Cochrane Library

Cochrane Database of Systematic Reviews

Adapting shift work schedules for sleep quality, sleep duration, and sleepiness in shift workers (Review)

Hulsegge G, Coenen P, Gascon GM, Pahwa M, Greiner B, Bohane C, Wong IS, Liira J, Riera R, Pachito DV

Hulsegge G, Coenen P, Gascon GM, Pahwa M, Greiner B, Bohane C, Wong IS, Liira J, Riera R, Pachito DV.
Adapting shift work schedules for sleep quality, sleep duration, and sleepiness in shift workers.
Cochrane Database of Systematic Reviews 2023, Issue 9. Art. No.: CD010639.
DOI: 10.1002/14651858.CD010639.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Intervention Review]

Adapting shift work schedules for sleep quality, sleep duration, and sleepiness in shift workers

Gerben Hulsegge[1], Pieter Coenen[2], Gregg M Gascon[3,4], Manisha Pahwa[5,6], Birgit Greiner[7], Ciarán Bohane[8], Imelda S Wong[9], Juha Liira[10], Rachel Riera[11,12,13], Daniela V Pachito[14]

[1]The Netherlands Organization for Applied Scientific Research, TNO, Leiden, Netherlands. [2]Department of Public and Occupational Health, Amsterdam Public Health Research Institute, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. [3]OhioHealth, Columbus, Ohio, USA. [4]Biomedical Informatics, College of Medicine, The Ohio State University, Columbus, Ohio, USA. [5]Occupational Cancer Research Centre, Ontario Health, Toronto, Canada. [6]Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Canada. [7]School of Public Health, University College Cork, Cork, Ireland. [8]Medmark Occupational Healthcare, Limerick, Ireland. [9]Division of Science Integration, National Institute for Occupational Safety and Health, Cincinnati, OH, USA. [10]Department of Occupational Health, University of Turku, Turku, Finland. [11]Cochrane Brazil Rio de Janeiro, Cochrane, Petrópolis, Brazil. [12]Center of Health Technology Assessment, Hospital Sírio-Libanês, São Paulo, Brazil. [13]Núcleo de Ensino e Pesquisa em Saúde Baseada em Evidência, Avaliação Tecnológica e Ensino em Saúde (NEP-Sbeats), Universidade Federal de São Paulo, São Paulo, Brazil. [14]Prossono Centro de Diagnóstico e Medicina do Sono, Ribeirão Preto, São Paulo, Brazil

Contact: Gerben Hulsegge, gerben.hulsegge@tno.nl.

Editorial group: Cochrane Work Group.
Publication status and date: New, published in Issue 9, 2023.

Citation: Hulsegge G, Coenen P, Gascon GM, Pahwa M, Greiner B, Bohane C, Wong IS, Liira J, Riera R, Pachito DV. Adapting shift work schedules for sleep quality, sleep duration, and sleepiness in shift workers. *Cochrane Database of Systematic Reviews* 2023, Issue 9. Art. No.: CD010639. DOI: 10.1002/14651858.CD010639.pub2.

ABSTRACT

Background

Shift work is associated with insufficient sleep, which can compromise worker alertness with ultimate effects on occupational health and safety. Adapting shift work schedules may reduce adverse occupational outcomes.

Objectives

To assess the effects of shift schedule adaptation on sleep quality, sleep duration, and sleepiness among shift workers.

Search methods

We searched CENTRAL, PubMed, Embase, and eight other databases on 13 December 2020, and again on 20 April 2022, applying no language restrictions.

Selection criteria

We included randomised controlled trials (RCTs) and non-RCTs, including controlled before-after (CBA) trials, interrupted time series, and cross-over trials. Eligible trials evaluated any of the following shift schedule components.

- Permanency of shifts
- Regularity of shift changes
- Direction of shift rotation
- Speed of rotation
- Shift duration
- Timing of start of shifts

• Distribution of shift schedule
• Time off between shifts
• Split shifts
• Protected sleep
• Worker participation

We included studies that assessed sleep quality off-shift, sleep duration off-shift, or sleepiness during shifts.

Data collection and analysis

Two review authors independently screened the titles and abstracts of the records recovered by the search, read through the full-text articles of potentially eligible studies, and extracted data. We assessed the risk of bias of included studies using the Cochrane risk of bias tool, with specific additional domains for non-randomised and cluster-randomised studies. For all stages, we resolved any disagreements by consulting a third review author. We presented the results by study design and combined clinically homogeneous studies in meta-analyses using random-effects models. We assessed the certainty of the evidence with GRADE.

Main results

We included 11 studies with a total of 2125 participants. One study was conducted in a laboratory setting and was not considered for drawing conclusions on intervention effects. The included studies investigated different and often multiple changes to shift schedule, and were heterogeneous with respect to outcome measurement.

Forward versus backward rotation

Three CBA trials (561 participants) investigated the effects of forward rotation versus backward rotation. Only one CBA trial provided sufficient data for the quantitative analysis; it provided very low-certainty evidence that forward rotation compared with backward rotation did not affect sleep quality measured with the Basic Nordic Sleep Questionnaire (BNSQ; mean difference (MD) −0.20 points, 95% confidence interval (CI) −2.28 to 1.89; 62 participants) or sleep duration off-shift (MD −0.21 hours, 95% CI −3.29 to 2.88; 62 participants). However, there was also very low-certainty evidence that forward rotation reduced sleepiness during shifts measured with the BNSQ (MD −1.24 points, 95% CI −2.24 to −0.24; 62 participants).

Faster versus slower rotation

Two CBA trials and one non-randomised cross-over trial (341 participants) evaluated faster versus slower shift rotation. We were able to meta-analyse data from two studies. There was low-certainty evidence of no difference in sleep quality off-shift (standardised mean difference (SMD) −0.01, 95% CI −0.26 to 0.23) and very low-certainty evidence that faster shift rotation reduced sleep duration off-shift (SMD −0.26, 95% CI −0.51 to −0.01; 2 studies, 282 participants). The SMD for sleep duration translated to an MD of 0.38 hours' less sleep per day (95% CI −0.74 to −0.01). One study provided very low-certainty evidence that faster rotations decreased sleepiness during shifts measured with the BNSQ (MD −1.24 points, 95% CI −2.24 to −0.24; 62 participants).

Limited shift duration (16 hours) versus unlimited shift duration

Two RCTs (760 participants) evaluated 80-hour workweeks with maximum daily shift duration of 16 hours versus workweeks without any daily shift duration limits. There was low-certainty evidence that the 16-hour limit increased sleep duration off-shift (SMD 0.50, 95% CI 0.21 to 0.78; which translated to an MD of 0.73 hours' more sleep per day, 95% CI 0.30 to 1.13; 2 RCTs, 760 participants) and moderate-certainty evidence that the 16-hour limit reduced sleepiness during shifts, measured with the Karolinska Sleepiness Scale (SMD −0.29, 95% CI −0.44 to −0.14; which translated to an MD of 0.37 fewer points, 95% CI −0.55 to −0.17; 2 RCTs, 716 participants).

Shorter versus longer shifts

One RCT, one CBA trial, and one non-randomised cross-over trial (692 participants) evaluated shorter shift duration (eight to 10 hours) versus longer shift duration (two to three hours longer). There was very low-certainty evidence of no difference in sleep quality (SMD −0.23, 95% CI −0.61 to 0.15; which translated to an MD of 0.13 points lower on a scale of 1 to 5; 2 studies, 111 participants) or sleep duration off-shift (SMD 0.18, 95% CI −0.17 to 0.54; which translated to an MD of 0.26 hours' less sleep per day; 2 studies, 121 participants). The RCT and the non-randomised cross-over study found that shorter shifts reduced sleepiness during shifts, while the CBA study found no effect on sleepiness.

More compressed versus more spread out shift schedules

One RCT and one CBA trial (346 participants) evaluated more compressed versus more spread out shift schedules. The CBA trial provided very low-certainty evidence of no difference between the groups in sleep quality off-shift (MD 0.31 points, 95% CI −0.53 to 1.15) and sleep duration off-shift (MD 0.52 hours, 95% CI −0.52 to 1.56).

Authors' conclusions

Forward and faster rotation may reduce sleepiness during shifts, and may make no difference to sleep quality, but the evidence is very uncertain. Very low-certainty evidence indicated that sleep duration off-shift decreases with faster rotation. Low-certainty evidence indicated that on-duty workweeks with shift duration limited to 16 hours increases sleep duration, with moderate-certainty evidence for minimal reductions in sleepiness. Changes in shift duration and compression of workweeks had no effect on sleep or sleepiness, but the evidence was of very low-certainty. No evidence is available for other shift schedule changes. There is a need for more high-quality studies (preferably RCTs) for all shift schedule interventions to draw conclusions on the effects of shift schedule adaptations on sleep and sleepiness in shift workers.

PLAIN LANGUAGE SUMMARY

Changing shift worker's schedules to improve sleep quality and duration and reduce sleepiness

Key messages

- There is limited evidence that changes in shift schedules improve sleep quality, increase sleep duration, or reduce sleepiness.
- More studies are needed to draw stronger conclusions about shift schedule changes on sleep and sleepiness.

What can be done to improve shift workers' sleep?

Shift work often leads to insufficient sleep that can compromise worker alertness, with ultimate effects on health and work safety. Changing shift work schedules is one method that may reduce the unwanted effects of shift work.

What did we want to find out?

We wanted to find out which shift schedule adaptations improve sleep on rest days and reduce sleepiness at work.

What did we do?

We searched for studies that evaluated the following features of shift schedules.

- Whether shift schedules changed (rotated) or stayed the same
- Whether shift changes were regular or irregular
- Direction of shift rotation (morning to afternoon to night or night to afternoon to morning)
- Speed of rotation
- Shift duration
- Timing of start of shifts
- Distribution of shift schedule (fewer shifts with more hours or more shifts with fewer hours)
- Time off between shifts
- Split (interrupted) shifts
- Whether workers had on-call shifts
- Whether workers were involved in organising the shift schedule

What did we find?

We included 11 studies, with 2125 participants. One study was conducted in a laboratory; we disregarded the results of this study when drawing conclusions. Most studies investigated a change in one feature of the shift schedule, while some investigated changes in two features. Four studies investigated the effect of changes in direction of shift rotation, three studies speed of rotation, five studies changes in shift duration, and one study changes in the distribution of days off.

Forward rotation compared to backward rotation may have no effect on sleep duration or sleep quality on rest days, but may reduce sleepiness at work. However, all of these results are very uncertain.

Faster shift rotation compared to slower shift rotation may have no effect on sleep quality on rest days. Faster rotation may reduce sleep duration on rest days, but may also reduce sleepiness at work; however, the evidence for both results is very uncertain.

Two studies investigated 80-hour workweeks among doctors. They found that a schedule with shifts of no more than 16 hours, compared with a schedule with unlimited shift duration (including shifts of 24 to 28 hours), may increase sleep duration on rest days and probably results in a slight reduction in sleepiness at work.

Shorter shift duration (eight or 10 hours) compared to longer shift duration (two to three hours longer) may have no effect on sleep quality or sleep duration on rest days, but the results are very uncertain. The effects of shift duration on sleepiness differed across studies.

Changes in the distribution shift schedules (e.g. two days versus four days off in a row) may have no effect on sleep quality or sleep duration on rest days, but the results are very uncertain.

We found no studies investigating other changes in shift schedules.

Overall, there is a need for more high-quality studies to draw firm conclusions on the effects of shift schedule changes on sleep and sleepiness. Currently, we cannot draw useful conclusions from the available evidence.

Main limitations of the evidence

Too few of the included studies allocated workers to the schedule change at random. In addition, many studies included few workers and lacked reliable measurements of sleep and sleepiness.

How up-to-date is this review?

The evidence is up-to-date to 13 December 2020.

Summary of findings 1. Summary of findings table - Forward rotation compared to backward rotation for improving sleep and reducing sleepiness among shift workers

Forward rotation compared to backward rotation for improving sleep and reducing sleepiness among shift workers

Patient or population: shift workers
Setting: airline company
Intervention: forward rotation
Comparison: backward rotation

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with backward rotation	Risk with forward rotation				
Sleep quality off-shift Assessed with: BNSQ	The mean sleep quality off-shift was **2.45** points	**MD0.2 points lower** (0.28 lower to 1.89 higher)	—	62 (1 non-randomised trial)	⊕⊖⊖⊖ **Very low**[a]	The evidence is very uncertain about the effect of forward rotation compared to backward rotation on sleep quality off-shift.
Sleep length off-shift	The mean sleep length off-shift was **7.36** hours	**MD0.21 hours fewer** (3.29 fewer to 2.88 more)	—	62 (1 non-randomised trial)	⊕⊖⊖⊖ **Very low**[a,b]	The evidence is very uncertain about the effect of forward rotation compared to backward rotation on sleep length off-shift.
Sleepiness during shifts Assessed with: BNSQ	The mean sleepiness during shifts was **1.90** points	**MD1.24 points lower** (2.24 lower to 0.24 lower)	—	62 (1 non-randomised trial)	⊕⊖⊖⊖ **Very low**[a,b]	The evidence is very uncertain about the effect of forward rotation compared to backward rotation on sleepiness during shifts.
Secondary outcomes (number of staff; number of hours worked; overtime; staff costs)	—	—	—	—	—	No studies evaluated these outcomes.

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

BNSQ: Basic Nordic Sleep Questionnaire; **CI:** confidence interval; **MD:** mean difference.

GRADE Working Group grades of evidence
High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.
Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.
Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

[a] With the GRADE approach, the initial level of evidence from non-randomised studies is low-certainty. We downgraded the certainty of evidence by one level (to very low) for imprecision due to small sample size (62 participants).
[b] We would have downgraded the certainty of evidence by one more level for high risk of bias due to the subjective measurement of this outcome.

Summary of findings 2. Summary of findings table - Faster rotation compared to slower rotation for shift workers

Faster rotation compared to slower rotation for shift workers

Patient or population: shift workers
Setting: police departments and airline company
Intervention: faster rotation
Comparison: slower rotation

Outcomes	Anticipated absolute effects*(95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with slower rotation	Risk with faster rotation				
Sleep quality off-shift Assessed with: KSQ and BNSQ	—	**SMD 0.01 SD lower** (0.26 lower to 0.23 higher)	—	282 (2 non-randomised trials)	⊕⊕⊖⊖ **Low**[a]	The SMD calculated back to −0.04 points (95% CI −1.09 to 0.96) on the BNSQ. Faster change may not increase sleep quality off-shift when compared to slower change.
Sleep length off-shift	—	**SMD 0.26 SD lower** (0.51 lower to 0.01 lower)	—	282 (2 non-randomised trials)	⊕⊖⊖⊖ **Very low**[b]	The SMD calculated back to an MD of −0.38 hours per day (95% CI −0.74 to −0.01). The evidence is very uncertain about the effect of faster change compared to slower change on sleep length off-shift.
Sleepiness during shifts Assessed with: BNSQ	The mean sleepiness during shifts was **1.90** points	**MD 1.24 points lower** (2.24 lower to 0.24 lower)	—	62 (1 non-randomised trial)	⊕⊖⊖⊖ **Very low**[c]	The evidence is very uncertain about the effect of faster change compared to slower change on sleepiness during shifts due to the absence of randomised controlled trials.
Secondary outcomes (number	—	—	—	—	—	No studies evaluated these outcomes.

of staff; number of hours worked; overtime; staff costs)						

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
BNSQ: Basic Nordic Sleep Questionnaire;**CI:** confidence interval; **KSQ:** Karolinska Sleep Questionnaire; **MD:** mean difference; **SD:** standard deviation; **SMD:** standardised mean difference.

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

[a] With the GRADE approach, the initial level of evidence from non-randomised studies is low-certainty.
[b] With the GRADE approach, the initial level of evidence from non-randomised studies is low-certainty. We downgraded the certainty of evidence by one level (to very low) for inconsistency (i.e. one study found an effect, while the other did not).
[c] With the GRADE approach, the initial level of evidence from non-randomised studies is low-certainty. We downgraded the certainty of evidence by one level (to very low) for high risk of bias due to the subjective measurement of sleepiness. We would have downgraded the certainty of evidence by one more level for imprecision (small sample size: 62 participants).

Summary of findings 3. Summary of findings table - Shift duration of no more than 16 hours compared to shift duration of 24 to 28 hours for improving sleep and reducing sleepiness among shift workers

Shift duration of no more than 16 hours compared to shift duration of 24 to 28 hours for improving sleep and reducing sleepiness among shift workers

Patient or population: shift workers
Setting: hospital wards and intensive care units
Intervention: shift duration of no more than 16 hours
Comparison: shift duration of 24 to 28 hours

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with shift duration of 24 to 28 hours	Risk with shift duration of no more than 16 hours				
Sleep quality off-shift	—	—	—	—	—	No studies evaluated this outcome.

Sleep length off-shift	—	**SMD 0.5 SD higher** (0.21 higher to 0.78 higher)	—	760 (2 RCTs)	⊕⊕⊖⊖ **Low**[a,b]	The SMD calculated back to an MD of 0.73 hours per day (95% CI 0.30 to 1.13), which is a small but clinically relevant effect: shifts of no more than 16 hours may result in a slight increase in sleep length off-shift compared to 24- to 28-hour shifts.
Sleepiness during shifts Assessed with: KSS	—	**SMD 0.29 SD lower** (0.44 lower to 0.14 lower)	—	716 (2 RCTs)	⊕⊕⊕⊖ **Moderate**[b]	The SMD calculated back to an MD of −0.37 (95% CI −0.55 to −0.17) on the KSS, which is a small but clinically irrelevant effect.
Number of hours worked	The mean number of hours worked was **68.4** hours per week	**MD 6.5 hours per week fewer** (7.89 fewer to 5.11 fewer)	—	318 (1 RCT)	⊕⊖⊖⊖ **Very low**[c]	The evidence is very uncertain about the effect of on-duty shifts of no more than 16 hours compared to 24- to 28-hour shifts on work hours.
Other secondary outcomes (number of staff; overtime; staff costs)	—	—	—	—	—	No studies evaluated these outcomes.

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

CI: confidence interval; **KSS:** Karolinska Sleepiness Scale; **MD:** mean difference; **SD:** standard deviation; **SMD:** standardised mean difference.

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

[a] Downgraded one level for inconsistency (I^2 value suggested moderate heterogeneity, although both studies showed the same direction of effect).
[b] Downgraded one level for indirectness (the intervention included changes in the shift schedule in addition to maximum shift length).
[c] Downgraded two levels for indirectness (the intervention included changes in the shift schedule in addition to maximum shift length, and the effect of these on-duty interventions on work hours will be different in each context) and by one level for imprecision (small sample size: 318 participants).

Summary of findings 4. Summary of findings table - Shorter shifts (8 or 10 hours) compared to shifts lasting 2 to 3 hours longer for improving sleep and reducing sleepiness among shift workers

Shorter shifts (8 or 10 hours) compared to shifts lasting 2 to 3 hours longer for improving sleep and reducing sleepiness among shift workers

Patient or population: shift workers
Setting: power plant and police departments
Intervention: shorter shifts (8 or 10 hours)
Comparison: shifts lasting 2 to 3 hours longer

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with shifts lasting 2 to 3 hours longer	Risk with shorter shifts (8 or 10 hours)				
Sleep quality off-shift	—	**SMD 0.23 SD lower** (0.61 lower to 0.15 higher)	—	111 (2 non-randomised trials)	⊕⊖⊖⊖ **Very low**[a,b]	The SMD calculated back to an MD of −0.13 points (95% CI −0.34 to 0.08) on a 5-point sleep quality scale. The evidence is very uncertain about the effects of shorter shifts compared to longer shifts on sleep quality off-shift.
Sleep length off-shift	—	**SMD 0.18 SD higher** (0.17 lower to 0.54 higher)	—	121 (2 non-randomised trials)	⊕⊖⊖⊖ **Very low**[c]	The SMD calculated back to an MD of 0.26 hours per day (95% CI −0.25 to 0.78). The evidence is very uncertain about the effects of shorter shifts compared to longer shifts on sleep length off-shift.
Sleepiness during shifts Assessed with: KSQ and PVT	—	**—**	—	121 (2 non-randomised trials)	⊕⊖⊖⊖ **Very low**[a,b,d]	The evidence is very uncertain about the effects of shorter shifts compared to longer shifts on sleepiness. We were unable to meta-analyse the data owing to substantial heterogeneity. One study showed an effect of −1.06 (95% CI −1.59 to −0.52) on the KSQ, while the other study found no significant effect using a PVT test (−0.06, 95% CI −0.57 to 0.45).
Overtime	The mean overtime was **7.17** hours per week	**MD 1.22 hours per week more** (0.94 more to 1.5 more)	—	59 (1 non-randomised trial)	⊕⊖⊖⊖ **Very low**[a,e]	The evidence is very uncertain about the effect of shorter shifts compared to longer shifts on overtime.
Other secondary outcomes (number of staff; num-	—	—	—	—	—	No studies evaluated these outcomes.

ber of hours worked; staff costs)

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

CI: confidence interval; **KSQ:** Karolinska Sleep Questionnaire; **MD:** mean difference; **PVT:** psychomotor vigilance task; **SD:** standard deviation; **SMD:** standardised mean difference.

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

[a] With the GRADE approach, the initial level of evidence from non-randomised studies is low-certainty. We downgraded the certainty of evidence by one level (to very low) owing to insufficient statistical adjustment for confounding or cluster allocation.
[b] We would have downgraded the certainty of evidence by one more level for imprecision (small sample size: 111–121 participants) and wide confidence intervals that included the null effect.
[c] With the GRADE approach, the initial level of evidence from non-randomised studies is low-certainty. We downgraded the certainty of evidence by one level (to very low) owing to the subjective measurement of sleep length and insufficient statistical adjustment for confounding or cluster allocation. We would have downgraded the certainty of evidence by one more level for imprecision due to small sample size (121 participants) and wide confidence intervals.
[d] We would have downgraded the certainty of evidence by one more level for inconsistency, as one study showed a favourable effect and the other did not.
[e] We would have downgraded the certainty of evidence by one more level for indirectness, as there were other changes in the shift schedule in addition to shift length (e.g. number of shifts and change in shift length of morning and evening shifts), and by another level for imprecision caused by small sample size (i.e. 59 participants).

Summary of findings 5. Summary of findings table - More compressed schedules compared to less compressed schedules for shift workers

More compressed schedules compared to less compressed schedules for shift workers

Patient or population: shift workers
Setting: police departments
Intervention: more compressed schedules
Comparison: less compressed schedules

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with less compressed schedules	Risk with more compressed schedules				

Sleep quality off-shift assessed with: VAS	The mean sleep quality off-shift was **5.49** points	**MD 0.31 points higher** (0.53 lower to 1.15 higher)	—	25 (1 non-randomised trial)	⊕⊖⊖⊖ **Very low**[a]	The evidence is very uncertain about the effect of more compressed compared to more spread out schedules on sleep quality off-shift.
Sleep length off-shift	The mean sleep length off-shift was **7.18** hours	**MD 0.52 hours higher** (0.52 lower to 1.56 higher)	—	30 (1 non-randomised trial)	⊕⊖⊖⊖ **Very low**[a]	The evidence is very uncertain about the effect of more compressed compared to more spread out schedules on sleep length off-shift.
Sleepiness during shifts	—	—	—	—	—	No studies evaluated this outcome.
Secondary outcomes (number of staff; number of hours worked; overtime; staff costs)	—	—	—	—	—	No studies evaluated these outcomes.

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

CI: confidence interval; **MD:** mean difference;**VAS:** visual analogue scale.

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

[a] With the GRADE approach, the initial level of evidence from non-randomised studies is low-certainty. We downgraded the certainty of evidence by two levels due to incomplete outcome data and no adjustment for cluster allocation. We would have downgraded the level of evidence with two more levels due to imprecision caused by very small sample size (i.e. 25–30 participants) and wide CIs that include a null effect.

BACKGROUND

An estimated 15% to 25% of workers in Europe and North America are employed by an institution that uses some type of shift system (IARC 2010; Wong 2011). Shift work is frequently associated with circadian misalignment and self-reported poor or insufficient sleep, with adverse effects on occupational health and safety (Kecklund and Axelsson 2016; Landrigan 2004). Nonstandard schedules more than double the risk for work-related injuries and safety-critical events (Wagstaff 2011; Wong 2011). Studies have associated long-term disrupted or shortened sleep and chronic circadian disruption with gastrointestinal illness, cardiovascular diseases (Torquati 2018), impaired mental health (Torquati 2019), metabolic disorders (Gao 2020), and cancer (IARC 2020; Moreno 2019; Wu 2021).

Work schedules can be categorised by individual shifts or by sequences of shifts. Individual shifts are often defined by shift duration (e.g. eight-hour shift) or shift type, which refers to the time of the day when most of the shift occurs (e.g. morning, afternoon, or night). Sequences of shifts, or cumulative shifts in a row, can be defined by the distribution of working hours over a given number of days (i.e. fewer days with more hours or more days with fewer hours) or in terms of rotation (switching between shift types). Rotation can be further categorised by direction or speed. Forward rotation refers to the progression of day to night shifts and follows the clockwise rotation of our national biological cycles. Backwards rotation describes rotation from nights to days, and is counter to our natural circadian rhythms (Åkerstedt 2003; Knauth 1995). Rotation speed refers to how quickly the shifts rotate through the schedule. For example, two days followed by two nights is a faster rotation than a schedule that switches from days to nights on a weekly basis. The combination of all these scheduling options can affect timing and duration of sleep, with further effects on daytime sleepiness. Although there is evidence to show that shift work has negative health effects, it is necessary in some occupations, such as emergency services (IARC 2020; Moreno 2019; Torquati 2018; Torquati 2019; Wu 2021). As such, it is important to evaluate possible shift schedule alterations (within and between shifts and sequences of shifts) that could reduce sleepiness during shifts and improve sleep off-shift, with ultimate positive effects on worker health and safety (Wong 2019).

Description of the condition

Shift work scheduling permits work outside regular daytime working hours, and can be adapted to fit the many needs of different workplaces (ILO 2004). While there is no standard definition of shift work (Stevens 2011), the International Labour Organization (ILO) has described it as "[...] a method of organization of working time in which workers succeed one another at the workplace so that the establishment can operate longer than the hours of work of individual workers." (ILO 1990), while Costa and colleagues define it simply as work that occurs outside a traditional daytime schedule (Costa 2003). Many shift workers work during normal sleeping hours and sleep during daylight hours.

Researchers commonly define sleep in terms of quality or quantity, and there are various validated self-report questionnaires and objective methods for measuring these outcomes. The current definition of insufficient sleep for healthy adults is fewer than the recommended seven hours per night (Consensus Conference Panel 2015). Sleep quality lacks an established definition but generally refers to a collection of sleep measures such as sleep onset latency, degree of fragmentation, sleep efficiency, and sleep-disruptive events (Andrew 2008).

Sleepiness, defined simply as the tendency to fall asleep, can be attributed to insufficient or impaired sleep (Lerman 2012; Shen 2006). While there is no clear consensus on what constitutes sleepiness, it is a complex, multi-faceted construct that can be assessed over various time frames with differing results.

Description of the intervention

This review focused on shift schedule adaptations aimed at minimising negative health and safety effects. Areas such as health care and protective services commonly use rotating shift schedules, with alternate day and night shifts, as opposed to fixed schedules (Rajaratnam 2011). Further possible alterations to rotating schedules involve rotation speed and direction (Knauth 1995). Shift duration can vary widely, with some shifts lasting as long as 24 or 48 hours, as in firefighting work (Choi 2014).

In nursing and manufacturing, common shift lengths are eight hours or 12 hours (Ball 2015). Rotation can be fast or slow. One of the slowest rotations is in the offshore petroleum industry, where employees may work 14 consecutive days or nights, followed by several weeks off (Parkes 2007). In firefighting and in medicine, 24-hour shifts are common (Choi 2014; Nasca 2010). Split shifts, where workers routinely work more than one short shift in a 24-hour period (e.g. four hours in the morning and four hours in the evening) are common in service occupations such as in the restaurant industry (ILO 2004).

How the intervention might work

The ILO provided the following recommendations for shift work schedules to minimise the adverse health and safety effects of shift work and increase worker satisfaction (ILO 2004).

- Use a short cycle period with regular rotations.
- Limit the number of successive night shifts for individual workers.
- Give individual workers some free weekends with at least two full days off.
- Avoid short intervals between shifts.
- Allow workers some flexibility regarding shift change times and shift duration.

It is unclear if these recommendations are based on a systematic review of the evidence (Knauth 2003). The recommendations regarding free weekends and flexibility aim to decrease the social isolation associated with certain shift systems. Others are common recommendations to avoid sleepiness and sleep disturbances. Short rest periods between shifts, many successive night shifts, and long shifts are all associated with sleepiness and sleep disturbances or adverse health effects (Akerstedt 1998; Bambra 2008a; Cotter 2011; Driscoll 2007; Erren 2010; Li 2011). Therefore, shift systems with shorter shifts, faster rotation, and longer periods of rest may minimise the adverse effects of shift work. Some experts believe that forward rotation is better than backward rotation (Driscoll 2007; Knauth 1995). In line with these hypotheses, a study published in 2020 recommended that shift schedules include three or fewer consecutive night shifts, with shift intervals of 11 or more hours and shift duration of no more than nine hours (Garde 2020). Wong and colleagues also found that shift intervals

of 11 or more hours were beneficial (Wong 2019). The International Agency for Research on Cancer (IARC) Working Group and the American Academy of Sleep Medicine and the Sleep Research Society reported that shift work studies should consider different features of the shift work system in relation to health, including start time of shift, number of hours per day, rotating or permanent shifts, speed and direction of rotating systems, time off between shifts, and regularity of schedules (Gurubhagavatula 2021; Stevens 2011).

Why it is important to do this review

Prior systematic reviews examined the effects of shift work schedules on work-related health and safety (Baltes 1999), neurobehavioral and physiological outcomes (Driscoll 2007), and health outcomes (Bambra 2008a). These reviews also examined sleep, fatigue, and alertness outcomes. While Bambra 2008a concluded that forward rotation and fast rotation are beneficial for health, Driscoll 2007 reported that the evidence on these shift patterns was inconclusive.

Driscoll 2007 and Bambra 2008a included observational cohort studies and before-after trials with only one before and one after measurement and without control groups. Bias due to time trends is common in such trials; reviews can reduce this bias by excluding trials that do not report several measures of the outcomes before and after the intervention. In addition, neither of the reviews used Cochrane's rigorous methodology for the literature search, study selection, or contacting of study authors. The most recent of these reviews was published in 2008, so they may be outdated. Commonly available scientific outlets tend not to publish shift work trials, and searching grey literature may yield additional trials. From previous experience in the field, we know that it may be difficult to identify eligible trials by their abstracts alone. Searching in duplicate and contacting study authors for additional information may improve the quality of data for meta-analysis.

OBJECTIVES

To assess the effects of shift schedule adaptation on sleep quality, sleep duration, and sleepiness among shift workers.

METHODS

Criteria for considering studies for this review

Types of studies

We considered the following types of studies for inclusion.

- Randomised controlled trials (RCTs)
- Randomised cross-over trials (trials that randomly allocate participants to one of two groups; one group receives the active intervention and then the control intervention, and the other group receives first the control intervention and then the active intervention (see Higgins 2011, section 16.4))
- Cluster-RCTs
- Controlled before-after (CBA) trials (non-randomised trials in which the intervention takes place in one group but not in another, which serves as a control group; the outcomes are measured once before and once after the intervention)
- Non-randomised cross-over trials (same as randomised cross-over trials, but the allocation is not random)
- Interrupted time series (uncontrolled before-after trials that measure outcomes at least three times before and three times after the intervention)

All of the above trial types were eligible for meta-analysis.

We also aimed to include laboratory trials, where participants receive the intervention in a laboratory setting rather than in their actual workplace. We presented data from laboratory studies in tables and used the data for comparison in the Discussion section; we did not present the findings in the Results section.

We considered the inclusion of studies published as full-text articles, those published as abstract only, and those that provided unpublished data.

Types of participants

We included adult workers (aged 18 years or older) with shift work schedules that included night shift work, irrespective of job title, country, or comorbidities. We defined a night shift as a shift including three or more hours of work between 00:00 and 05:00 (Stevens 2011).

Types of interventions

Study interventions included changes in shift work schedules.

We assessed interventions according to the different components of the shift systems, as follows.

- Permanency of shifts: fixed versus any rotation
- Regularity of shift changes: regular versus irregular changes
- Direction of shift rotation: forward versus backward rotation
- Speed of rotation: faster versus slower rotation
- Duration of shifts: shorter versus longer shifts
- Timing of start of shifts: earlier versus later start
- Distribution of shift schedule: more compressed versus more spread out
- Time off between shifts: longer versus shorter rest
- Split shifts: non-interrupted versus interrupted shifts
- Protected sleep: no on-call duties versus on-call duties
- Worker participation: participative versus non-participative scheduling

Types of outcome measures

The aim of this review was to evaluate the effectiveness of adapting shift schedules for improving sleep quality, sleep duration, and sleepiness. We included subjective and objective measurements of sleep quality off-shift, duration of sleep off-shift, and sleepiness during shifts.

Primary outcomes

Sleep-wake disturbance associated with shift work is a core health problem of shift workers. To characterise sleep-wake disturbance, we included studies that reported intervention effects with the following outcome measures.

- Sleep quality off-shift: measured with a validated questionnaire such as the Bergen Insomnia scale (Pallesen 2008), Pittsburgh Sleep Quality Index (Buysse 1989), Basic Nordic Sleep Questionnaire (Partinen 1995), Jenkins Sleep Questionnaire

(Lallukka 2011), or relevant questions in the Standard Shift Work Index. We also accepted sleep quality as reported in sleep diaries. If available, we used sleep quality after night shifts. Otherwise, we used the average sleep quality across shifts or across the week. We also included wrist-worn actigraphy-based measures as objective measures of sleep quality.

- Sleep duration off-shift: average duration of sleep was based on the relevant questions from validated questionnaires (see examples in previous bullet point), sleep diaries, or wrist-worn actigraphy. If available, we used sleep duration off-shift, a measure that excludes naps and sleep during shifts but includes sleep after a shift and during days off. If sleep duration off-shift was unavailable, we used the 24-hour sleep duration.
- Sleepiness during shifts: we considered sleepiness during night shifts. If studies did not provide this information, we used the average sleepiness across shifts or across the week. We considered the following measures of sleepiness.
  - Self-rated (subjective) sleepiness measured with a validated questionnaire such as the Karolinska Sleepiness Scale (Kaida 2006), Stanford Sleepiness Scale (Herscovitch 1981; Hoddes 1972), relevant questions in the Standard Shift Work Index (Barton 1995), or other visual analogue scales
  - Physiological sleepiness measured with electrophysiological methods during work (e.g. electroencephalogram or electro-oculogram measurement such as PERCLOS (percentage of eyelid closure; Dinges 1998; Sommer 2010) or by standardised physiological tests of sleepiness such as the Multiple Sleep Latency Test (Carskadon 1986), the Maintenance of Wakefulness Test (Mitler 1982) or pupillometric assessment
  - Behavioural sleepiness measured as performance in a validated vigilance test such as a psychomotor vigilance test (e.g. Basner 2011; Thorne 2005), the Mackworth Clock Test (Mackworth 1950), or single or multiple choice reaction time tests
  - Behavioural sleepiness measured as characteristics of overt behaviour subjectively rated by a researcher through video recording methods such as Observer Rating of Drowsiness (e.g. Wierwille 1994)
  - Fatigue usually refers to exhaustion or tiredness due to long-lasting exertion. However, fatigue can be considered a symptom or outcome related to sleepiness at work. Therefore, we included studies that measured fatigue at any moment during the shift, as a self-reported outcome measured with a validated questionnaire or interview if sleepiness was not measured with one of the above-mentioned validated measures of sleepiness.

Table 1 provides an overview of the measurement tools for sleep outcomes considered in this review, including the score range and the minimum clinically important difference (MCID), if available.

Secondary outcomes

In studies that reported primary outcomes of this review, we examined the following secondary outcomes that could be affected by changes in shift schedules.

- Number of staff
- Number of hours worked
- Overtime
- Staff costs

A full cost-effectiveness analysis was beyond the scope of this review, as it would require information not only on our primary outcomes and their 'value' (e.g. willingness to pay), but also of potential adverse effects of shift systems such as errors or injuries and their costs and 'values'.

Search methods for identification of studies

Electronic searches

We conducted a systematic search to identify randomised and non-randomised studies. We searched databases from inception to 13 December 2020, then updated the search on 20 April 2022. Eligible studies retrieved through the second search were marked as 'awaiting classification' and will be fully incorporated in the first update of this review. We searched the following databases.

1. PubMed (searched 20 April 2022; Appendix 1)
2. Embase (via Elsevier; 1946 to 20 April 2022; Appendix 2)
3. Cochrane Central Register of Controlled Trials (CENTRAL) in the Cochrane Library (searched 20 April 2022; Appendix 3)
4. Scopus (searched 20 April 2022; Appendix 4)
5. PsycINFO (APA, from 1806 to 20 April 2022; Appendix 5)
6. NIOSHTIC (OSH-UPDATE; searched 20 April 2022; Appendix 6)
7. NIOSHTIC-2 (OSH-UPDATE; from 1971 to 20 April 2022; Appendix 6)
8. HSELINE (OSH-UPDATE; searched 20 April 2022; Appendix 6)
9. CISDOC (OSH-UPDATE; searched 20 April 2022; Appendix 6)
10. LILACS (Literatura Latino-Americana e do Caribe em Ciências da Saúde, via Bibliotea Virtual em Saúde (BVS); from 1982 to 20 April 2022; Appendix 7)
11. System for Information on Grey Literature in Europe (www.opengrey.eu; searched 20 April 2022; Appendix 8)

Since the search term 'shift' alone leads to a very high number of citations, we applied many combinations of the term 'shift' with other relevant terms. Examples are 'shift work', 'night shift', 'shift schedule', and 'graveyard shift'. We also accounted for terms that describe shift work without including the word shift, such as 'duty time' or 'duty hours' (e.g. transport industry), 'rota' (medicine), or the four-day week alias 'compressed workweek', used to denote a series of 12-hour shifts. We limited the search using terms for different outcomes or types of interventions. We imposed no restriction on the language of publication.

Searching other resources

Seven review authors checked the reference lists of all primary studies and reviewed other relevant articles for additional references (GH, PC, GG, MP, CB, IW, DP). Two review authors (IW, DP) contacted experts in the field to identify additional unpublished materials and searched the conference proceedings of the biennial Symposium on Shiftwork and Working Time. Two review authors (RR, DP) searched the World Health Organization International Clinical Trials Registry Platform (WHO ICTRP; trialsearch.who.int), the US National Institutes of Health Ongoing Trials Register ClinicalTrials.gov (clinicaltrials.gov/), and the EU Clinical Trials Register (www.clinicaltrialsregister.eu/).

Data collection and analysis

Selection of studies

Two review authors (out of GH, PC, GG, MP, CB, IW, DP) independently screened the titles and abstracts of the records, coding them as 'retrieve' (i.e. eligible or potentially eligible/unclear) or 'do not retrieve' (i.e. clearly ineligible). Two review authors (out of GH, PC, GG, MP, CB, IW, DP) independently assessed the retrieved full-text articles against our eligibility criteria and recorded reasons for excluding ineligible studies. We resolved any disagreements by consulting a third review author (out of BG, RR, DP). We excluded duplicates and collated multiple reports of the same study so that each study, rather than each report, was the unit of interest in the review. We recorded the selection process in sufficient detail to complete a PRISMA flow diagram (Page 2021) and Characteristics of excluded studies table. We also contacted the study authors when a paper had insufficient information for us to reach a decision on eligibility.

Data extraction and management

Two review authors per study (out of GH, PC, GG, BG, IW, DP) independently extracted data from the included studies. We extracted the following information and presented it in the review.

- Methods: type of study, allocation, inclusion criteria, statistical analysis
- Basic information: country, dates of study (beginning and end of allocation or study), duration of study, number of participants, number of participants evaluated, information about shift schedules
- Basic information about the participants: age, sex, occupation, chronotype (morningness-eveningness score or similar)
- Intervention: details of interventions being compared, cointerventions
- Outcome measures
- Sources of funding and notable conflicts of interest of study authors

Assessment of risk of bias in included studies

Two review authors (out of GH, PC, MP, BG, DP) independently assessed the risk of bias of each included study using Review Manager 5 (RevMan 5; RevMan 2012). We resolved disagreements by consulting a third review author (out of PC, BG, RR). Where studies provided insufficient information for us to evaluate the methodological criteria, we contacted the study authors. If we received no response, we judged the relevant domain at unclear risk of bias. Where possible, we used quotes from the text to support our judgements.

We used the Cochrane risk of tool (RoB 1) for all study types, with additional domains for CBA trials, cross-over trials, and trials with cluster allocation (Higgins 2011).

Randomised controlled trials

For RCTs, we evaluated the following domains (taken directly or modified where applicable from Higgins 2011).

- Random sequence generation
  - We considered trials at low risk of bias if they described a random element in sequence generation, such as:
    - a random number table;
    - a computer random number generator;
    - coin tossing;
    - shuffling cards or envelopes;
    - throwing dice;
    - drawing lots; or
    - minimisation.
  - We considered trials at high risk of bias if they described sequence generation using:
    - odd or even date of birth; or
    - a rule based on, for example, work record number.
- Allocation concealment
  - We considered trials at low risk of bias if they used:
    - central allocation (including telephone and web-based); or
    - sequentially numbered, opaque, sealed envelopes.
  - We considered trials at high risk of bias if they used:
    - an open random allocation schedule (e.g. a list of random numbers);
    - assignment envelopes without appropriate safeguards (e.g. if envelopes were unsealed or non-opaque or not sequentially numbered);
    - alternation or rotation;
    - date of birth;
    - record number; or
    - any other explicitly unconcealed procedure.
- Blinding of participants and personnel (evaluated for objective and subjective measures separately): since blinding of participants and personnel is not possible in the context of interventions involving shift work scheduling, we assigned an unclear risk of bias judgement for subjective measures (e.g. responses to questionnaires and scales), as participants' beliefs may have influenced responses; and a low risk of bias judgement for objective measures, such as a psychomotor vigilance test.
- Blinding of outcome assessors (evaluated for objective and subjective measures separately)
  - We considered trials at low risk of bias if:
    - there was no blinding of outcome assessment, but the outcome measurement was unlikely to be influenced by lack of blinding; or
    - the investigators ensured blinding of outcome assessment, and it was unlikely that the blinding could have been broken.
  - We considered trials at high risk of bias if:
    - there was no blinding of outcome assessment, and the outcome measurement was likely to be influenced by lack of blinding; or
    - the investigators blinded outcome assessment, but blinding could have been broken, and the outcome measurement was likely to be influenced by lack of blinding.
- Incomplete outcome data (evaluated for each outcome separately)
  - We considered trials at low risk of bias if:
    - there were no missing outcome data;
    - reasons for missing outcome data were unlikely to be related to the outcome;

- missing outcome data were balanced in number across intervention groups, with similar reasons for missing data across groups;
- in dichotomous outcome data, the proportion of missing outcomes compared with the observed event risk was not large enough to have a clinically relevant impact on the intervention effect estimate;
- in continuous outcome data, a plausible effect size (mean difference (MD) or standardised mean difference (SMD)) among missing outcomes was not large enough to have a clinically relevant impact on observed effect size; or
- missing data were imputed using appropriate methods.

  - We considered trials at high risk of bias if:
    - the reason for missing outcome data was likely related to the outcome, with either imbalance in numbers or reasons for missing data across intervention groups;
    - in dichotomous outcome data, the proportion of missing outcomes compared with observed event risk was large enough to induce clinically relevant bias in the intervention effect estimate;
    - in continuous outcome data, the plausible effect size (MD or SMD) among missing outcomes was large enough to induce clinically relevant bias in observed effect size;
    - the trial authors performed as-treated analysis with substantial departure from the number of participants assigned at randomisation (or beginning of the trial)
    - there was potentially inappropriate application of simple imputation;
    - in cluster-randomised trials, loss of full clusters was likely to introduce bias.
- Selective outcome reporting
  - We considered trials at low risk of bias if:
    - the study protocol was available and the study measured and reported all prespecified (primary and secondary) outcomes of interest, using the prespecified methods; or
    - the study protocol was unavailable, but the published reports clearly included all expected outcomes, including those prespecified.
  - We considered trials at high risk of bias if:
    - not all prespecified primary outcomes were reported;
    - one or more primary outcomes were reported using measurements, analysis methods, or subsets of the data (e.g. subscales) that were not prespecified;
    - one or more reported primary outcomes were not prespecified (unless clear justification for their reporting is provided, such as an unexpected adverse effect);
    - one or more outcomes of interest in the review were reported incompletely and could not be entered into a meta-analysis; or
    - the study report failed to include results for a key outcome that would normally be included in such a study.
- Reliable or objective measurement (for each of the outcomes relevant to the review)
  - We considered trials at low risk of bias if they measured the outcome objectively (e.g. psychomotor vigilance test) or the agreement between two or more raters was above 90% or at least kappa 0.8.
  - We considered trials at high risk of bias if the agreement between two or more raters was below 90% or below kappa 0.8.
- Other sources of bias: we assessed whether there were any other potential sources of bias.

Randomised cross-over trials

We assessed all domains for RCTs in addition to the following domain.

- Applicability of the design for every outcome
  - We considered trials at low risk of bias if:
    - there was no statistically significant interaction between the order of interventions and the outcome; or
    - we believed the outcome to be independent of the order of treatments (e.g. if the washout period was sufficiently long).
  - We considered trials at high risk of bias if:
    - there was an interaction between the order of interventions and the outcome; or
    - we judged the outcome measure to be influenced by the order of treatments (e.g. if the washout period was very short).

Cluster-randomised trials

We assessed all items for RCTs in addition to the following domains.

- Recruitment bias (e.g. Puffer 2003). Recruitment of individuals to different clusters after randomisation may occur. This may lead to different types of participants being recruited to the different clusters.
  - We considered trials at low risk of bias if:
    - the trial reported no or minimal recruitment after randomisation; or
    - the effect of recruitment to different clusters after randomisation was unlikely to influence the outcome.
  - We considered trials at high risk of bias if recruitment after randomisation occurred and we considered it might have changed the interpretation of the results.
- Appropriate statistical analyses. Cluster-randomised trials do not always take the cluster effect into account.
  - We considered trials at low risk of bias if:
    - the analysis took the cluster effect into account (see section 16.3 of Higgins 2011); or
    - we were able to correct the analysis using the following information: number of clusters (or groups) randomised to each intervention group or the average (mean) size of each cluster; the outcome data ignoring the cluster design for the total number of individuals (e.g. number or proportion of individuals with events, or means and standard deviations (SDs)); and an estimate of the intracluster (or intraclass) correlation coefficient (ICC). See section 16.3.4 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).
  - We considered trials at high risk of bias if the analysis did not take the cluster effect into account and we were unable to correct the analysis.

Interrupted time series

Ramsay 2003 describes a risk of bias assessment method for interrupted time series. We planned to assess all items included in Ramsay 2003, three of which are also included in RoB 1 (blinding of outcome assessors, reliable outcome measure, and incomplete data). We also added selective reporting. Therefore, in addition to blinding of outcome assessors, incomplete outcome data, selective reporting, and reliable or objective measurement, we planned to assess the following items specific to interrupted time series.

- Intervention independent of other changes over time
  - We considered studies at low risk of bias if the intervention appeared to be independent of other changes over time.
  - We considered studies at high risk of bias if the intervention did not appear to be independent of other changes over time.
- Intervention unlikely to affect data collection
  - We considered studies at low risk of bias if the intervention was unlikely to affect data collection (e.g. same sources and methods of data collection before and after the intervention).
  - We considered studies at high risk of bias if data collection was likely affected by the intervention.
- Shape of the intervention effect prespecified
  - We considered studies at low risk of bias if the study authors provided a rational explanation for the shape of the intervention effect.
  - We considered studies at high risk of bias if study authors provided no explanation or an inadequate explanation for the shape of the intervention effect.
- Rationale for the number and spacing of data points
  - We considered studies at low risk of bias if the study authors provided an adequate rationale for the number of points (e.g. monthly data for 12 months postintervention were used because the anticipated effect was expected to decay) or if sample size calculation based on reasonable assumptions influenced the study design.
  - We considered studies at high risk of bias if the study authors provided no adequate rationale for the number of points and performed no or inadequate sample size calculation.
- Appropriate statistical analyses
  - We considered studies at low risk of bias if:
    - they used autoregressive integrated moving average (ARIMA) models;
    - they used time series regression models to analyse the data and adjusted or tested for serial correlation; or
    - we were able to correct the analyses.
  - We considered studies at high risk of bias if none of the above criteria applied.

Controlled before-after trials

We assessed blinding of outcome assessors, incomplete outcome data, selective outcome reporting, reliable or objective measurement, other sources of bias, intervention done independently of other changes over time, and intervention unlikely to affect data collection, in addition to the following two domains specific to CBA trials.

- Baseline differences between groups. We considered the variables type and place of work, age, sex, and chronotype.
  - We considered studies at low risk of bias if all four variables were similar across groups.
  - We considered studies at high risk of bias if any of these variables differed enough to introduce bias.
- Appropriate statistical analyses
  - We considered studies at low risk of bias if the analysis was adequate and the study authors had adequately controlled for baseline differences and other confounders.
  - We considered studies at high risk of bias if the analysis was inadequate, for example if:
    - it did not report baseline data and changes from baseline for both study groups; or
    - confounding was not adequately addressed in the analysis.

Non-randomised cross-over trials

For non-randomised cross-over trials, we assessed all risk of bias domains for randomised cross-over trials and the additional items listed for CBA trials.

Laboratory studies

We did not assess risk of bias of laboratory studies, considering them at high risk by design.

Assessment of bias in conducting the systematic review

We conducted the review according to a published protocol and reported any deviations from it in the Differences between protocol and review section of the systematic review (Erren 2013).

Measures of treatment effect

We entered outcome data for each study into the data tables in RevMan 5 (RevMan 2012). We reported the mean and SD for continuous outcomes then calculated the MD if studies used the same measurement scale, or the SMD if studies used different measurement scales (Higgins 2011). We contacted study authors to obtain additional data where necessary. If studies only reported effect estimates and their 95% confidence intervals (CIs) or standard errors (SEs), we entered these data into RevMan 5 using the generic inverse variance method. Whenever we were unable to enter results in the software, we described them in the Results section. We reversed the scoring of scales if needed, so that a high score for sleep duration and a low score for sleep quality and sleepiness denoted good/beneficial outcomes. For CBA trials, we considered outcome measurements of changes from baseline to ensure that baseline imbalances were taken into account.

Unit of analysis issues

For studies that employed a cluster-randomised design and had sufficient data to be included in the meta-analysis, but did not make an allowance for the design effect, we calculated the design effect based on the methods described in section 16.3.6 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). If we identified no reliable estimate of the ICC, we included the trial using an ICC of 0.1.

Dealing with missing data

We attempted to contact study authors for missing information to assess risk of bias or outcome measures relevant to this systematic review. We read all reports of trials in an effort to find missing data.

For each included trial, where data were available, we noted the level of attrition for each group, per outcome or group of outcomes.

Assessment of heterogeneity

Within each comparison of interventions and each outcome, we assessed the homogeneity of the results of included studies based on similarity of interventions, exact outcome definitions, outcome timing, and follow-up.

We combined studies if they reported the same outcome measure (e.g. sleepiness) irrespective of measurement time points, although sleepiness had to be measured during shifts, and sleep quality and duration had to be measured off-shift. We reported differences in type and timing of measurements and in outcome definitions. Where there were variations of validated measures for a single outcome across studies, we calculated SMDs. When a study reported an outcome in more than one way, we prioritised the following measures for each primary outcome in the main analysis.

- Sleep quality: measures reported in sleep diaries
- Sleep duration: objective measures
- Sleepiness: subjective measures

We deviated from the protocol, as we decided to include additional interventions related to shift schedules that emerged during the data extraction phase. These were regularity of shift changes, distribution of shift schedules, duration of rest between shifts, split shifts, protected sleep, and worker participation.

We tested for statistical heterogeneity by means of the Chi^2 test as implemented in the forest plot in RevMan 5 software (RevMan 2012). We considered a significance level of $P < 0.1$ indicative of heterogeneity. Moreover, we quantified the degree of heterogeneity using the I^2 statistic, where an I^2 value of 25% to 50% indicated a low degree of heterogeneity, 50% to 75% a moderate degree of heterogeneity, and more that 75% a high degree of heterogeneity (Higgins 2011).

Assessment of reporting biases

We reduced the effect of reporting bias by including unpublished studies. To avoid the introduction of duplicate data (i.e. two articles could represent duplicate publications of the same study), we attempted to detect duplicate reports and, if more than one article reported data from the same study, we extracted data only once (Cho 2000). We prevented location bias by searching across multiple databases. We prevented language bias by including studies published in any language. This review included too few trials to formally assess publication bias using funnel plots.

Data synthesis

First, we presented results separately for randomised studies and non-randomised studies. For each component of the shift system, we pooled data from studies within the same category (e.g. randomised and non-randomised studies) using RevMan 5 (RevMan 2012). Data from the laboratory study were not included in narrative or quantitative synthesis (Cruz 2003). Whenever possible, we combined SMDs. To make the pooled SMDs more readily interpretable, we transformed them to MDs by multiplying them by the median SD taken from studies that used the scale in question. The median SDs for the preferred scales were as follows.

- Basic Nordic Sleep Questionnaire (sleep quality): median SD 4.18
- Hours of sleep per day (sleep duration): median SD 1.45
- Karolinska Sleepiness Scale (sleepiness): median SD 1.26

Owing to heterogeneity across studies in interventions, outcomes, and follow-up times, we used a random-effects model for meta-analysis. All estimates included a 95% CI.

Subgroup analysis and investigation of heterogeneity

We planned to conduct subgroup analyses based on the following characteristics (Erren 2013).

- Chronotype
- Shift schedule details (e.g. different types of rotation)
- Occupational setting, branch, or industry
- Different ways of measuring the same outcome
- Age

Because few studies evaluated each comparison, we did not undertake any subgroup analyses.

Sensitivity analysis

We had planned to conduct the following sensitivity analyses.

- Excluding studies with a high risk of bias
- Different methods for measuring the same outcome (e.g. self-reporting versus physiological sleepiness)
- Different assumptions for imputation of missing data and different proportions of missing data
- Different assumptions for intra-class correlation (for cluster-randomised trials)

We did not undertake any of these sensitivity analyses due to the small numbers of included studies per comparison.

Summary of findings and assessment of the certainty of the evidence

Where we were able to meta-analyse any of our primary or secondary outcomes for a comparison, we reported the results in a summary of findings table. We created a separate table for each included comparison using GRADEpro software (GRADEpro 2008), and in each table we reported all seven prespecified outcomes (see Types of outcome measures). We used the five GRADE domains (study limitations, consistency of effect, imprecision, indirectness, and publication bias) to assess the certainty of the body of evidence based on the studies that contributed data to the outcome. We followed methods and recommendations described in Chapter 8 and Chapter 12 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).

Two review authors (GH and PC) independently assessed the certainty of the evidence for each outcome. We considered evidence from RCTs as high certainty to begin with. For the risk of bias domain, we downgraded the certainty of the evidence by one level if any contributing study had an 'unclear risk' judgement in any domain and no studies had any 'high risk' judgements. We downgraded the certainty of the evidence by two levels if any contributing study had a 'high risk' judgement in any domain, or if there were 'unclear risk' judgements for multiple domains that

substantially lowered our confidence in the results. We considered evidence from non-randomised studies to be low certainty to begin with, downgrading for the same reasons as for RCTS, and upgrading if studies had a large sample size, showed dose-response relations, or had opposing plausible residual bias and confounding. All decisions to downgrade or upgrade the certainty of the evidence were justified in the footnotes of the summary of findings tables.

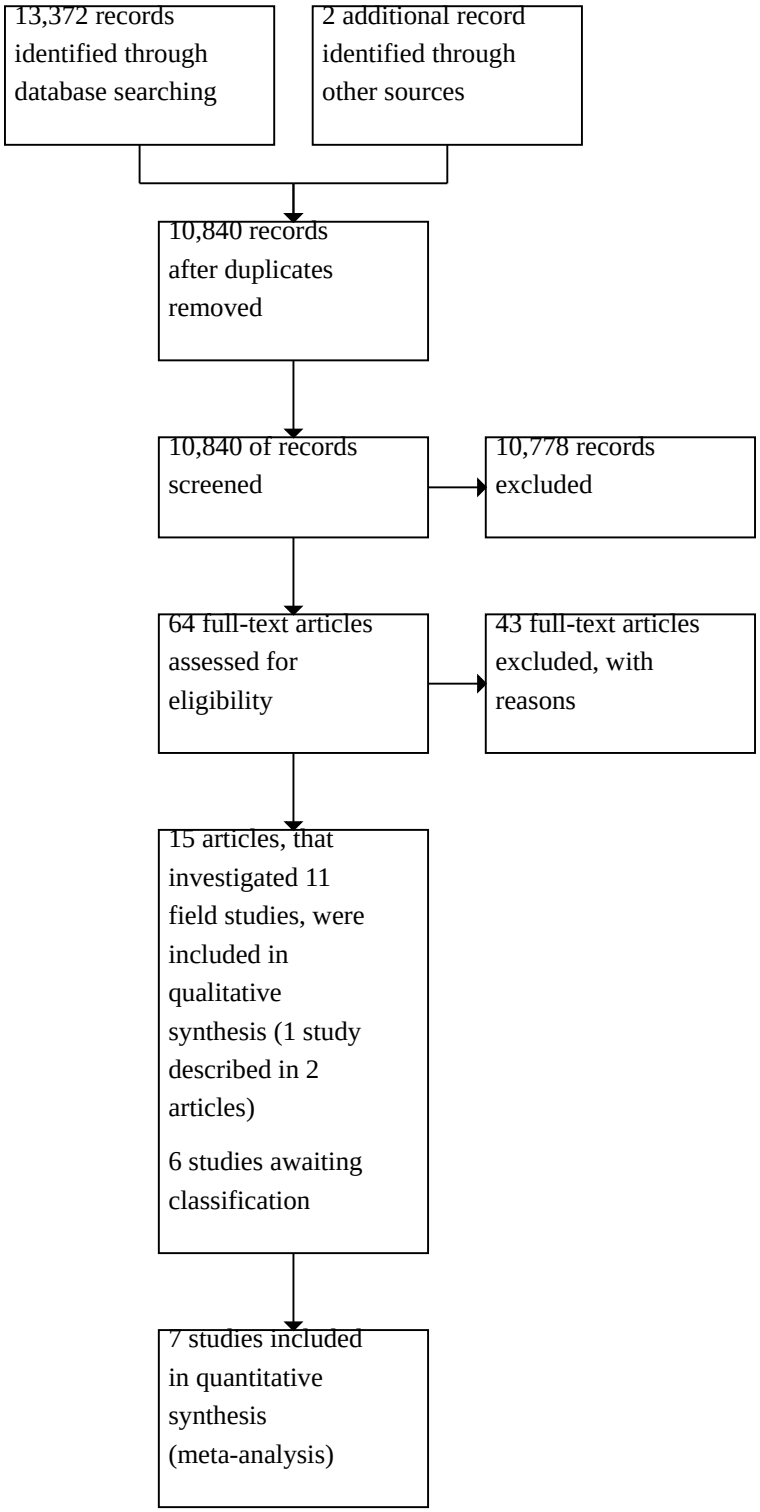
RESULTS

Description of studies

Results of the search

Our systematic search conducted on 13 December 2020, and updated on 20 April 2022, yielded 13,372 references. After removing 2534 duplicates, we screened the titles and abstracts of the remaining 10,840 records and selected 64 articles for full-text review. Eleven studies (reported in 15 articles and including a total of 2125 participants) met our eligibility criteria (Amendola 2011; Axelsson 1998; Barger 2019a; Barton 1994; Basner 2019; Bell 2015; Cruz 2003; Garde 2020; Knauth 1998; Totterdell 1992; Viitasalo 2008). The results of one study were reported in two papers (Barger 2019a). We were unable to retrieve the full text of two potentially eligible studies, which are awaiting classification (Chevreau 2012; Toussaint 2003). Four other studies are awaiting classification because they were published after 13 December 2020, and we retrieved them in the updated search (Cori 2021; Hakola 2021; Puttonen 2022; Rahman 2021). In the Characteristics of excluded studies table, we recorded the studies excluded at full-text review stage. Reasons for exclusion were lack of validated methods for the outcome assessment, wrong type of intervention, and wrong study design. Figure 1 shows the study selection process in a PRISMA diagram.

Figure 1. Study flow diagram.



Included studies

Study designs

Of the 11 included studies, three were cluster-RCTs (Amendola 2011; Barger 2019a; Basner 2019), five were CBA trials (Barton 1994; Bell 2015; Knauth 1998; Totterdell 1992; Viitasalo 2008), two were non-randomised cross-over trials (Axelsson 1998; Garde 2020), and one was a laboratory study (Cruz 2003). The RCTs employed a parallel design and stratified randomisation according to site of work and time of day of shift (i.e., day, evening, and night shifts). Three of the CBA trials allocated participants in clusters (Bell 2015; Knauth 1998; Totterdell 1992), and two allocated volunteer workers individually (Barton 1994; Viitasalo 2008). All CBA trials applied a parallel design.

Due to generalisation issues, we presented the characteristics and results of the laboratory study only in tables and not in this section of the review (Cruz 2003).

Type of settings and participants

Four studies were conducted in the USA (Amendola 2011; Barger 2019a; Basner 2019; Bell 2015), and six studies were conducted in Europe: two in the UK (Barton 1994; Totterdell 1992), and one each in Finland (Viitasalo 2008), Sweden (Axelsson 1998), Denmark (Garde 2020), and Germany (Knauth 1998). Worker populations included car manufacturing employees (Barton 1994), law enforcement officers (Amendola 2011; Bell 2015; Garde 2020; Totterdell 1992), medical residents or interns (Barger 2019a; Basner 2019), line maintenance workers at an airline company (Viitasalo 2008), power plant workers (Axelsson 1998), and steelworkers (Knauth 1998). Most studies included only male workers or had mostly male participants; the exceptions were the two RCTs conducted in healthcare settings (Barger 2019a; Basner 2019).

The three RCTs enroled 275 workers (Amendola 2011), 302 workers (Barger 2019a), and 398 workers (Basner 2019). We did not include Amendola 2011 in any meta-analyses due to insufficient data (no effect estimates, means, or measures of variability were available). Sample sizes in CBA trials ranged significantly (293 participants in Barton 1994, 386 in Bell 2015, 179 in Knauth 1998, 71 in Totterdell 1992, and 89 in Viitasalo 2008). We did not include Knauth 1998 or Barton 1994 in any meta-analyses due to insufficient available data: Knauth 1998 presented no data on relevant outcomes, while Barton 1994 did not report any measures of variability, such as SDs. The non-randomised cross-over trials included 31 participants (Axelsson 1998) and 73 participants (Garde 2020).

Interventions

Shift schedule changes are usually complex and involve multiple components of the shift system. Shift schedule components evaluated in the studies were direction of rotation (forward or backward; Barton 1994; Knauth 1998; Viitasalo 2008), speed of rotation (slow, fast, or very fast; Garde 2020; Knauth 1998; Viitasalo 2008), duration of shift (eight hours versus 12 hours in Axelsson 1998, 10 hours versus 12 hours in Amendola 2011, 10 hours versus 13 hours and 20 minutes in Bell 2015, and 16 hours versus 24 to 28 hours in Barger 2019a and Basner 2019); and compressed versus more spread out schedules (Bell 2015; Knauth 1998; Totterdell 1992). No studies evaluated the effects of changing permanency of shifts (fixed versus any type of rotation), regularity of shift changes (predictable versus unpredictable changes), timing of start (earlier versus later start), time off between shifts (longer versus shorter rest between shifts), split shifts (non-interrupted versus interrupted shifts), protected sleep (no-on call duties versus on-call duties), or worker participation (participative versus non-participative scheduling).

Six studies changed more than one component of the shift schedule (Amendola 2011; Axelsson 1998; Bell 2015; Knauth 1998; Totterdell 1992; Viitasalo 2008). Of the studies that examined the effects of changes in only one scheduling component, Barton 1994 focused on shift rotations, Garde 2020 on rotation speed, and Barger 2019a and Basner 2019 on shift duration. Table 2 shows the components embedded in shift systems in each study. We compared the components of the shift systems that we considered could affect sleep duration and quality and sleepiness. Table 3 presents detailed descriptions of all shift schedules.

Direction of rotation of shifts: forward versus backward rotation

Three CBA trials compared shift systems with different direction of shift rotation (Barton 1994; Knauth 1998; Viitasalo 2008). In total, the three studies included 561 shift workers.

Barton 1994 had one intervention group and two control groups. One control group comprised daytime workers and was not included in this review. Participants in the second control group worked eight-hour shifts with a backward rotation and weekly changes. The intervention group emulated the control group in all aspects but used forward shift rotation.

Viitasalo 2008 had two intervention groups and one control group. One intervention group had worked a shift schedule with forward rotation and the control group had backward rotation. The other intervention arm was not relevant for this review.

Knauth 1998 had two intervention groups that worked in forward-rotating shift schedules and two control groups that worked in backward-rotating shift schedules.

Speed of rotation: faster versus slower rotation

Two CBA trials and one non-randomised cross-over trial (341 participants) compared faster shift rotation with slower shift rotation.

In Viitasalo 2008, one intervention group had very fast rotation, with shift changes after every shift, while the control group had fast rotation, with changes every three days.

In Knauth 1998, the intervention groups had fast or intermediate rotation schedules, with changes after every two or three shifts, while the control groups had very slow rotation schedules, with rotations after six or seven shifts.

Garde 2020 (the cross-over trial) exposed participants to three different work schedules with varying numbers of consecutive night shifts: two night shifts followed by two recovery days, four night shifts followed by four recovery days, and seven night shifts followed by seven recovery days. Recovery days could be day/morning shifts or days off.

Shift duration: shorter versus longer duration

All three RCTs investigated shift duration (Amendola 2011; Barger 2019a; Basner 2019), as did two CBA trials (Axelsson 1998; Bell 2015). In total, these five studies included 1392 participants.

Amendola 2011 had three study arms, with different duration of shifts. The first intervention group worked 10-hour shifts and the second intervention group worked three consecutive 12-hour shifts, or four consecutive shifts with the first three lasting 12 hours and the fourth lasting eight hours. The control group worked eight-hour shifts.

In Barger 2019a, the intervention group worked a four- to five-day rotation schedule with two 11- to 15-hour day shifts followed by an overnight shift lasting 16 hours, and the control group worked a four- to five-day rotation schedule with two 12-hour shifts and one overnight shift lasting 24 to 28 hours.

Basner 2019 compared two 80-hour work weeks in which the intervention group did not have any shift duration limits or mandatory time off between shifts, and the control group had duty-hour limits (i.e. 16 hours for normal shifts and 24 hours for in-house duty) and included a minimum number of hours rest between shifts.

In Axelsson 1998, the analyses were separated into day and night shifts. For day shifts, the study authors compared the first three eight-hour morning shifts with the first three 12-hour morning shifts. For night shifts, they compared the first three 12-hour shifts with the first two and the fourth eight-hour shift.

In Bell 2015, the intervention group worked shifts lasting 13 hours and 20 minutes (two shifts per week), and the control group worked 10-hour shifts (three shifts per week).

Distribution of work schedule: more compressed versus more spread out

Although Amendola 2011, Bell 2015, and Knauth 1998 compared more compressed and more spread-out schedules, their primary aim was to compare shorter and longer shifts. Totterdell 1992 (71 participants) studied the effects of a specific type of shift system, called the Ottawa Shift System, in which morning and afternoon shifts were longer (10 hours) than night shifts (8.5 to 9 hours). The control group worked a more spread out schedule with standard eight-hour morning, afternoon, and night shifts.

Outcomes

Sleep quality off-shift

Seven studies reported sleep quality off-shift (Amendola 2011; Axelsson 1998; Barton 1994; Bell 2015; Garde 2020; Totterdell 1992; Viitasalo 2008). In four studies, participants recorded perceived quality of sleep in sleep diaries every day (Amendola 2011; Axelsson 1998; Barton 1994; Garde 2020). Bell 2015 assessed self-reported sleep quality using the Pittsburgh Sleep Quality Index at baseline and three and six months after the shift schedule change. Totterdell 1992 used a 10-cm visual analogue scale ranging from "worst" to "best". Viitasalo 2008 did not report sleep quality off-shifts in the published article, but the study authors provided us with data based on responses to the first four questions of the Basic Nordic Sleep Questionnaire.

Three studies reported sleep quality after night shifts (Axelsson 1998; Barton 1994; Garde 2020), while four studies used the average sleep quality across all shifts (Amendola 2011; Bell 2015; Totterdell 1992; Viitasalo 2008).

Table 4 summarises the methods of measuring and reporting sleep quality in the different studies.

Sleep duration off-shift

Ten studies measured sleep duration off-shift (Amendola 2011; Axelsson 1998; Barger 2019a; Barton 1994; Basner 2019; Bell 2015; Garde 2020; Knauth 1998; Totterdell 1992; Viitasalo 2008). Three studies used actigraphy (Barger 2019a; Basner 2019; Garde 2020), and three studies used sleep diaries (Amendola 2011; Barton 1994; Axelsson 1998). Participants in Bell 2015 reported hours of sleep with the Pittsburgh Sleep Quality Index at baseline and at three and six months after the intervention. Totterdell 1992 asked participants to report their usual sleep start and end times before the first shift and between shifts for morning, afternoon, and night shifts. Viitasalo 2008 assessed sleep duration off-shift by asking participants "How many hours a day do you usually sleep, including naps?". Knauth 1998 used a questionnaire where participants could indicate their average sleeping time during different schedules.

Five studies reported sleep duration after night shifts (Axelsson 1998; Barton 1994; Basner 2019; Garde 2020; Totterdell 1992), while the other five studies did not (Amendola 2011; Barger 2019a; Bell 2015; Knauth 1998; Viitasalo 2008).

Table 5 shows how the different studies measured and reported sleep duration off-shift.

Sleepiness during shifts

Seven studies reported sleepiness during shifts (Amendola 2011; Axelsson 1998; Barger 2019a; Basner 2019; Bell 2015; Totterdell 1992; Viitasalo 2008). Amendola 2011 and Bell 2015 used both objective and subjective methods. The objective methods in Amendola 2011 were a psychomotor vigilance test and an optical tracker, applied at baseline and at six months, while subjective assessment involved a self-assessed alertness log derived from the Karolinska Sleepiness Scale and a composite measure of items from the Harvard Study of Work Hours and the Epworth Sleepiness Scale. The objective method in Bell 2015 was the three-minute version of the psychomotor vigilance test, administered during the last hour of the last shift of the week, at one and five months. For the subjective assessment, participants completed the 'Daytime dysfunction due to sleepiness' item of the Pittsburgh Sleep Quality Index at baseline and at three and six months.

Five studies used only subjective methods to assess sleepiness (Axelsson 1998; Barger 2019a; Basner 2019; Totterdell 1992; Viitasalo 2008). Axelsson 1998, Barger 2019a, and Basner 2019 used the Karolinska Sleepiness Scale. Totterdell 1992 used a 10-cm visual analogue scale ranging from "drowsy" to "alert"; participants recorded how alert they typically felt at specified two-hour intervals during a morning, afternoon, and night shift. Viitasalo 2008 assessed sleepiness with the Basic Nordic Sleep Questionnaire and the Epworth Sleepiness Scale.

Three studies reported sleepiness for night shifts separately from other shifts (Axelsson 1998; Basner 2019; Totterdell 1992), and four studies reported average sleepiness scores across shifts or days (Amendola 2011; Barger 2019a; Bell 2015; Viitasalo 2008). No studies measured only fatigue without measuring sleepiness.

Table 6 provides an overview of how the different studies measured and reported sleepiness.

Secondary outcomes

Only three studies assessed secondary outcomes: Amendola 2011 and Bell 2015 assessed overtime between different shift systems, and Barger 2019a assessed the number of hours worked. No studies reported number of staff or staff costs.

Funding

Seven of the 10 studies reported sources of funding (Amendola 2011; Axelsson 1998; Barger 2019a; Basner 2019; Bell 2015; Garde 2020; Viitasalo 2008), while three studies did not report whether they received funding (Axelsson 1998; Barton 1994; Knauth 1998). More information on funding sources can be found in the Characteristics of included studies table.

Excluded studies

We excluded 44 studies during the full-text review for the following reasons.

- Interventions unrelated to the shift system (NCT03813654; Garde 2011; McPherson 1993; Rosa 1996)
- Irrelevant outcomes for this review (Landrigan 2020; Grewal 2022; Tucker 2021)
- Non-validated methods (Akersted 1978; Levin 2014)
- Focus on the reduction of working hours rather than shift systems (Cappuccio 2009)
- Evaluation of biological responses after just one shift (Dutheil 2012)
- Reporting of differences related to the time of day for the same shift system (Eriksen 2006)
- Ineligible control group (day workers; Duchon 1994).
- Focus on different schedules for sleep opportunities (Jackson 2014)
- Evaluation of human responses to confined conditions (Chiles 1968), forced desynchrony (Kosmadopoulos 2014), or sleep deprivation (D'Amico 1985)
- Non-randomised studies conducted in laboratory settings that did not include a baseline measurement and did not qualify as a CBA trial (Kudielka 2007; Rosa 1993; Skornyakov 2017; Smith 1998)
- Interrupted time series with fewer than three measures before and after the implementation of the intervention (Duplessis 2007; Harris 2010; Hossain 2004; Ng-A-Tham 1993; Rosa 1989; Shattuck 2015b; Waage 2012; Williamson 1986; Williamson 1994)
- Non-experimental studies (Chang 2021; Cheng 2021; Costa 2014; Fischer 2021; ISRCTN17016944; Hong 2021; Pavageau 2006; Seibt 1990; Shattuck 2015a; van de Ven 2021)
- Combination of four different interventions in the analyses, precluding the isolated effect estimates of each intervention (Knauth 1987)
- Conference abstracts or protocols of included studies (Barger 2019b; Blackwell 2019; Shea 2018)

Risk of bias in included studies

Figure 2 and Figure 3 summarise the results of the risk of bias assessment.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies. Empty spaces represent domains not applicable for at least one study. We did not assess laboratory studies; risk of bias is considered high in such studies.

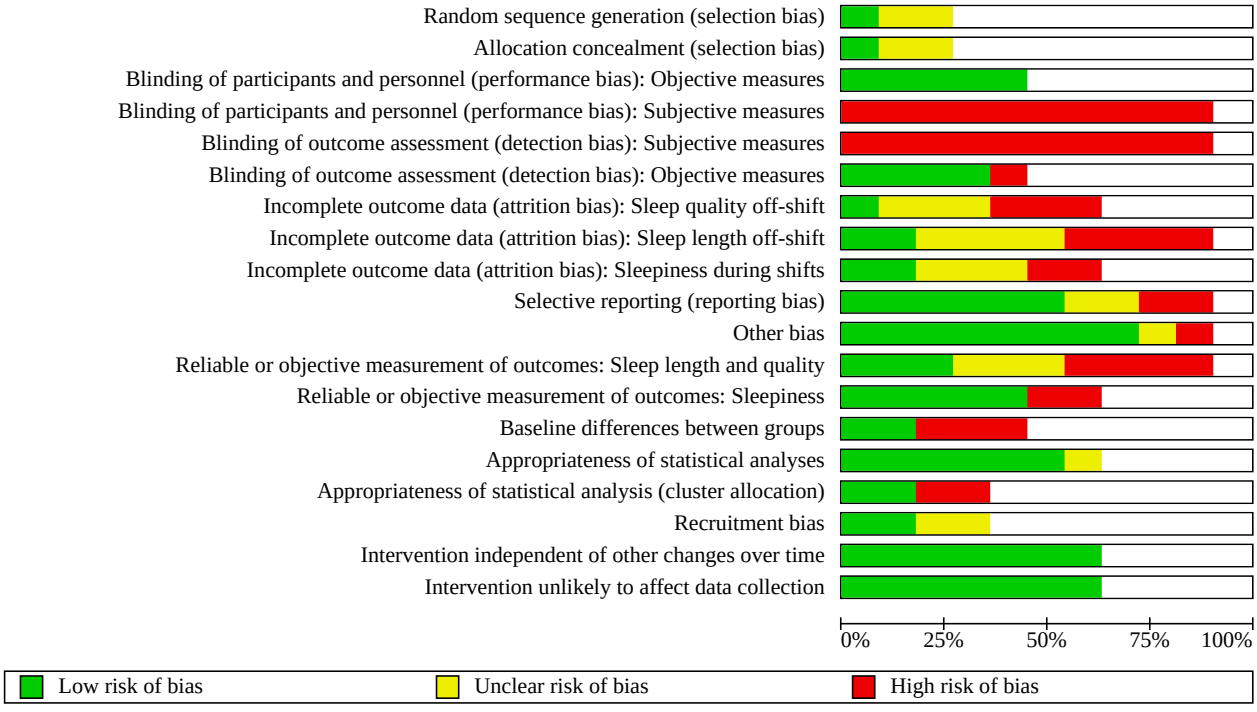


Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study. Empty boxes represent domains not relevant to the study design or outcomes not assessed or reported. We did not assess laboratory studies; risk of bias of laboratory studies is considered high by design.

Allocation

The selection bias assessment only applied to the randomised studies. Two RCTs did not fully report the methods used to generate the allocation sequence (Amendola 2011; Barger 2019a). Amendola 2011 first stratified participants by work site and current shift, then randomised participants within each stratification. However, it was unclear how the random sequence was generated, so we judged the study at unclear risk of bias for this domain. The third RCT was at low risk of bias, as it used statistical software (SAS version 9.3) to determine random assignment (Basner 2019).

We judged Amendola 2011 and Barger 2019a at unclear risk of bias for allocation concealment, as they did not describe any method of ensuring allocation concealment. Basner 2019 was at low risk.

Blinding

Blinding of participants and personnel

We associated subjective measures with high risk of performance bias, as they were self-reported and participants may have been influenced by their beliefs and attitudes when reporting sleep outcomes. We associated objective measures with low risk of performance bias, as it would be difficult to influence the results.

All studies used at least one subjective measure for sleep duration, sleep quality (Amendola 2011; Axelsson 1998; Barton 1994; Bell 2015; Garde 2020; Knauth 1998; Totterdell 1992) or sleepiness (Axelsson 1998; Barger 2019a; Basner 2019; Totterdell 1992; Viitasalo 2008). For these outcomes, we rated the studies at high risk of bias. Five studies used objective measures to measure sleep duration (Barger 2019a; Basner 2019; Garde 2020) or sleepiness (Amendola 2011; Bell 2015). For these outcomes, we rated the studies at low risk of bias.

Blinding of outcome assessors

We associated objective measures with low risk of detection bias, as lack of blinding is unlikely to result in biased results. For subjective measures, we considered risk of bias unclear when a researcher had assessed the outcome, and high for self-reported instruments. All subjective outcomes were rated at high risk of bias: sleep duration and sleep quality in seven studies (Amendola 2011; Axelsson 1998; Barton 1994; Bell 2015; Garde 2020; Knauth 1998; Totterdell 1992), and sleepiness in five studies (Axelsson 1998; Barger 2019a; Basner 2019; Totterdell 1992; Viitasalo 2008).

Barger 2019a, Basner 2019 and Garde 2020 used objective measurements of sleep duration. We judged Barger 2019a and Basner 2019 at a low risk of bias as data interpretation of the actigraphy measurements were blinded. However, we considered Garde 2020 at high risk of bias because analyses of actigraphy data were not blinded and may be subject to bias. Amendola 2011 and Bell 2015 assessed sleepiness during shifts by objective measures, so were at low risk of bias.

Incomplete outcome data

Since attrition impacted the assessment of the outcomes equally, we described the risk of attrition bias for all outcomes combined.

We considered three studies at high risk of attrition bias (Amendola 2011; Barton 1994; Totterdell 1992), five studies at unclear risk (Axelsson 1998; Barger 2019a; Bell 2015; Garde 2020; Knauth 1998), and two studies at low risk (Basner 2019; Viitasalo 2008).

In Amendola 2011, substantial proportions of randomised participants did not receive the intervention (group 1: 28/105 (27%); group 2: 10/109 (9%); group 3: 13/108 (12%)) or did not complete the study (group 1: 13/105, group 2: 18/109, group 3: 19/108). In group 1 (eight-hour shift), the main reason for voluntary dropout was preference for an alternative schedule. A substantial proportion of participants did not provide data in Barton 1994 (23% of the intervention group and 30% of the control group). In Totterdell 1992, 48.7% of participants completed the questionnaires.

Axelsson 1998, Barger 2019a, Bell 2015, Garde 2020, and Knauth 1998 did not clearly report the number of participants completing the study.

Viitasalo 2008 reported low dropout rates in all study arms; losses were justified and unlikely related to the interventions. In Basner 2019, missing outcome data were limited and reasonably balanced; the study authors used single imputation to account for missing data in the analysis.

Selective reporting

We judged six studies at low risk of reporting bias: Axelsson 1998 reported three primary outcomes, Barger 2019a and Basner 2019 reported the outcomes as described in their protocol, and for Amendola 2011, Barton 1994, and Bell 2015, we considered the reported outcomes were likely preplanned. We judged two studies at unclear risk (Garde 2020; Knauth 1998). Totterdell 1992 did not report the findings related to sleep quality as planned. Viitasalo 2008 did not present all results for the Basic Nordic Sleep Questionnaire. We therefore considered these two studies at high risk of reporting bias.

Other potential sources of bias

We considered eight studies at low risk of bias related to other sources (Barger 2019a; Barton 1994; Basner 2019; Bell 2015; Garde 2020; Knauth 1998; Totterdell 1992; Viitasalo 2008). Amendola 2011 did not adjust the analysis for potential socio-demographic differences between the groups, such as age or years of service, so was at unclear risk. Axelsson 1998 had a small sample size and did not report a power analysis, so was at high risk.

Reliable or objective measurement of outcomes

Sleep duration and sleep quality off-shift

Overall, we rated studies at a high risk of bias (Amendola 2011; Barton 1994; Totterdell 1992; Viitasalo 2008) or unclear risk of bias (Bell 2015; Axelsson 1998; Knauth 1998), owing to the absence of validated measures for sleep duration and quality. Amendola 2011 assessed sleep quality with a single question in the sleep diary. In Barton 1994, participants completed sleep records, with usual sleep onset and wake-up times associated with morning, afternoon, and night shifts. For sleep quality, they answered questions derived from the sleep quality index (Åkerstedt 1994). Totterdell 1992 assessed sleep duration by asking participants to recollect their usual sleep start and end times, and to rate their sleep quality on a visual analogue scale. Viitasalo 2008 used the Basic Nordic Sleep Questionnaire to measure sleep quality. Bell

2015 assessed sleep duration and quality with the Pittsburgh Sleep Quality Index (a validated instrument) but did not use objective measures such as actigraphy. It is unclear whether Knauth 1998 used a validated questionnaire to measure sleep duration. Axelsson 1998 used a sleep diary but it is not clear how reliable and complete the answers were.

We rated three studies at low risk of bias, as they measured sleep duration using validated objective instruments (Barger 2019a; Basner 2019; Garde 2020).

Sleepiness during shifts

Totterdell 1992 assessed sleepiness during shifts by asking respondents to record how alert they normally felt at specified two-hour intervals during a typical morning, afternoon, and night shift. For each two-hour interval, there was a 10-cm visual analogue scale ranging from "drowsy" to "alert". Viitasalo 2008 employed the Basic Nordic Sleep Questionnaire and the Epworth Sleepiness Scale. As these instruments are sensitive to transitory fluctuations, we judged Totterdell 1992 and Viitasalo 2008 at a high risk of bias. Two studies employed objective methods for the assessment of sleepiness during shifts: a 10-minute psychomotor vigilance test in Amendola 2011 and a three-minute psychomotor vigilance test in Bell 2015. We considered these studies at low risk of bias for this outcome. Three studies used the Karolinska Sleepiness Scale (Axelsson 1998; Barger 2019a; Basner 2019), which we also considered sufficiently reliable to warrant a low risk of bias judgement.

Baseline differences between groups

We judged three CBA trials at high risk of bias in relation to baseline differences (Barton 1994; Totterdell 1992; Viitasalo 2008), and two CBA trials at low risk (Bell 2015; Knauth 1998). In Barton 1994, the intervention and control groups experienced different sleep onset time at baseline, which may be related to the personal characteristics of the participants, such as age and chronotype. In Totterdell 1992, the intervention and control groups had different sleep start times and sleep duration, which may reflect differences related to chronotype and sleep pattern. In Viitasalo 2008, study groups were heterogeneous in age, occupational position, years of shift work, alcohol intake, and smoking; the study authors did not adjust for these differences in the statistical analyses. In Bell 2015, the study arms were balanced in terms of age, sex, number of children, number of children living in the home, age of the youngest child living in the home, ethnicity, and medications. In Knauth 1998, there were no important differences in age, marital status, or number of children.

Appropriateness of statistical analyses

Controlled before-after trials and non-randomised cross-over trials

We considered three of the five CBA trials at low risk of bias for this domain, as all employed appropriate methods for statistical analysis, either by using baseline measures as covariates (Bell 2015) or by using repeated-measures analysis (Barton 1994; Totterdell 1992; Viitasalo 2008). Knauth 1998 provided insufficient data for us to judge risk of bias.

Of the non-randomised cross-over trials, Garde 2020 used repeated-measures analysis of variance (ANOVA) with a random intercept for each individual to account for within-subject variation, and Axelsson 1998 used repeated-measures ANOVA with two, three, or four within-group factors. We judged both of these studies at low risk of bias for this domain.

Studies with cluster allocation

Two studies that enroled participants by clusters did not take the unit of analysis error into account (Bell 2015; Totterdell 1992); we judged both studies at high risk of bias for this domain. Two of the RCTs did adjust for the clustering effect so were at low risk of bias (Barger 2019a; Basner 2019).

Recruitment bias

Studies with cluster allocation

Bell 2015 and Totterdell 1992 did not report methods for the choice of allocation of clusters into the study arms. We considered both studies at unclear risk of bias for this domain. We judged Barger 2019a and Basner 2019 at low risk of bias.

Intervention independent of other changes over time

This domain applied to non-randomised trials. We judged all CBA trials and cross-over trials at low risk of bias for this domain, as no other changes were mentioned in the reports, and most follow-up periods were relatively short, meaning no major changes were expected (Axelsson 1998; Barton 1994; Bell 2015; Garde 2020; Knauth 1998; Totterdell 1992; Viitasalo 2008).

We judged all CBA trials (Barton 1994; Bell 2015; Knauth 1998; Totterdell 1992; Viitasalo 2008) and cross-over trials (Axelsson 1998; Garde 2020) at low risk of bias for this domain, as there were no changes reported in the data collection methods during the studies.

Intervention unlikely to affect data collection

This domain was applied to non-randomised trials. We judged all CBA trials and cross-over trials at low risk of bias, because there were no changes in data collection between measurements in any of the studies (Axelsson 1998; Barton 1994; Bell 2015; Garde 2020; Knauth 1998; Totterdell 1992; Viitasalo 2008).

Effects of interventions

See: **Summary of findings 1** Summary of findings table - Forward rotation compared to backward rotation for improving sleep and reducing sleepiness among shift workers; **Summary of findings 2** Summary of findings table - Faster rotation compared to slower rotation for shift workers; **Summary of findings 3** Summary of findings table - Shift duration of no more than 16 hours compared to shift duration of 24 to 28 hours for improving sleep and reducing sleepiness among shift workers; **Summary of findings 4** Summary of findings table - Shorter shifts (8 or 10 hours) compared to shifts lasting 2 to 3 hours longer for improving sleep and reducing sleepiness among shift workers; **Summary of findings 5** Summary of findings table - More compressed schedules compared to less compressed schedules for shift workers

The organisation of shift systems encompasses distinct components, such as the direction of shift changes, shift durations, and intervals between shifts. Below, we describe the findings of each distinct component of the shift system, combined and separately.

Permanency of shifts: fixed versus any rotation

No studies evaluated the permanency of shifts.

Regularity of shift changes: regular (predictable) changes versus irregular (unpredictable) changes

No studies evaluated the regularity of shift changes.

Direction of shift rotation: forward versus backward rotation

Three CBA studies investigated the direction of rotation in relation to a range of sleep outcomes (Barton 1994; Knauth 1998; Viitasalo 2008).

Barton 1994 compared two groups of shift workers: one group who changed from a backward to a forward rotation schedule, and a control group that kept the original schedule with backward rotation. We did not consider the third study arm, comprising daytime workers.

Viitasalo 2008 employed two study arms relevant for this comparison: one intervention group with shift changes forward and no days off between shift changes, and the control group with a backward shift system and days off between shift changes. A third arm involved the same backward shift system with a certain level of flexibility for scheduling; we did not use data from this trial arm. The investigators took measurements five to six months before and seven to eight months after implementation of the intervention.

Knauth 1998 had two intervention groups with forward rotation schedules and two control groups with backward rotation schedules. The investigators took measurements before and 10 months after implementation of the intervention.

Sleep quality off-shift

Randomised trials

We found no randomised trials for this comparison.

Non-randomised trials

All three non-randomised studies measured sleep quality off-shift (Barton 1994; Knauth 1998; Viitasalo 2008). Barton 1994 found no clear difference between the groups, but we could not pool the data from this study in the meta-analysis as it did not report SDs. Knauth 1998 found no difference but provided no data to support this finding. Very low-certainty evidence from Viitasalo 2008 suggested that forward rotation reduces sleep quality (MD −0.20 points (on a scale of 1 to 5), 95% CI −2.28 to 1.89; Analysis 1.1). We downgraded the certainty of the evidence for imprecision.

Sleep duration off-shift

Randomised trials

We found no randomised trials for this comparison.

Non-randomised trials

All three non-randomised studies measured sleep duration off-shift (Barton 1994; Knauth 1998; Viitasalo 2008). Barton 1994 showed that sleep duration was reduced in the forward rotation group from 7.34 hours at baseline to 7.14 hours six months later, while the backward rotation exhibited similar reductions (7.53 hours to 7.25 hours). Knauth 1998 found no statistically significant differences in sleep duration between the groups, but provided no supporting data. As contact with the study authors to obtain additional information was unsuccessful, we could not include Knauth 1998 in the meta-analysis.

Very low-certainty evidence from Viitasalo 2008 showed no difference in sleep duration amongst participants in the forward rotation system compared to those with backward rotations (MD −0.21 hours, 95% CI −3.29 to 2.88; Analysis 1.2). We downgraded the certainty of the evidence for imprecision and high risk of bias.

Sleepiness during shifts

Randomised trials

We found no randomised trials for this comparison.

Non-randomised trials

Only Viitasalo 2008 measured sleepiness during shifts, using items of the Basic Nordic Sleep Questionnaire (scale of 1 to 5). Very low-certainty evidence suggests that shift workers with forward rotations have lower levels of sleepiness during shifts compared to those with backward rotations (MD −1.24 points, 95% CI −2.24 to −0.24; Analysis 1.3). We downgraded the certainty of the evidence for imprecision and high risk of bias. In addition, the mean number of days per week when workers reported sleepiness decreased amongst those in the forward-rotating shift system, from 2.89 days (SD 2.09) to 2.08 days (SD 1.77), a change of −0.81 days (SD 0.58); whereas participants in the control group experienced sleepiness on more days per week after implementation of the intervention (mean 2.33 days, SD 2.11) than before the intervention (mean 1.90 days, SD 1.59), with a change of 0.43 days (SD 0.36).

Secondary outcomes

Randomised trials

We found no randomised trials for this comparison.

Non-randomised trials

No non-randomised trials reported any of our secondary outcomes.

Speed of rotation: faster versus slower rotation

Three non-randomised studies evaluated faster versus slower shift rotation (Garde 2020; Knauth 1998; Viitasalo 2008).

In Garde 2020, all participants were exposed to three different work schedules: two night shifts followed by two recovery days (fast rotation), four night shifts followed by four recovery days (slow rotation), and seven night shifts followed by seven recovery days (very slow rotation). Recovery days could be days off or day shifts. We compared the fast rotation with slow rotation and with very slow rotation.

Knauth 1998 had two intervention groups with faster rotation schedules and two control groups with slower rotation schedules.

Viitasalo 2008 compared very fast forward rotation (rotation after every shift with no days off between shift changes) with fast backward rotation (rotation every three shifts with two days off between shift changes).

Sleep quality off-shift

Randomised trials

We found no randomised trials for this comparison.

Non-randomised trials

Garde 2020 and Viitasalo 2008 reported sleep quality off-shift. Low-certainty evidence suggests that faster rotation (every one or two days) compared with slower rotation (every three, four, or seven days) has no effect on sleep quality (SMD −0.01, 95% CI −0.26 to 0.23; Analysis 2.1).

Sleep duration off-shift

Randomised trials

We found no randomised trials for this comparison.

Non-randomised trials

All three non-randomised trials reported sleep duration off-shift (Garde 2020; Knauth 1998; Viitasalo 2008).

Knauth 1998 found no statistically significant differences in sleep duration between the control and intervention groups, but provided no supporting data. As contact with the study authors to obtain additional information was unsuccessful, we could not include Knauth 1998 in the meta-analysis.

Very low-certainty evidence resulting from the meta-analysis of data from Garde 2020 and Viitasalo 2008 suggests that faster rotations reduce sleep duration compared with slower rotations (SMD −0.26; 95% CI −0.51 to −0.01; Analysis 2.2). We downgraded the certainty of the evidence for inconsistency.

Sleepiness during shifts

Randomised trials

We found no randomised trials for this comparison.

Non-randomised trials

Only Viitasalo 2008 reported sleepiness during trials. Very low-certainty evidence suggests that faster rotations reduce sleepiness during shifts compared with slower rotations (MD −1.24 (on a scale of 1 to 5), 95% CI −2.24 to −0.24; Analysis 2.3). We downgraded the certainty of the evidence for high risk of bias and imprecision. In addition, the mean number of days per week when workers reported sleepiness decreased amongst those with very fast rotation, from a mean of 2.08 days (SD 1.77) to 2.89 days (SD 2.09), a change of −0.81 days (SD 0.58); whereas participants in the control group experienced sleepiness on more days per week after implementation of the intervention (mean 2.33 days, SD 2.11) than before the intervention (mean 1.90 days, SD 1.59), with a change of 0.43 days (SD 0.36).

Secondary outcomes

Randomised trials

We found no randomised trials for this comparison.

Non-randomised trials

No non-randomised trials reported any of our secondary outcomes.

Shift duration: shorter versus longer

All three RCTs (Amendola 2011; Barger 2019a; Basner 2019) and two non-randomised studies (Axelsson 1998; Bell 2015) compared shift systems with different shift duration. Amendola 2011 compared the effects of three shift durations (eight hours, 10 hours, 12 hours). Barger 2019a and Basner 2019 compared long overnight shifts of up to 28 hours to shift schedules with a maximum shift duration of 16 hours. The intervention group in Basner 2019 also had a minimum number of rest hours between shifts (at least 14 rest hours after a 24-hour in-house duty and at least eight rest hours after a regular shift). Axelsson 1998 compared 12-hour night shifts with eight-hour night shifts, and Bell 2015 compared three consecutive shifts lasting 13 hours 20 minutes with four consecutive shifts lasting 10 hours.

No more than 16 hours versus 24 to 28 hours

Sleep quality off-shift

No studies reported sleep quality of shift.

Sleep duration off-shift

Randomised trials

Barger 2019a and Basner 2019 reported sleep duration off-shift. Participants who worked overnight shifts lasting no longer than 16 hours slept more hours per week (in Barger 2019a) or during (over)night shifts (in Basner 2019) than those who worked overnight shifts of 24 to 28 hours. Low-certainty evidence resulting from the meta-analysis of data from these studies suggests that working no more than 16 hours in an overnight shift, compared with working 24 to 28 hours, increases sleep duration off-shift (SMD 0.50, 95% CI 0.21 to 0.78; Analysis 3.1). We downgraded the certainty of the evidence for inconsistency and indirectness.

Non-randomised trials

No non-randomised trials reported sleep duration off-shift.

Sleepiness during shifts

Randomised trials

Barger 2019a and Basner 2019 reported sleepiness during shifts. Participants with shorter overnight shifts had lower sleepiness ratings than those who worked very long overnight shifts. However, moderate-certainty evidence resulting from the meta-analysis showed no clear difference between the groups (SMD −0.29, 95% CI −0.44 to −0.14; Analysis 3.2). We downgraded the certainty of the evidence for indirectness.

Non-randomised trials

No non-randomised trials reported sleepiness during shifts.

Secondary outcomes

Randomised trials

Participants in Barger 2019a who worked overnight shifts of no more than 16 hours had fewer working hours per week than the control group (MD −6.50 hours, 95% CI −7.73 to −5.27; Analysis 3.3), though this evidence is of very low certainty. We downgraded the certainty of evidence for indirectness and imprecision.

Non-randomised trials

No non-randomised trials reported any of our secondary outcomes.

Shorter shifts (eight hours or 10 hours) versus shifts lasting two to three hours longer

Sleep quality off-shift

Randomised trials

Amendola 2011 reported sleep quality off-shift. Participants indicated their average quality of sleep as "good". The study authors found little variation and no significant differences between the groups (Cohen's f = 0.09). However, the report did not include SDs for all outcome measures, and our attempts to contact the corresponding author were unsuccessful.

Non-randomised trials

Axelsson 1998 and Bell 2015 reported sleep quality off-shift. Axelsson 1998 found no difference in sleep quality between the participants who worked eight-hour shifts and those who worked 12-hour shifts. Bell 2015 found no significant differences between the groups at month three (mean 1.67 points (SD 0.13) in those who worked shifts of 13 hours 20 minutes versus mean 1.11 points (SD 0.11) in those who worked shifts of 10 hours; P = 0.512; $F_{1,348}$ = 0.431). At month six, Bell 2015 reported that the participants working longer shifts exhibited worse sleep quality compared to those who worked shorter shifts, but this did not take into account the unit-of-analysis error due to clustering of participants. We meta-analysed the six-month follow-up data from Bell 2015 and the results of Axelsson 1998. Very low-certainty evidence suggests there is no clear difference between shorter and longer shifts in terms of sleep quality off-shift (SMD −0.23, 95% CI −0.61 to 0.15; Analysis 4.1). We downgraded the certainty of the evidence for insufficient statistical adjustment for confounding or cluster allocation.

Sleep duration off-shift

Randomised trials

Amendola 2011 reported sleep duration off-shift. Participants working 10-hour shifts exhibited higher average hours of sleep (7.86 hours) compared to those allocated to eight-hour shifts (7.35 hours) and 12-hour shifts (7.55 hours), after controlling for the effect of the average number of hours of sleep at baseline. The strength of the association of the effect on shift duration using the Cohen's f effect size index indicated a small to medium effect (f = 0.19). Amendola 2011 reported that there was no statistically significant difference in sleep duration between participants working 12-hour shifts compared to those allocated to eight-hour shifts.

Non-randomised trials

Axelsson 1998 and Bell 2015 reported sleep duration off-shift. Axelsson 1998 found no significant difference in off-shift sleep duration between the eight-hour and 12-hour groups. Bell 2015 reported that participants working shifts of 13 hours 20 minutes obtained significantly fewer hours of sleep than those working 10-hour shifts, but the study authors did not take into account the unit-of-analysis error. Very low-certainty evidence resulting from the meta-analysis indicated no clear difference between shorter shifts and longer shifts in terms of sleep duration off-shift (SMD 0.18, 95% CI −0.17 to 0.54; Analysis 4.2). We downgraded the certainty of the evidence for imprecision and insufficient statistical adjustments for confounding or cluster allocation.

Sleepiness during shifts

Randomised trials

Amendola 2011 reported sleepiness during shifts. There were no significant differences across groups for the measures of the psychomotor vigilance test and the ocular tracker, but the study did find a significant effect of shift duration on sleepiness reported with the Karolinska Sleepiness Scale. Mean level of alertness was significantly lower in participants working 12-hour shifts than in those working eight-hour shifts (6.11 points versus 6.74 points, P = 0.012). However, there was no significant difference with the 10-hour shift (mean 6.31 points).

Non-randomised trials

Axelsson 1998 and Bell 2015 reported sleepiness during shifts. In Axelsson 1998, sleepiness was higher in participants on the 12-hour night shift than in those on the eight-hour night shift (SMD −1.06 95% CI −1.59 to −0.52; Analysis 4.3). Bell 2015 applied an objective and a subjective measure of sleepiness. Without taking the unit-of-analysis error into account, the study authors found that both scores were significantly higher in the group working shifts of 13 hours 20 minutes compared to those working 10-hour shifts. After correcting for unit-of-analysis error, we found no effect (SMD for objective measure −0.06, 95% CI −0.57 to −0.45; Analysis 4.3). We did not perform meta-analysis owing to substantial heterogeneity (Tau = 0.43, P = 0.008; I^2 = 86%). The certainty of the evidence from both trials was very low. We downgraded the certainty of the evidence for imprecision, inconsistency, and insufficient statistical adjustments for confounding or cluster allocation.

Secondary outcomes

Randomised trials

Amendola 2011 assessed overtime over a six-month period. The study authors reported that participants working eight-hour shifts did significantly more overtime hours (5.75 hours) compared with group working 10-hour shifts (mean 5.75 hours versus mean 0.97 hours, P = 0.000) and the group working 12-hour shifts (mean 5.75 hours versus mean 1.89 hours, P = 0.000).

Non-randomised trials

Bell 2015 compared overtime hours worked during a six-month study period with hours worked during the same six-month period in the previous year. Participants in the 10-hour shift group worked significantly more overtime during the six-month study period compared with participants who worked shifts of 13 hours 20 minutes (mean 9.02 hours versus mean 6.89 hours; MD 1.22, 95% CI 0.94 to 1.50; Analysis 4.4). This evidence was of very low certainty; we downgraded for imprecision, indirectness, and insufficient statistical adjustments for confounding or cluster allocation. The participants also worked more overtime during the year prior to the intervention.

Timing of start: earlier versus later start

No studies evaluated start time of shifts.

Distribution of shift schedule: more compressed versus more spread out

One RCT (Amendola 2011) and one non-randomised study (Totterdell 1992) evaluated more compressed shift schedules versus more spread out shift schedules. Amendola 2011 compared

three groups: intervention group 1 worked four consecutive 10-hour days, followed by three days off; intervention group 2 had a more compressed schedule, with three consecutive 12-hour days followed by four days off; and the control group followed a more traditional, less compressed schedule, with five consecutive eight-hour days followed by two days off. Totterdell 1992 evaluated a compressed shift system with 10-hour morning and afternoon shifts and 8.5-to-9-hour night shifts versus a standard, more spread out shift system with eight-hour shifts.

Sleep quality off-shift

Randomised trials

Amendola 2011 measured sleep quality though sleep diaries, where participants recorded their perceived quality of sleep on a daily basis. There were no statistically significant differences between groups. All participants indicated their average quality of sleep as "good", with little variation. The study provided no further results.

Non-randomised trials

Totterdell 1992 observed no difference in sleep quality between the intervention and control group (MD 0.31 points (on a scale of 0 to 10), 95% CI −0.53 to 1.15; Analysis 5.1). This is very low-certainty evidence; we downgraded for imprecision, high risk of bias, and no adjustment for cluster allocation.

Sleep duration off-shift

Randomised trials

Amendola 2011 reported no statistically significant difference in sleep duration off-shift between the least compressed eight-hour shift system (mean 7.35 hours) and the most compressed 12-hour shift system (mean 7.55 hours).

Non-randomised trials

In Totterdell 1992, participants working the more compressed work schedule slept longer between night shifts, but the analysis did not take clustering of participants into account ($F_{1,59}$ = 4-24; P = 0.05). After adjusting for clustering, we found no difference between the groups (MD 0.52 hours, 95% CI −0.52 to 1.56; Analysis 5.2). This was very low-certainty evidence; we downgraded for imprecision, high risk of bias, and lack of adjustment for cluster allocation. Totterdell 1992 found no difference between the groups in terms of sleep duration between the morning and the afternoon shifts.

Sleepiness during shifts

Randomised trials

Amendola 2011 found no statistically significant differences across groups for the objective measures of sleepiness (psychomotor vigilance test and ocular tracker). For the subjective measure (Karolinska Sleepiness Scale), participants with the more compressed work schedule reported lower levels of alertness than the participants with the least compressed work schedule (mean 6.11 points versus mean 6.74 points, P = 0.012).

Non-randomised trials

Totterdell 1992 applied an alertness rating before the intervention and six months into the intervention. Without taking clustering into account, the study authors found that the more compressed work schedule was associated with higher alertness levels at the end of a shift and also at the start of the early shift. We were unable to include data from Totterdell 1992 in a meta-analysis owing to missing variance measures.

Secondary outcomes

Randomised trials

Amendola 2011 evaluated the effect of less and more compressed work schedules on overtime hours and reported that overtime hours were significantly greater amongst participants with the least compressed schedule, who worked eight-hour shifts, than amongst participants with the more compressed schedules, who worked 10- and 12-hour shifts (5.75 hours versus 0.97 hours versus 1.89 hours; P < 0.001).

Non-randomised trials

Totterdell 1992 did not reported any of our secondary outcomes.

Time off between shifts: longer versus shorter rest

No studies evaluated time off between shifts.

Split shifts: non-interrupted versus interrupted shifts

No studies evaluated split shifts.

Protected sleep: no on-call duties versus on-call duties

No studies evaluated protected sleep.

Worker participation: participative versus non-participative scheduling

No studies evaluated worker participation.

DISCUSSION

Summary of main results

We found very low-certainty evidence that forward rotation reduces sleepiness, and that it has no effect on sleep quality and sleep duration. One CBA trial observed clinically relevant lower sleepiness during shifts for forward rotation in combination with faster rotation, reported with the Basic Nordic Sleep Questionnaire on a scale of one to five points (MD −1.24 points, 95% CI −2.24 to −0.24).

We found very low-certainty evidence that faster rotation reduces sleep duration off-shift, but may also reduce sleepiness during shifts. There was very low-certainty evidence that rotation speed has no effect on sleep quality. The pooled results of one CBA trial and one non-randomised cross-over trial associated faster rotation with less sleep (SMD −0.26, 95% CI −0.51 to −0.01; which translated to an MD of 0.38 hours less per day, 95% CI −0.74 to −0.01). The CBA trial also showed that faster rotation in combination with forward rotation reduced sleepiness during shifts, reported with the Basic Nordic Sleep Questionnaire on a scale of one to five points (MD −1.24 points, 95% CI −2.24 to −0.24).

We found low-certainty evidence from two RCTs that on-duty 80-hour workweeks with shift duration limited to 16 hours led to clinically relevant increases in sleep duration compared to workweeks with no limits on sleep duration (SMD 0.50, 95% CI 0.21 to 0.78). This translated to an MD of 0.73 hours per day (95% CI 0.30 to 1.13; 95% CI based on an SD of 1.45). To determine clinical relevance, we consulted meta-analyses and high-quality cohort studies based on epidemiological studies on the (dose-response)

relationship between sleep duration and major health outcomes (e.g. Cepeda 2016; Von Ruesten 2012; Wang 2016). After discussion with the authors of these reviews, we concluded that an increase of 0.73 hours for short sleepers would have a clinically relevant beneficial effect on a range of outcomes. In the comparison group shift schedule of the RCTs, participants worked shifts of up to 24 hours or 28 hours. We found moderate-certainty evidence from the same RCTs that on-duty workweeks with shift duration limited to 16 hours had a small and clinically irrelevant effect on sleepiness during shifts (SMD −0.29, 95% CI −0.44 to −0.14). This translated to an MD of 0.37 points (95% CI −0.55 to −0.17; 95% CI based on an SD of 1.56) on the Karolinska Sleepiness Scale (range 0 to 9 points).

We found one RCT, one CBA trial, and one non-randomised cross-over trial comparing shorter shift duration (eight hours to nine hours) to longer shift duration (10 hours to 13 hours). These studies provided very low-certainty evidence of no effect on sleep quality and sleep duration. The effect on sleepiness was inconsistent across the studies. The RCT and the non-randomised cross-over study found reduced sleepiness in shorter shifts, while the CBA trial found no effect.

One RCT and one CBA trial provided very low-certainty evidence of no effect of compressed shift schedules compared with more spread out shift schedules on sleep quality and duration off-shift.

No studies investigated the effect of other shift schedule changes (i.e. fixed versus any rotation, regular versus irregular changes, earlier versus later start of shifts, longer versus shorter rest between shifts, interrupted versus non-interrupted shifts, no on-call duties versus on-call duties, or participative versus non-participative scheduling) on sleep or sleepiness. Nor did we find much evidence on fatigue or any of the secondary outcomes (i.e. staff levels, overtime hours, staffing costs).

Overall completeness and applicability of evidence

This review sought to establish evidence to support changes to shift schedules for improving sleep and decreasing sleepiness. We searched for evidence for eleven shift schedule components that experts consider could have a positive effect on sleep and sleepiness, but found mainly very low-certainty evidence for only four of those components. We found only two high-quality RCTs with sufficient data to include in our meta-analysis (Barger 2019a; Basner 2019). They investigated a very specific shift system: an 80-hour workweek with or without limits to the maximum shift duration. Although they provided moderate-certainty evidence that limiting maximum shift duration reduces sleepiness during shifts, this finding may only be relevant for the healthcare sector, and for a few countries (including the USA, where the trial took place) where workers have such demanding workweeks.

The evidence for direction and speed of rotation, as well as distribution of shift schedules, is based on non-randomised studies; we found no published RCTs on this topic. Moreover, we found no trials evaluating permanency of shifts, regularity of shift changes, timing of shift start, time off between shifts, split shifts, on-call duties, and worker participation in relation to sleep and sleepiness. This indicates that current recommendations to adapt shift schedules to improve sleep and reduce sleepiness, based on expert opinion and observational cross-sectional and cohort studies, cannot be substantiated with high-certainty evidence (Driscoll 2007; Garde 2020; ILO 2004; Knauth 1995).

Most studies included (almost) only men, and focused on a few areas of work (i.e. car manufacturing, police force, maintenance unit of an airline company, power plant, and steel industry; Amendola 2011; Axelsson 1998; Bell 2015; Garde 2020; Knauth 1998; Totterdell 1992; Viitasalo 2008). Only the two RCTs investigating the 80-hour workweeks included a substantial number of women and were conducted in the healthcare sector (Barger 2019a; Basner 2019). Research in other sectors with shift work (e.g. security, hospitality, construction, transportation, and storage) is currently lacking.

Quality of the evidence

This review included three RCTs and seven non-randomised trials. We could only include two RCTs and five non-randomised trials in the meta-analysis. Two RCTs provided moderate-certainty evidence that limiting shift duration to 16 hours affected sleepiness in the specific context of an 80-hour workweek. For other shift components, we only found low- or very low-certainty evidence. The main reasons for downgrading the certainty of the evidence were imprecision (low sample sizes) and subjective measurements of the outcomes.

The nature of the intervention itself imposed difficulties in blinding personnel and participants. All studies employed subjective measurement of the outcomes of interest, which were supplemented by objective measures in only four studies. We considered that self-reported scales introduced high risk of bias, as participants' awareness of which group they were in might have influenced their responses. Five studies were at high risk of bias, and five were at unclear risk, in relation to completeness of data. Most studies had no reporting bias or other sources of bias, used validated measures to measure sleepiness, conducted appropriate statistical analysis, and conducted the intervention independently of other changes over time. These items were mostly judged as having a low risk of bias.

Two CBA trials used cluster allocation, but did not correct for the unit-of-analysis error (Bell 2015; Totterdell 1992). In our meta-analysis, we performed this correction and calculated effective sample sizes, resulting in less statistical power and highly imprecise results. Some outcomes reported as statistically significant in the study reports became non-significant after correction in our analysis.

Potential biases in the review process

Shift systems are complex and consist of multiple components. We compared the components that experts hypothesise have a beneficial effect, and that are in line with current recommendations. However, changes in one component of a shift system most often occurs simultaneously with other changes in the shift system, which may either be beneficial or adverse. This is often unavoidable and may have influenced the results of our review. If multiple changes occur, it is difficult to disentangle what shift schedule component has caused the particular effect.

We chose to include CBA trials and non-randomised cross-over trials, as we expected there would be few RCTs in the literature (and we found only three). The inclusion of non-randomised studies affected the assessment of the certainty of the evidence, as GRADE ratings for non-randomised studies start at low certainty.

We were unable to retrieve the full text of two studies to assess their relevance for inclusion. We had planned to investigate selective reporting by constructing and analysing funnel plots and perform the Egger test (Egger 1997). Owing to the low number of studies, we were unable to do so, and cannot draw strong conclusions regarding reporting bias. However, we do not have strong indications that it exists.

Sleepiness is a complex, multi-dimensional phenomenon that includes temporal elements (Shen 2006). This is evident in the included studies, which measured outcomes that ranged from immediate states of sleepiness on an interval scale (e.g. visual analogue scale or Karolinska Sleepiness Scale) to a broader time frame such as the usual chance of falling asleep during typical tasks (e.g. Epworth Sleepiness Scale). Measurement methods also ranged from subjective self-assessments to objective measures such as those with reaction tests. Given the heterogeneity of sleepiness measures, it is difficult to combine results of different studies and draw definitive conclusions or recommendations. As such, more future studies are needed to allow for comparison of individual domains of sleepiness.

Owing to the limited available data, we were unable to perform subgroup and sensitivity analyses with regard to chronotype, occupational setting, different ways of measuring the outcome, age, and each domain of risk of bias assessment as planned.

Agreements and disagreements with other studies or reviews

We found no systematic reviews with the same research question to compare our results with. Previous (mostly non-systematic) reviews on similar topics identified no RCTs and based their findings on non-randomised trials and observational (cross-sectional and cohort) studies (Akerstedt 1998; Bambra 2008a; Bambra 2008b; Driscoll 2007; Hanifah 2021; Harris 2015; Sallinen 2010). These reviews indicated beneficial effects on some sleep outcomes of faster rotation compared to slower rotation and forward rotation compared to backward rotation (Bambra 2008a; Driscoll 2007; Hanifah 2021Sallinen 2010). Their conclusions were based on a few observational studies only. Previous non-systematic reviews also indicated that observational studies consistently suggest that short rests between shifts (less than 11 hours) are detrimental to sleep (Akerstedt 1998; Sallinen 2010). In line with this review, most previous reviews found no consistent effects of shift duration and compressed workweeks on sleep outcomes (Bambra 2008a; Bambra 2008b; Driscoll 2007; Harris 2015).

We intended to include laboratory studies, and found only one controlled laboratory study that met our eligibility criteria (Cruz 2003). This study assigned 28 participants to one of two interventions: a forward-rotating shift schedule or a backward-rotating shift schedule (Cruz 2003). This study found no significant difference in subjective sleep duration based on sleep logs across morning shifts, afternoon shifts, night shifts, and recovery days between forward and backward rotation. Similarly, objectively measured sleep duration did not differ significantly between rotation conditions. However, the study did not provide details of statistical testing. The study measured sleepiness during shifts with the Stanford Sleepiness Scale. Individuals in the backward rotation group reported significantly higher sleepiness at the end of the midnight shift during the first of the two weeks of shift work than individuals in the forward rotation group ($F_{1,26}$ = 4.8, P < 0.05). The findings of Cruz 2003 are similar to the findings of this review, as our meta-analyses also suggested that forward rotation may reduce sleepiness, but may have no effect on sleep quality or sleep duration.

AUTHORS' CONCLUSIONS

Implications for practice

The available evidence indicates that some changes to shift systems may have a small effect on sleepiness and sleep, but evidence for effects on sleep quality is absent or uncertain. The literature provides very low-certainty evidence that forward rotation compared with backward rotation, and faster rotation compared with slower rotation, results in a clinically relevant reduction in sleepiness during shifts. Although evidence from intervention studies is still limited, no harmful effects were reported in interventions associated with rapid forward rotation of night shifts. Due to lack of clear evidence, it is particularly important for organisations that change their shift schedule to involve employees and to evaluate the effects on employees.

Moderate-certainty evidence indicates that setting shift limits at 16 hours during 80-hour workweeks reduces sleepiness in resident physicians and interns in medical units, but this effect was small and not clinically relevant. For these shift schedule adaptations, there was also low-certainty evidence of a clinically relevant increase in sleep after night shifts. Based upon all available evidence on sleep and sleepiness, organisations should be extremely careful with long workweeks and shift duration exceeding 16 hours.

We found very low-certainty evidence that changes in shift duration or compression of workweeks have no effect on sleep and sleepiness. For the other shift schedule changes we found no evidence at all. As such, we cannot provide additional shift schedule recommendations.

When implementing or changing a shift schedule, it is important to consider the effects of shift schedules on other outcomes such as fatigue, mental health, and cardiometabolic parameters, as well as work productivity and feasibility.

Implications for research

Most evidence in the literature was of low or very low certainty due to the scarcity of randomised controlled trials (RCTs) and other methodological limitations of the studies, including small sample sizes and subjective measurements of sleep and sleepiness. This implies that the true effect of shift schedule interventions could be substantially different from our estimates. High-quality cluster-RCTs conducted with objective and validated measurements of the outcomes and reported in line with CONSORT Cluster are needed to establish how shift schedules can be adapted to promote sleep and reduce sleepiness (Campbell 2012). These trials should be carried out systematically across work sectors (e.g. health, transportation) and geographic regions, documenting worker chronotype, age, and autonomy over working times to address potential interactions.

The studies included in this review used a wide range of instruments to measure sleep and sleepiness, and assessed these outcomes at different times across the control and intervention period. To better synthesise future research and make evidence more useful, researchers should use standard measurement

instruments and protocols. We consider that shift work research would benefit from the development of a core outcome set.

There was a great variability regarding shift across studies. The lack of standardisation of working hours in the intervention and comparator arms was anticipated, considering real-life variability of work schedules across economic sectors and geographic regions, but it hampered a broader assessment of the impact of shift duration on the outcomes of interest. Future research should aim to employ standardised shift durations that could provide more robust and meaningful evidence.

ACKNOWLEDGEMENTS

We thank all authors of the previously published protocol.

The Cochrane Work review group supported the authors in the development of this review. The following people conducted the editorial process for this article.

- Sign-off Editor (final editorial decision): Dr Jan Hoving, Cochrane Work Co-ordinating Editor, and Assistant Professor at Amsterdam UMC, Amsterdam
- Handling Editor (content expertise review): Dr Jos Verbeek, Senior Researcher for Cochrane Work, Amsterdam UMC, Finland
- Managing Editor (selected peer reviewers, collated peer-reviewer comments, provided editorial guidance to authors, edited the article, prepared for Production): Julitta Boschman, Amsterdam UMC, Amsterdam; and Jessica Thomas, Amsterdam UMC, UK
- Information Specialist (searching support): Joost Daams, Amsterdam UMC, Amsterdam. We thank Cochrane Work's former Information Specialists for revising search strategies.
- Copy-editing (initial copy-edit and proofread): Julia Turner, c/o Cochrane Central Production Service

We thank the following peer reviewers for their comments on the manuscript.

- Teddy Oosterhuis, Netherlands Society of Occupational Medicine (Consumer Review)
- Claudia RC Moreno, University of Sao Paulo, Brazil (Clinical Review)

We also thank authors of included studies for having kindly provided requested data, and Jolanda ter Laak for kindly conducting a search update.

REFERENCES

References to studies included in this review

Amendola 2011 *{published data only}*

Amendola KL, Weisburd D, Hamilton EE, Jones G, Slipka M. An experimental study of compressed work schedules in policing: advantages and disadvantages of various shift lengths. *Journal of Experimental Criminology* 2011;**7**(4, SI):407-42.

Axelsson 1998 *{published data only}*

Axelsson J, Kecklund G, Akerstedt T, Lowden A. Effects of alternating 8- and 12-hour shifts on sleep, sleepiness, physical effort and performance. *Scandinavian Journal of Work, Environment & Health* 1998;**24**:62-8.

Barger 2019a *{published data only}*

Barger LK, Sullivan JP, Blackwell T, O'Brien CS, St Hilaire MA, Rahman SA. Effects on resident work hours, sleep duration, and work experience in a randomized order safety trial evaluating resident-physician schedules(ROSTERS). *Sleep* 2019;**42**(8):1-10.

Rahman SA, Sullivan JP, Barger LK, St Hilaire MA, O'Brien CS, Stone KL. Extended work shifts and neurobehavioral performance in resident-physicians. *Pediatrics* 2021;**147**(3):e2020009936.

Barton 1994 *{published data only}*

Barton J, Folkard S, Smith L, Poole C JM. Effects on health of a change from a delaying to an advancing shift system. *Occupational and Environmental Medicine* 1994;**51**(11):749-55.

Basner 2019 *{published data only}*

Basner M, Asch DA, Shea JA, Bellini LM, Carlin M, Ecker AJ, et al. Sleep and alertness in a duty-hour flexibility trial in internal medicine. *New England Journal of Medicine* 2019;**380**:915-23. [DOI: 10.1056/NEJMoa1810641]

Bell 2015 *{published data only}*

Bell LB, Virden TB, Lewis DJ, Cassidy BA. Effects of 13-hour 20-minute work shifts on law enforcement officers' sleep, cognitive abilities, health, quality of life, and work performance: the Phoenix study. *Police Quarterly* 2015;**18**(3):293-337.

Cruz 2003 *{published data only}*

Cruz C, Detwiler C, Nesthus T, Boquet A. Clockwise and counterclockwise rotating shifts: effects on sleep duration, riming, and quality. *Aviation, Space, and Environmental Medicine* 2003;**74**(6):597-605.

Garde 2020 *{published data only}*

Garde AH, Nabe-Nielsen K, Jensen MA, Kristiansen J, Sørensen JK, Hansen ÅM. The effects of the number of consecutive night shifts on sleep duration and quality. *Scandinavian Journal of Work, Environment & Health* 2020;**46**(4):446-53. [DOI: https://doi.org/10.5271/sjweh.3885]

Knauth 1998 *{published data only}*

* Knauth P, Hornberger S. Changes from weekly backward to quicker forward rotating shift systems in the steel industry. *International Journal of Industrial Ergonomics* 1998;**21**(3):267-73.

Totterdell 1992 *{published data only}*

Totterdell P, Smith L. Ten-hour days and eight-hour nights: can the Ottawa shift system reduce the problems of shiftwork? *Work & Stress* 1992;**6**(2):139-52.

Viitasalo 2008 *{published data only}*

Viitasalo K, Kuosma E, Laitinen J, Härmä M. Effects of shift rotation and the flexibility of a shift system on daytime alertness and cardiovascular risk factors. *Scandinavian Journal of Work, Environment and Health, Supplement* 2008;**34**(3):198-205.

References to studies excluded from this review

Akersted 1978 *{published data only}*

Akerstedt T, Torsvall L. Experimental changes in shift schedules: their effects on well-being. *Ergonomics* 1978;**21**(10):849-56.

Barger 2019b *{published data only}*

Barger LK, Sullivan JP, Blackwell T, O'Brien CS, St Hilaire MA, Rahman S. Resident physician work hours decreased and sleep duration increased following elimination of scheduled extended duration shifts. In: Sleep. Vol. 42. 2019:A390.

Blackwell 2019 *{published data only}*

Blackwell T, Kriesel DR, Vittinghoff E, Brien Cs, Sullivan JP, Viyaran NC. Design and recruitment of the randomized order safety trial evaluating resident-physician schedules (ROSTERS) study. *Contemporary Clinical Trials* 2019;**80**:22-33.

Cappuccio 2009 *{published data only}*

Cappuccio FP, Bakewell A, Taggart FM, Ward G, Ji C, Sullivan JP, et al. Implementing a 48 h EWTD-compliant rota for junior doctors in the UK does not compromise patients safety: assessor-blind pilot comparison. *QJM-AN International Journal of Medicine* 2009;**102**(4):271-82.

Chang 2021 *{published data only}*

Chang WP, Wang CH. Influence of sleep fragmentation and fatigue on turnover of female nurses working rotating shifts. *Journal of Clinical Nursing* 2022;**31**:23-4.

Cheng 2021 *{published data only}*

Cheng WJ, Puttonen S, Vanttola P, Koskinen A, Kivimäki M, Härmä M. Association of shift work with mood disorders and sleep problems according to chronotype: a 17-year cohort study. *Chronobiology International* 2021;**38**(4):518-25.

Chiles 1968 *{published data only}*

Chiles WD, Alluisi EA, Adams OS. Work schedules and performance during confinement. Human factors 1968.

Costa 2014 *{published data only}*

Giovanni C, Matteo MA, Giovanna C, Silvia F, Luca N. Stress and sleep in nurses employed in "38" and "212" fast rotating shift

schedules. *Chronobiology International* 2014;**31**:1169-78. [DOI: 10.3109/07420528.2014.957309]

D'Amico 1985 *{published data only}*

D'Amico A. A simulation study of the effects of sleep deprivation, time of watch and length of time on watch on watchstanding effectiveness. National Maritime Research Center 1985;(7-B):2338.

Duchon 1994 *{published data only}*

Duchon JC, Keran CM, Smith TJ. Extended workdays in an underground mine: a work performance analysis. *Human Factors* 1994;**36**(2):258-68.

Duplessis 2007 *{published data only}*

Duplessis CA, Miller JC, Crepeau LJ, Osborn CM, Dyche J. Submarine watch schedules: underway evaluation of rotating (contemporary) and compressed (alternative) schedules. *Undersea & Hyperbaric Medicine* 2007;**34**(1):21-33.

Dutheil 2012 *{published data only}*

Dutheil F, Boudet G, Perrier C, Lac G, Ouchchane L, Chamoux A, et al. JOBSTRESS study: comparison of heart rate variability in emergency physicians working a 24-hour shift or a 14-hour night shift - a randomized trial. *International Journal of Cardiology* 2012;**158**(2):322-5.

Eriksen 2006 *{published data only}*

Eriksen CA, Gillberg M, Vestergren P. Sleepiness and sleep in a simulated "six hours on/six hours off" sea watch system. *Chronobiology International* 2006;**23**(6):1193-202.

Fischer 2021 *{published data only}*

Fischer D, Roenneberg T, Vetter C. Chronotype-specific sleep in two versus four consecutive shifts. *Journal of Biological Rhythms* 2021;**36**(4):395-409.

Garde 2011 *{published data only}*

Garde AH, Nabe-Nielsen K, Aust B. Influence on working hours among shift workers and effects on sleep quality: an intervention study. *Applied Ergonomics* 2011;**42**(2):238-43.

Grewal 2022 *{published data only}*

Grewal A, Holt KE. Comparison of 12-hour biweekly versus 6-hour weekly shifts on fatigue in student nurses. *Worldviews on Evidence-Based Nursing* 2022;**19**(2):160-5.

Harris 2010 *{published data only}*

Harris A, Waage S, Ursin H, Hansen AM, Bjorvatn B, Eriksen HR. Cortisol, reaction time test and health among offshore shift workers. *Psychoneuroendocrinology* 2010;**35**(9):1339-47.

Hong 2021 *{published data only}*

Hong J, Kim M, Suh EE, Cho S, Jang S. Comparison of fatigue, quality of life, turnover intention, and safety incident frequency between 2-shift and 3-shift Korean nurses. *International Journal of Environmental Research and Public* 2021;**18**(15):7953.

Hossain 2004 *{published data only}*

Hossain JL, Reinish LW, Heslegrave RJ, Hall GW, Kayumov L, Chung SA. Subjective and objective evaluation of sleep and performance in daytime versus nighttime sleep in extended-hours shift-workers at an underground mine. *Journal of Occupational and Environmental Medicine* 2004;**46**:212-26.

ISRCTN17016944 *{published data only}*

ISRCTN17016944. How can you sleep well and enjoy your lifestyle when you are a night worker? trialsearch.who.int/Trial2.aspx?TrialID=ISRCTN17016944 (first received 12 June 2019).

Jackson 2014 *{published data only}*

Jackson ML, Banks S, Belenky G. Investigation of the effectiveness of a split sleep schedule in sustaining sleep and maintaining performance. *Chronobiology International* 2014;**31**(10):1218-30.

Knauth 1987 *{published data only}*

Knauth P, Kiesswetter E. A change from weekly to quicker shift rotations: a field study of discontinuous three-shift workers. *Ergonomics* 1987;**30**(9):1311-21. [DOI: 10.1080/00140138708966025]

Kosmadopoulos 2014 *{published data only}*

Kosmadopoulos A, Sargent C, Darwent D, Zhou X, Dawson D, Roach GD. The effects of a split sleep-wake schedule on neurobehavioural performance and predictions of performance under conditions of forced desynchrony. *Chronobiology International* 2014;**31**(10):1209-17.

Kudielka 2007 *{published data only}*

Kudielka BM, Buchtal J, Uhde A, Wüst S. Circadian cortisol profiles and psychological self-reports in shift workers with and without recent change in the shift rotation system. *Biological Psychology* 2007;**74**(1):92-103.

Landrigan 2020 *{published data only}***10.1056/NEJMoa1900669**

Landrigan CP, Rahman SA, Sullivan JP, Vittinghoff E, Barger LK, Sanderson AL. Effect on patient safety of a resident physician schedule without 24-hour shifts. *New England Journal of Medicine* 2020;**382**(26):2514-23. [DOI: 10.1056/NEJMoa1900669]

Stone KL, Landrigan CP, Blackwell T, Rahman SA, Kriesel DR, Vittinghoff E, et al. In: Sleep. Vol. 42. 2019:A391.

Levin 2014 *{published data only}*

Levin H, Lynch TP, Seabrook J, Poonai N, Sangha G. Improving shiftwork: are casino shifts the answer? *Canadian Journal of Emergency Medicine* 2014;**16**:S19-S114.

McPherson 1993 *{published data only}*

McPherson G. Shiftwork and the offshore worker. *Occupational Health* 1993;**45**(7):237-9.

NCT03813654 *{published data only}*

NCT03813654. Sleep and Circadian Treatments for Shift Workers. clinicaltrials.gov/ct2/show/NCT03813654 (first received 23 January 2019). [CLINICALTRIALS.GOV: https://clinicaltrials.gov/ct2/show/NCT03813654]

Ng-A-Tham 1993 *{published data only}*

Ng-A-Tham JE, Thierry H. An experimental change of the speed of rotation of the morning and evening shift. *Ergonomics* 1993;**36**(1):51-7.

Pavageau 2006 *{published data only}*

Pavageau Pierre. The effects of work content and working hours on the health of prison guards. [Les effets conjoints du travail et des horaires alternants sur la santé des agents de surveillance des établissements pénitentiaires]. *Pistes: Perspectives Interdisciplinaires sur le Travail et la Santé* 2006;**8**(2):no pagination.

Rosa 1989 *{published data only}*

Rosa RR, Colligan MJ, Lewis P. Extended workdays: effects of 8-hour and 12-hour rotating shift schedules on performance, subjective alertness, sleep patterns, and psychosocial variables. *Work & Stress* 1989;**3**(1):21-32.

Rosa 1993 *{published data only}*

Rosa RR, Bonnet MH. Performance and alertness on 8 h and 12 h rotating shifts at a natural gas utility. *Ergonomics* 1993;**36**(10):1177-93.

Rosa 1996 *{published data only}*

Rosa RR, Härmä M, Pulli K, Mulder M, Näsman O. Rescheduling a three shift system at a steel rolling mill: effects of a one hour delay of shift starting times on sleep and alertness in younger and older workers. *Occupational and Environmental Medicine* 1996;**53**(10):677-85.

Seibt 1990 *{published data only}*

Seibt A, Geist HW, Tannenhauer J, Friedrichsen G. Assessment of continuous shifts including 12-hr shifts in terms of occupational medicine [Évaluation d'horaires postés continus avec postes de douzes heures en fin de semaine du point de vue de la médecine du travail]. *Le Travail Humain: A Bilingual and Multi-Disciplinary Journal in Human Factors* 1990;**53**(3):253-64.

Shattuck 2015a *{published data only}*

Shattuck NL, Matsangas P. A comparison of sleep and performance of US navy sailors on four different shiftwork schedules. *Sleep* 2015;**38**:A130.

Shattuck 2015b *{published data only}*

Shattuck NL, Matsangas P, Eriksen E, Kulubis S. Comparison of two watch schedules for personnel at the White House Military Office president's Emergency Operations Center. *Human Factors* 2015;**57**(5):864-78.

Shea 2018 *{published data only}*

Shea JA, Silber JH, Desai SV, Dinges DF, Bellini LM, Tonascia J, et al. Development of the individualised Comparative Effectiveness of Models Optimizing Patient Safety and Resident Education (iCOMPARE) trial: a protocol summary of a national cluster-randomised trial of resident duty hour policies in internal medicine. *BMJ OPEN* 2018;**8**(9):e021711. [DOI: 10.1136/bmjopen-2018-021711]

Skornyakov 2017 *{published data only}*

Skornyakov E, Shattuck NL, Winser MA, Matsangas P, Sparrow AR, Layton ME, et al. Sleep and performance in simulated Navy watch schedules. *Accident Analysis and Prevention* 2017;**99**:422-7.

Smith 1998 *{published data only}*

Smith L, Hammond T, Macdonald I, Folkard S. 12-h shifts are popular but are they a solution? *International Journal of Industrial Ergonomics* 1998;**21**(3-4):323-31.

Smith PA, Wright BM, Mackey RW, Milsop HW, Yates SC. Change from slowly rotating 8-hour shifts to rapidly rotating 8-hour and 12-hour shifts using participative shift roster design. *Scandinavian Journal of Work, Environment & Health* 1998;**24**(Suppl 3):55-61.

Tucker 2021 *{published data only}*

Tucker P, Härmä M, Ojajärvi A, Kivimäki M, Leineweber C, Oksanen T, et al. Association of rotating shift work schedules and the use of prescribed sleep medication: a prospective cohort study. *Journal of Sleep Research* 2021;**30**(6):e13349.

van de Ven 2021 *{published data only}*

van de Ven HA, Hulsegge G, Zoomer T, de Korte EM, Burdorf A, Oude Hengel KM. The acute effects of working time patterns on fatigue and sleep quality using daily measurements of 6195 observations among 223 shift workers. *Scandinavian Journal of Work, Environment & Health* 2021;**47**(6):446-55.

Waage 2012 *{published data only}*

Waage S, Harris A, Pallesen S, Saksvik IB, Moen BE, Bjorvatn B. Subjective and objective sleepiness among oil rig workers during three different shift schedules. *Sleep Medicine* 2012;**13**(1):64-72.

Williamson 1986 *{published data only}*

Williamson AM, Sanderson JW. Changing the speed of shift rotation: a field study. *Ergonomics* 1986;**29**(9):1085-95.

Williamson 1994 *{published data only}*

Williamson AM, Gower CGI, Clarke BC. Changing the hours of shiftwork: a comparison of 8- and 12-hour shift rosters in a group of computer operators. *Ergonomics* 1994;**37**(2):287-98.

References to studies awaiting assessment

Chevreau 2012 *{published data only}*

Chevreau A. Follow up of an experimentation on twelve hours shiftwork in three departments at the University Hospital of Bordeaux. Thesis 2012.

Cori 2021 *{published data only}*

Cori JM, Downey LA, Sletten TL, Beatty CJ, Shiferaw BA, Soleimanloo SS, et al. The impact of 7-hour and 11-hour rest breaks between shifts on heavy vehicle truck drivers' sleep, alertness and naturalistic driving performance. *Accident Analysis and Prevention* 2021;**159**:106224.

Hakola 2021 *{published data only}*

Hakola T, Niemelä P, Rönnberg S, Ropponen A. Longer work shifts, faster forward rotation-more sleep and more alert in aircraft inspection. *International Journal of Environmental Research and Public Health* 2021;**18**(15):8105.

Puttonen 2022 *{published data only}*

Puttonen S, Karhula K, Ropponen A, Hakola T, Sallinen M, Härmä M. Sleep, sleepiness and need for recovery of industrial employees after a change from an 8- to a 12-hour shift system. *Industrial Health* 2022;**60**(2):146-53.

Rahman 2021 *{published data only}*

Rahman SA, Sullivan JP, Barger LK, St Hilaire MA, O'Brien CS, Stone KL. Extended work shifts and neurobehavioral performance in resident-physicians. *Pediatrics* 2021;**147**(3):e2020009936.

Toussaint 2003 *{published data only}*

Touissant de Quievrecourt AA. Alertness and quality of life in shiftworkers and permanent night workers (effects of work schedule on the sleep quality and the alertness). Thesis 2003.

Additional references

Akerstedt 1998

Akerstedt T. Is there an optimal sleep-wake pattern in shift work? *Scandinavian Journal of Work, Environment & Health* 1998;**24**(Suppl 3):18-27.

Andrew 2008

Andrew K, Edinger JD. Measuring sleep quality. *Sleep Medicine* 2008;**9**:S10-7.

Ball 2015

Ball J, Maben JE, Murrells TJ, Day TL, Griffiths P. 12 hour shifts: prevalence, views and impact. King's College London, 2015.

Baltes 1999

Baltes BB, Briggs TE, Huff JW, Wright JA, Neuman GA. Flexible and compressed workweek schedules: a meta-analysis of their effects on work-related criteria. *Journal of Applied Psychology* 1999;**84**:496-513.

Bambra 2008a

Bambra CL, Whitehead MM, Sowden AJ, Akers J, Petticrew MP. Shifting schedules. The health effects of reorganising shift work. *American Journal of Preventive Medicine* 2008;**34**(5):427-34.

Bambra 2008b

Bambra C, Whitehead M, Sowden A, Akers J, Petticrew M. "A hard day's night?" The effects of Compressed Working Week interventions on the health and work-life balance of shift workers: a systematic review. *Journal of Epidemiology & Community Health* 2008;**62**(9):764-77.

Barton 1995

Barton J, Spelten E, Totterdell P, Smith L, Folkard S, Costa G. The Standard Shiftwork Index - a battery of questionnaires for assessing shiftwork-related problems. *Work and Stress* 1995;**9**(1):4-30.

Basner 2011

Basner M, Mollicone D, Dinges DF. Validity and sensitivity of a Brief Psychomotor Vigilance Test (PVT-B) to total and partial sleep deprivation. *Acta Astronautica* 2011;**69**:949-59.

Buysse 1989

Buysse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research* 1989;**28**(2):193-213.

Campbell 2012

Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ* 2012;**345**:e5661. [DOI: 10.1136/bmj.e5661]

Carskadon 1986

Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep* 1986;**9**(4):519-24.

Cepeda 2016

Cepeda MS, Stang P, Blacketer C, Kent JM, Wittenberg GM. Clinical relevance of sleep duration: results from a cross-sectional analysis using NHANES. *Journal of Clinical Sleep Medicine* 2016;**12**(6):813-9.

Cho 2000

Cho BK, Rosenfeldt F, Turina MI, Karp RB, Ferguson TB, Bodnar E, et al. Joint statement on redundant (duplicate) publication by the editors of the undersigned cardiothoracic journals. *Annals of Thoracic Surgery* 2000;**69**(2):663.

Choi 2014

Choi B, Schnall PL, Dobson M, Garcia-Rivas J, Kim H, Zaldivar F, et al. Very long (> 48 hours) shifts and cardiovascular strain in firefighters: a theoretical framework. *Annals of Occupational and Environmental Medicine* 2014;**26**(1):1-9.

Consensus Conference Panel 2015

Consensus Conference Panel, Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, et al. Joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society on the recommended amount of sleep for a healthy adult: methodology and discussion. *Journal of Clinical Sleep Medicine* 2015;**11**(8):931-52. [DOI: 10.5664/jcsm.4950]

Costa 2003

Costa G. Shift work and occupational medicine: an overview. *Occupational Medicine* 2003;**53**(2):83-8.

Cotter 2011

Cotter S. Seeking shift work solutions. Shift work can harm sleep, and sleep loss can lead to fatigue and errors – how is that impacting you and your patients? *EMS World* 2011;**40**(4):39-44.

Crook 2019

Crook S, Sievi NA, Bloch KE, Stradling JR, Frei A, Puhan MA, et al. Minimum important difference of the Epworth Sleepiness Scale

in obstructive sleep apnoea: estimation from three randomised controlled trials. *Thorax* 2019;**74**:390-6.

Dinges 1998

Dinges DF, Mallis M, Maislin G, Powell JW. Evaluation of techniques for ocular measurement as an index of fatigue and as the basis for alertness management. National Highway Traffic Safety Administration 1998.

Driscoll 2007

Driscoll TR, Grunstein RR, Rogers NL. A systematic review of the neurobehavioural and physiological effects of shiftwork systems. *Sleep Medicine Reviews* 2007;**11**:179-94.

Egger 1997

Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**(7109):629-34. [DOI: 10.1136/bmj.315.7109.629]

Erren 2010

Erren TC. Shift work, cancer and "white-box" epidemiology: association and causation. *Epidemiologic Perspectives & Innovations* 2010;**7**:11.

Gao 2020

Gao Y, Gan T, Jiang L, Yu l, Tang D, Wang Y, et al. Association between shift work and risk of type 2 diabetes mellitus: a systematic review and dose-response meta-analysis of observational studies. *Chronobiology International* 2020;**37**(1):29-46. [DOI: 10.1080/07420528.2019.1683570]

GRADEpro 2008 [Computer program]

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

Gurubhagavatula 2021

Gurubhagavatula I, Barger LK, Barnes CM, Basner M, Boivin DB, Dawson D, et al. Guiding principles for determining work shift duration and addressing the effects of work shift duration on performance, safety, and health: guidance from the American Academy of Sleep Medicine and the Sleep Research Society. *Journal of Clinical Sleep Medicine* 2021;**17**(11):2283-306.

Hanifah 2021

Hanifah AMS, Ismail N. Fatigue and its associated risk factors among shift workers: a systematic review. *Human Factors and Ergonomics in Manufacturing & Service Industries* 2021;**31**(2):208-22.

Harris 2015

Harris R, Sims S, Parr J, Davies N. Impact of 12 h shift patterns in nursing: a scoping review. *International Journal of Nursing Studies* 2015;**52**(2):605-34.

Herscovitch 1981

Herscovitch J, Broughton R. Sensitivity of the Stanford Sleepiness Scale to the effects of cumulative partial sleep-deprivation and recovery over-sleeping. *Sleep* 1981;**4**(1):83-92.

Higgins 2011

Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from training.cochrane.org/handbook/archive/v5.1/.

Hoddes 1972

Hoddes E, Zarcone V, Dement W. Development and use of Stanford Sleepiness Scale (SSS). *Psychophysiology* 1972;**9**(1):150.

IARC 2010

International Agency for Research on Cancer. Painting, firefighting, and shiftwork. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Vol. **98**. Lyon, France: WHO/IARC, 2010.

IARC 2020

IARC Working Group on the Identification of Carcinogenic Hazards to Humans. Night shift work. *Lyon: IRAC monographs on the identification of Carcinogenic Hazards to Humans* 2020;**124**:1–371.

ILO 1990

International Labour Organization. C171 - Night Work Convention. www.ilo.org/dyn/normlex/en/f?p=NORMLEXPUB:12100:0::NO::P12100_INSTRUMENT_ID:312316 1990.

ILO 2004

International Labour Office Geneva. Information Sheet No. WT-8: Shift work. ILO, 2004. www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---travail/documents/publication/wcms_170713.pdf (accessed 2 July 2013).

Johns 1992

Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 1992;**15**(4):376-81.

Kaida 2006

Kaida K, Takahashi M, Akerstedt T, Nakata A, Otsuka Y, Haratani T, et al. Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology* 2006;**117**(7):1574-81.

Kecklund 1992

Kecklund G, Åkerstedt T. The psychometric properties of the Karolinska Sleep Questionnaire. *Journal of Sleep Research* 1992;**1**:113.

Kecklund and Axelsson 2016

Kecklund G, Axelsson J. Health consequences of shift work and insufficient sleep. *BMJ* 2016;**355**:i5210. [DOI: 10.1136/bmj.i5210]

Knauth 1995

Knauth P. Speed and direction of shift rotation. *Journal of Sleep Research* 1995;**4**:41-6.

Knauth 2003

Knauth P, Hornberger S. Preventive and compensatory measures for shift workers. *Occupational Medicine* 2003;**53**(2):109-16.

Lallukka 2011

Lallukka T, Dregan A, Armstrong D. Comparison of a sleep item from the General Health Questionnaire-12 with the Jenkins Sleep Questionnaire as measures of sleep disturbance. *Journal of Epidemiology* 2011;**21**(6):474-80.

Landrigan 2004

Landrigan CP, Rothschild JM, Cronin JW, Kaushal R, Burdick E, Katz JT, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. *New England Journal of Medicine* 2004;**351**(18):1838-48.

Lerman 2012

Lerman SE, Eskin E, Flower DJ, George EC, Gerson B, Hartenbaum N, et al. Fatigue risk management in the workplace. *Journal of Occupational Environmental Medicine* 2012;**54**(2):231-58. [DOI: 10.1097/JOM.0b013e318247a3b0]

Li 2011

Li Y, Sato Y, Yamaguchi N. Shift work and the risk of metabolic syndrome: a nested case-control study. International Journal of Occupational & Environmental Health 2011;**17**(2):154-60.

Mackworth 1950

Mackworth NH. Researches on the measurement of human performance. *Medical Research Council Special Report Series* 1950;**268**:156.

Mitler 1982

Mitler MM, Gujavarty KS, Browman CP. Maintenance of wakefulness test: a polysomnographic technique for evaluation treatment efficacy in patients with excessive somnolence. *Electroencephalography and Clinical Neurophysiology* 1982;**53**(6):658-61.

Moreno 2019

Moreno C, Marqueze EC, Sargent C, Wright KP, Ferguson SA, et al. Working Time Society consensus statements: evidence-based effects of shift work on physical and mental health. *Industrial Health* 2019;**57**(2):139-57. [DOI: https://doi.org/10.2486/indhealth.SW-1]

Nasca 2010

Nasca TJ, Day SH, Amis ES. The new recommendations on duty hours from the ACGME Task Force. *New England Journal of Medicine* 2010;**363**(2):e3(1-6).

Page 2021

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;**372**:n71.

Pallesen 2008

Pallesen S, Bjorvatn B. A new scale for measuring insomnia: the Bergen insomnia scale. *Perceptual and Motor Skills* 2008;**107**:691-701.

Parkes 2007

Parkes, KR. Working hours in the offshore petroleum industry Current knowledge and research needs regarding extended work periods and shift work offshore. In: Petroleum Safety Authority, Stavanger, Norway. 2007.

Partinen 1995

Partinen M, Gislason T. Basic Nordic Sleep Questionnaire (BNSQ) - a quantitated measure of subjective sleep complaints. *Journal of Sleep Research* 1995;**4**:150-5.

Patel 2017

Patel S, Kon SS, Nolan CM, Simonds AK, Morrell MJ, Man WDC, et al. The Epworth Sleepiness Scale: minimum clinically important difference in obstructive sleep apnea. *European Respiratory Journal* 2017;**50**:PA330. [DOI: 10.1183/1393003.]

Puffer 2003

Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* 2003;**327**:785.

Rajaratnam 2011

Rajaratnam SM, Barger LK, Lockley SW, Shea SA, Wang W, Landrigan CP, et al. Sleep disorders, health, and safety in police officers. Journal of the American Medical Association 2011;**306**(23):2567-78.

Ramsay 2003

Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. *International Journal of Technology Assessment in Health Care* 2003;**19**:613-23.

RevMan 2012 [Computer program]

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

Sallinen 2010

Sallinen M, Kecklund G. Shift work, sleep, and sleepiness – differences between shift schedules and systems. *Scandinavian Journal of Work, Environment & Health* 2010;**36**(2):121-33.

Shen 2006

Shen J, Barbera J, Shapiro CM. Distinguishing sleepiness and fatigue: focus on definition and measurement. *Sleep Medicine Reviews* 2006;**10**(1):63-76.

Sommer 2010

Sommer D, Golz M. Evaluation of PERCLOS based current fatigue monitoring technologies. In: Conference Procedings IEEE English Medical and Biological Society. 2010:4456-9.

Stevens 2011

Stevens RG, Hansen J, Costa G, Haus E, Kauppinen T, Aronson KJ, et al. Considerations of circadian impact for defining 'shift work' in cancer studies: IARC Working Group Report. *Occupational and Environmental Medicine* 2011;**68**(2):154-62. [DOI: 10.1136/oem.2009.053512]

Thorne 2005

Thorne DR, Johnson DE, Redmond DP, Sing HC, Belenky G, Shapiro JM. The Walter Reed palm-held psychomotor vigilance test. *Behavior Research Methods* 2005;**37**:111-8.

Torquati 2018

Torquati L, Mielke GI, Brown WJ, Kolbe-Alexander T. Shift work and the risk of cardiovascular disease. A systematic review and meta-analysis including dose–response relationship. *Scandinavian Journal of Work, Environment & Health* 2018;**44**(3):229-38.

Torquati 2019

Torquati L, Mielke GI, Brown WJ, Burton NW, Kolbe-Alexander TL. Shift work and poor mental health: a meta-analysis of longitudinal studies. *American Journal of Public Health* 2019;**109**:E13-20. [DOI: 10.2105/AJPH.2019.305278]

Von Ruesten 2012

Von Ruesten A, Weikert C, Fietze I, Boeing H. Association of sleep duration with chronic diseases in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study. *PLOS One* 2012;**7**(1):e30972.

Wagstaff 2011

Wagstaff AS, Sigstad Lie J-A. Shift and night work and long working hours – a systematic review of safety implications. Scandinavian Journal of Work, Environment & Health 2011;**37**(3):173-85.

Wang 2016

Wang D, Li W, Cui X, Meng Y, Zhou M, Xiao L, et al. Sleep duration and risk of coronary heart disease: a systematic review and meta-analysis of prospective cohort studies. *International Journal of Cardiology* 2016;**219**:231-9.

Wierwille 1994

Wierwille WW, Ellsworth LA, Wreggit SS, Fairbanks RJ, Kirn CL. Research on vehicle based driver status performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness. National Highway Traffic Safety Administration, Tech Report DOT HS808 247 1994.

Wong 2011

Wong IS, McLeod CB, Demers PA. Shift work trends and risk of work injury among Canadian workers. *Scandinavian Journal of Work, Environment & Health* 2011;**37**(1):54-61.

Wong 2019

Wong IS, Popkin S, Folkard S. Working Time Society consensus statements: a multi-level approach to managing occupational sleep-related fatigue. *Industrial Health* 2019;**57**(2):228-44.

Wu 2021

Wu QJ, Sun H, Wen ZY, Zhang M, Wang HY, He XH, et al. Shift work and health outcomes: an umbrella review of systematic reviews and meta-analyses of epidemiological studies. *Journal of Clinical Sleep Medicine* 2022;**18**(2):653-62. [DOI: 10.5664/jcsm.9642]

Åkerstedt 1994

Åkerstedt T, Hume K, Minors D, Waterhouse J. The meaning of good sleep: a longitudinal study of polysomnography and subjective sleep quality. *Journal of Sleep Research* 1994;**3**:152-158.

Åkerstedt 2003

Åkerstedt, T. Shift work and disturbed sleep/wakefulness. *Occupational Medicine* 2003;**53**(2):89-94.

References to other published versions of this review

Erren 2013

Erren TC, Herbst C, Koch MS, Fritschi L, Foster RG, Driscoll TR, et al. Adaptation of shift work schedules for preventing and treating sleepiness and sleep disturbances caused by shift work. *Cochrane Database of Systematic Reviews* 2013, Issue 7. Art. No: CD010639. [DOI: 10.1002/14651858.CD010639]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

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

Amendola 2011

Study characteristics	
Methods	Study design: cluster-RCT with random allocation
	Statistical analysis: block-Randomized ANCOVA
	Study dates: January 2007–June 2009
Participants	Setting: 2 police departments
	Occupation: police officers
	Number of participants: 257

Amendola 2011 *(Continued)*

	Age: 48% aged 18–34 years Sex: 77% men
Interventions	Intervention 1: 10-hour shifts Intervention 2: 12-hour shifts Control: 8-hour shifts
Outcomes	• Sleep quality off-shift and sleep length off-shift were subjectively assessed using a sleep diary. • Sleepiness during shifts was objectively assessed by psychomotor vigilance test (PVT) and subjectively assessed by the Karolinska Sleepiness Study and a composite measure compound by items from the Harvard Study of Work Hours, the assessment made by a sleep specialist, and the Epworth Sleepiness Scale. The outcomes were assessed at baseline and at 6 months.
Notes	Contact with the corresponding author to obtain outcome data was unsuccessful. The study was supported by the U.S. Department of Justice, National Institute of Justice.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	The detail of the randomisation procedure is unclear. Volunteers were first stratified by district and by current shift. Within each block, a separate randomisation procedure was conducted; however, it is unclear how the random sequence was generated. Quote: "In each study site, we obtained a complete list of officers who were willing to volunteer for the study. Officers willing to participate were told that they may be assigned to a different shift, which would be assigned randomly (i.e. not based on seniority or preference). All officers on the volunteer lists were stratified by their respective assigned patrol district (six districts in Detroit and four districts in Arlington) and shift schedule (day, evening, and midnight) prior to the random assignment sequence. We conducted separate randomization procedures within each block (agency and time of shift)."
Allocation concealment (selection bias)	Unclear risk	Methods for ensuring allocation concealment were not described.
Blinding of participants and personnel (performance bias) Objective measures	Low risk	Objectively measured outcome was alertness, assessed by optical tracker and PVT. Blinding of participants and personnel was not possible due to the nature of the intervention; however, it is unlikely that beliefs and attitudes towards the best shift system affected the performance of these tests.
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Subjectively measured outcomes were sleep length off-shift, sleep quality off-shift, and sleepiness during shifts. Blinding of participants and personnel to the group allocation was not possible. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep. Quote: "Sleep diary and alertness log These instruments were put together by the Police Foundation under the direction of Dr. Anneke Heitmann, a sleep and fatigue expert. These booklets were completed by officers during the two-week period prior to the administration of performance measures at Time 1 and Time 2."

Amendola 2011 *(Continued)*

Blinding of outcome assessment (detection bias) Subjective measures	High risk	Subjective measures were self-assessed by sleep diary and alertness logs. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of these measures. Quote: "In addition to the simulations, each participant was asked to complete a series of surveys and other instruments including: (1) a sleep diary and alertness log that were completed for a two-week period prior to the laboratory simulations; and (2) a survey entitled the Law Enforcement Officer Survey of Work Attitudes, Personal Characteristics, Health, Safety, and Quality of Life (presented in a Scantron® booklet) that was generally completed in advance of the simulations or the same day."
Blinding of outcome assessment (detection bias) Objective measures	Low risk	It is unclear whether the outcome assessors of the PVT were blinded to the group allocation. However, the lack of blinding is unlikely to result in biased outcome measurement, as the measures were objective. Quote: "We used the PVT (Dinges and Powell 1985) to assess reaction time for each participant. We used an adapted version that was developed by researchers at the Walter Reed Army Institute of Research for use on a hand-held PDA (Thorne et al. 2005)."
Incomplete outcome data (attrition bias) Sleep quality off-shift	High risk	Substantial proportions of randomised participants did not receive treatment (group 1: 28/105 (27%); group 2: 10/109 (9%), group 3: 13/108 (12%)) or did not complete the study (group1: 13/105; group 2: 18/109; group 3: 19/108). In group 1 (8-hour shift), the main reason for voluntary dropout was mostly preference for an alternative schedule. The effect of this systematic attrition of the outcomes is unclear and no ITT analyses were performed. Quote: "In the present study, there were barriers to some officers' continued participation in the treatment groups, such as family conflicts, promotions, and transfers of assignments (i.e., no longer on patrol). As such, there was both voluntary and involuntary attrition in our study."
Incomplete outcome data (attrition bias) Sleep length off-shift	High risk	Substantial proportions of randomised participants did not receive treatment (group 1: 28/105 (27%); group 2: 10/109 (9%), group 3: 13/108 (12%)) or did not complete the study (group1: 13/105, group 2: 18/109, group 3: 19/108). In group 1 (8-hour shift), the main reason for voluntary dropout was mostly the preference for an alternative schedule. The effect of this systematic attrition of the outcomes is unclear and no ITT analyses were performed. Quote: "In the present study, there were barriers to some officers' continued participation in the treatment groups, such as family conflicts, promotions, and transfers of assignments (i.e. no longer on patrol). As such, there was both voluntary and involuntary attrition in our study."
Incomplete outcome data (attrition bias) Sleepiness during shifts	High risk	Substantial proportions of randomised participants did not receive treatment (group 1: 28/105 (27%); group 2: 10/109 (9%), group 3: 13/108 (12%)) or did not complete the study (group1: 13/105, group 2: 18/109, group 3: 19/108). In group 1 (8-hour shift), the main reason for voluntary dropout was mostly the preference for an alternative schedule. The effect of this systematic attrition of the outcomes is unclear and no ITT analyses were performed. Quote: "In the present study, there were barriers to some officers' continued participation in the treatment groups, such as family conflicts, promotions, and transfers of assignments (i.e. no longer on patrol). As such, there was both voluntary and involuntary attrition in our study."
Selective reporting (reporting bias)	Low risk	We found no evidence of selective reporting.

Amendola 2011 *(Continued)*

Other bias	Unclear risk	Study did not adjust for potential sociodemographic differences between the groups, such as age or years of service.
Reliable or objective measurement of outcomes Sleep length and quality	High risk	Sleep quantity and quality off-shift were assessed only by subjective measures. Participants rated sleep quality from very poor to very good in the sleep diary. Both measures may be subject to under- or overestimation. Quote: "In addition to the simulations, each participant was asked to complete a series of surveys and other instruments including: (1) a sleep diary and alertness log that were completed for a two-week period prior to the laboratory simulations; and (2) a survey entitled the Law Enforcement Officer Survey of Work Attitudes, Personal Characteristics, Health, Safety, and Quality of Life (presented in a Scantron® booklet) that was generally completed in advance of the simulations or the same day."
Reliable or objective measurement of outcomes Sleepiness	Low risk	Comment: Sleepiness during shifts was assessed by optical tracker and PVT, which are validated measures. Quote 1: "The FIT® is a pupil-response test that is short and noninvasive. This assessment tool, developed by PMI, Inc., measures involuntary eye movements and serves as an optical tracker and recording system in order to detect human impairment related to fatigue as well as ingestion of substances (e.g., medications, drugs, or alcohol)." Quote 2: "We used the PVT (Dinges and Powell 1985) to assess reaction time for each participant. We used an adapted version that was developed by researchers at the Walter Reed Army Institute of Research for use on a hand-held PDA (Thorne et al. 2005). The PVT measures the participant's ability to sustain attention and respond in a timely manner to salient signals (the random appearance of a graphic target/bulls-eye)."

Axelsson 1998

Study characteristics	
Methods	Study design: non-randomised cross-over trial Statistical analysis: a repeated-measures ANOVA using 2, 3, or 4 within-group factors. The main effects were shift type, shift length, shift sequence, and time of day (when appropriate). Study dates: unclear
Participants	Setting: power plant Occupation: control room operators, shift engineers, machinists, and shift supervisors Number of participants: 31 Age: mean age of 38 years for males and 29 years for females Sex: 87% men
Interventions	All participants worked 23 shifts during six weeks in which there were 3 12-hour shifts and 4 8-hour night shifts and morning shifts. 2/6 teams started with the long nights and 4/6 teams started with the long morning shifts. The analyses compared the first 3 8-hour and 12-hour morning shifts, and, for night shifts, the analyses compared the 3 12-shifts with the first 2 8-hour shifts and the fourth 8-hour shift.
Outcomes	• Total sleep time for morning and night shifts (main sleep episode, naps not included) and total sleep time for the last day off (self-reported)

Axelsson 1998 *(Continued)*

- Sleep quality measured using a sleep diary: sleep quality index (validated in their previous study): "sleep quality" ("How was your sleep?"), "ease of falling asleep", "calm sleep" and "slept throughout"
- Sleepiness during shifts measured with the KSS

Notes	The study was supported by the Swedish Work Environment Fund.

Risk of bias

Bias	Authors' judgement	Support for judgement
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Blinding of participants and personnel to the group allocation was not possible. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep.
Blinding of outcome assessment (detection bias) Subjective measures	High risk	Outcome assessor not blinded.
Incomplete outcome data (attrition bias) Sleep quality off-shift	Unclear risk	Reasons for missing outcome data are related to the outcome though not the cause. The size of the effects of missing data is unclear. No data for 18 of 49 participants. Quote: "Of the remaining 49 subjects another 18 had to be excluded from the analysis because of incomplete data or because of too many deviations from the schedule (changed shifts with colleagues, much overtime, holidays and sick leave)."
Incomplete outcome data (attrition bias) Sleep length off-shift	Unclear risk	Reasons for missing outcome data are related to the outcome though not the cause. The size of the effects of missing data is unclear. No data for 18 of 49 participants. Quote: "Of the remaining 49 subjects another 18 had to be excluded from the analysis because of incomplete data or because of too many deviations from the schedule (changed shifts with colleagues, much overtime, holidays and sick leave)."
Incomplete outcome data (attrition bias) Sleepiness during shifts	Unclear risk	Reasons for missing outcome data are related to the outcome though not the cause. The size of the effects of missing data is unclear. No data for 18 of 49 participants.
Selective reporting (reporting bias)	Low risk	Unlikely that any relevant outcomes were not reported, as the study reported all three primary outcomes of interest for this review.
Other bias	High risk	Small sample size. Reported no power analysis.
Reliable or objective measurement of outcomes Sleep length and quality	Unclear risk	No clear data on reliability, and data on agreement and kappa were unavailable. Quote: "The sleep diary was collected daily after each main sleep period, and it had questions about bed times, wake-up times, napping, and different aspects of sleep quality. A sleep quality index was computed (as a mean across items), containing the items "sleep quality" (phrased "How was your sleep?"), "ease of falling asleep", "calm sleep" and "slept throughout". In previous studies the sleep quality index showed a significant covariation with objective measures of sleep."
Reliable or objective measurement of outcomes Sleepiness	Low risk	Reliable and validated questionnaire.

Axelsson 1998 *(Continued)*

		Quote: "In the wake diary the subjects were instructed to rate their sleepiness and physical effort every 2nd hour on the Karolinska Sleepiness Scale (KSS) (9) and on the Borg CR-10 scale (10) both during work and free time."
Appropriateness of statistical analyses	Low risk	Repeated-measures ANOVA using 2, 3 or 4 within-group factors. Quote: "The data were analyzed with a repeated-measures analysis of variance (ANO VA) using 2, 3 or 4 within-group factors. The main effects were shift type (N shifts versus M shifts), shift length (8-hour shifts versus 12- hour shifts), shift sequence (1st, 2nd, and 3rd shift in a row) and, when appropriate, time of day."
Intervention independent of other changes over time	Low risk	No other changes mentioned in the paper. It is a relatively short follow-up period, no major changes expected.
Intervention unlikely to affect data collection	Low risk	No changes in data collection between interventions in control period.

Barger 2019a

Study characteristics	
Methods	Study design: cluster-randomised cross-over clinical trial Statistical analysis: for sleep quality and duration, the trial authors performed linear/logistic mixed models, adjusting for study site, randomisation order, and unbalanced baseline characteristics. For subjective sleepiness (KSS scores), they compared the 2 shift schedules using a repeated-measures log-link negative binomial model. The regression models were adjusted for the log number of tests taken by each individual. Study dates: July 2013–March 2017.
Participants	Setting: PICUs in six medical centres in the USA. Academic medical centres were not eligible to participate, as they had previously eliminated shifts schedules longer than 16 hours for resident physicians working in their PICUs. Occupation: senior resident physicians (PGY2 and higher) with a minimum of 14 actigraphy or eDiary measurements Number of participants: 302 for sleep quality and duration; 294 for sleepiness Age: Mean age of 29.4 years Sex: 38% men
Interventions	Intervention group: Rapid Cycling Work Roster that limited scheduled work shifts to ≤ 16 consecutive hours, including regular overnight shifts. Sequence of shifts in a repeating 4- or 5-day cycle. The approximate schedule was 2 day shifts (lasting 11–15 hours) and 1 overnight shift (16 hours) that started in the evening and ended the next morning. Control group: Extended Duration Work Roster, with regularly scheduled extended-duration work shifts (lasting 24–28 hours). 4- or 5-day rotation schedule consisting of 2 day shifts (lasting approximately 12 hours), followed by 1 overnight shift that started in the morning one day and ended in the morning the next day (about 24–28 hours).
Outcomes	• Sleep duration measured with Actigraphy (participants wore wrist motion logger actigraphs continuously) • Sleepiness measured every 5 hours during a shift with the KSS

Barger 2019a *(Continued)*

Notes	Funding: the study was supported by National Heart, Lung, and Blood Institute and the National Institute of Occupational Safety and Health. Conflicts of interest: the study authors mentioned several potential conflicts of interest, but stated that this did not influence the submitted work.

Risk of bias

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	No clear description of the randomisation procedure. Quote: "Residency programs underwent cluster randomization to a schedule following the 2011 ACGME standards, or a schedule that permitted more flexible duty hours (removing the 16 h restriction on shift length)."
Allocation concealment (selection bias)	Unclear risk	No clear description of the randomisation procedure. Quote: "Residency programs underwent cluster randomization to a schedule following the 2011 ACGME standards, or a schedule that permitted more flexible duty hours (removing the 16-hour restriction on shift length)."
Blinding of participants and personnel (performance bias) Objective measures	Low risk	Objective measurements using actigraphy. Quote: "During the rotation, resident-physician volunteers continuously wore wrist Motionlogger actigraphs (Ambulatory Monitoring, Inc., Ardsley, NY) to collect rest/activity patterns. The Motionlogger is a battery-operated device and is the size of a watch. Participants were instructed to wear it on the wrist of their nondominant hand. Sleep was estimated for each day using the Action-W version 2.0 software (Ambulatory Monitoring, Inc., Ardsley, NY; UCSD algorithm with rescoring)."
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Blinding of participants and personnel to the group allocation was not possible. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep. Quote: "Resident physicians completed daily sleep/wake electronic logs (“eDiary”) as part of their morning routine."
Blinding of outcome assessment (detection bias) Subjective measures	High risk	The outcome assessment was not blinded.
Blinding of outcome assessment (detection bias) Objective measures	Low risk	Blinding of outcome assessment was ensured, and it was unlikely that the blinding could have been broken.
Incomplete outcome data (attrition bias) Sleep length off-shift	Unclear risk	No clear reason for dropouts provided.
Incomplete outcome data (attrition bias) Sleepiness during shifts	Unclear risk	No clear reason for dropouts provided.
Selective reporting (reporting bias)	Low risk	All outcomes mentioned in the protocol were reported either in Barger 2019a or Rahman 2021.
Other bias	Low risk	We identified no other potential sources of bias.

Barger 2019a *(Continued)*

Reliable or objective measurement of outcomes Sleep length and quality	Low risk	Objective measure of sleep length.
Reliable or objective measurement of outcomes Sleepiness	Low risk	KSS used to measure sleepiness. Quote: "Resident-physicians completed the Karolinska Sleepiness Scale (KSS) before each PVT.
Appropriateness of statistical analysis (cluster allocation)	Low risk	The study adjusted for cluster effect. Quote: "We used generalized linear models to estimate the effects of schedule. Fixed effects included schedule, site, and randomization order, as well as baseline characteristics found to be unbalanced by schedule."
Recruitment bias	Low risk	Trial reported minimal recruitment after randomisation.

Barton 1994

Study characteristics	
Methods	Study design: CBA trial with non-random individual allocation Statistical analysis: multivariate ANOVA, with age as a covariate Study dates: unclear
Participants	Setting: car manufacturing Occupation: car manufacturing workers Number of participants: 293 Age: mean age of 30.1 years in the intervention group and 30.2 in the control group Sex: 93% men
Interventions	Intervention group: forward rotation of 8-hour shifts with weekly rotations Control group: backward rotation and otherwise similar conditions A third study group, comprised by daytime workers, was not considered.
Outcomes	• Sleep quality off-shift and sleep duration off-shift assessed by the Standard Shift Work Index
Notes	The study authors did not report whether they received funding.

Risk of bias

Bias	Authors' judgement	Support for judgement
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Considering the nature of the intervention itself, blinding of participants was not feasible. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep
Blinding of outcome assessment (detection bias)	High risk	Blinding of participants was not feasible, considering the nature of the intervention. All outcomes were subjectively measured and self-reported.

Barton 1994 *(Continued)*

Subjective measures		Quote: "The measures included in the questionnaire were chosen to cover the main problems commonly reported by shift workers. All the scales were taken from, or derived from the standard shift work index (SSI)".
Incomplete outcome data (attrition bias) Sleep quality off-shift	High risk	High attrition rate (23% of the intervention group and 30% of the control group did not complete the study). Quote: "A total of 363 people took part in the study, 120 in the experimental group, 173 in the control three shift group, and 70 in the control day group. Table 1 shows the demographic details of the three groups. Of these, 248 (68%) took part in the second stage of the study; 92 (77%) of the experimental group,121 (70%) of the control three shift group, and 35 (50%) of the control day workers."
Incomplete outcome data (attrition bias) Sleep length off-shift	High risk	High attrition rate (23% of the intervention group and 30% of the control group did not complete the study). Quote: "A total of 363 people took part in the study, 120 in the experimental group, 173 in the control three shift group, and 70 in the control day group. Table 1 shows the demographic details of the three groups. Of these, 248 (68%) took part in the second stage of the study; 92 (77%) of the experimental group,121 (70%) of the control three shift group, and 35 (50%) of the control day workers."
Selective reporting (reporting bias)	Low risk	No evidence of selective reporting. Results were probably reported as planned.
Other bias	Low risk	We identified no additional sources of bias.
Reliable or objective measurement of outcomes Sleep length and quality	High risk	Outcomes were only assessed by subjective measures. No objective measures such as actigraphy. Quote: "We used a four-item measure of sleep quality associated with morning, afternoon, and night shifts, and rest days."
Baseline differences between groups	High risk	Study groups differed at baseline in relation to time to sleep and to sleep quality. Quote: "There were only five significant differences between the two three shift groups at time 1.Those on the delaying system (experimental group) reported more satisfaction with their shift system and social life, yet reported less enjoyment at work than the advancing (control three shift) group. Also, they reported going to sleep earlier and having fewer sleep difficulties between afternoon shifts. More detailed analyses of the items that comprised the sleep difficulties scale showed that the delaying group had less difficulty falling asleep (F1,281) = 16&14, P < 0-001), and slept better (F1,281) = 15-11, P < 0-001) on the afternoon shifts."
Appropriateness of statistical analyses	Low risk	Data analysed using change from baseline. Quote: "CHANGE FROM TIME 1 TO TIME 2. Only those people who had taken part in the study at both time 1 and time 2 were included in this stage of the analyses.'"
Intervention independent of other changes over time	Low risk	No other changes mentioned in the paper.
Intervention unlikely to affect data collection	Low risk	No changes in data collection between the measurements.

Basner 2019

Study characteristics	
Methods	Study design: cluster-RCT Statistical analysis: linear mixed effects models with random program intercepts, and adjusting for age and sex. Study dates: November 2015–May 2016
Participants	Setting: general medicine, cardiology, or critical care units Occupation: interns of the respective departments Number of participants: 457 participants included and 398 evaluated Age: mean age 27.8 years in control group and 27.9 years in intervention group Sex: 48% men in control group and 54% men in intervention group
Interventions	Intervention group: 80-hour workweek without limits on shift duration or mandatory time off between shifts Control group: standard 80-hour workweek with the following limits on shift duration and mandatory time off between shifts: • Duty-hour periods must not exceed 16 hours. • Duty-hour periods must not exceed 24 hours, with an additional 4 hours permitted for transitions in care. • All residents must have ≥ 14 hours off after 24 hours of in-house duty and ≥ 8 hours off after a regular shift.
Outcomes	• Sleep duration: average sleep time per 24 hours measured with wristwatch-like accelerometer (model wGT3X-BT) • Sleepiness: KSS
Notes	The study was supported by National Heart, Lung, and Blood Institute and the American Council for Graduate Medical Education.

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Valid random sequence generation. Quote: "The Data Coordinating Center generated the random treatment assignment schedule using SAS version 9.3. The randomization schedule was designed to yield an expected assignment ratio of 1:1 for Curr and Flex and employed a permuted block design, with block sizes documented at the Data Coordinating Center. Documentation of all these processes are retained at the DCC and are accessible only to authorized personnel."
Allocation concealment (selection bias)	Low risk	Adequate allocation concealment, Quote: "The Data Coordinating Center generated the random treatment assignment schedule using SAS version 9.3 The randomization schedule was designed to yield an expected assignment ratio of 1:1 for Curr and Flex and employed a permuted block design, with block sizes documented at the Data Co-

Basner 2019 *(Continued)*

		ordinating Center. Documentation of all these processes are retained at the DCC and are accessible only to authorized personnel."
Blinding of participants and personnel (performance bias) Objective measures	Low risk	Actigraph measurement of sleep length. Quote: "After obtaining written informed consent, coordinators scheduled interns for a single 14-day measurement period, commencing on a Monday, during which the intern underwent continuous sleep–wake measurement by means of actigraphy (a wristwatch-like accelerometer; model wGT3X-BT, ActiGraph)."
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Blinding of participants and personnel to the group allocation was not possible. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep.
Blinding of outcome assessment (detection bias) Subjective measures	High risk	No blinding.
Blinding of outcome assessment (detection bias) Objective measures	Low risk	Objective measurements were performed blinded. Quote: "Sleep times will be extracted from the wrist actigraph and sleep survey data by Pulsar staff who are blind to Curr and Flex conditions."
Incomplete outcome data (attrition bias) Sleep length off-shift	Low risk	Missing data are limited and reasonably balanced. Single imputation used.
Incomplete outcome data (attrition bias) Sleepiness during shifts	Low risk	Missing data limited and reasonably balanced. Single imputation used. Quote: "For epochs with an unknown sleep–wake state (mean, 0.76 days per intern in the flexible group and 0.64 days per intern in the standard group, out of 13 expected days), we used single imputation with stratification according to program (standard or flexible), shift type reported by the intern (e.g., day, night, or off), and time of day (1440 periods of 1 minute each)"
Selective reporting (reporting bias)	Low risk	All outcomes prespecified in the protocol were reported.
Other bias	Low risk	We identified no other potential sources of bias.
Reliable or objective measurement of outcomes Sleep length and quality	Low risk	Objective measurement.
Reliable or objective measurement of outcomes Sleepiness	Low risk	Validated questionnaire. Quote: "Each day between 6 a.m. and 9 a.m., interns were asked to complete a brief survey on the smartphone that included a question on the shift that the intern was working, a sleep log (in which they recorded sleep periods during the past 24 h), a score for sleep quality (on a five-point scale, from 1 [bad] to 5 [good]), a question on the experience of periods of excessive sleepiness during the past 24 h (with instructions to check all that apply: none, 12 a.m. to 6 a.m., 6 a.m. to 12 p.m., 12 p.m. to 6 p.m., and 6 p.m. to 12 a.m.), and the score on the Karolinska Sleepiness Scale".
Appropriateness of statistical analysis (cluster allocation)	Low risk	Random intercepts for programmes and interns (clustered within programmes) used to adjust for cluster allocation.

Basner 2019 *(Continued)*

		Quote: "Linear mixed-effects models with random program intercepts were used for noninferiority analyses."
Recruitment bias	Low risk	Trial reported no recruitment after randomisation.

Bell 2015

Study characteristics

Methods	Study design: CBA trial with non-random cluster allocation Statistical analysis: ANCOVA was used to determine the main effects between the control and intervention group. Study dates: December 2011–August 2012
Participants	Setting: Phoenix Police Department, AZ, USA Occupation: police officers (first responders, patrol sergeants, and patrol lieutenants) Number of participants: 386 Age: mean age 37.2 years in control group and 37.6 years in intervention group Sex: 87% men in control group and 86% men in intervention group
Interventions	Intervention group: 3 consecutive 13-hour 20-minute shifts per week in 2 shifts: daytime (Shift 1) and nighttime (Shift 2). Daytime shifts started at 05:00 to 06:00 and nighttime shifts started at 17:00 to 18:00 Control group: 4 consecutive 10-hour shifts per week in 3 shifts (daytime, evening, nighttime). Daytime shifts started at 05:00 to 06:00, evening shifts started at 13:30 to 14:30 PM, and nighttime shifts started at 20:00 to 21:00
Outcomes	• Sleep quality off-shift assessed by PSQI • Sleep duration off-shift assessed by PSQI • Sleepiness during shifts assessed by PVT • Overtime assessed through official records
Notes	The study was supported by the Phoenix Police Department and Midwestern University, College of Health Sciences. Stoelting Publishers provided the STROOP test for this study.

Risk of bias

Bias	Authors' judgement	Support for judgement
Blinding of participants and personnel (performance bias) Objective measures	Low risk	Sleepiness during shifts assessed by the 3-minute version of the PVT, which is an objective and validate measure. Quote: "Reaction time and attention/vigilance was determined using the computerized Psychomotor Vigilance Test (PVT; Dinges & Powell, 1985)".
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Sleep length and quality were self-reported through the Pittsburgh Sleep Quality Index questionnaire. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep. Quote: "Hours of sleep per night, time to get to sleep (i.e., sleep latency), subjective sleep quality (i.e., “very good,” “fairly good,” “fairly bad,” or “very bad”),

Bell 2015 *(Continued)*

		habitual sleep efficiency (i.e., total time asleep/time spent in bed), sleep disturbances (i.e., having trouble sleeping due to waking in the middle of the night, having to use the bathroom, coughing, snoring or having difficulty breathing, feeling too hot or cold, having bad dreams, or being in pain), frequency of taking medications to fall asleep, and daytime dysfunction due to sleepiness (i.e., having trouble staying awake while driving, eating, or during social engagements or having enthusiasm to get things done) were measured using the Pittsburgh Sleep Quality Index (PSQI)."
Blinding of outcome assessment (detection bias) Subjective measures	High risk	Sleep length and quality were self-reported through the Pittsburgh Sleep Quality Index questionnaire. Quote: "The PSQI is an established, 19-item, self-report inventory of sleep quality (Buysse, Reynolds, Monk, Berma, & Kupfer, 1989)"
Blinding of outcome assessment (detection bias) Objective measures	Low risk	Sleepiness during shifts assessed by an objective and validated method (3-minute version of the PVT). Unclear who assessed test results, but unlikely that lack of blinding to participant allocation could have influenced results.
Incomplete outcome data (attrition bias) Sleep quality off-shift	Unclear risk	There was a clear restriction regarding the change of the work schedule. However, the number of participants completing the study was not clearly reported. Quote 1: "Officers assigned to the control and experimental precincts remained in their respective precincts for the duration of the study. They were not allowed to switch precincts or remove themselves from the experimental precinct." Quote 2: "All assessments were conducted during officer briefings at the beginning of their shift or during the last hour of their shift when they would return to the precinct for testing. Not all officers completed all of the assessments because their field work did not allow them to return to the precinct during the last hour of their shift".
Incomplete outcome data (attrition bias) Sleep length off-shift	Unclear risk	There was a clear restriction regarding the change of the work schedule. However, the number of participants completing the study was not clearly reported. Quote 1: "Officers assigned to the control and experimental precincts remained in their respective precincts for the duration of the study. They were not allowed to switch precincts or remove themselves from the experimental precinct." Quote 2: "All assessments were conducted during officer briefings at the beginning of their shift or during the last hour of their shift when they would return to the precinct for testing. Not all officers completed all of the assessments because their field work did not allow them to return to the precinct during the last hour of their shift".
Incomplete outcome data (attrition bias) Sleepiness during shifts	Unclear risk	There was a clear restriction regarding the change of the work schedule. However, the number of participants completing the study was not clearly reported. Quote 1: "Officers assigned to the control and experimental precincts remained in their respective precincts for the duration of the study. They were not allowed to switch precincts or remove themselves from the experimental precinct." Quote 2: "All assessments were conducted during officer briefings at the beginning of their shift or during the last hour of their shift when they would return to the precinct for testing. Not all officers completed all of the assessments be-

Bell 2015 *(Continued)*

		cause their field work did not allow them to return to the precinct during the last hour of their shift".
Selective reporting (reporting bias)	Low risk	Outcomes were reported as stated in the methods session. We found no reason to suspect selective reporting.
Other bias	Low risk	We identified no other potential sources of bias.
Reliable or objective measurement of outcomes Sleep length and quality	Unclear risk	Sleep quality and sleep length assessed by the Pittsburgh Sleep Quality Index, which is a validated instrument. However, objective measures, such as actigraphy, were not used. It is unclear whether the inclusion of objective methods would have influenced the results. Quote: "Hours of sleep per night, time to get to sleep (i.e., sleep latency), subjective sleep quality (i.e., “very good,” “fairly good,” “fairly bad,” or “very bad”), habitual sleep efficiency (i.e., total time asleep/time spent in bed), sleep disturbances (i.e., having trouble sleeping due to waking in the middle of the night, having to use the bathroom, coughing, snoring or having difficulty breathing, feeling too hot or cold, having bad dreams, or being in pain), frequency of taking medications to fall asleep, and daytime dysfunction due to sleepiness (i.e., having trouble staying awake while driving, eating, or during social engagements or having enthusiasm to get things done) were measured using the Pittsburgh Sleep Quality Index (PSQI)".
Reliable or objective measurement of outcomes Sleepiness	Low risk	Sleepiness assessed by the 3-minute version of the PVT. Quote: "Reaction time and attention/vigilance was determined using the computerized Psychomotor Vigilance Test".
Baseline differences between groups	Low risk	The characteristics of age, sex, number of children, number of children living in the home, age of youngest child living in the home, ethnicity, and medications taken were balanced across groups.
Appropriateness of statistical analyses	Low risk	Multivariate analysis of variance was conducted, using baseline scores as the covariate.
Appropriateness of statistical analysis (cluster allocation)	High risk	Analyses were not adjusted for the unit of analysis issues imposed by cluster allocation.
Recruitment bias	Unclear risk	Methods for allocating clusters into intervention or control groups were not explained.
Intervention independent of other changes over time	Low risk	No other changes mentioned in the paper. Relatively short follow-up period; no major changes expected.
Intervention unlikely to affect data collection	Low risk	No changes in data collection.

Cruz 2003

Study characteristics	
Methods	Study design: non-randomised laboratory study Statistical analysis: linear model (GLM) for repeated measures

Cruz 2003 *(Continued)*

	Study dates: unclear
Participants	Setting: laboratory Occupation: general population Number of participants: 28 Age: mean age 41.2 years Sex: 43% men
Interventions	Intervention group: clockwise rapidly rotating shift work schedule Control group: counterclockwise rapidly rotating shift work schedules
Outcomes	• Sleep duration subjectively measured using daily logbooks and objectively using a wrist activity monitor • Sleepiness measured using SSS
Notes	The laboratory study was not taken into account in the decision-making for the final conclusions. We did not assess risk of bias of laboratory studies, considering them at high risk by design.

Garde 2020

Study characteristics	
Methods	Study design: non-randomised cross-over trial. Statistical analysis: repeated measures ANOVA using the PROC MIXED procedure with a random intercept for each individual to account for within subject variation. Study dates: April–June 2013 and September–November 2013.
Participants	Setting: police departments from 5 districts Occupation: police officers Number of participants: 73 Age: mean age 38 years Sex: 100% men
Interventions	Study exposed the participants to 3 different work schedules • 2 night shifts followed by 2 recovery days (day shift or day off) (2 + 2) • 4 night shifts followed by 4 recovery days (day shift or day off) (4 + 4) • 7 night shifts followed by 7 recovery days (day shift or day off) (7 + 7)
Outcomes	• Sleep quality: Karolinska Sleep Diary and sleep efficiency with actiwatch (ActiGraph wGT3X-BT from ActiGraph FL, USA), worn on the non-dominant wrist during all 26 data collection days • Sleep duration: actiwatches (ActiGraph wGT3X-BT from ActiGraph FL, USA) worn on the non-dominant wrist during all 26 data collection days
Notes	The study was supported by the Danish Working Environment Research Fund and a PhD grant from Copenhagen University.

Garde 2020 *(Continued)*

Risk of bias

Bias	Authors' judgement	Support for judgement
Blinding of participants and personnel (performance bias) Objective measures	Low risk	Objective measurements using actigraphy. Quote: "Actiwatches (ActiGraph wGT3X-BT from ActiGraph FL, USA) were worn on the non-dominant wrist during all 26 data collection days. Data were collected with a sampling rate of 30 Hz and 1-minute epochs were used to score sleep. Data were analyzed with ActiGraph Sleep Analysis (ActiGraph, FL, USA)."
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Blinding of participants and personnel to the group allocation was not possible. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep.
Blinding of outcome assessment (detection bias) Subjective measures	High risk	No blinding.
Blinding of outcome assessment (detection bias) Objective measures	High risk	No blinding mentioned, but actigraph interpretation is subject to bias. Quote: "Actiwatches (ActiGraph wGT3X-BT from ActiGraph FL, USA) were worn on the non-dominant wrist during all 26 data collection days. Data were collected with a sampling rate of 30 Hz and 1-minute epochs were used to score sleep. Data were analyzed with ActiGraph Sleep Analysis (ActiGraph, FL, USA)."
Incomplete outcome data (attrition bias) Sleep quality off-shift	Unclear risk	121 participants showed interest, 73 were analysed. Unclear whether this selection was random.
Incomplete outcome data (attrition bias) Sleep length off-shift	Unclear risk	121 participants showed interest, 73 were analysed. Unclear whether this selection was random. Quote: "total of 121 police officers showed interest in participating in the study. Of these, a total of 73 received individual, detailed information about the project either face-to-face or on the phone and completed at least one of the three work schedules, and 64 completed all three work schedules. Reasons for dropping out were holidays or other fixed duties, change to a job without night shift work or family considerations."
Selective reporting (reporting bias)	Unclear risk	Unclear as no protocol is available. Study measured sleep quality and length but not sleepiness.
Other bias	Low risk	We identified no other potential sources of bias.
Reliable or objective measurement of outcomes Sleep length and quality	Low risk	Reliable objective measurements of sleep length and a reliable questionnaire on sleep quality. Quote: "Sleep was scored using a modified version of the Karolinska Sleep Diary (KSD) (18, 19). In total, seven items were used: premature awakening, difficulty falling asleep, difficulty awakening, nonrefreshing sleep, disturbed sleep, number of awakenings, and overall sleep quality."
Appropriateness of statistical analyses	Low risk	Appropriate analyses with a random intercept for each individual to account for within-subject variation.

Garde 2020 *(Continued)*

		Quote: "Unless otherwise stated, we performed repeated measures ANOVA using the PROC MIXED procedure with a random intercept for each individual to account for within subject variation"
Intervention independent of other changes over time	Low risk	No other changes mentioned in the paper. Relatively short follow-up period; no major changes expected.
Intervention unlikely to affect data collection	Low risk	No changes in data collection.

Knauth 1998

Study characteristics	
Methods	Study design: CBA trial with non-random cluster-allocation Statistical analysis: repeated measures ANOVA Study dates: unclear
Participants	Setting: steel industry in Germany Occupation: workers in the steel industry Number of participants: 179 Age: mean age of in control groups 39.8 years and 35.8 years, mean age in intervention groups 35.6 years and 34.1 years Sex: 100% men
Interventions	Intervention group 1: forward quick rotations, with a maximum of 3 nights in a row and up to 4 days off in a row Control group 1: backward slow rotations Intervention group 2: forward quick rotations, with a maximum of 2 nights in a row and up to 3 days off in a row Control group 2: backward slow rotations, with 7 nights in a row and up to 3 days off in a row Shift length in all groups was 8 hours.
Outcomes	• Sleep duration measured with questionnaire before and 10 months after implementation of the intervention
Notes	Data related to the outcomes of interest were not reported on the retrieved publication. Contact with the corresponding author was unsuccessful. The study authors did not report whether they received funding.

Risk of bias

Bias	**Authors' judgement**	**Support for judgement**
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Blinding of participants and personnel to the group allocation was not possible. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep.

Knauth 1998 *(Continued)*

Blinding of outcome assessment (detection bias) Subjective measures	High risk	Outcomes were self-reported.
Incomplete outcome data (attrition bias) Sleep length off-shift	High risk	Self-reported.
Selective reporting (reporting bias)	Unclear risk	No protocol provided and not all relevant outcomes related to sleep were reported.
Other bias	Low risk	We identified no other potential sources of bias.
Reliable or objective measurement of outcomes Sleep length and quality	Unclear risk	Study reports that their sleep questionnaire was validated, but unclear how reliable the questionnaire was. Quote: "In the questionnaire, the subjects had to indicate how often they had personal, social or sleeping problems on days with morning, evening, and night shifts (12 items per shift; Knauth and Kiesswetter, 1987). They also noted the average sleeping time and average leisure time on specific days."
Baseline differences between groups	Low risk	No large differences in age, marital status and number of children. Quote: "The mean age was 35.6 yr (range 21 55, SD = 9.19) in the group El, 39.8 yr (range 22-55, SD = 10.42) in the group C1, 34.1 yr (range 21 to 54, SD = 8.86) in the group E2 and 35.8 yr (range 19 54, SD = 9.69) in the group C2. Furthermore, the groups did not differ significantly (c~ = 0.05) regarding the family status: 83% were married in the group E1 compared to 87% in the group C1, 86% in the group E2 compared to 91% in the group C2. Finally, no significant differences (~ = 0.05) were found concerning the number of children in the family: the mean number of children was 1.6 in group E1 (one family with nine children), 0.8 in the group C1 (maximum two children), 1.1 in the group E2 (maximum four children) and 1.0 in the group C2 (maximum three children)."
Appropriateness of statistical analyses	Unclear risk	Insufficient data to judge statistical analysis. Quote: "The data were analysed with the help of variance analyses for repealed measurements. The software used was the Statistical Analysis System (SAS)."
Intervention independent of other changes over time	Low risk	No other changes mentioned in the paper.
Intervention unlikely to affect data collection	Low risk	No changes in data collection.

Totterdell 1992

Study characteristics	
Methods	Study design: CBA trial with cluster allocation Statistical analysis: multivariate analysis of variance and multivariate parametric analysis with age as a covariate Study dates: March–October (year is unclear)

Totterdell 1992 *(Continued)*

Participants	Setting: Ottawa Police Force in Canada Occupation: police officers Number of participants: 71 Age: mean age 29.5 years in the control group and 34.2 years in the intervention group Sex: 100% men
Interventions	Intervention group: 10-hour morning and afternoon shifts, 8-hour night shift. Blocks of 3 or 4 morning or afternoon shifts followed by 2 rest days, block of 6 night shifts followed by 6 rest day. Control group: all shifts 8 hours. 7 night shifts followed by 2 rest days, 2 afternoon and 5 morning shifts followed by 2 rest days, 5 afternoon and 2 morning shifts followed by 3 rest days.
Outcomes	• Sleep duration, assessed by asking the respondents to recollect their usual sleep start and end times. Sleep quality was assessed by using a 10-cm VAS with the 2 ends labelled 'worst' and 'best'. • Sleepiness, assessed by asking respondents to record how alert they normally felt at specified 2-hour intervals during a typical morning, afternoon, and night shift. For each 2-hour interval there was a 10-cm VAS with the 2 ends labelled 'drowsy' and 'alert'.
Notes	The study authors did not report whether they received funding.

Risk of bias

Bias	Authors' judgement	Support for judgement
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Sleep length and quality were self reported. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep. Quote: "Survey questionnaires (n = 150) were sent to both the control group and the Ottawa group one month prior to the adoption of the Ottawa system, in March."
Blinding of outcome assessment (detection bias) Subjective measures	High risk	Outcomes were assessed by self-reported questionnaires and participants were aware of the type of the shift system they were assigned to. Quote: "Survey questionnaires (n = 150) were sent to both the control group and the Ottawa group one month prior to the adoption of the Ottawa system, in March."
Incomplete outcome data (attrition bias) Sleep quality off-shift	High risk	150 questionnaires were sent to participants in the control and in the intervention group, with completion rates of 48.7%. Quote: "The results reported are based solely on the results of those officers from whom the authors received survey questionnaires both before and after the change (41 in the control group and 32 in the Ottawa group)"
Incomplete outcome data (attrition bias) Sleep length off-shift	High risk	150 questionnaires were sent to participants in the control and in the intervention group, with completion rates of 48.7%. Quote: "the results reported are based solely on the results of those officers from whom the authors received survey questionnaires both before and after the change (41 in the control group and 32 in the Ottawa group)"
Incomplete outcome data (attrition bias) Sleepiness during shifts	High risk	150 questionnaires were sent to participants in the control and in the intervention group, with completion rates of 48.7%.

Totterdell 1992 *(Continued)*

		Quote: "the results reported are based solely on the results of those officers from whom the authors received survey questionnaires both before and after the change (41 in the control group and 32 in the Ottawa group)"
Selective reporting (reporting bias)	High risk	Sleep quality not reported, although mentioned among prespecified outcomes in the methods section.
Other bias	Low risk	We identified no other potential sources of bias.
Reliable or objective measurement of outcomes Sleep length and quality	High risk	Sleep length and sleep quality were not assessed by objective measures. Quote: "Sleep behaviour was assessed by asking the respondents to recollect their usual sleep start and end times, and their usual quality of sleep: before the first morning shift, first afternoon shift, first night shift and first rest day, and between morning shifts, between afternoon shifts, between night shifts and between rest days, for the previous shift cycle. For example, 'Between my night shifts I usually went to sleep at... and woke up at.. .'. Quality of sleep was assessed using 10 cm visual analogue scales with the two ends labelled 'worst' and 'best'."
Reliable or objective measurement of outcomes Sleepiness	High risk	Sleepiness was not assessed by objective measures. Quote: "Alertness was assessed by asking respondents to record how alert they normally felt at specified 2 h intervals during a typical morning, afternoon and night shift. For each 2 h interval there was a 10 cm visual analogue scale with the two ends labelled 'drowsy' and 'alert'"
Baseline differences between groups	High risk	Intervention and control groups were different in relation sleep duration at baseline. Quote: "Figure 4 shows that the Ottawa group were getting more sleep per 24 h on the Ottawa system than on their previous shift system; however, the control group were getting more sleep than the Ottawa group before the intervention."
Appropriateness of statistical analyses	Low risk	Analyses were conducted using multivariate analysis of variance, with repeated measures. Quote: "There should be a significant and positive change in the results of the Ottawa group at T2. In a multivariate analysis of variance (MANOVA) this will appear as a significant interaction between the two factors group and time, where time is the within subjects repeated measure."
Appropriateness of statistical analysis (cluster allocation)	High risk	Analyses were not adjusted for the unit of analysis issues imposed by the allocation of participants by clusters.
Recruitment bias	Unclear risk	The choice of allocation of clusters into the intervention or control groups was not explained.
Intervention independent of other changes over time	Low risk	No other changes mentioned in the paper.
Intervention unlikely to affect data collection	Low risk	Same questionnaire used at baseline and follow-up.

Viitasalo 2008

Study characteristics	
Methods	Study design: CBA trial with non-random individual allocation Statistical analysis: repeated measures ANOVA Study dates: October/November 2004–November/December 2005
Participants	Setting: airline company Occupation: maintenance workers Number of participants: 89 Age: mean age 44 years in control group and 47 years in intervention group Sex: 100% men
Interventions	Intervention group: rapidly forward-rotating shift system Control group: backward-rotating system
Outcomes	• Sleep quality off-shift assessed with BNSQ • Sleepiness during shift assessed with BNSQ and ESS
Notes	The study was supported by a grant from the Mutual Pension Insurance Company Ilmarinen.

Risk of bias

Bias	Authors' judgement	Support for judgement
Blinding of participants and personnel (performance bias) Subjective measures	High risk	Blinding of participants not feasible for this situation. Participants may have been influenced by their beliefs and attitudes towards the best shift system in their perception and reporting of their sleep. Quote: "The study criteria were described in the advertisements. Altogether 89 men volunteered for the study, all of whom were also eligible for participation."
Blinding of outcome assessment (detection bias) Subjective measures	High risk	Outcomes were assessed with self-reported questionnaires, and participants were not blinded. Quote: "The self-administered questionnaire contained questions on diseases diagnosed by a physician, regular medication, lifestyle, and dietary factors. Leisure-time physical activity was assessed with the use of a modified version of the International Physical Activity Questionnaire (IPAQ) (28), which requested the number of physical activity sessions per week and minutes per session. The intensity of leisure-time physical activity was enquired about with the aid of examples of strenuous, moderately strenuous, and light physical activity. Daytime sleepiness, the frequency of sleep disturbances, and the probability of falling asleep at work were studied with the Basic Nordic Sleep Questionnaire (BNSQ) (29) and the Epworth Sleepiness Scale (ESS)."
Incomplete outcome data (attrition bias) Sleep quality off-shift	Low risk	Low attrition (0 from group 1, 3 from group 2, 2 from group 3). Loss to follow up was unbalanced across groups, but was justified and probably unrelated to the intervention. Quote: "Before the study was completed, one participant died accidentally, one changed to fixed night work, one could not take part in the follow-up sur-

Viitasalo 2008 *(Continued)*

		vey because of sick leave due to a leisure-time injury, and two were on a leave of several months at the follow-up time."
Incomplete outcome data (attrition bias) Sleep length off-shift	Low risk	Low attrition (0 from group 1, 3 from group 2, 2 from group 3). Loss to follow up was unbalanced across groups, but was justified and probably unrelated to the intervention. Quote: "Before the study was completed, one participant died accidentally, one changed to fixed night work, one could not take part in the follow-up survey because of sick leave due to a leisure-time injury, and two were on a leave of several months at the follow-up time."
Incomplete outcome data (attrition bias) Sleepiness during shifts	Low risk	Low attrition (0 from group 1, 3 from group 2, 2 from group 3). Loss to follow up was unbalanced across groups, but was justified and probably unrelated to the intervention. Quote: "Before the study was completed, one participant died accidentally, one changed to fixed night work, one could not take part in the follow-up survey because of sick leave due to a leisure-time injury, and two were on a leave of several months at the follow-up time."
Selective reporting (reporting bias)	High risk	Data from the BNSQ were not clearly reported.
Other bias	Low risk	We identified no other potential sources of bias.
Reliable or objective measurement of outcomes Sleep length and quality	High risk	Sleep length was not assessed by objective measures, such as actigraphy. These outcomes were not reported in the study publication, but were provided to us by the corresponding author.
Reliable or objective measurement of outcomes Sleepiness	High risk	Sleepiness was not assessed by validated methods such as PVT. Quote: "Daytime sleepiness, the frequency of sleep disturbances, and the probability of falling asleep at work were studied with the Basic Nordic Sleep Questionnaire (BNSQ) (29) and the Epworth Sleepiness Scale (ESS)"
Baseline differences between groups	High risk	Study groups differed in terms of age, occupational position, years of shift work, alcohol intake, and smoking. Quote: "The workers who started in the rapidly forward-rotating shift system were about 10 years older than those who started in the flexible shift system. The levels of CVD risk factors and health habits at baseline are shown in tables 2 and 3 for each study group. Alcohol intake at baseline was the most frequent in the group starting in the rapidly forward-rotating shift system."
Appropriateness of statistical analyses	Low risk	Repeated-measures ANOVA was performed, using time (before and after) as one of the variables. Quote: "A repeated-measures analysis of variance (ANOVA) was used for each outcome separately. The variables in the models were time (before and after), shift system (rapidly forward-rotating, flexible, and old shift systems), interaction between time and shift system, age (< 45, ≥ 45- years) at baseline, smoking (yes, no) at baseline, and alcohol consumption (≤ 1, 2, ≥ 3 alcohol doses daily) at baseline and at the end of the study."
Intervention independent of other changes over time	Low risk	No other changes mentioned in the paper.
Intervention unlikely to affect data collection	Low risk	No changes in data collection during the study.

ANCOVA: analysis of covariance; ANOVA: analysis of variance; BNSQ: Basic Nordic Sleep Questionnaire; CBA: controlled before-after; ESS: Epworth Sleepiness Scale; ITT: intention-to-treat; KSS: Karolinska Sleepiness Scale; PGY: postgraduate year; PICU: paediatric intensive care unit; PSQI: Pittsburgh Sleep Quality Index; PVT: Psychomotor Vigilance Test; RCT: randomised controlled trial; SSS: Stanford Sleepiness Scale; VAS: visual analogue scale.

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

Study	Reason for exclusion
Akersted 1978	None of the methods prespecified in the protocol were used to assess sleep quality and quantity (validated scales, actigraphy, or sleep diary).
Barger 2019b	Conference abstract of already included paper (Barger 2019a).
Blackwell 2019	Protocol of already included paper (Barger 2019a).
Cappuccio 2009	The study intervention consisted of reduction of working hours rather than organisation of the shift system.
Chang 2021	Wrong intervention and observational study design.
Cheng 2021	Wrong study design.
Chiles 1968	This is a report of 8 studies to examine human responses in confined conditions (to mimic conditions that US Air Force military would face). The purpose was to see how far they could push their workers under extreme conditions (e.g. little sleep, exposure to noise and radiation). Little information provided about each of the study details such as participant description (e.g. age, sex), intervention details, control condition and outcome measures (i.e. definitions). Results are descriptive and there is little information about statistical analyses.
Costa 2014	The study has a first phase (cross-sectional), and an experimental phase in which some sort of propensity score matching was applied to select participants to the experimental protocol. Ineligible design.
D'Amico 1985	Laboratory study on sleep deprivation and its effect on performance measures. Interventions of work schedules were not directly assessed.
Duchon 1994	Non-randomised study in which the control group was daytime workers.
Duplessis 2007	Interrupted time series with < 3 measures.
Dutheil 2012	RCT with a cross-over design to assess heart rate variability after 1 shift of 14 hours compared to 1 shift of 24 hours.
Eriksen 2006	The comparison involved the same shift schedule for both study arms, beginning at different periods of the day. Different types of shift schedule were not compared.
Fischer 2021	Wrong study design.
Garde 2011	The study included 3 different types of interventions, namely self-rostering, education and/or policy for working hours, meetings for discussion. Interventions of work schedules were not directly assessed.
Grewal 2022	No measurements of sleep or sleepiness.
Harris 2010	Interrupted time series with < 3 measurements before and after the intervention. Although reactionary tests were performed several times along a period of 4 weeks to yield reliable mean values,

Study	Reason for exclusion
	we did not consider that this procedure could be regarded as three different time point measurements.
Hong 2021	Observational study.
Hossain 2004	Uncontrolled before-after trial with too few measurements.
ISRCTN17016944	Study design (part on shift schedule is observational).
Jackson 2014	Study intervention consisted of providing different schedules for sleep. Interventions of work schedules were not directly assessed.
Knauth 1987	Intervention group consisted of 4 subgroups working different schedules. The analyses were done combining all of them, precluding the isolated effect estimates for each of the tested interventions.
Kosmadopoulos 2014	Experiment conducted under forced desynchrony. Interventions of work schedules were not directly assessed.
Kudielka 2007	Non-randomised comparative study that assessed cortisol and sleep quality and quantity after the implementation of new shift schedules. Measurements for the pre-intervention period were not performed.
Landrigan 2020	Outcome of interest not measured.
Levin 2014	Neither of the review outcomes were assessed by validated methods. Overall sleep routine and re-establishment of sleep routine following night shifts by participants using a 5-point Likert scale.
McPherson 1993	No intervention or comparison on shift work schedule (also wrong study design, cohort study).
NCT03813654	Protocol study, results expected in June 2023. Intervention seems irrelevant.
Ng-A-Tham 1993	Interrupted time series with < 3 measures before and after the intervention.
Pavageau 2006	Non-comparative, observational study.
Rosa 1989	Interrupted time series with < 3 measures before and after the intervention.
Rosa 1993	Non-randomised comparative study that assessed performance and alertness after the implementation of new shift schedules. Measurements for the pre-intervention period were not performed.
Rosa 1996	Shift schedule at the control site similar to that employed at the intervention site.
Seibt 1990	Non-comparative, observational study.
Shattuck 2015a	Observational study.
Shattuck 2015b	Interrupted time series with < 3 measures before and after the intervention.
Shea 2018	Protocol of included study (Basner 2019).
Skornyakov 2017	Laboratory study in which participants were not randomised and without measures before the implementation of the intervention.
Smith 1998	Non-randomised study with 3 intervention arms simultaneously instituted, with no control group under the original work schedule.

Study	Reason for exclusion
Tucker 2021	No measurement of sleep or sleepiness.
van de Ven 2021	Observational study design.
Waage 2012	Interrupted time series with < 3 measures before and after the intervention.
Williamson 1986	Interrupted time series with < 3 measures before and after the intervention.
Williamson 1994	Interrupted time series with < 3 measures before and after the intervention.

RCT: randomised controlled trial.

Characteristics of studies awaiting classification *[ordered by study ID]*

Chevreau 2012

Methods	Unknown
Participants	Unknown
Interventions	Unknown
Outcomes	Unknown
Notes	

Cori 2021

Methods	Quasi-experimental cross-over study
Participants	Truck drivers
Interventions	7-hour and 11-hour rest breaks between shifts
Outcomes	• Sleep • Alertness • Naturalistic driving performance
Notes	

Hakola 2021

Methods	Trial
Participants	Aircraft inspectors
Interventions	Faster rotation schedule
Outcomes	• Sleep • Alertness

Hakola 2021 *(Continued)*

	• Work ability
Notes	

Puttonen 2022

Methods	Quasi-experimental controlled intervention design
Participants	Industrial employees
Interventions	Change from an 8-hour to a 12-hour shift system
Outcomes	• Sleep • Sleepiness • Need for recovery
Notes	

Rahman 2021

Methods	Cluster-RCT
Participants	Senior resident physicians (PGY2 and higher) working in paediatric intensive care units
Interventions	Intervention group: Rapid Cycling Work Roster that limited scheduled work shifts to ≤ 16 consecutive hours, including regular overnight shifts. Sequence of shifts in a repeating 4- or 5-day cycle. The approximate schedule was 2 day shifts (lasting 11 to 15 hours) and 1 overnight shift (16 hours) that started in the evening and ended the next morning. Control group: Extended Duration Work Roster, with regularly scheduled extended-duration work shifts lasting 24 to 28 hours. 4- or 5-day rotation schedule consisting of 2 day shifts (lasting approximately 12 hours), followed by 1 overnight shift that started in the morning one day, ending in the morning the next day (about 24 to 28 hours).
Outcomes	• Daily sleep • Work log • 10-minute PVT and KSS
Notes	

Toussaint 2003

Methods	Unknown
Participants	Unknown
Interventions	Unknown
Outcomes	Unknown

Toussaint 2003 *(Continued)*

Notes

KSS: Karolinska Sleepiness Scale; PGY: postgraduate year; PVT: Psychomotor Vigilance Test; RCT: randomised controlled trial.

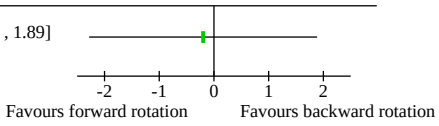
DATA AND ANALYSES

Comparison 1. Direction of rotation of shifts: forward versus backward rotation (non-randomised studies)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Sleep quality off-shift	1		Mean Difference (IV, Fixed, 95% CI)	Totals not selected
1.2 Sleep length off-shift	1		Mean Difference (IV, Fixed, 95% CI)	Totals not selected
1.3 Sleepiness during shift	1		Mean Difference (IV, Fixed, 95% CI)	Totals not selected

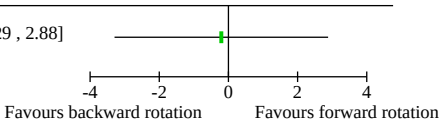
Analysis 1.1. Comparison 1: Direction of rotation of shifts: forward versus backward rotation (non-randomised studies), Outcome 1: Sleep quality off-shift

Study or Subgroup	Forward rotation Mean	SD	Total	Backward rotation Mean	SD	Total	Mean Difference IV, Fixed, 95% CI	Mean Difference IV, Fixed, 95% CI
Viitasalo 2008	-0.05	5.157131	40	0.1478	3.197866	22	-0.20 [-2.28 , 1.89]	



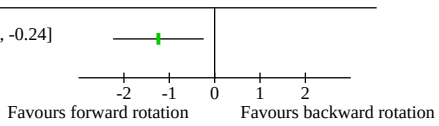
Analysis 1.2. Comparison 1: Direction of rotation of shifts: forward versus backward rotation (non-randomised studies), Outcome 2: Sleep length off-shift

Study or Subgroup	Forward rotation Mean	SD	Total	Backward rotation Mean	SD	Total	Mean Difference IV, Fixed, 95% CI	Mean Difference IV, Fixed, 95% CI
Viitasalo 2008	-0.2511	6.996072	40	-0.0454	5.262188	22	-0.21 [-3.29 , 2.88]	



Analysis 1.3. Comparison 1: Direction of rotation of shifts: forward versus backward rotation (non-randomised studies), Outcome 3: Sleepiness during shift

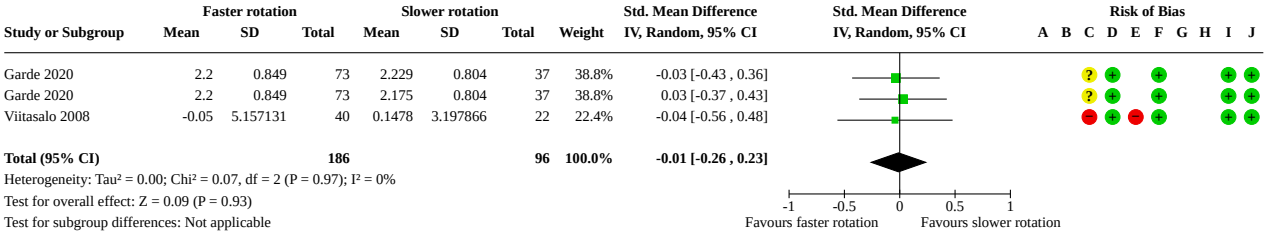
Study or Subgroup	Forward rotation Mean	SD	Total	Backward rotation Mean	SD	Total	Mean Difference IV, Fixed, 95% CI	Mean Difference IV, Fixed, 95% CI
Viitasalo 2008	-0.81	1.948333	40	0.43	1.905833	22	-1.24 [-2.24 , -0.24]	



Comparison 2. Speed of rotation: faster shift rotation (1 to 2 shifts in a row) versus slower shift rotation (3 to 7 shifts in a row) (non-randomised studies)

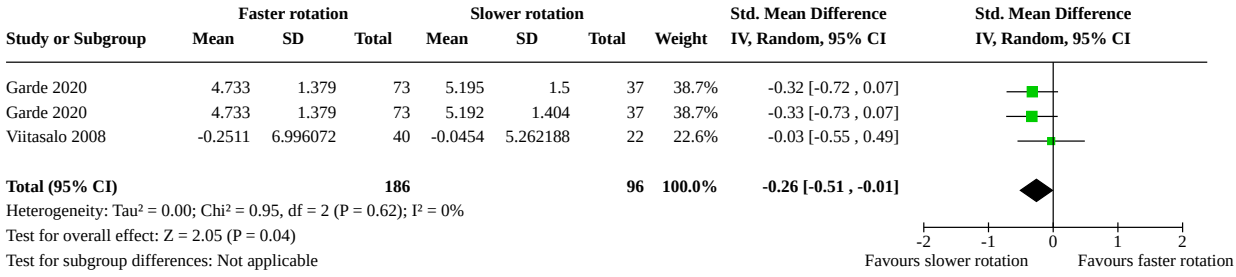
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 Sleep quality off-shift	2	282	Std. Mean Difference (IV, Random, 95% CI)	-0.01 [-0.26, 0.23]
2.2 Sleep length off-shift	2	282	Std. Mean Difference (IV, Random, 95% CI)	-0.26 [-0.51, -0.01]
2.3 Sleepiness during shift	1		Mean Difference (IV, Fixed, 95% CI)	Totals not selected

Analysis 2.1. Comparison 2: Speed of rotation: faster shift rotation (1 to 2 shifts in a row) versus slower shift rotation (3 to 7 shifts in a row) (non-randomised studies), Outcome 1: Sleep quality off-shift



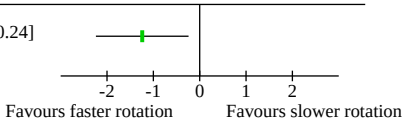
Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Selective reporting (reporting bias)
(D) Other bias
(E) Baseline differences between groups
(F) Appropriateness of statistical analyses
(G) Appropriateness of statistical analysis (cluster allocation)
(H) Recruitment bias
(I) Intervention independent of other changes over time
(J) Intervention unlikely to affect data collection

Analysis 2.2. Comparison 2: Speed of rotation: faster shift rotation (1 to 2 shifts in a row) versus slower shift rotation (3 to 7 shifts in a row) (non-randomised studies), Outcome 2: Sleep length off-shift



Analysis 2.3. Comparison 2: Speed of rotation: faster shift rotation (1 to 2 shifts in a row) versus slower shift rotation (3 to 7 shifts in a row) (non-randomised studies), Outcome 3: Sleepiness during shift

Study or Subgroup	Faster rotation Mean	SD	Total	Slower rotation Mean	SD	Total	Mean Difference IV, Fixed, 95% CI	Mean Difference IV, Fixed, 95% CI
Viitasalo 2008	-0.81	1.948333	40	0.43	1.905833	22	-1.24 [-2.24 , -0.24]	



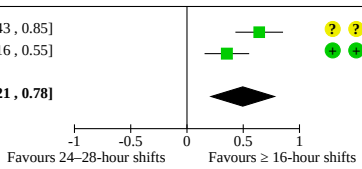
Comparison 3. Shift duration: no more than 16 hours versus 24- to 28-hours (randomised studies)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 Sleep duration off-shift	2	760	Std. Mean Difference (IV, Random, 95% CI)	0.50 [0.21, 0.78]
3.2 Sleepiness during shift	2	716	Std. Mean Difference (IV, Random, 95% CI)	-0.29 [-0.44, -0.14]
3.2.1 Sleepiness	2	716	Std. Mean Difference (IV, Random, 95% CI)	-0.29 [-0.44, -0.14]
3.3 Work hours	1	318	Mean Difference (IV, Fixed, 95% CI)	-6.50 [-7.89, -5.11]

Analysis 3.1. Comparison 3: Shift duration: no more than 16 hours versus 24- to 28-hours (randomised studies), Outcome 1: Sleep duration off-shift

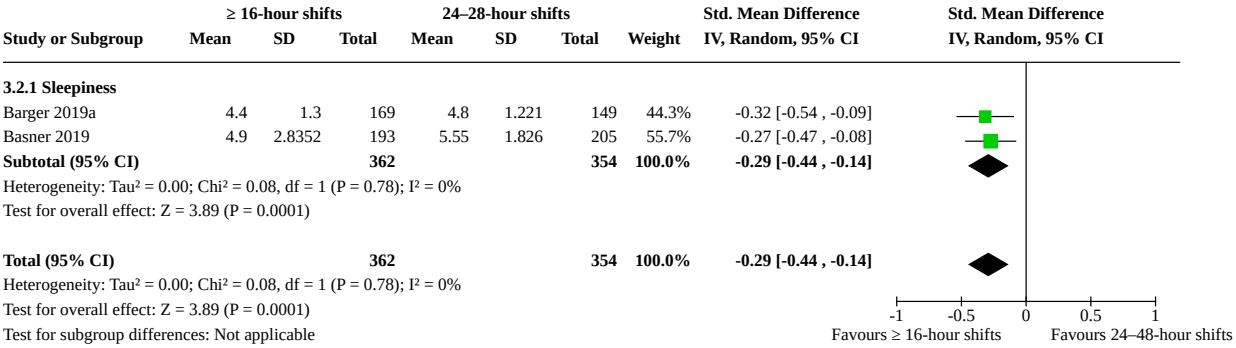
Study or Subgroup	≥ 16-hour shifts Mean	SD	Total	24–28-hour shifts Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	Std. Mean Difference IV, Random, 95% CI	Risk of Bias A B C D E F G H I J
Barger 2019a	7.557	0.857	191	7.014	0.829	171	49.1%	0.64 [0.43 , 0.85]		? ? + + + +
Basner 2019	7.35	3.4022	193	6.36	2.045	205	50.9%	0.35 [0.16 , 0.55]		+ + + + + +
Total (95% CI)			**384**			**376**	**100.0%**	**0.50 [0.21 , 0.78]**		

Heterogeneity: Tau² = 0.03; Chi² = 3.78, df = 1 (P = 0.05); I² = 74%
Test for overall effect: Z = 3.45 (P = 0.0006)
Test for subgroup differences: Not applicable

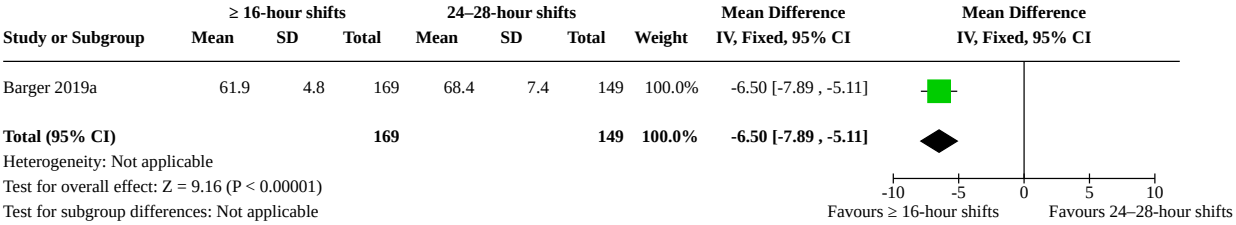


Risk of bias legend
(A) Random sequence generation (selection bias)
(B) Allocation concealment (selection bias)
(C) Selective reporting (reporting bias)
(D) Other bias
(E) Baseline differences between groups
(F) Appropriateness of statistical analyses
(G) Appropriateness of statistical analysis (cluster allocation)
(H) Recruitment bias
(I) Intervention independent of other changes over time
(J) Intervention unlikely to affect data collection

Analysis 3.2. Comparison 3: Shift duration: no more than 16 hours versus 24- to 28-hours (randomised studies), Outcome 2: Sleepiness during shift



Analysis 3.3. Comparison 3: Shift duration: no more than 16 hours versus 24- to 28-hours (randomised studies), Outcome 3: Work hours



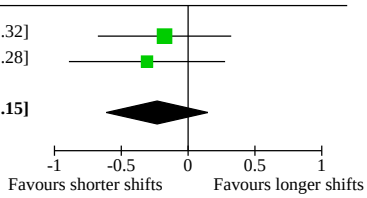
Comparison 4. Shorter shifts (8 or 10 hours) versus shifts lasting 2 to 3 hours longer

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.1 Sleep quality off-shift	2	111	Std. Mean Difference (IV, Random, 95% CI)	-0.23 [-0.61, 0.15]
4.2 Sleep length off-shift	2	121	Std. Mean Difference (IV, Random, 95% CI)	0.18 [-0.17, 0.54]
4.3 Sleepiness during shift	2		Std. Mean Difference (IV, Random, 95% CI)	Totals not selected
4.4 Overtime	1	59	Mean Difference (IV, Fixed, 95% CI)	1.22 [0.94, 1.50]

Analysis 4.1. Comparison 4: Shorter shifts (8 or 10 hours) versus shifts lasting 2 to 3 hours longer, Outcome 1: Sleep quality off-shift

Study or Subgroup	Shorter shifts Mean	SD	Total	Longer shifts Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Axelsson 1998	-4.1	0.56	31	-4	0.56	31	57.8%	-0.18 [-0.68 , 0.32]
Bell 2015	10.54	6.708	31	12.6	6.38	18	42.2%	-0.31 [-0.89 , 0.28]
Total (95% CI)			**62**			**49**	**100.0%**	**-0.23 [-0.61 , 0.15]**

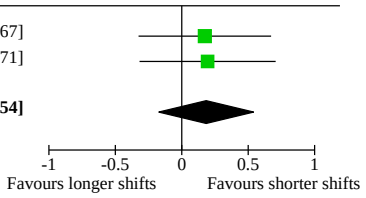
Heterogeneity: Tau² = 0.00; Chi² = 0.11, df = 1 (P = 0.74); I² = 0%
Test for overall effect: Z = 1.20 (P = 0.23)
Test for subgroup differences: Not applicable



Analysis 4.2. Comparison 4: Shorter shifts (8 or 10 hours) versus shifts lasting 2 to 3 hours longer, Outcome 2: Sleep length off-shift

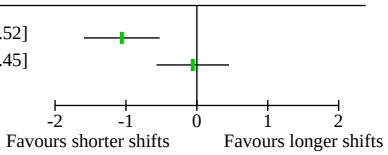
Study or Subgroup	Shorter shifts Mean	SD	Total	Longer shifts Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Axelsson 1998	5.48	1.11	31	5.3	0.93	31	51.3%	0.17 [-0.33 , 0.67]
Bell 2015	6.35	2.75	31	5.833	2.477	28	48.7%	0.19 [-0.32 , 0.71]
Total (95% CI)			**62**			**59**	**100.0%**	**0.18 [-0.17 , 0.54]**

Heterogeneity: Tau² = 0.00; Chi² = 0.00, df = 1 (P = 0.95); I² = 0%
Test for overall effect: Z = 1.01 (P = 0.31)
Test for subgroup differences: Not applicable



Analysis 4.3. Comparison 4: Shorter shifts (8 or 10 hours) versus shifts lasting 2 to 3 hours longer, Outcome 3: Sleepiness during shift

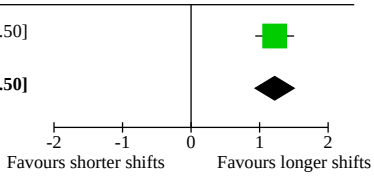
Study or Subgroup	Shorter shifts Mean	SD	Total	Longer shifts Mean	SD	Total	Std. Mean Difference IV, Random, 95% CI
Axelsson 1998	5.1	0.56	31	5.7	0.56	31	-1.06 [-1.59 , -0.52]
Bell 2015	272.01	459.512	31	291.75	67.283	28	-0.06 [-0.57 , 0.45]



Analysis 4.4. Comparison 4: Shorter shifts (8 or 10 hours) versus shifts lasting 2 to 3 hours longer, Outcome 4: Overtime

Study or Subgroup	Shorter shifts Mean	SD	Total	Longer shifts Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Bell 2015	0.94	0.6565	31	-0.28	0.4422	28	100.0%	1.22 [0.94 , 1.50]
Total (95% CI)			**31**			**28**	**100.0%**	**1.22 [0.94 , 1.50]**

Heterogeneity: Not applicable
Test for overall effect: Z = 8.44 (P < 0.00001)
Test for subgroup differences: Not applicable



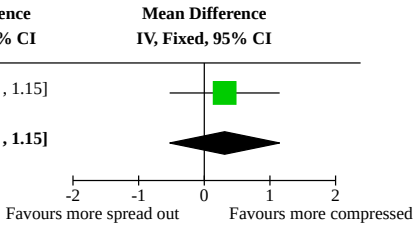
Comparison 5. Distribution of shift schedule: more compressed versus more spread out (non-randomised studies)

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5.1 Sleep quality off-shift	1	25	Mean Difference (IV, Fixed, 95% CI)	0.31 [-0.53, 1.15]
5.2 Sleep length off-shift	1	30	Mean Difference (IV, Fixed, 95% CI)	0.52 [-0.52, 1.56]

Analysis 5.1. Comparison 5: Distribution of shift schedule: more compressed versus more spread out (non-randomised studies), Outcome 1: Sleep quality off-shift

Study or Subgroup	More compressed Mean	SD	Total	More spread out Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Totterdell 1992	0.56	1.0442	10	0.25	1.0508	15	100.0%	0.31 [-0.53 , 1.15]
Total (95% CI)			**10**			**15**	**100.0%**	**0.31 [-0.53 , 1.15]**

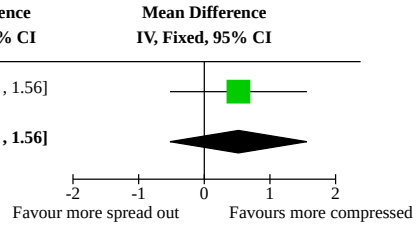
Heterogeneity: Not applicable
Test for overall effect: Z = 0.73 (P = 0.47)
Test for subgroup differences: Not applicable



Analysis 5.2. Comparison 5: Distribution of shift schedule: more compressed versus more spread out (non-randomised studies), Outcome 2: Sleep length off-shift

Study or Subgroup	More compressed Mean	SD	Total	More spread out Mean	SD	Total	Weight	Mean Difference IV, Fixed, 95% CI
Totterdell 1992	0.29	1.206	12	-0.23	1.707	18	100.0%	0.52 [-0.52 , 1.56]
Total (95% CI)			**12**			**18**	**100.0%**	**0.52 [-0.52 , 1.56]**

Heterogeneity: Not applicable
Test for overall effect: Z = 0.98 (P = 0.33)
Test for subgroup differences: Not applicable



ADDITIONAL TABLES

Table 1. Measurement tools used for sleep outcomes considered in this review

	Outcome	Score	MCID	Reference for MCID
Basic Nordic Sleep Questionnaire (Partinen 1995)	Sleep quality; frequency of sleep complaints in preceding 3 months	27 items in 21 main questions. 13 questions are rated on 5-point scale (1–5) reporting the frequency of sleep complaints in nights per week (1 = never or very rarely, 5 = every night/day or almost every night/day).	Not available	Not applicable
Bergen Insomnia scale (Pallesen 2008)	Sleep quality; frequency of insomnia symptoms over 1 week	6 questions in the scale; range from 0 to 7 for individual measures, where scoring ≥ 3 indicates the presence of insomnia; total score range from 0 to 42 as a continuous measure for combined items, where higher scores indicate more frequent sleep problems.	Criteria for insomnia: difficulty maintaining/initiating sleep present for ≥ 3 nights per week for ≥ 1 month	Based on criteria from the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edi-

Table 1. Measurement tools used for sleep outcomes considered in this review *(Continued)*

				tion, Text Revision (DSM-IV-TR)
Pittsburgh Sleep Quality Index (Buysse 1989)	Sleep quality	19 questions, each rated on a 0–3 scale. These are grouped into 7 components (sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medications, and daytime dysfunction), which can be combined into a global score (range 0–21). Higher scores indicate worse sleep quality.	Not available	Not applicable
Karolinska Sleep Questionnaire (Kecklund 1992)	Sleep quality	7 items to measure overall sleep quality, with a score ranging from 1 to 5 with higher score representing poorer sleep.	Not available	Not applicable
Standard Shift Work Index (Barton 1995)	Sleep quality	8 questions on a Likert scale ranging from 1 to 5. Combined responses are used to create a global score, where higher scores indicate greater sleep disturbance.	Not available	Not applicable
Sleep duration (in hours)	Sleep duration	≤ 7 hours of sleep is associated with adverse health and safety outcomes. It is uncertain if sleeping > 9 hours for most healthy adults is associated with health and safety risk.	Adults should sleep ≤ 7 hours per night on a regular basis to promote optimal health.	Consensus Conference Panel 2015
Epworth Sleepiness Scale (Johns 1992**)**	Sleepiness	Total score can range from 0 to 24 (the sum of 8 items with score of 0 to 3). Higher scores represent higher average sleep propensity in daily life, or higher 'daytime sleepiness'.	MCID is estimated at 2–3 points	Crook 2019; Patel 2017
Jenkins Sleep Questionnaire (Lallukka 2011)	Sleepiness/sleep disturbance over preceding 4 weeks	4 questions rated on 6-point scale based on frequency of sleep disturbances/sleepiness. Responses were dichotomised and coded 1 if respondents reported any sleep disturbances on at least 15 nights or 0 if not.	Criteria for insomnia: ≥ 15 nights of sleep disturbances in past 4 weeks	Based on criteria from the Diagnostic and Statistical Manual of Mental Disorder, Fourth Edition, Text Revision (DSM-IV-TR)
Karolinska Sleepiness Scale (Kaida 2006**)**	Sleepiness	Single-item survey based on a 9-point scale ranging from 1 (very alert) to 9 (very sleepy, great effort to keep awake).	Not available	Not applicable
MackWorth Clock Test (Mackworth 1950)	Sleepiness	Response to visual/audio vigilance task; better scores (greater alertness) reflected more frequent and accurate scores.	Not available	Not applicable
Maintenance of Wakefulness Test (Mitler 1982)	Sleepiness	Polysomnography to evaluate treatment efficacy in people with excessive somnolence. Measures include the elapsed time before sleep onset (range 0–20 minutes, lower scores indicate greater sleepiness) and frequency of REM sleep (higher scores indicate greater sleepiness).	Not available	Not applicable

Table 1. Measurement tools used for sleep outcomes considered in this review *(Continued)*

Multiple Sleep Latency Test (Carskadon 1986)	Sleepiness	Polysomnography measuring time to sleep latency and time to REM sleep onset. Lower scores indicate greater sleepiness.	Daily score > 5 minutes indicates the pathological level of daytime sleepiness	Carskadon 1986
Observer Rating of Drowsiness (Wierwille 1994)	Sleepiness	Descriptive Graphics Scale consists of 5 descriptors: not drowsy, slightly drowsy, moderately drowsy, very drowsy, and extremely drowsy.	Not available	Not applicable
PERCLOS (percentage of eyelid closure) (Dinges 1998; Sommer 2010)	Sleepiness	Proportion of time that eyes are 80% closed over a 1-minute interval. Higher scores represent greater sleepiness.	Not available	Not applicable
Psychomotor Vigilance Test (Basner 2011; Thorne 2005)	Sleepiness	2 main measures: response speed (slower speeds indicate greater sleepiness) and number of lapses (higher indicates greater sleepiness).	Not available	Not applicable
Stanford Sleepiness Scale (Herscovitch 1981; Hoddes 1972)	Sleepiness	Range from 1 to 7, with lower scores indicating better results.	Not available	Not applicable

MCID: minimal clinically important difference; REM: rapid eye movement.

Table 2. Multi-component framework of shift systems

Study	Study arm	Shift Components							
		Fixed vs any rotation	Regular (predictable) changes vs irregular (unpredictable) changes	Forward vs Backward	Faster shift rotation vs slower shift rotation	Shorter vs longer shift	Earlier start vs later start	More compressed vs more spread out	Quick shift turnaround
Amendola 2011	I1	Fixed	NA	NA	NA	Long (10 h)	Unclear	Compressed	No
	I2	Fixed	NA	NA	NA	Long (12 h)	Unclear	Compressed	No
	C	Fixed	NA	NA	NA	Short (8 h)	Unclear	Spread out	No
Axelsson 1998	C	Rotation	Predictable changes	Forward	Intermediate (3 to 4 shifts in a row)	Short (8 h)	07:00	NA (8-h shifts rotate into 12-h shifts)	No
	I	Rotation	Predictable changes	Forward	Intermediate (3 to 4 shifts in a row)	Long (12 h)	07:00	NA (8-h shifts rotate into 12-h shifts)	No
Barger 2019a	C	Rotation	Predictable changes	Forward	Fast	Very long (24 to 28 h)	Differs per site	Spread out	Less than I but not explicitly part of intervention
	I	Rotation	Predictable changes	Forward	Fast	Long (max 16 h)	Differs per site	Spread out	More than C but not explicitly part of intervention
Barton 1994	I	Rotation	Predictable	Forward	Slow	8 h	06:00	Spread out	No
	C	Rotation	Predictable	Backward	Slow	8 h	06:00	Spread out	No
Basner 2019	C	Rotation	Unpredictable changes	Unclear (may differ over time)	Unclear (may differ over time)	Max 16 h	Unclear	Unclear	Unclear

Table 2. Multi-component framework of shift systems *(Continued)*

	I	Rotation	Unpredictable changes	Unclear (may differ over time)	Unclear (may differ over time)	No restriction	Unclear	Unclear	Unclear
Bell 2015	I	Probably fixed	Probably NA	Probably NA	Probably NA	Long (13 h 20 min)	05:00 to 06:00	Compressed	No
	C	Probably fixed	Probably NA	Probably NA	Probably NA	Short (10 h)	05:00 to 06:00	Spread out	No
Cruz 2003	I	Rotation	Predictable	Forward	Fast	8 h	06:00	Spread out	No
	C	Rotation	Predictable	Backward	Fast	8 h	06:00	Spread out	No
Garde 2020	C	Rotation	Predictable changes	NA	Fast (2 night shifts in a row)	Short (8 h)	07:00	Unclear	No
	I1	Rotation	Predictable changes	NA	Slow (4 night shifts in a row)	Short (8 h)	07:00	Unclear	No
	I2	Rotation	Predictable changes	NA	Very slow (7 night shifts in a row)	Short (8 h)	07:00	Spread out	No
Knauth 1998	I1	Rotation	Predictable changes	Forward	Intermediate (3 shifts in a row)	Short (8 h)	05:50	Spread out	No
	C1	Rotation	Predictable changes	Backward	Slow	Short (8 h)	05:50	More spread out	No
	I2	Rotation	Predictable changes	Forward	Fast	Short (8 h)	05:50	Spread out	No
	C2	Rotation	Predictable changes	Backward	Slow	Short (8 h)	05:50	More spread out	No
Totterdell 1992	I	Rotation	Predictable changes	Backward and forward rotation with days off between each change	Both types for both groups	Long (8.5 to 10 h)	07:00	Compressed	No
	C	Rotation	Predictable changes	Backward and forward rotation with or without	Both types for both groups	Short (8 h)	07:00	Spread out	No

Table 2. Multi-component framework of shift systems *(Continued)*

				days off between changes					
Viitasalo 2008	I	Rotation	Predictable changes	Forward with no days off between changes	Very fast change	Long	Earlier start	Spread out	No
	C	Rotation	Predictable changes	Backward with days off between changes	Fast change	Short	Later start	Spread out	No

On-call duty and interruption of shifts not presented as there were no differences between the control and intervention situations.
C: control; I: intervention; NA: not applicable.

Table 3. Shift systems in each study

Study	Study arm	Shift system
Amendola 2011	C	WWWWW – –
		Shift duration: 8 h
		Start time: not reported
		Cycle duration: 7 days
	I1	WWWW – – –
		Shift duration: 10 h
		Start time: not reported
		Cycle duration: 7 days
	I2	Week 1: WWW – – – –
		Week 2: WWWW* – – –
		Shift duration: 12 h (except for W* = 8 h)
		Start time: not reported
		Cycle duration: 14 days
Axelsson 1998	C	3 or 4 morning and night shifts in a row, for example:
		DDDDD––
		AAAA---
		MMMM**NNN**
		––––**MMM**
		NNNN–––

		Shift duration: bold indicates 12-h shifts, non-bold indicates 8-h shifts
		Control measurements obtained for 8-h shifts.
	I	3 or 4 morning and night shifts in a row, for example:
		DDDDD––
		AAAA---
		MMMM**NNN**
		––––**MMM**
		NNNN–––

		Shift duration: *bold indicates 12-h shifts, non-bold indicates 8-h shifts
		Intervention measurements obtained for 12-h shifts.

Table 3. Shift systems in each study *(Continued)*

Barger 2019a	C	Extended Duration Work Roster, with regularly scheduled 24–28 h extended-duration work shifts working in a 4- or 5-day rotation schedule consisting of 2 day shifts of approximately 12 h, followed by 1 overnight shift that started in the morning one day and ended in the morning the next day (24–28 h long).
	I	Rapid Cycling Work Roster that limited resident physicians' scheduled work shifts to no more than 16 consecutive hours, including regular overnight shifts. Resident-physicians were scheduled to work in a sequence of shifts in a repeating 4 or 5 day cycle. The approximate schedule was 2 day shifts (11 h to 15 h long) and one 16-h overnight shift that started in the evening and ended the next morning.
Barton 1994	I	Week 1: MMMMM– – Week 2: AAAAA – – Week 3: NNNNN – – Shift duration: 8 h Start time: M: 06:00; A: 14:00; N: 22:00 Cycle duration: 21 days
	C	Week 1: NNNNN – – Week 2: AAAAA – – Week 3: MMMMM– – Shift duration: 8 h Start time: M: 06:00; A: 14:00; N: 22:00 Cycle duration: 21 days
Basner 2019	C	80-h workweek with the following limits on shift duration and mandatory time off between shifts • Duty-hour periods must not exceed 16 h • Duty-hour periods must not exceed 24 h, with an additional 4 h permitted for transitions in care • All residents must have ≥ 14 h off after 24 h of in-house duty and ≥ 8 h off after a regular shift
	I	80-h workweek without limits on shift duration or mandatory time off between shifts. Participants worked in shifts of up to 28 h.
Bell 2015	C	SSSS – – – Shift duration: 10 h Start time: D (shift 1): 05:00/06:00; E (shift 2): 13:30/14:30; N(shift 3): 20:00/21:00 Cycle duration: 7 days if no rotation
	I	SSS– – – – Shift duration: 13 h 20 min Start time: D: 05:00/06:00; N: 17:00/18:00

Table 3. Shift systems in each study *(Continued)*

		Cycle duration: 7 days if no rotation
Cruz 2003	I	MMAAN--
		Shift duration: 8 h
		start time: M: 06:00; A: 14:00; N: 22:00
		Cycle duration: 7 days
	C	AAMMN
		Shift duration: 8 h
		start time: M: 06:00; A: 14:00; N: 22:00
		Cycle duration: 7 days
Garde 2020	C	2 night shifts followed by 2 recovery days (day shift or day off)
	I1	4 night shifts followed by 4 recovery days (day shift or day off)
	I2	7 night shifts followed by 7 recovery days (day shift or day off)
Knauth 1998	I1	Week 1: –––MMM–
		Week 2: MMMEEE–
		Week 3: NN––––
		Week 4: EENNN–
		Shift duration: 8 h
		Start time: M: 05:50; E: 13:50; N: 21:50
	I2	Week 1: MEENN––
		Week 2: –MMEENN
		Week 3: –––MMEE
		Week 4: NN–––MM
		Week 5: EENN–––
		Week 6: MMEENN–
		Week 7: ––MMEEN
		Week 8: N–––MME
		Week 9: ENN–––M
		Shift duration: 8 h
		Start time: M: 05:50; E: 13:50; N: 21:50
	C1	Week 1: NNNNNN–
		Week 2: EEEEEE–
		Week 3: MMMMMM–
		Shift duration: 8 h

Table 3. Shift systems in each study *(Continued)*

		Start time: M: 05:50; E: 13:50; N: 21:50
	C2	Week 1: NNNNNNN Week 2: ––EEEE Week 3: EE––MMM Week 4: MMMM––– Shift duration: 8 h Start time: M: 05:50; E: 13:50; N: 21:50
Totterdell 1992	I	Week 1: - - MMM - - Week 2: AAA - - MM Week 3: MM – – NNN Week 4: NNNN - - - Week 5: - - - AAAA Shift duration: M and A: 10 h; N: 8.5 or 9 h Start time: M: 07:00; A: 14:00 (Sun to Wed) or 17:00; N: 22.30 (Sat to Thu) or 22:00 (Fri) Cycle duration: 35 days
	C	Week 1: NNNNNNN Week 2: – – AAMMM Week 3: MM – – AAA Week 4: AAMM– – – Shift duration: M and A: 8 h; N: 8 h Start time: M: 07:00; A: 15:00; N: 23:00 (Sat to Thu) Cycle duration: 28 days
Viitasalo 2008	C	AAA – – MMM – – NNN – – Shift duration: 8 h Start time: M: 07:00; A: 15:00; N: 23:00 Cycle duration: 15 days
	I1	MAN – – Shift duration: M and A: 10 h; N: 9 h Start time: M: 06:00; A: 15:00; N: 21:00 Cycle duration: 5 days
	I2	AAA – – – MMM – – – NNN – – –

Table 3. Shift systems in each study *(Continued)*

	Flexible (roster) determined by the employer for the third or fourth weeks of the cycle and in return participative scheduling adopted on the basis of mutual consent. Shift duration: M: 10–13 h, depending on operational needs; A: 6–13 h, depending on operational needs; N: 7 h Start time: M: 06:00 or 08:00; N: 13:00 or 14:00; N: 23:30 Cycle duration: 18 days

A: afternoon; C: control; D: daytime; E: evening; I: intervention; M: morning; N: night; S: shift, W: work.

Table 4. Measurement of sleep quality

Study	Measurement instrument	Definition of sleep quality	Measurement time points	Number of measurements	Timing of measurements	Reported for night shifts	Reported for all shifts
Amendola 2011	Sleep diaries with 1 question on general sleep quality and answer options ranging from very poor to very good	1 item on sleep quality with answers ranging from very poor to very good for each of the sleep periods recorded	Baseline and 6 months after implementation of intervention	2 weeks	Twice per day	No (only the average of all days including rest days)	No (only the average of all days including rest days)
Axelsson 1998	Sleep diary using the mean score of answers to the questions: 1) "How was your sleep?"; 2) "ease of falling asleep"; 3) "calm sleep"; and 4) "slept throughout".	Self-designed questions on sleep quality: "sleep quality" (phrased "How was your sleep?"), "ease of falling asleep", "calm sleep" and "slept throughout"	During the intervention	23 shifts over 6 weeks	Daily after each main sleep period	Yes (mean sleep for all night shifts reported for the control and intervention period)	Each shift reported separately (not combined)
Barger 2019a	NA	NA	NA	NA	NA	NA	NA
Barton 1994	Questionnaire using 1 question to measure general sleep quality/ sleep difficulties	4 items to measure sleep quality associated with morning, afternoon, and night shifts, and rest days.	2 months before and 6 months after implementation of the intervention	2 (before and after)	No specific time (retrospective questionnaire)	Yes (each night shift reported separately)	Each shift reported separately (not combined)
Basner 2019	NA	NA	NA	NA	NA	NA	NA
Bell 2015	Questionnaire with 1 question on general sleep quality with answer options ranging from very good to very bad	19 items of PSQI combined	Baseline and 3 and 6 months after implementation of the intervention	3	No specific time (retrospective questionnaire)	No (only the average of all days including rest days)	No (only the average of all days including rest days)
Cruz 2003	Logbooks with sleep quality ratings	4 items related to falling asleep, depth of sleep, difficulties arising and feeling rested	Daily	3 weeks	Unclear	Yes (each type of shift reported separately)	Each shift reported separately
Garde 2020	7 items of the Karolinska Sleep Diary	The first sleep episode after the night shift or as sleep dur-	During intervention	26	Once a day upon awaken-	Yes (each night shift reported separately)	Each shift reported sepa-

Table 4. Measurement of sleep quality *(Continued)*

		ing the night after recovery days.			ing from their primary sleep		rately (not combined)
Knauth 1998	NA	NA	NA	NA	NA	NA	NA
Totterdell 1992	Questionnaire using 10-cm VAS ranging from "worst" to "best"	10-cm VAS ranging from "worst" to "best"	1 month before and 6 months after implementation of the intervention	2 (before and after)	No specific time (retrospective questionnaire)	No (only the average of all days including rest days)	No (only the average of all days including rest days)
Viitasalo 2008	BNSQ and ESS. We could not include the ESS data in this review.	Sum of points of 4 questions of the BNSQ (provided by study authors on request)	5–6 months before and 7–8 months after implementation of the intervention	2 (before and after)	No specific time (retrospective questionnaire)	No (only the average of all days including rest days)	No (only the average of all days including rest days)

BNSQ: Basic Nordic Sleep Questionnaire; ESS: Epworth Sleepiness Scale; KSS: Karolinska Sleepiness Scale; NA: not applicable; PSQI: Pittsburgh Sleep Quality Index.

Table 5. Measurement of sleep duration off-shift

Study	Measurement instrument	Definition of sleep period	Time point of measurement	Number of measurements	Timing of measurements	Reported separately for night shifts	Reported for all shifts combined (without rest days)
Amendola 2011	Sleep diaries	Total sleep over 24 hours	Baseline and 6 months after implementation of the intervention	Twice daily for 2 weeks	Unclear	No (only the average of all days including rest days)	No (only the average of all days including rest days)
Axelsson 1998	Sleep diary	Main sleep episode (naps not included)	During the intervention	23 shifts over 6 weeks	Daily after each main sleep period	Yes (mean sleep after all night shifts reported for intervention and control)	Each shift reported separately (not combined)
Barger 2019a	Actiwatch	Weekly average including all sleep period including naps	During the intervention (1 month)	NA	Continuously	No (only the average of all days including rest days)	No (only the average of all days including rest days)
Barton 1994	Questionnaire	Average sleep duration after shift	2 months before and 6 months after implementation of the intervention	2 (before and after)	No specific time (retrospective questionnaire)	Yes (for each night shift)	Each shift reported separately (not combined)

Table 5. Measurement of sleep duration off-shift *(Continued)*

Basner 2019	Actiwatch	Average sleep time per 24 hours	During the intervention (14 days)	NA	Continuously	Yes (night shift for the control group and two extended overnight shifts for the intervention group)	No (only the average of all days including rest days)
Bell 2015	Questionnaire	Hours of sleep per night according to PSQI	Baseline and 3 and 6 months after implementation of the intervention	3	No specific time (retrospective questionnaire)	No (only the average of all days including rest days)	No (only the average of all days including rest days)
Cruz 2003	Questionnaire and wrist activity sensor	Hours of sleep per 24 hours and nap time	Daily for 3 weeks	NA	Continuously	Yes	Each shift reported separately (not combined)
Garde 2020	Actiwatch	The first sleep episode after the night shift or as sleep during the night after recovery days.	During the intervention (26 days)	NA	Continuously	Yes (each night shift reported separately)	Each shift reported separately (not combined)
Knauth 1998	Questionnaire	Average sleep time per 24 hours	Before and 10 months after implementation of the intervention	2 (before and after)	No specific time (retrospective questionnaire)	No (only the average of all shifts combined)	Yes
Totterdell 1992	Questionnaire	Sleep duration per 24-h averaged over shift cycle.	1 month before and 6 months after implementation of the intervention	2 (before and after)	No specific time (retrospective questionnaire)	Yes (mean sleep duration for night shifts)	No (only the average of all days including rest days)
Viitasalo 2008	Questionnaire	Total sleep over 24 hours	5–6 months before and 7–8 months after implementation of the intervention	2 (before and after)	No specific time (retrospective questionnaire)	No (only the average of all days including rest days)	No (only the average of all days including rest days)

NA: not applicable; PSQI: Pittsburgh Sleep Quality Index.

Table 6. Measurement of sleepiness

Study	Measurement instrument	Time point of measurement	Number of measurements during shift	Number of days measured	Timing of measurements	Reported for night shifts	Reported for all shifts
Amendola 2011	PVT and optical tracker (objective); KKS and ESS (subjective)	Baseline + 6 months after implementation of intervention	KSS every hour during work time; PVT 5 times before and after intervention	KSS: 2 times on each workday for 2 weeks PVT: 1 day before and after the intervention	KSS every hour during work time + PVT 5 times pre- and postintervention	No, all shifts combined	Yes, all shifts combined
Axelsson 1998	KSS (subjective)	During the intervention	Every 2 hours	23 shifts over 6 weeks	Every 2 hours during work and in free time	Yes, mean sleepiness over all night shifts reported for control and intervention period	Not combined
Barger 2019a	KSS (subjective)	During the intervention	Once a week	1 month	Approximately every 5 hours during a shift, including the beginning and end of the shift, once per week	No, all shifts combined	Yes, all shifts combined
Barton 1994	NA	NA	NA	NA	NA	NA	NA
Basner 2019	KSS (subjective)	During the intervention	1	14 days	Between 06:00 and 09:00	Yes (night shift for the control group and 2 extended overnight shifts for the intervention group)	No (only the average of all days including rest days)
Bell 2015	3-minute version of the PVT (objective); 'Daytime dysfunction due to sleepiness' item of the PSQI (subjective)	PVT at 1 and 5 months after implementation of intervention; PSQI at baseline and 3 and 6 months after implementation of intervention,	1	PVT twice; PSQI 3 times	PVT at last hour of last shift of the week; PSQI not specified	No, all shifts combined	Yes, all shifts combined

Table 6. Measurement of sleepiness *(Continued)*

Cruz 2003	SSS	During the intervention	4	3 weeks	00:00, 02:45, 04:45, and 07:30 h into the shift	Yes	Not combined
Garde 2020	NA	NA	NA	NA	NA	NA	NA
Knauth 1998	NA	NA	NA	NA	NA	NA	NA
Totterdell 1992	10-cm VAS ranging from "drowsy" to "alert" (subjective)	1 month before and 6 months after implementation of the intervention	Measured once (participants retrospectively recalled how they normally felt at specified 2-hourly intervals during each shift)	2 (before and after)	No specific time (retrospective questionnaire)	Yes, each shift reported separately	Not combined
Viitasalo 2008	BNSQ and ESS. ESS results could not be included in this review	5–6 months before and 7–8 months after implementation of the intervention	1	2 (before and after)	No specific time (retrospective questionnaire)	No, all days combined	No, all days combined

BNSQ: Basic Nordic Sleep Questionnaire; ESS: Epworth Sleepiness Scale; KSS: Karolinska Sleepiness Scale; PSQI: Pittsburgh Sleep Quality Index; PVT: psychomotor vigilance test; SSS: Stanford Sleepiness Scale

APPENDICES

Appendix 1. PubMed search strategy

(((((((((("rota*"[Title/Abstract] OR "system*"[Title/Abstract] OR "schedul*"[Title/Abstract] OR "hours"[Title/Abstract] OR "time"[Title/Abstract] OR "pattern*"[Title/Abstract] OR "cycle"[Title/Abstract] OR "extend*"[Title/Abstract] OR "evening"[Title/Abstract] OR "late"[Title/Abstract] OR "roster"[Title/Abstract] OR "early"[Title/Abstract] OR "weekend"[Title/Abstract] OR "twilight"[Title/Abstract] OR "graveyard"[Title/Abstract] OR "night*"[Title/Abstract] OR "split"[Title/Abstract] OR "non-standard"[Title/Abstract] OR "non-standard"[Title/Abstract] OR "flex*"[Title/Abstract] OR "turnaround"[Title/Abstract] OR "continuous"[Title/Abstract] OR "rotat*"[Title/Abstract]) AND "shift*"[Title/Abstract]) OR ("day"[Title/Abstract] AND "schedule*"[Title/Abstract])) AND ("sleep"[Title/Abstract] OR "sleepiness"[Title/Abstract] OR "circadian"[Title/Abstract] OR "vigilance"[Title/Abstract] OR "alertness"[Title/Abstract] OR "alert"[Title/Abstract] OR "wakefulness"[Title/Abstract] OR "drowsiness"[Title/Abstract] OR "fatigue"[Title/Abstract] OR "insomnia"[Title/Abstract] OR "hypersomnolence"[Title/Abstract] OR "dyssomnia"[Title/Abstract] OR "eveningness"[Title/Abstract] OR "morningness"[Title/Abstract] OR "concentration difficulties"[Title/Abstract] OR "attentiveness"[Title/Abstract] OR "arousal"[Title/Abstract] OR "performance"[Title/Abstract] OR "vigilant"[Title/Abstract] OR "nap"[Title/Abstract] OR "napping"[Title/Abstract] OR "rest"[Title/Abstract] OR "resting"[Title/Abstract])) OR ("errors"[Title/Abstract] OR "incidents"[Title/Abstract] OR "accidents"[Title/Abstract] OR "mistakes"[Title/Abstract] OR "safety"[Title/Abstract] OR "death, sudden, cardiac"[MeSH Terms] OR "death*"[Title/Abstract] OR "costs and cost analysis"[MeSH Terms] OR "costs"[Title/Abstract] OR "chronotherapy"[Title/Abstract] OR "light"[Title/Abstract] OR "daylight"[Title/Abstract] OR "dark"[Title/Abstract] OR "darkness"[Title/Abstract] OR "sleep disorders, intrinsic"[MeSH Terms] OR "sleep initiation and maintenance disorders"[MeSH Terms]) OR ("sleep"[Title/Abstract] OR "sleepiness"[Title/Abstract] OR "circadian"[Title/Abstract] OR "vigilance"[Title/Abstract] OR "alertness"[Title/Abstract] OR "alert"[Title/Abstract] OR "wakefulness"[Title/Abstract] OR "drowsiness"[Title/Abstract] OR "fatigue"[Title/Abstract] OR "insomnia"[Title/Abstract] OR "hypersomnolence"[Title/Abstract] OR "dyssomnia"[Title/Abstract] OR "eveningness"[Title/Abstract] OR "morningness"[Title/Abstract] OR "concentration difficulties"[Title/Abstract] OR "attentiveness"[Title/Abstract] OR "arousal"[Title/Abstract] OR "performance"[Title/Abstract] OR "vigilant"[Title/Abstract] OR "nap"[Title/Abstract] OR "napping"[Title/Abstract] OR "rest"[Title/Abstract] OR "resting"[Title/Abstract]) OR ("errors"[Title/Abstract] OR "incidents"[Title/Abstract] OR "accidents"[Title/Abstract] OR "mistakes"[Title/Abstract] OR "safety"[Title/Abstract] OR "death, sudden, cardiac"[MeSH Terms] OR "death*"[Title/Abstract] OR "costs and cost analysis"[MeSH Terms] OR "costs"[Title/Abstract] OR "chronotherapy"[Title/Abstract] OR "light"[Title/Abstract] OR "daylight"[Title/Abstract] OR "dark"[Title/Abstract] OR "darkness"[Title/Abstract] OR "sleep disorders, intrinsic"[MeSH Terms] OR "sleep initiation and maintenance disorders"[MeSH Terms])) AND ((("backward"[Title/Abstract] OR "forward"[Title/Abstract] OR "rapid"[Title/Abstract] OR "slow"[Title/Abstract] OR "slowly"[Title/Abstract] OR "advancing"[Title/Abstract] OR "delaying"[Title/Abstract]) AND ("rotation"[Title/Abstract] OR "rotate"[Title/Abstract] OR "rotating"[Title/Abstract])) OR ("day week"[Title/Abstract] OR "flexitime"[Title/Abstract] OR "hours of work"[Title/Abstract] OR "shiftwork*"[Title/Abstract]))) OR (("work*"[Title/Abstract] OR "duty"[Title/Abstract]) AND (("rota*"[Title/Abstract] OR "system*"[Title/Abstract] OR "schedul*"[Title/Abstract] OR "hours"[Title/Abstract] OR "time"[Title/Abstract] OR "pattern*"[Title/Abstract] OR "cycle"[Title/Abstract] OR "extend*"[Title/Abstract] OR "evening"[Title/Abstract] OR "late"[Title/Abstract] OR "roster"[Title/Abstract] OR "early"[Title/Abstract] OR "weekend"[Title/Abstract] OR "twilight"[Title/Abstract] OR "graveyard"[Title/Abstract] OR "night*"[Title/Abstract] OR "split"[Title/Abstract] OR "non-standard"[Title/Abstract] OR "non-standard"[Title/Abstract] OR "flex*"[Title/Abstract] OR "turnaround"[Title/Abstract] OR "continuous"[Title/Abstract] OR "rotat*"[Title/Abstract]) AND "shift*"[Title/Abstract]))) AND ("clinical trial"[Publication Type] OR "meta analysis"[Publication Type] OR "randomized controlled trial"[Publication Type]) AND "humans"[MeSH Terms]) OR (("sleep wake disorders"[MeSH Terms] OR "chronobiology disorders"[MeSH Terms] OR "circadian rhythm"[MeSH Terms] OR "wounds and injuries"[MeSH Terms] OR "occupational"[Title/Abstract]) AND ("injuries"[Title/Abstract] OR "occupational injuries"[Title/Abstract]))) AND ((("rota*"[Title/Abstract] OR "system*"[Title/Abstract] OR "schedul*"[Title/Abstract] OR "hours"[Title/Abstract] OR "time"[Title/Abstract] OR "pattern*"[Title/Abstract] OR "cycle"[Title/Abstract] OR "extend*"[Title/Abstract] OR "evening"[Title/Abstract] OR "late"[Title/Abstract] OR "roster"[Title/Abstract] OR "early"[Title/Abstract] OR "weekend"[Title/Abstract] OR "twilight"[Title/Abstract] OR "graveyard"[Title/Abstract] OR "night*"[Title/Abstract] OR "split"[Title/Abstract] OR "non-standard"[Title/Abstract] OR "non-standard"[Title/Abstract] OR "flex*"[Title/Abstract] OR "turnaround"[Title/Abstract] OR "continuous"[Title/Abstract] OR "rotat*"[Title/Abstract]) AND "shift*"[Title/Abstract]) OR ("day"[Title/Abstract] AND "schedule*"[Title/Abstract]))

Appendix 2. EMBASE search strategy

#1 ((rotat* adj3 (backward or forward or rapid or slow or rapidly or slowly or advancing or delaying)) and (shift* or work* or schedule or time or duty or hours or rota or roster)).ab,ti.

#2 (shift* adj2 (rota or system* or schedul* or hours or time or pattern* or cycle or extend* or evening or late or roster or early or weekend or twilight or graveyard or night* or split or non-standard or "non standard" or flex* or turnaround or continuous or rotat*)).ab,ti.

#3 (shift* adj3 (backward or forward or rapid or slow or rapidly or slowly or advancing or delaying or roster or rota)).ab,ti.

#4 (nightshift* or shiftwork*).ab,ti.

#5 exp sleep disorder/ or circadian rhythm/ or occupational accident/ or exp chronobiology/ or occupational injury.ab,ti.

#6 (sleep or sleepiness or circadian or vigilance or alertness or alert or wakefulness or drowsiness or fatigue or insomnia or hypersomnolence or dyssomnia or eveningness or morningness or "concentration difficulties" or attentiveness or arousal or performance).ab,ti.

#7 ((cross adj1 sectional) or compared or compares or cohort or cross-sectional or case-control or study or survey or surveys or diary or diaries or questionnaire* or groups or comparison* or multivariate or risk factor* or effectiveness).ab,ti.

#8 1 or 2 or 3 or 4

#9 5 or 6

#10 8 and 9 and 7

Appendix 3. Cochrane CENTRAL

#1 ((rotat* NEAR/2 (backward OR forward OR rapid OR slow OR rapidly OR slowly OR advancing OR delaying)) AND (shift* OR work* OR schedule OR time OR duty OR hours OR rota OR roster)):kw,ab,ti

#2 (shift$ NEAR/2 (rota OR system$ OR schedul* OR hours OR time OR pattern* OR cycle OR extend* OR evening OR late OR roster OR early OR weekend OR twilight OR graveyard OR night$ OR split OR non-standard OR "non standard" OR flex$ OR turnaround OR continuous OR rotat*)):kw,ab,ti

#3 (shift* NEAR/2 (backward OR forward OR rapid OR slow OR rapidly OR slowly OR advancing OR delaying) OR (roster OR rota) OR "day week"):kw,ab,ti

#4 (nightshift* OR shiftwork*):kw,ab,ti

#5 (Sleep Disorders OR Circadian Rhythm OR Sleep Phase Chronotherapy OR Chronotherapy OR Chronobiology Disorders OR Occupational Accident OR Occupational injury):kw,ab,ti

#6 (sleep OR sleepiness OR circadian OR vigilance OR alertness OR alert OR wakefulness OR drowsiness OR fatigue OR insomnia OR hypersomnolence OR dyssomnia OR eveningness OR morningness OR "neurocognitive performance" OR "concentration difficulties" OR attentiveness OR arousal OR performance OR vigilant OR nap OR napping OR rest OR resting):kw,ab,ti

#7 #1 OR #2 OR #3 OR #4

#8 #5 OR #6

#9 #7 AND #8

limited to CENTRAL

Appendix 4. Scopus

((((((TITLE-ABS-KEY (work W/2 hour*)) OR (TITLE-ABS-KEY (work W/2 week)) OR (TITLE-ABS-KEY (shift W/2 work*)) OR (TITLE-ABS-KEY (day W/2 schedule*)) OR ((TITLE-ABS-KEY (nightshift) OR TITLE-ABS-KEY (shiftwork*))) OR ((TITLE-ABS-KEY (rotat* W/1 backward) OR TITLE-ABS-KEY (rotat* W/1 forward) OR TITLE-ABS-KEY (rotat* W/1 rapid*) OR TITLE-ABS-KEY (rotat* W/1 slow*) OR TITLE-ABS-KEY (rotat* W/1 advancing) OR TITLE-ABS-KEY (rotat* W/1 delaying)) AND ((TITLE-ABS-KEY (shift*) OR TITLE-ABS-KEY (work*) OR TITLE-ABS-KEY (schedule) OR TITLE-ABS-KEY (time) OR TITLE-ABS-KEY (duty) OR TITLE-ABS-KEY (hours) OR TITLE-ABS-KEY (rota) OR TITLE-ABS-KEY (roster))))) OR (TITLE-ABS-KEY (shift* W/1 rota) OR TITLE-ABS-KEY (shift* W/1 system*) OR TITLE-ABS-KEY (shift* W/1 schedul*) OR TITLE-ABS-KEY (shift* W/1 hours) OR TITLE-ABS-KEY (shift* W/1 pattern*) OR TITLE-ABS-KEY (shift* W/1 cycle) OR TITLE-ABS-KEY (shift* W/1 extend*) OR TITLE-ABS-KEY (shift* W/1 evening) OR TITLE-ABS-KEY (shift* W/1 late) OR TITLE-ABS-KEY (shift* W/1 roster) OR TITLE-ABS-KEY (shift* W/1 early) OR TITLE-ABS-KEY (shift* W/1 weekend) OR TITLE-ABS-KEY (shift* W/1 twilight) OR TITLE-ABS-KEY (shift* W/1 graveyard) OR TITLE-ABS-KEY (shift* W/1 night*) OR TITLE-ABS-KEY (shift* W/1 split) OR TITLE-ABS-KEY (shift* W/1 non-standard) OR TITLE-ABS-KEY (shift* W/1 "non standard") OR TITLE-ABS-KEY (shift* W/1 flex*) OR TITLE-ABS-KEY (shift* W/1 turnaround) OR TITLE-ABS-KEY (shift* W/1 continuous) OR TITLE-ABS-KEY (shift* W/1 rotat*)) OR ((TITLE-ABS-KEY (roster) OR TITLE-ABS-KEY (rota) OR TITLE-ABS-KEY ("day week"))) OR (TITLE-ABS-KEY (shift* W/1 backward) OR TITLE-ABS-KEY (shift* W/1 forward) OR TITLE-ABS-KEY (shift* W/1 rapid*) OR TITLE-ABS-KEY (shift* W/1 slow*) OR TITLE-ABS-KEY (shift* W/1 advancing) OR TITLE-ABS-KEY (shift* W/1 delaying))) AND (((TITLE-ABS-KEY (sleep) OR TITLE-ABS-KEY (sleepiness) OR TITLE-ABS-KEY (circadian) OR TITLE-ABS-KEY (vigilance) OR TITLE-ABS-KEY (alertness) OR TITLE-ABS-KEY (alert) OR TITLE-ABS-KEY (wakefulness) OR TITLE-ABS-KEY (drowsiness) OR TITLE-ABS-KEY (fatigue) OR TITLE-ABS-KEY (insomnia) OR TITLE-ABS-KEY (hypersomnolence) OR TITLE-ABS-KEY (dyssomnia) OR TITLE-ABS-KEY (eveningness) OR TITLE-ABS-KEY ("neurocognitive performance") OR TITLE-ABS-KEY ("concentration difficulties") OR TITLE-ABS-KEY (attentiveness) OR TITLE-ABS-KEY (arousal) OR TITLE-ABS-KEY (performance) OR TITLE-ABS-KEY (vigilant))) OR ((TITLE-ABS-KEY (nap) OR TITLE-ABS-KEY (napping) OR TITLE-ABS-KEY (rest) OR TITLE-ABS-KEY (resting) OR TITLE-ABS-KEY (errors) OR TITLE-ABS-KEY (accidents) OR TITLE-ABS-KEY (incidents) OR TITLE-ABS-KEY (mistakes) OR TITLE-ABS-KEY (safety) OR TITLE-ABS-KEY (death*) OR TITLE-ABS-KEY (mortality) OR TITLE-ABS-KEY (injur*) OR TITLE-ABS-KEY (chronotherapy) OR TITLE-ABS-KEY (ligth) OR TITLE-ABS-KEY (daylight) OR

TITLE-ABS-KEY (dark) OR TITLE-ABS-KEY (darkness) OR TITLE-ABS-KEY (econom*) OR TITLE-ABS-KEY (cost*))))) AND (((TITLE-ABS-KEY ("randomized controlled trial") OR TITLE-ABS-KEY ("controlled clinical trail") OR TITLE-ABS-KEY ("controlled trial") OR TITLE-ABS-KEY (random) OR TITLE-ABS-KEY (double-blind) OR TITLE-ABS-KEY ("double blind") OR TITLE-ABS-KEY (single-blind) OR TITLE-ABS-KEY ("single blind") OR TITLE-ABS-KEY ("clinical trial*"))) OR (((TITLE-ABS-KEY (singl*) OR TITLE-ABS-KEY (doubl*) OR TITLE-ABS-KEY (trebl*) OR TITLE-ABS-KEY (tripl*))) AND ((TITLE-ABS-KEY (mask*) OR TITLE-ABS-KEY (blind*)))) OR ((TITLE-ABS-KEY ("latin square") OR TITLE-ABS-KEY (placebo*) OR TITLE-ABS-KEY ("research design") OR TITLE-ABS-KEY ("comparative stud*") OR TITLE-ABS-KEY ("evaluation stud*") OR TITLE-ABS-KEY ("follow-up stud*") OR TITLE-ABS-KEY ("prospective stud*") OR TITLE-ABS-KEY ("cross-over stud*") OR TITLE-ABS-KEY (volunteer*))))) AND NOT (TITLE-ABS-KEY (animal*))) AND NOT (TITLE-ABS-KEY (human*))

Appendix 5. PsycINFO

#1 sleep.ti. or sleep.ab. or sleepiness.ab. or sleepiness.ti. or circadian.ab. or cicardian.ti. or vigilance.ab. or vigilance.ti. or alertness.ab. or alertness.ti. or alert.ab. or alert.ti. or wakefulness.ab. or wakefullness.ti. or drowsiness.ab. or drowsiness.ti. or fatigue.ab. or fatigue.ti. or insomnia.ab. or insomnia.ti.

#2 hypersomnolence.ti. or hypersomnolence.ab. or dyssomnia.ab. or dyssomnia.ti. or eveningness.ab. or eveningness.ti. or morningness.ab. or morningness.ti. or "neurocognitive performance".ab. or "neurocognitive performance".ti. or "concentration difficulties".ab. or "concentration difficulties".ti. or attentiveness.ab. or attentiveness.ti. or arousal.ab. or arousal.ti. or performance.ab. or performance.ti. or vigilant.ab. or vigilant.ti.

#3 1 or 2

#4 work scheduling.mp.

#5 workday shift.mp. or exp Workday Shifts/

#6 3 or 4 or 5

#7 work hour*.ti. or "work hour*".ab. or "shift work".ab. or "shift work*".ti. or "work* week".ti. or "work* week".af. or nightshift*.ab. or nightshift*.ti. or shiftwork*.ab. or shiftwork*.ti. or "day schedule".ab. or "day schedule*".ti.

#8 work*.ti. or work*.ab. or schedule.ab. or schedule.ti. or time.ti. or time.af. or duty.ab. or duty.ti. or hours.ab. or hours.ti. or rota.ab. or rota.ti. or roster.ti. or roster.ab.

#9 rotat*.ti. or rotat*.ab.

#10 8 and 9

#11 rota.ti. or rota.ab. or systems.ab. or systems.ti. or schedul*.ti. or schedul*.ab. or hours.ab. or hours.ti. or pattern*.ab. or pattern*.ti. or time.ab. or time.ti. or cycle.ti. or cycle.ab. or extend*.ab. or extend*.ti. or evening.ab. or evening.ti.

#12 late.ti. or late.ab. or early.ab. or early.ti. or roster.ti. or roster.ab. or weekend.ab. or weekend.ti. or twilight.ab. or twilight.ti. or graveyard.ab. or graveyard.ti. or night*.ti. or night*.ab. or split.ab. or split.ti. or non-standard.ab. or non-standard.ti. or "non standard".ab. or "non standard".ti. or flex*.ab. or flex*.ti. or turnaround.ab. or turnaround.ti. or continuous.ab. or continuous.ti. or rotat*.ab. or rotat*.ti.

#13 backward.ti. or backward.ab. or forward.ab. or forward.ti. or rapid.ti. or rapid.ab. or slow.ab. or slow.ti. or rapidly.ab. or rapidly.ti. or slowly.ab. or slowly.ti. or advancing.ti. or advancing.ab. or delaying.ab. or delaying.ti. or roster.ab. or roster.ti. or rota.ab. or rota.ti. or "day week".ab. or "day week".ti.

#14 shift*.ti. or shift*.ab.

#15 11 or 12 or 13

#16 14 and 15

#17 6 or 7 or 10 or 16

#18 4 or 5

#19 7 or 10 or 16 or 18

#20 exp Human Biological Rhythms/

#21 exp Sleepiness/

#22 exp Sleep Deprivation/

#23 sleep wake disorders.mp.

#24 exp Sleep/

#25 exp Physiological Arousal/

#26 exp Fatigue/

#27 exp performance/

#28 exp Occupational Safety/

#29 exp Napping/

#30 exp Job Performance/

#31 exp Wakefulness/

#32 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31

#33 10 or 32

#34 19 and 33

#35 exp Risk Factors/

#36 control.ti. or control.ab. or "cross sectional".ab. or "cross sectional".ti. or compared.ti. or compared.ab. or compares.ab. or compares.ti. or cohort.ab. or cohort.ti. or cross-sectional.ab. or cross-sectional.ti. or "case control".ti. or "case control".ab. or study.ab. or study.ti.

#37 survey*.ti. or survey*.ab. or diary.ab. or diary.ti. or diaries.ti. or diaries.ab. or questionnaire*.ab. or questionnaire*.ti. or evaluation.ab. or evaluation.ti. or evaluate.ab. or evaluate.ti. or groups.ti. or groups.ab. or comparison*.ab. or comparison.ti.

#38 effectiveness.ti. or effectiveness.ab. or random*.ab. or random*.ti. or allocation.ti. or allocation.ab. or allocate.ab. or allocate.ti. or allocated.ab. or allocated.ti.

#39 35 or 36 or 37 or 38

#40 34 and 39

#41 exp Work Rest Cycles/ or work.mp. or exp Work Scheduling/

#42 occupation.mp. or exp Occupations/

#43 41 or 42

#44 40 and 43

#45 limit 44 to (human and english language and adulthood <18+ years>)

Appendix 6. OSH UPDATE

#1 GW{sleep OR sleepiness OR circadian OR vigilance OR alertness OR alert OR wakefulness OR drowsiness OR fatigue OR insomnia OR hypersomnolence OR dyssomnia OR eveningness OR morningness OR neurocognitive performance OR concentration difficult* OR attentiveness OR arousal OR performance OR vigilant OR nap OR napping OR rest OR resting OR errors OR incidents OR accidents OR mistakes OR safety OR deaths OR death OR mortality OR injury OR injuries OR chronotherapy OR light OR daylight OR dark OR darkness}

#2 GW{nightshift* OR night shift* OR shiftwork OR shift work OR rotating shift* OR roster OR rota OR work schedule OR work system* OR shift system* OR hours of work OR work hour*}

#3 GW{random* or trial* or control* or blind*}

#4 #1 AND #2 AND #3

#5 DC{OUHSEL OR OUCISD OR OUNIOC OR OUNIOS}

#6

#4 AND #5

Appendix 7. LILACS

(tw:((tw:(MH:C10.886.425.800.800$ OR MH:F03.870.400.800.800$ OR Insomnia OR "Disorders of Initiating and Maintaining Sleep" OR Sleeplessness OR Insomnio OR "Trastornos de la Mantención e Inicio del Sueño" OR "Trastornos de la Iniciación y Mantención del Sueño" OR "Falta de Sueño" OR Insônia)) OR (tw:(MH:C10.281$ OR "Chronobiology Disorders" OR "Circadian Rhythm Disorders" OR "Trastornos Cronobiológicos" OR "Trastornos del Ritmo Circadiano" OR "Transtornos Cronobiológicos" OR "Transtornos do Ritmo Circadiano")) OR (tw:(MH:G07.180.562.190$ OR "Circadian Rhythm" OR "Diurnal Rhythm" OR "Nyctohemeral Rhythm" OR "Twenty-Four Hour Rhythm" OR "Ritmo Circadiano" OR "Ritmo Diurno" OR "Ritmo Nictohemeral" OR "Ritmo de Veinticuatro Horas" OR "Ritmos Circadianos" OR "Ritmo de Vinte e Quatro Horas")) OR (tw:(chronotherapy OR light OR daylight OR darkness OR cronoterapia OR luz OR oscuridad OR escuridao)) OR (tw:(sleep OR sueno OR sono OR sleepiness OR drownsiness OR somnolencia OR sonolencia OR circadian OR circadian* OR alertness OR alert OR alerta OR wakefulness OR vigilancia OR fatigue OR fadiga OR insomnia OR insomnio OR insonia)) OR (tw:(nap OR napping OR siesta OR cochilo OR rest OR resting OR descanso OR repouso)))) AND (tw:((tw:(((shift OR shifts) n1 (rota OR system OR systems OR schedul* OR hours OR time OR pattern* OR roster OR twilight OR graveyard OR night*)))) OR (tw:(((turno OR turnos) n2 (rotat* OR sistema OR regime OR horario OR programa OR noturno)))) OR (tw:((rota OR roster OR nightshift OR shiftwork OR "horario de trabajo" OR "plan de trabajo" OR "escala de trabalho" OR "trabalho noturno"))))) AND (tw:(((PT:"randomized controlled trial" OR PT:"controlled clinical trial" OR PT:"multicenter study" OR MH:"randomized controlled trials as topic" OR MH:"controlled clinical trials as topic" OR MH:"multicenter study as topic" OR MH:"random allocation" OR MH:"double-blind method" OR MH:"single-blind method") OR ((ensaio$ OR ensayo$ OR trial$) AND (azar OR acaso OR placebo OR control$ OR aleat$ OR random$ OR enmascarado$ OR simpleciego OR ((simple$ OR single OR duplo$ OR doble$ OR double$) AND (cego OR ciego OR blind OR mask))) AND clinic$)) AND NOT (MH:animals OR MH:rabbits OR MH:rats OR MH:primates OR MH:dogs OR MH:cats OR MH:swine OR PT:"in vitro")))

Appendix 8. OPEN GREY

(((work NEAR/2 hour*) OR (shift NEAR/2 work*) OR (work* NEAR/2 week) OR nightshift* OR shiftwork* OR (day NEAR/2 schedule) OR ((rotat* NEAR/1 (backward OR forward OR rapid OR slow OR rapidly OR slowly OR advancing OR delaying) AND (shift* OR work* OR schedule OR time OR duty OR hours OR rota OR roster)) OR (shift$ NEAR/1 (rota OR system$ OR schedul* OR hours OR time OR pattern* OR cycle OR extend* OR evening OR late OR roster OR early OR weekend OR twilight OR graveyard OR night$ OR split OR non-standard OR "non standard" OR flex$ OR turnaround OR continuous OR rotat*)) OR (shift* NEAR/2 (backward OR forward OR rapid OR slow OR rapidly OR slowly OR advancing OR delaying OR roster OR rota OR “day week”)) AND (sleep OR sleepiness OR circadian OR vigilance OR altertness OR alert OR wakefulness OR drowsiness OR fatigue OR insomnia OR hypersomnolence OR dyssomnia OR eveningness OR morningness OR "neurocognitive performance" OR "concentration difficulties" OR attentiveness OR arousal OR performance OR vigilant OR nap OR napping OR rest OR resting OR errors OR incidents OR accidents OR mistakes OR safety OR deaths OR death OR mortality OR injury OR injuries OR chronotherapy OR light OR daylight OR dark OR darkness OR econom$ OR cost OR costs OR light OR dark OR darkness OR goggles OR exercise))

HISTORY

Protocol first published: Issue 7, 2013

CONTRIBUTIONS OF AUTHORS

Screening titles and abstracts: GH, PC, GG, MP, CB, IW, DP
Third person to solve disagreements: BG, RR, DP
Data extraction: GH, PC, GG, BG, IW, DP
Risk of bias assessment: GH, PC, MP, BG, DP
Third person to solve disagreements in risk of bias assessment: BG, RR, DP
Check references of all primary studies and review studies: PC, JL, CB, DP
Contact experts in the field: IW, DP

DECLARATIONS OF INTEREST

GH: none known
PC: none known
GG: none known
MP: none known
BG: none known
CB: is an occupational health physician at Medmark; otherwise no known conflicts of interest
IW: none known
JL: none known
RR: none known
DP: none known

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention.

SOURCES OF SUPPORT

Internal sources

- The Netherlands Organization for Applied Scientific Research, TNO, Netherlands

  The study was supported with internal funding of TNO.

External sources

- None, Other

  No external sources of support have been used for this review.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

At the review stage, we defined different components of shift systems to use as a structure for establishing comparisons and presenting results. When developing this structure, we took into account the multi-component nature of shift systems. We deviated from the protocol (Erren 2013) as we decided to include additional interventions related to shift schedules that emerged during the data extraction phase as important comparisons. For that reason, we also included regularity of shift changes, distribution shift schedules, rest time between shifts, split shifts, protected sleep, and worker participation.

We removed the subgroup analysis based on shift schedule details as we structured the comparisons based on a framework that could account for multi-components of shift systems.

Blinding of participants and personnel: we judged this domain differently as it was not possible to blind participants or organising personnel to different shift schedules. We considered low risk of bias for objective measures and high risk of bias for subjective measures.

We changed the title from 'Adaptation of shift work schedules for preventing and treating sleepiness and sleep disturbances caused by shift work' to 'Adapting shift work schedules for sleep quality, sleep duration and sleepiness in shift workers' based on peer reviewers' and editorial comments.

We planned to assess only the sleep duration off-shift, however for a study in which the intervention involved extended shift duration with the possibility of resting during shift, we considered the reported 24-hour sleep duration for the outcome sleep duration off-shift.

We did not search the database ProQuest Dissertations and Theses.