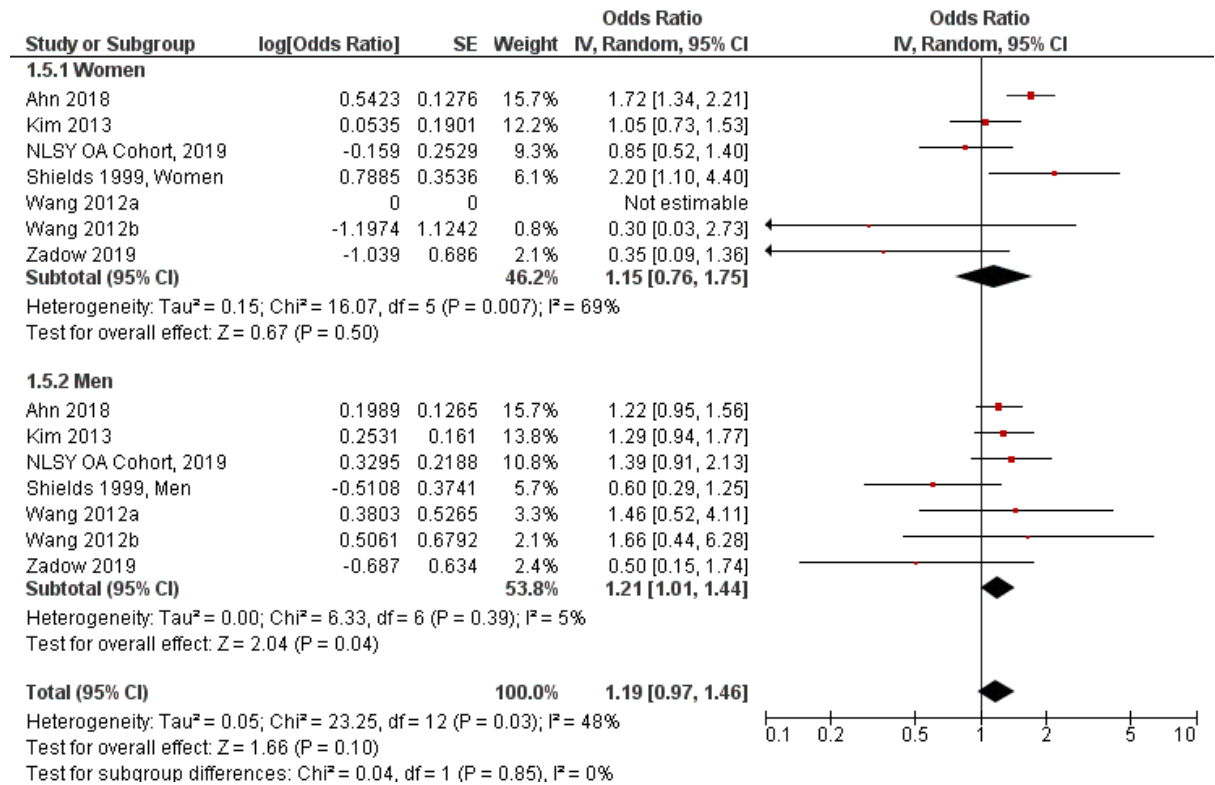


1
2
3
4
5
6
7
8
9

A8.2 By sex

We did not find a statistical significant difference between women and men (test for subgroup differences $p=0.85$) (Fig. A8.2).

Fig. A8.2 Subgroup analysis by sex, Acquired depression, worked ≥ 55 hours/week compared with worked 35-40 hours/week



10
11

1

2 **A.8.3. By age group**

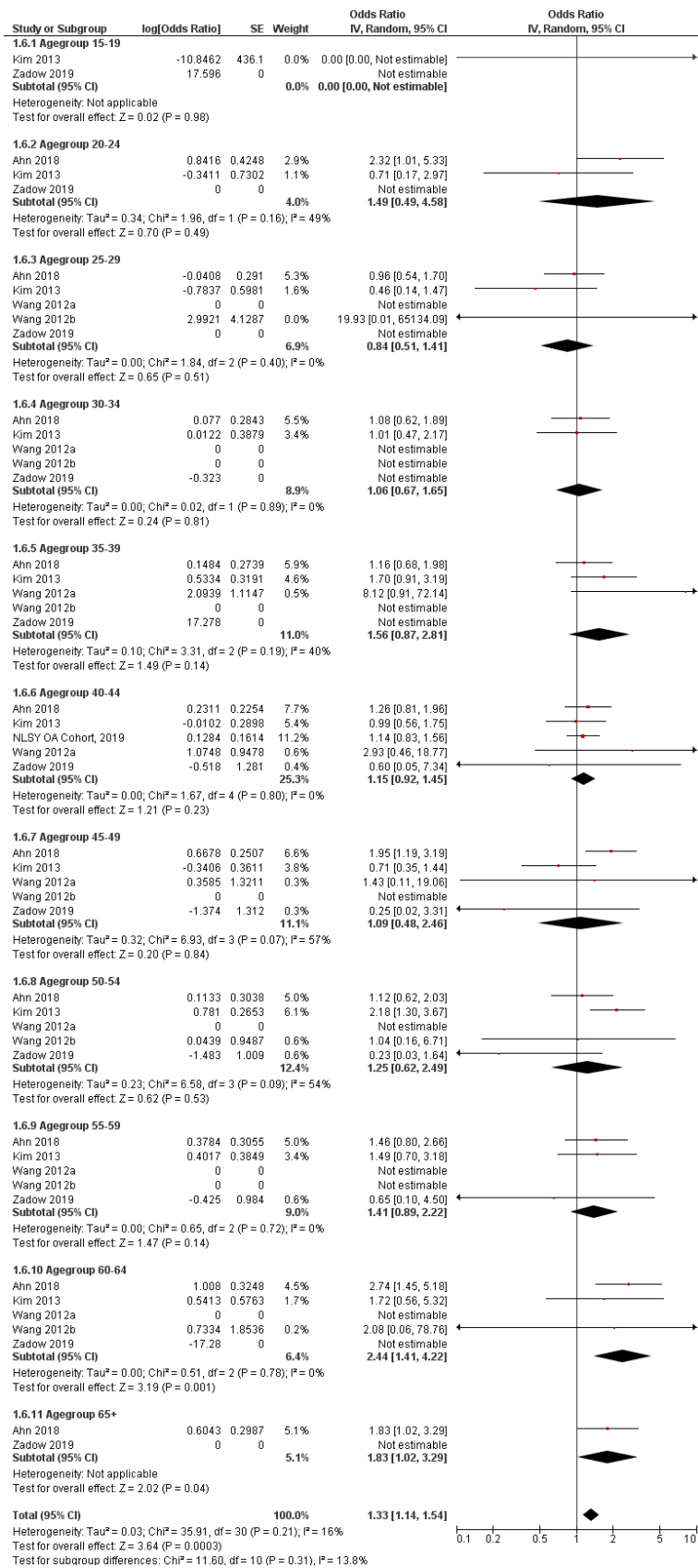
3

4 We did not find a statistical significant difference between the age groups under study (test for
5 subgroup differences $p=0.31$) (Fig. A8.3).

6

1 *Fig. A8.3 Subgroup analysis by age groups, Acquired depression, worked ≥ 55 hours/week compared*
 2 *with worked 35-40 hours/week*

3



4

5

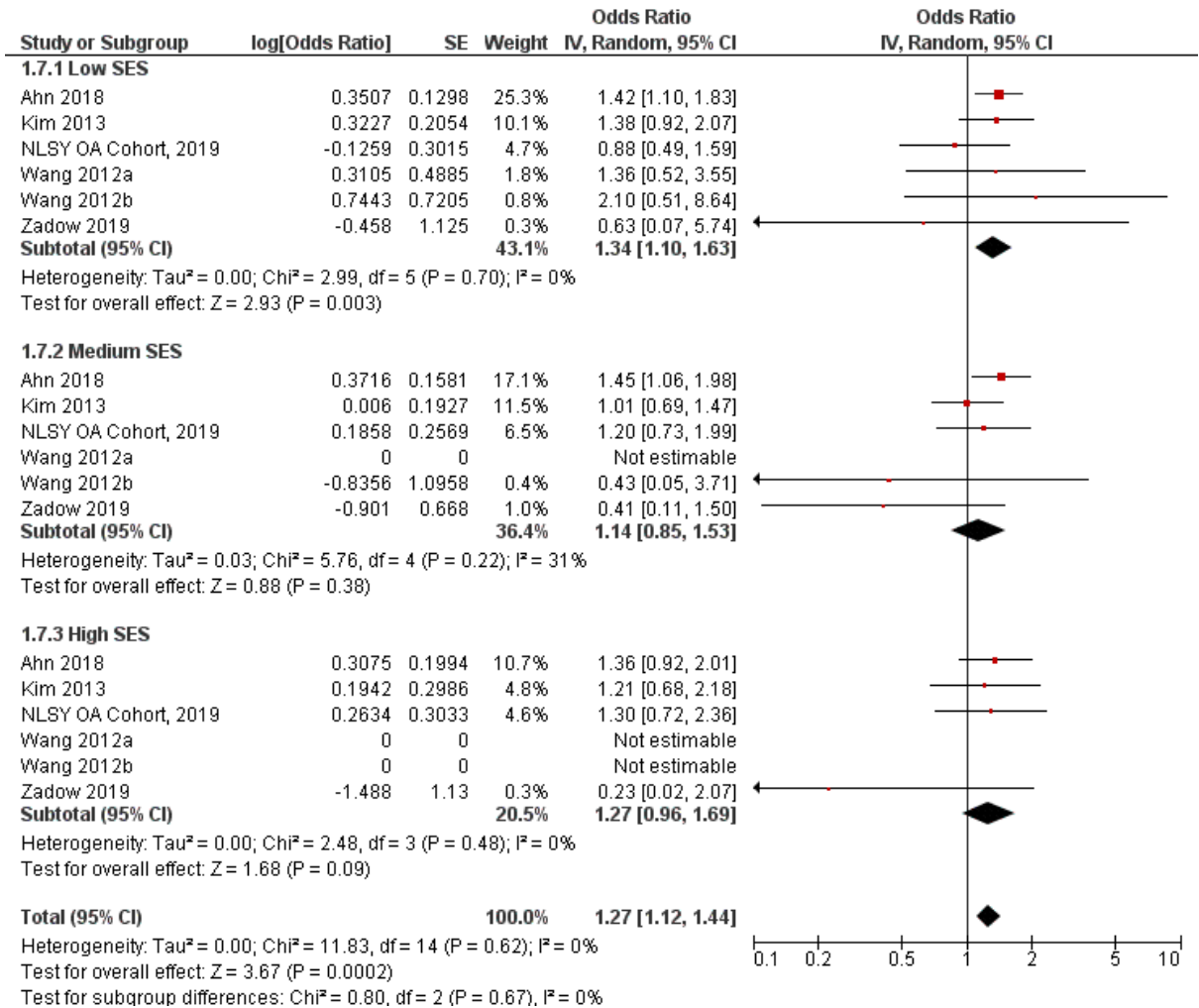
1
2
3
4
5
6
7
8
9

10
11
12
13

A8.4. By SES

We did not find a statistical significant difference between the three SES groups under study (test for subgroup differences $p=0.67$) (Fig. A8.4).

Fig. A8.4 Subgroup analysis by SES, Acquired depression, worked ≥ 55 hours/week compared with worked 35-40 hours/week



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

A8.5. By industrial sector

No studies provided effect estimates disaggregated by industrial sector, and we could therefore not assess differences in effect estimates by industrial sector.

A8.6. By occupation

No studies provided effect estimates disaggregated by occupation, and we could therefore not assess differences in effect estimates by occupation.

A8.7. By formality of economy

No studies provided effect estimates disaggregated by formality of economy, and we could therefore not assess differences in effect estimates by formality of economy.

1 *Appendix 9 Pre-defined sensitivity analyses for type of depression assessment*
2 *and risk of bias*

3

4

5 Sensitivity analyses were conducted to compare pooled estimates for studies

- 6 • that assessed depression by a clinical diagnostic interview versus other assessment methods
- 7 • with “low”/”probably low” risk of bias versus “high”/”probably high” risk of bias

8

9 We conducted the sensitivity analyses for the comparison of ≥ 55 hours/week versus 35-40 hours per
10 week.

11

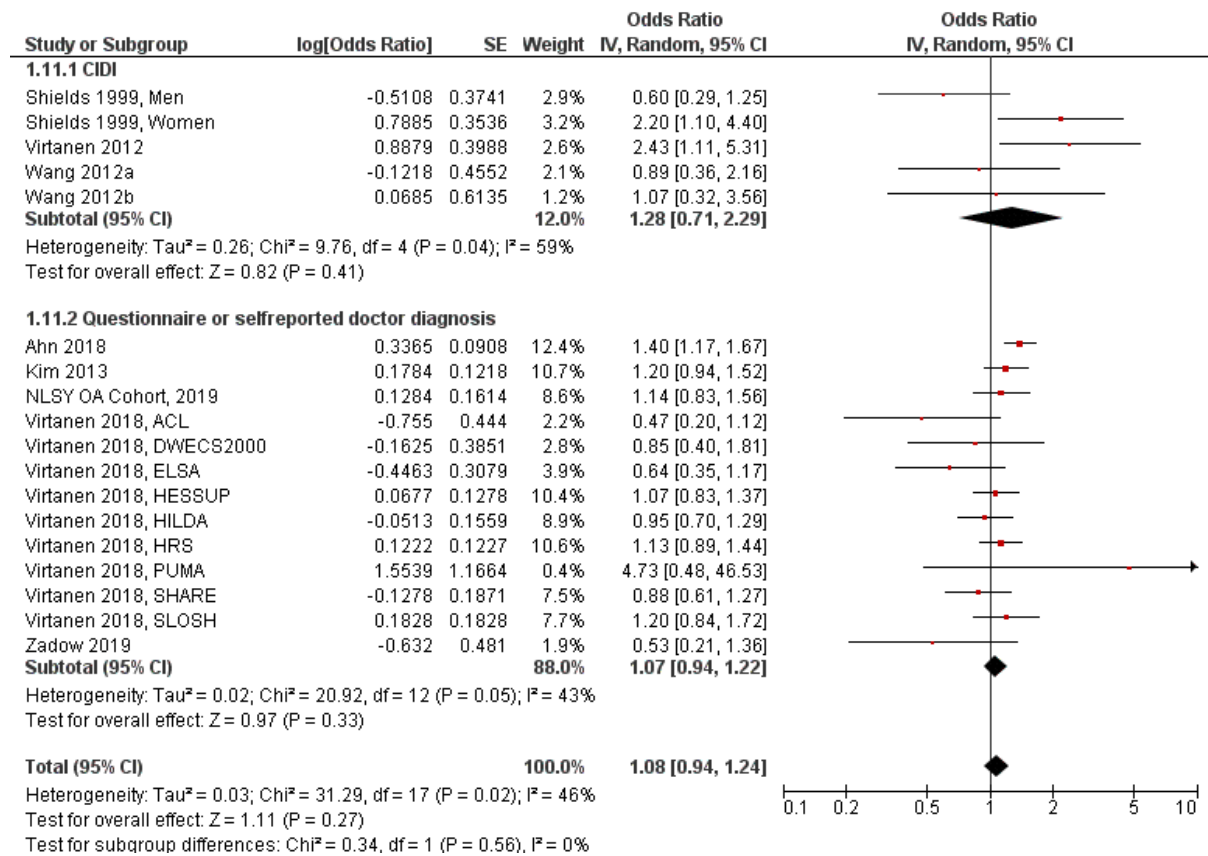
12

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

A9.1. Studies with a clinical diagnostic interview versus studies that used other assessment methods

There were similar estimates for the four studies using a clinical diagnostic interview and the 13 studies using other assessment methods (11 studies using a validated self-administered rating scale, 1 study using a comprehensive interview to assess previous and current diagnosed and treated depression and 1 study using one question on doctor-diagnosed depression), test for subgroup differences $p=0.56$ (Fig. A9.1).

Fig. A9.1 Pre-defined sensitivity analysis, Acquired depression (clinical diagnostic interview vs. "other assessment methods), worked ≥ 55 hours/week compared with worked 35-40 hours/week

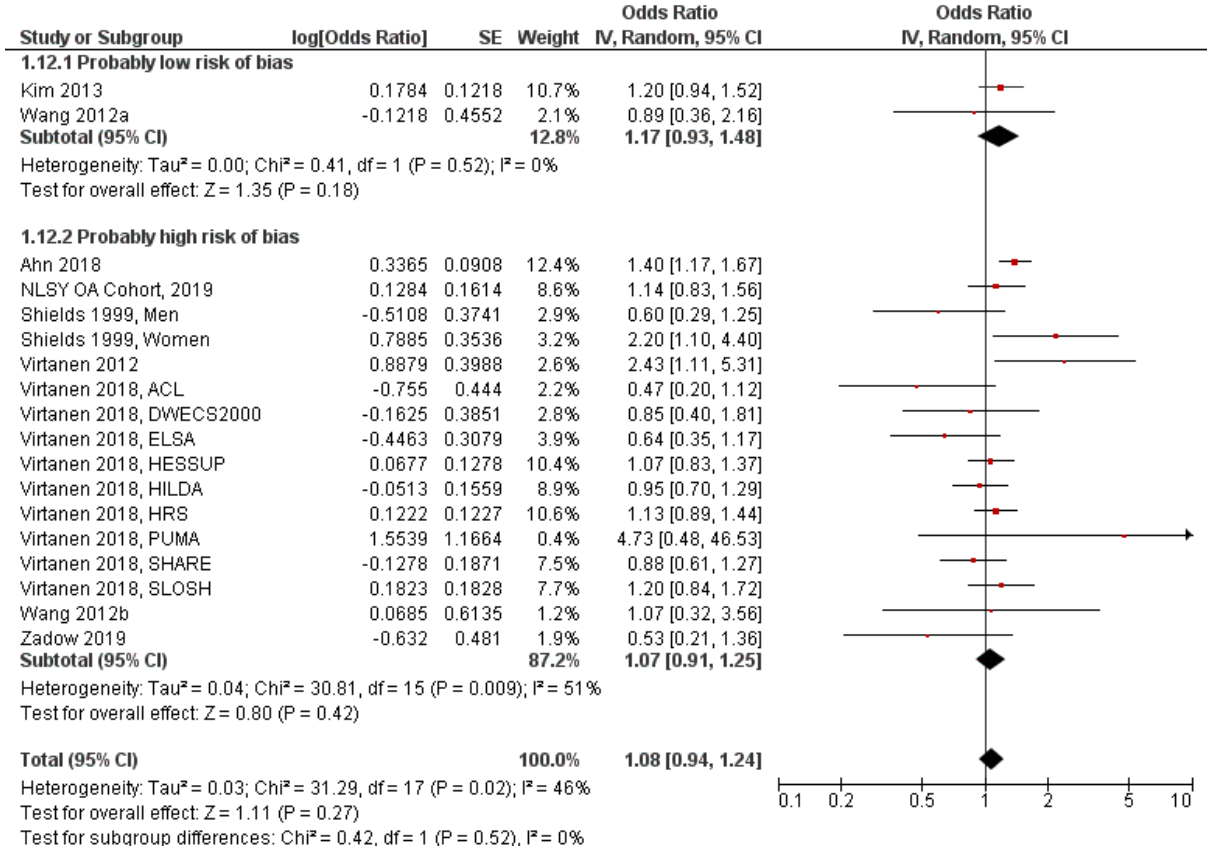


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

A9.2. Studies with low/probably low risk of bias versus studies with high/probably high risk of bias

There were no noticeable differences between the two studies with “low”/“probably low” risk of bias in all RoB domains and the 11 studies with at least one rating of “high” or “probably high” in any RoB domain s “ (test for subgroup differences p=0.52) (Fig. A9.2).

Fig A9.2 Pre-defined sensitivity analysis, Acquired depression (“low”/“probably low” risk of bias vs. “high”/“probably high” risk of bias), worked ≥55 hours/week compared with worked 35-40 hours/week



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36

Appendix 10. Supplementary information on strength of evidence: Bradford Hill considerations

Given the restriction to observational epidemiologic investigations the standard Navigation Guide methodology (Lam et al. 2016c) to rate the strength of the evidence could only partially be applied (see 4.7). The following remarks provide additional information according to relevant considerations proposed by Bradford Hill.

Temporal sequence: Of the 22 studies, 20 studies met this consideration, as was these studies had identified and subsequently excluded participants with prevalent depression at baseline.

Strength of association: Overall, the pooled estimates were weak (1.05, 1.06, 1.08, respectively) for all three comparisons (41-48, 49-54, ≥ 55 hours/week compared to 35-40 hours/week) and the confidence intervals included unity in all comparisons. Thus, the consideration “strengths of association” was not met.

Consistency of associations: Although heterogeneity, determined by I^2 , was only moderate in the three main meta-analyses, the study-specific estimates showed inconsistency. Of the 18 estimates (from 17 studies) in the meta-analysis of working hours ≥ 55 hours/week compared to 35-40 hours/week, eight estimates indicated that long working hours were associated with a lower risk of depression (with point estimates ranging from 0.47 to 0.95) whereas 10 estimates indicated that long working hours were associated with a higher risk of depression (with point estimates ranging from 1.07 to 4.73). Thus, the consideration “consistency of association” was not met.

Dose-response relationship: The pooled odds ratios from the meta-analyses of 41-48 hours/week, 49-54 hours/week and ≥ 55 hours week compared to 35-40 hours were 1.05, 1.06, 1.08, respectively. As the confidence intervals overlapped and included unity in all analyses, we do not regard this as a convincing indication of a dose-response relationship.

Confounding: All estimates in the meta-analyses were adjusted for all pre-defined tier 1 confounders, that is sex, age and a measure of SES. Further, some estimates were additionally adjusted for other potential confounders, such as industry, job group, marital status or children at home. With the exception of one study, estimate were not adjusted for variables that we considered potential mediators. Thus, the consideration “confounding” was met.

1 **Biological plausibility:** There are several plausible mechanisms, both psycho-emotional, health-
2 behaviour related and psycho-physiological that may explain an association between long working
3 hours and risk of depression (see Fig. 1 in section 1.4). However, to our knowledge, research has not
4 advanced yet to test these mechanisms in either prospective cohort studies or experimental studies.
5 Thus, although the mechanisms are plausible, there is no strong evidence yet, that these mechanisms
6 may explain a possible association between long working hours and risk of depression and therefore
7 we conclude that the consideration “biological plausibility” was not met.

8
9
10

1 **PRISMA checklist**

2

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	4
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	7-9
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	10-12
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	10-15
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	17
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	20-23
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	18-20
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 1, page 1-13
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	20

Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	23-24
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	23-24
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	24-25
Summary measures	13	State the principal summary measures (e.g., RR, difference in means).	25-26
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	25-26
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	27
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	26-27
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	31-33
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	34-47
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	50-59
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and CIs, ideally with a forest plot.	59-65
Synthesis of results	21	Present results of each meta-analysis done, including CIs and measures of consistency.	59-65
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	68-70
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	65-68
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	71-77

Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	78-80
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	80-81
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	82

1