Published in final edited form as:

Appl Ergon. 2017 January ; 58: 386–397. doi:10.1016/j.apergo.2016.07.018.

Job Rotation Designed to Prevent Musculoskeletal Disorders and Control Risk in Manufacturing Industries: A Systematic Review

Rosimeire Simprini Padula, PhD^{1,*}, Maria Luiza Caires Comper, PhD¹, Emily H. Sparer, PhD², and Jack T Dennerlein, PhD^{3,4}

¹Department of Physical Therapy. Masters and Doctoral Programs in Physical Therapy, Universidade Cidade de São Paulo, São Paulo, Brazil

²Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States

³Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States

⁴Department of Physical Therapy, Movement, and Rehabilitation Sciences Bouvé College of Health Sciences, Northeastern University, Boston, Massachusetts, United States

Abstract

To better understand job rotation in the manufacturing industry, we completed a systematic review asking the following questions: 1) How do job-rotation programs impact work-related musculoskeletal disorders (MSDs) and related risk control for these MSDs, as well as psychosocial factors? and 2) How best should the job rotation programs be designed? We searched MEDLINE, EMBASE, Business Source Premier, ISI Web of Knowledge, CINAHL, PsycINFO, Scopus, and SciELO databases for articles published in peer-reviewed journals. Eligible studies were examined by two independent reviewers for relevance (population of manufacturing workers, outcomes of musculoskeletal disease, physical factors, psychosocial factors, and strategies used in job-rotation implantation) and methodological quality rating. From 10,809 potential articles, 71 were read for full text analysis. Of the 14 studies included for data extraction, two were non-randomized control trial studies, one was a case-control study, and 11 were cross-sectional comparisons. Only one, with a case-control design, was scored with good methodological quality. Currently, weak evidence exists supporting job rotation as a strategy for the prevention and control of musculoskeletal disorders. Job rotation did not appear to reduce the exposure of physical risk factors; yet, there are positive correlations between job rotation and higher job satisfaction. Worker training has been described as a crucial component of a successful job-rotation program. The studies reported a range of parameters used to implement and measure job-rotation programs. More rigorous studies are needed to better understand the full impact of job rotation on production and health.

*Corresponding Author: Rosimeire Simprini Padula, rosipadula@gmail.com, Masters and Doctoral Programs in Physical Therapy, Universidade Cidade de São Paulo, Rua Cesário Galeno 475, 03071-000 - São Paulo-SP, Brazil. Tel.: +55 11 21781564.
PROSPERO Register: CRD420140133191.

Keywords

task rotation; ergonomic; industrial workers

Introduction

Job-rotation programs emerged in the 1980s and 1990s as organizational strategies aimed at increasing the performance and the flexibility of workers (Cristini and Pozzoli, 2010; Kernan and Sheahan, 2012). These programs have often been adopted by engineers and managers to reduce time and production costs (Azizi and Liang, 2013; Corominas et al., 2006; Moreira and Costa, 2013). The initial motivations for implementing job-rotation programs were part of a lean production system and total quality, focused on the need for more workers with more autonomy (Corominas et al., 2006; Cristini and Pozzoli, 2010). Job-rotation programs are also frequently recommended to mitigate continuous exposure to risk factors for musculoskeletal disorders (MSDs) (Comper and Padula, 2014; Leider et al., 2015a; Mathiassen, 2006).

The definitions for job rotation are many and vary according to the purpose for which this strategy is adopted. In terms of management, it can be defined as alternating workers between tasks and jobs that require different skills and responsibility (Huang, 1999). In terms of MSD risk control, job rotation is defined as a strategy for alternating workers between tasks with different exposure levels and occupational demands (Howarth et al., 2009; Jorgensen et al., 2005), which aims at avoiding overloading specific body parts (Mathiassen, 2006).

The specifics in planning and implementing job-rotation programs vary across each professional sector, which is essential to the success of the intervention (Frazer et al., 2003). The success depends also on training workers in several jobs and defining the specific parameters needed to generate the most effective job rotation in risk control and prevention of MSDs (Leider et al., 2015b). In order to be successful, the planning and implementing needs to identify physical, cognitive, and organizational demands; to determine exposure levels; and to evaluate and define how the job-rotation schedule will be created. All workers should then be trained in each job to develop competence and skills and to ensure process and product quality (Guimarães et al., 2012). The planning and implementation needs to also consider how other factors, such as psychosocial (e.g., job satisfaction, engagement) and environmental factors can affect worker health outcomes and the success of the health promotion and prevention programs (Ho et al., 2009; Park and Jang, 2010).

There are numerous reasons justified by the manufacturing industry and described in the literature for using job rotation as an ergonomic organizational strategy (Corominas et al., 2006; Jorgensen et al., 2005). As a result, studies that have evaluated the effects of job rotation do not always use the same criteria to evaluate the positive or negative aspects of this intervention, leading to challenges in practical application for practitioners and researchers.

The review presented here aims to identify evidence for the benefits of job rotation on improving musculoskeletal health and to investigate whether or not this evidence varies across different implementation designs. The definition of job rotation used in this study is the rotation of workers in tasks with different exposure levels and job demands for workers who have a daily work period of eight hours (480 minutes), with a break for lunch.

The specific questions addressed in this review are:

1. What is the effect of job rotation in manufacturing workers? In terms of:
 - a. specific work-related musculoskeletal issues (disorders, complaints, injury, pain, discomfort)
 - b. risk control for MSDs, specifically exposure to physical load (posture, force, biomechanics, fatigue, effort exertion)
 - c. psychosocial work factors (job satisfaction, stress, job control, engagement)
2. How should such job-rotation programs be designed?

2. Methods

2.1. Search Strategy

Independent searches were conducted in electronic databases: MEDLINE, CINAHL, EMBASE, Business Source Premier, ISI Web of Knowledge, PsyINFO, Scopus, and SciELO, in the English language, with no restriction on publication data. The search terms were defined based on the list of terms used in the systematic review studies of the Institute for Work and Health (IWH) and the National Institute for Occupational Safety and Health (NIOSH). The search terms were grouped into three categories according to the principles of PICO: *population*, *intervention*, *comparison*, and *outcomes*. The following includes examples of the group terms: *population* (*workers* or *employees*), *intervention* (*job rotation* or *task rotation*), and *outcomes* (*musculoskeletal disorders* or *work ability*). The groups of terms used for search strategies can be seen in Appendix 1. The last search was completed on October 17, 2014.

2.2. Inclusion and Exclusion Criteria

The eligible studies contained the following criteria: (1) the population of manufacturing workers; (2) exposures to known risk factors (biomechanical overload, repetitive tasks, fatigue, posture at work, force, etc.), MSDs (pain, discomfort, injury, absence from work, absenteeism), and psychosocial factors (job satisfaction, stress, job control, engagement); (3) written in English; (4) full text papers published in peer-reviewed journals; and (5) designed observational studies: cohort, case-control, randomized control trial, and cross-sectional. Excluded studies comprised the following: outcomes of productivity and costs only, outcomes that were assessed through qualitative methods only, studies in which the definition of job rotation was different from our aforementioned definition, and those that evaluated the variability of factors within a single given task without changing tasks. The design of job-rotation programs was reported independently of study designs and effect size.

It should be noted that the size of the effect of job rotation is best demonstrated in randomized control trials; however, we were not able to locate any studies with this design for this analysis. The information extracted to describe the implementation of job rotation was: (1) exposure (physical, organizational, and cognitive) factors targeted by the rotation including physical factors, work demands, worker autonomy, engagement in the job rotation program, worker performance, and level of the concentration; (2) protocols and intervals used to determine rotation schedules; (3) tools and a schedule for organizing job rotation (simple or equation); (4) prioritizing the order of the criteria (different work demands and exposure). The search results were exported to EndNote® X7 software, where duplicates were removed and extraction of data was obtained in full text.

2.3. Study Analysis

The review process consisted of first identifying studies to include through a review of the titles and reading the abstracts. If necessary, the full text was then reviewed. Two independent researchers (RP and ML) completed this review. In the cases of any disagreement between the two reviewers, a decision was reached by consensus, or a third researcher (ES) provided further review and the decision was made by arbitration.

2.4 Methodological Quality Assessment of Studies

Identified studies were assessed for their quality according to internal and external validity criteria (Sanderson et al., 2007). We used the quality assessment tools proposed by the United States' National Institutes of Health (NIH). These included a tool for the Quality Assessment of Controlled Intervention Studies (Appendix 2), a tool for Observational Cohort and Cross-Sectional Studies (Appendix 3), and a tool for Case-Control Studies (Appendix 4), the latter two tools for Observational Cohort and Cross-sectional Studies and the tool for Case Control Studies had 14 items and 12 items, respectively. Items in these tools included e.g: reporting, sample size, inclusion criteria, measures of exposure, assessment bias, and statistical analysis. Each of the study tools used in this study had an accompanying guidance document that was used to help keep the reviews objective. The guidance documents provided examples of outcomes evaluated, information on measurement accuracy, and program timing, as well as other criteria (randomization and participant allocation, sample blinding, dropout rate, and confounding). This information along with previous training, provided the framework and experience to evaluate the internal validity of each study (Padula et al., 2012).

The quality rating was classified as good, fair, or poor according to the general analysis of the evaluators considering all the items described above. The study quality was determined by the total number of positive (+) items and the rating classification adopted by Wong et al. (2008). The studies with 67% or more positive-item checks were an indication of good quality, studies with 34–66% positive checks were an indication of fair quality, and 33% or less were an indication of poor quality.

2.5. Data Extraction

Data extraction was performed on the final identified studies and included the following: (1) year of publication; (2) the manufacturing industry sector; (3) country where the study was

conducted; (4) study design type; (5) characteristics of participants; (6) type of task (e.g., repetitive with short cycles, handling load); (7) outcomes evaluated, such as musculoskeletal issue, biomechanical exposure factors, and psychosocial factors (e.g., injury, complaints and musculoskeletal symptoms, fatigue, level of risk, muscle activity, perceived job satisfaction, stress, job control); (8) follow-up; (9) job-rotation effects (results); (10) ergonomic parameters used to create job rotation (e.g., biomechanics, organizational, cognitive, motor-learning curves, psychosocial aspects); (11) protocol and interval (time) used between rotation (number of rotations per day); (12) tools to create rotation and organize job rotation (e.g., simple or using an equation) and daily schedule (from easy to difficult; high cognitive/low physical demands); and (13) priorities to make decisions (competences and training workers, physical or cognitive demands, biomechanical exposure).

The methods for creating job-rotation schedules were classified in one of two categories. We refer to the first category as “simple” because the job-rotation schedule was created according to one or more criteria (biomechanical risk rating, demand rate, complexity of task) and it was assessed one by one (Guimarães et al., 2012). We refer to the second category as “Equation” because the job-rotation schedules were created by a mathematical equation that addresses multiple factors at the same time to more precisely determine the best job-rotation schedules available to workers. The equation accounted for ergonomic aspects of the job and competence of the tasks (Asensio-Cuesta et al., 2012b).

The data extraction process and review were identical to the study-selection process with two independent reviewers extracting the data and, in cases of disagreement, consensus was reached through discussion between the reviewers or by arbitration with a third reviewer.

In the original proposal published in the PROSPERO register, it was indicated that a meta-analysis would be carried out. However, due to the large heterogeneity of the criteria and measuring instruments used in the studies, the analysis was not possible. Thus, data from this study are presented descriptively. The study followed the *PRISMA Statement* for reporting systematic reviews (Liberati et al., 2009).

3. Results

3.1. Characteristics of the Studies and Methodological Quality

The search resulted in 10,809 potentially eligible studies, of which 1,362 were duplicates and 9,376 were excluded based on the review of the titles and abstracts. An additional 57 studies were excluded after reviewing the full text, leaving a final 14 studies (Figure 1) for data extraction.

In the reviewed studies, the participants were of the both genders, of working age (18–65 years), with samples ranging from 11 to 957 workers, and with 4 to 25 workstations.

The final 14 studies were from seven countries, the majority of which (eight) featured and examined an assembly-line area in the automotive/automobile industry. Most studies (nine) proposed job-rotation schedules exclusively between tasks with short-cycle times and highly repetitive movements (Table 1).

Of the 14 studies included, only one was rated as good (case-control design) (Roquelaure et al., 1997), six as fair (Balogh et al., 2006; Dawal et al., 2009; Dawal and Taha, 2007; Fredriksson et al., 2001; Guimarães et al., 2012; Sato and Coury, 2009), and seven as poor methodological quality (Asensio-Cuesta et al., 2012a; Asensio-Cuesta et al., 2012b; Carnahan et al., 2000; Diego-Mas et al., 2009; Filus and Okimorto, 2012; Frazer et al., 2003; Tharmmaphornphilas and Norman, 2007a).

Of the final 14 studies, two were non-randomized control trials (Guimarães et al., 2012; Fredriksson et al., 2001), one had case-control design (Roquelaure et al., 1997), and 11 were of cross-sectional designs (Table 2).

In terms of outcomes, five studies targeted MSD prevention (outcomes included measures of disability, disorders, pain, discomfort) (Balogh et al., 2006; Guimarães et al., 2012; Fredriksson et al., 2001; Roquelaure et al., 1997; Sato and Coury, 2009), 11 studies targeted changes in exposure to physical risk factors (biomechanics, repetition, fatigue, effort exertion) (Asensio-Cuesta et al., 2012a; Asensio-Cuesta et al., 2012b; Balogh et al., 2006; Carnahan et al., 2000; Diego-Mas et al., 2009; Frazer et al., 2003; Fredriksson et al., 2001; Filus and Okimorto, 2012; Roquelaure et al., 1997; Sato and Coury, 2009; Tharmmaphornphilas and Norman, 2007a), and three targeted the job-rotation effect on psychosocial factors (job satisfaction) (Dawal et al., 2009; Dawal and Taha, 2007; Guimarães et al., 2012).

The indicators and tools used to evaluate the outcomes varied widely in the studies. The majority used tools that had either dichotomous (yes or no) or Likert-scale responses. The outcomes were evaluated immediately after job rotation occurred or in a follow-up period of between one and three and a half years (Table 3).

3.2. Effect of Job Rotation on MSDs

Overall, it appears that the evidence for the job-rotation effect is mixed. One study with good methodological quality, featuring a case-control design found a positive association (OR 6.3; $P < 0.05$) between no job-rotation implementation and carpal tunnel syndrome (Roquelaure et al., 1997). However, the non-randomized control trial studies with fair methodological quality showed the opposite. Guimarães et al. (2012) reported decreases of MSDs (from 7.00% to 1.41%; $P < 0.05$), absenteeism (from 6.63% to 3.60%; $P < 0.05$) and absence from work (from 2.22% to 0.44%; $P < 0.05$) in group-performed job rotation. Fredriksson et al. (2001) observed significant increases in musculoskeletal symptoms of the lower back ($P < 0.05$) after job-rotation implementation. Many studies have indicated that job rotation did not influence ($P > 0.05$) sick-leave amount (Fredriksson et al., 2001; Sato and Coury, 2009) or musculoskeletal complaints (Balogh et al., 2006) (Table 3).

3.3. Effects of Job Rotation on Exposure to Physical Factors

Workers in a one-year job-rotation intervention reported a statistically significant increase ($P < 0.05$) in their perception of exertion (from 6.9 to 8.4 points) when compared to workers without a job-rotation program (Fredriksson et al., 2001). Sato and Coury (2009) reported that exertion can be increased at a statistically significant rate ($P < 0.05$) during the workday when workers are exposed to a wide variety of tasks due to job rotation. Frazer et al. (2003)

reported a linear increase in the risk of lower-back injury by 57% (peak and cumulative tissue load) due to different demands in jobs (low and high) generated by job rotation. Some authors also predicted an increase of incidences and severity of lower-back issues as a result of job rotation (Carnahan, 2000; Tharmmaphornphilas and Norman, 2007a).

In addition, high-risk tasks and better recovery time from cumulative fatigue were observed when job rotation occurred every two hours (Asensio-Cuesta et al., 2012a; Filus and Okimorto, 2012), with a three-hour maximum allowable time at a given workstation (Asensio-Cuesta et al., 2012b). To reduce workload by movement, breaks of ten-minute durations have been suggested between rotations (Diego-Mas et al., 2009). Biomechanical risk factors related to motion repetitiveness (short circles of less than 10 seconds), little task variation, and the absence of breaks were observed to increase the chance of injury (OR 8.8; $P < 0.05$) (Roquelaure et al., 1997).

The physical load (trapezius and infraspinatus muscular activities, work posture, movement, and repetitiveness) was decreased significantly ($P < 0.05$) in assembly-line jobs with job-rotation programs. However, a job-rotation program does not appear to reduce musculoskeletal complaints (Balogh et al., 2006) (Table 3).

3.4 Effects of Job Rotation on Psychosocial Work Factors

Many studies have indicated that job satisfaction outcomes increase with the implementation of a job-rotation program (Dawal et al., 2009; Dawal and Taha, 2007; Guimarães et al., 2012). Job rotations were implemented in two automotive companies, with positive correlations (Auto 1. $R^2 = 0.804$; Auto 2. $R^2 = 0.667$) between job-rotation performance and job satisfaction (Dawal et al., 2009), with more than 70% of workers satisfied (Dawal and Taha, 2007).

Despite these positive findings, Fredriksson et al. (2001) reported workers having lower job control and, consequently, were less stimulated as a result of job rotation; however, this study did not rate job satisfaction directly (Table 3).

3.5 Design of Job-Rotation Programs

The parameters used to create job-rotation schedules varied between studies; however, the most used to evaluate working conditions were biomechanical (13 studies) and organizational factors (12 studies), with 10 studies using both parameters. Cognitive-mental aspects (seven studies), environmental and safety factors were reported in five studies. Only three studies (Asensio-Cuesta et al., 2012b; Diego-Mas et al., 2009; Guimarães et al., 2012) used all the parameters mentioned above.

In most of the articles, after the parameter description, the author(s) indicated that rotation occurred between different biomechanical risk levels or task complexity levels or both. There were six studies that used mathematical equations to organize the job-rotation schedules, and four studies organized the job rotation using the parameters previously defined, often with the support of the production teams, and health and safety (Asensio-Cuesta et al., 2012a; Asensio-Cuesta et al., 2012b; Carnahan et al., 2000; Diego-Mas et al., 2009; Fraser and Hvolby, 2010; Tharmmaphornphilas and Norman, 2007a).

Most of the studies describing job-rotation schedules were organized in four rotations per day, (three after two hours long and one after an hour of work) (Asensio-Cuesta et al., 2012b; Comper and Padula, 2014; Diego-Mas et al., 2009; Guimarães et al., 2012). Three studies proposed job rotation schedules by evaluating the best time between each schedule, considering aspects such as fatigue (Asensio-Cuesta et al., 2012a; Asensio-Cuesta et al., 2012b; Filus and Okimorto, 2012), characteristics of production and tasks (Asensio-Cuesta et al., 2012a; Carnahan et al., 2000; Tharmmaphornphilas and Norman, 2007a), and three studies did not report the time between each job rotation (Dawal et al., 2009; Dawal and Taha, 2007; Roquelaure et al., 1997).

Of all of the recommendations, it appeared that the most important was the training of workers according to their competences in the tasks. Parameters used for organizing and implementing the job-rotation schedule are reported in Table 4.

4. Discussion

The results of this study have demonstrated weak evidence for the effectiveness of job rotation when implemented in manufacturing industries, regardless of the target of rotation, to prevent and control work-related musculoskeletal disease and to reduce exposure to physical factors. Improvements in job satisfaction as a result of job rotation were observed in some studies, yet it should be noted that these studies were deemed to have only fair methodological quality.

Job rotation strategies varied from study to study. Most used biomechanical and organizational factors to identify the risk factors of the job-rotation schedules. Many studies emphasized the importance of training of workers according to their competences and defining clearly which job-rotation program was possible for each worker. Almost all of the studies showed an increase in exposure to physical load doing job rotation. There was also some similarity in the scheduling of job-rotation programs, as job-rotation shifts tended to vary between one and two hours at each workstation according to the fatigue analysis or defined by the context of production and breaks.

These mixed results concerning job rotation can be explained by the variability in study designs and outcome measures. The gold-standard level of evidence to evaluate interventions is randomized control trial design (Schelvis et al., 2015), yet despite an extensive search of the literature, we were not able to find any studies with this design. However, much can be learned from studies with other designs (Burns et al., 2011), such as the highly scored Roquelaure et al. (1997), which employed a cross-sectional design. We completed a systematic review to try to understand what the existing body of evidence says about the effects of job rotation

The only non-randomized control trials, proposed by Guimarães et al. (2012) and Fredriksson et al. (2001), had fair quality, indicating a high risk of bias. Guimarães et al. (2012) reported positive results associated with job rotation, while Fredriksson et al. (2001) determined that job rotation did not contribute to MSD reduction, although workers reported being more satisfied, as in the study by Guimarães et al. (2012). This is likely due to a

different understanding of the meaning of job rotation as a strategy for the prevention and control of MSDs. Fredriksson et al. (2001) used it as a way to explain the fragmentation of tasks and to increase the production rate. They evaluated workers of the car-body-sealing department at an automobile assembly plant and compared those who worked on cars in a line-out system (reference group) with workers on job-rotation schedules in a line system where each worker did part of the job (intervention group). The line system with job rotation had worse outcomes than the line-out system—the physical workload and MSDs increased significantly ($P < 0.05$). It is possible that by increasing the job rotation, workers had more exposure to biomechanical risk, which could reflect a lack of prioritization of ergonomic principles in the job rotation program risk.

When multiple workers perform tasks with high-biomechanical risk, as opposed to just a single worker, there is the potential for exposure bracketing and an increase in risk factors for other workers (Barrero, Katz, and Dennerlein, 2009). The principles of ergonomics have not previously been prioritized in decisions regarding how to implement job rotation, meaning that there are jobs with high exposure that thus need improvement in terms of safety so that no injury befalls workers. Allocating all workers to all tasks throughout the whole job rotation means all will be exposed to high risks and increased chances of MSDs. The combination of incorrect job-rotation principles leading to increased exposure to MSD risk factors with the low methodological quality of the studies may be the reason why we were not able to demonstrate job-rotation effectiveness.

Guimarães et al. (2012) prioritized the acquisition of skills by workers in job-rotation schedules and evaluated job-rotation effects after three and a half years, indicating a significant reduction in work-related musculoskeletal injuries, absenteeism, turnover, rework, and spoilage among those with job-rotation scheduling compared with workers not in a job rotation program. In addition, Roquelaure et al. (1997) found a positive association between carpal tunnel syndrome and workers who perform job rotation over a year and 11 months, and the perception of exertion increased significantly. However, the design of this study (case-control) had a potential risk of bias, especially because the data was collected retrospectively, making it difficult to assess exposure. Others studies have demonstrated that job rotation increases the perception of exertion, risks, and the severity of lower-back injury. This could be related to repetitiveness of the movement and the overload observed due to the variety of tasks, increasing the exposure to biomechanical risk factors.

There were three studies that reported positive results of job-rotation implementation increasing job satisfaction (Dawal and Taha, 2007; Dawal et al., 2009; Guimarães et al., 2012). However, while these results were not directly related to the prevention or control of MSDs, they surely influenced the acceptance of job rotation, which is also associated with health outcomes (Rissen et al., 2002). The studies by Dawal et al. (2009) and Dawal and Taha (2007) evaluated the effect of various organizational factors (work methods, training, job rotation) showing job-rotation programs were positively correlation with job satisfaction, independent of age, and that the job rotation program had approval rating of 70% from workers. Guimarães et al. (2012) observed that training workers in job rotation enabled them to carry out tasks with more confidence and they showed more satisfaction with their work, although the study's methodological quality was only fair and with a cross-sectional design.

We found a great number of parameters for the organization and implementation of job rotation. Biomechanical and organizational parameters were most used and most often included aspects related to posture, movement frequency, the level of exposure, and the level of task complexity. These findings were consistent with the ergonomic job-rotation purposes of alternating tasks of different complexity and biomechanical exertion (Guimarães et al., 2012; Mathiassen, 2006). On the other hand, interestingly, we found a large number of studies that proposed job rotation of tasks with high repeatability and no possibility of variation of production rate and movement (Asensio-Cuesta et al., 2012a; Asensio-Cuesta et al., 2012b; Balogh et al., 2006; Diego-Mas et al., 2009; Filus and Okimorto, 2012; Fraser and Hvolby, 2010; Fredriksson et al., 2001; Roquelaure et al., 1997). In these cases, the job rotation strategies increased worker exposure but might not have reduced overload, which could have exacerbated worker job satisfaction.

The implementation of job rotation in tasks with high repeatability seems to meet the problem caused by fragmented work on production lines. This result could be related to the fact that many companies implemented job-rotation programs with production, cost, and quality in mind, rather than employee health (Azizi and Liang, 2013; Costa and Miralles, 2009; Michalos et al., 2010; Moreira and Costa, 2013). Another related explanation could be that companies implemented a job-rotation program with the goal of skill expansion or of production rearrangement (Balogh et al., 2006; Guimarães et al., 2012; Fredriksson et al., 2001; Roquelaure et al., 1997; Sato and Coury, 2009).

Regarding job-rotation schedules, most studies used mathematical equations or algorithms to create proposals for switching tasks for situations where the working day was eight hours, with intervals of one or two hours (Asensio-Cuesta et al., 2012b; Carnahan et al., 2000; Comper and Padula, 2014; Diego-Mas et al., 2009; Guimarães et al., 2012; Fredriksson et al., 2001; Sato and Coury, 2009). The reasons for the choice of these intervals were hardly mentioned in any of the studies. Only one study, by Guimarães et al. (2012), found that a job-rotation schedule of every two hours rather than an average one and a half hours was better for the group. Job rotation occurred four times daily, each period lasting two hours until the worker acquired a skill for a large number of tasks, and then the time average between each job rotation was one and a half hours and the tasks could be decided by the worker him/herself (Guimarães et al., 2012).

Based on this review, we cannot recommend a standard job-rotation schedule. This is because, in general, the studies reported little about the strategies used to perform the rotation, indicating that job rotation was likely not implemented as a preventive measure for MSD reduction, rather a need for engineering the flexibility of workers. Many of the equations proposed to define job-rotation schedules relied on a variety of parameters (e.g., posture and movement, mental and cognitive capacities, strategies and policies, learning skills, responsibility, long-distance vision, color vision, audibly identifying direction, writing, and speaking. However, none of the studies described how they determined that these were the important factors. Thus, much remains to be studied on these equations to integrate theory and practice and thereby contribute to decision making in job-rotation implementation.

Job rotation can be an important intervention when ergonomic principles are employed, helping to reduce hazards, even if it is not possible to fully implement rotation (Comper and Padula, 2014). Industries that use job rotation as a solution for ergonomic concerns instead of finding the root of the problem and making major ergonomic improvements are failing to achieve injury prevention goals. Managers should be aware of job rotation as an organizational strategy at the administrative level. They should provide adequate training to workers in the various tasks and should not forget the ergonomic contexts of each task. Another important aspect was that many studies identified in this review had only poor or fair methodological quality due to insufficient levels of details regarding the study population, sample size, power, randomization, blinding, dropout rate, intervention, and outcomes. The problem of the poor quality of the studies could possibly be explained by the difficulty in implementing new health-and-safety programs in worksites.

Dempsey (2007) described the benefits, progress, and the barriers encountered in the design of research on the effectiveness of ergonomic interventions to prevent MSDs and found that there was negative bias in the evidence of the studies. Although expanded database research with a focus on manufacturing industries, there was weak evidence for the effectiveness of job rotation in reducing overload and preventing MSDs (Leider et al., 2015a). Further studies are required with better methodological quality, and it is also necessary to consider alternative research designs (Schelvis et al., 2015) taking into account, among other factors, the specificities of each manufacturing industry and the possibilities within different production organizations. Identifying facilitators or barriers in each organizational level could also contribute to job-rotation planning and implementation (Leider et al., 2015b).

Successful implementation of job-rotation programs depends on a number of important factors, such as the involvement and acceptance of managers, job characteristics, production methods, the number of employees, and those tasks involving the biomechanics of exposure level and participatory ergonomics. It is thus necessary to analyze those activities that will be part of the job rotation, to sort out complexity, to perform training (Guimaraes et al., 2012), and to determine how long each worker should remain doing each task (Tharmmaphornphilas and Norman, 2007b). Meeting these requirements is important for achieving positive results concerning worker health. In addition, we believe that scheduling should be selected based on the specifics of the job and individual characteristics of the workers. We must also not forget that industrial innovation, adjustments in production, and ergonomic improvements create different exposure scenarios and difficult intervention strategies (Dempsey, 2007).

5. Limitations

A limitation of this study relates to how the articles included in the review were selected, as only information in the title and abstract were reviewed initially. Also, the studies did not follow clear criteria for the description of their methods, which meant it was difficult to find information often located in another part of the text. Some studies were excluded because the results of different manufacturing industries were presented together, preventing the analysis of information. Others did not present central tendency and dispersion

measurements or statistical analysis. Finally, due to the heterogeneity of the included studies and lack of statistical data, it was not possible to carry out a meta-analysis.

6. Conclusion

We were only able to find a limited number of current studies on job rotation in manufacturing industries. Therefore, it was not possible to make any solid conclusions on job-rotation effectiveness with regard to prevention and control of MSDs. Weak evidence exists for the reduction of exposure to physical overload and for the influence of psychosocial factors. Although some studies have attempted to provide support for the advantages of a job-rotation program, the methodological quality was often poor and they had inappropriate designs for assessing outcomes.

Acknowledgments

The National Counsel of Technological and Scientific Development (CNPq), Brazil (473651/2013-0 and 249621/2013-4), is funding this study.

References

- Asensio-Cuesta S, Diego-Mas JA, Canós-Darós L, Andrés-Romano C. A genetic algorithm for the design of job rotation schedules considering ergonomic and competence criteria. *International Journal of Advanced Manufacturing Technology*. 2012b; 60:1161–1174.
- Asensio-Cuesta S, Diego-Mas JA, Cremades-Oliver LV, González-Cruz MC. A method to design job rotation schedules to prevent work-related musculoskeletal disorders in repetitive work. *International Journal of Production Research*. 2012a; 50:7467–7478.
- Azizi N, Liang M. An integrated approach to worker assignment, workforce flexibility acquisition, and task rotation. *Journal of the Operational Research Society*. 2013; 64:260–275.
- Balogh I, Ohlsson K, Hansson G-Å, Engström T, Skerfving S. Increasing the degree of automation in a production system: Consequences for the physical workload. *International Journal of Industrial Ergonomics*. 2006; 36:353–365.
- Barrero LH, Katz JN, Dennerlein JT. Validity of self-reported mechanical demands for occupational epidemiologic research of musculoskeletal disorders. *Scandinavian Journal of Work, Environment & Health*. 2009; 35:245–260.
- Burns PB, Rohrich RJ, Chung KC. The Levels of Evidence and their role in Evidence-Based Medicine. *Plastic and Reconstructive Surgery*. 2011; 128:305–310. [PubMed: 21701348]
- Carnahan BJ, Redfern MS, Norman B. Designing safe job rotation schedules using optimization and heuristic search. *Ergonomics*. 2000; 43:543–560. [PubMed: 10801086]
- Comper MLC, Padula RS. The effectiveness of job rotation to prevent work-related musculoskeletal disorders: Protocol of a cluster randomized clinical trial. *BMC Musculoskeletal Disorders*. 2014;15.
- Corominas A, Pastor R, Rodriguez E. Rotational allocation of tasks to multifunctional workers in a service industry. *International Journal of Production Economics*. 2006; 103:3–9.
- Costa AM, Miralles C. Job rotation in assembly lines employing disabled workers. *International Journal of Production Economics*. 2009; 120:625–632.
- Cristini A, Pozzoli D. Workplace practices and firm performance in manufacturing. *International Journal of Manpower*. 2010; 31:818–842.
- Dawal SZ, Taha Z, Ismail Z. Effect of job organization on job satisfaction among shop floor employees in automotive industries in Malaysia. *International Journal of Industrial Ergonomics*. 2009; 39:1–6.
- Dawal SZM, Taha Z. The effect of job organizational factors on job satisfaction in two automotive industries in Malaysia. *Immediately*. 2007; 36:63–68.
- Dempsey PG. Effectiveness of ergonomics interventions to prevent musculoskeletal disorders: Beware of what you ask. *International Journal of Industrial Ergonomics*. 2007; 37:169–173.

- Diego-Mas JA, Asensio-Cuesta S, Sanchez-Romero MA, Artacho-Ramirez MA. A multi-criteria genetic algorithm for the generation of job rotation schedules. *International Journal of Industrial Ergonomics*. 2009; 39:23–33.
- Filus R, Okimorto ML. The effect of job rotation intervals on muscle fatigue - lactic acid. *Work*. 2012; 41:1572–1581. [PubMed: 22316939]
- Fraser K, Hvolby HH. Effective teamworking: can functional flexibility act as an enhancing factor?: An Australian case study. *Team Performance Management*. 2010; 16:74–94.
- Frazer MB, Norman RW, Wells RP, Neumann WP. The effects of job rotation on the risk of reporting low back pain. *Ergonomics*. 2003; 46:904–919. [PubMed: 12775488]
- Guimaraes LB, Anzanello MJ, Renner JS. A learning curve-based method to implement multifunctional work teams in the Brazilian footwear sector. *Applied ergonomics*. 2012; 43:541–547. [PubMed: 21907970]
- Ho WH, Chang CS, Shih YL, Liang RD. Effects of job rotation and role stress among nurses on job satisfaction and organizational commitment. *BMC health services research*. 2009; 9:8. [PubMed: 19138390]
- Howarth SJ, Beach TA, Pearson AJ, Callaghan JP. Using sitting as a component of job rotation strategies: are lifting/lowering kinetics and kinematics altered following prolonged sitting. *Appl Ergon*. 2009; 40:433–439. [PubMed: 19081557]
- Huang HJ. Job rotation from the employees' point of view. *Research and Practice in Human Resource Management*. 1999; 7:75–85.
- Jorgensen M, Davis K, Kotowski S, Aedla P, Dunning K. Characteristics of job rotation in the Midwest US manufacturing sector. *Ergonomics*. 2005; 48:1721–1733. [PubMed: 16373313]
- Fredriksson C, Bildt G, Hagg Kilbom A. The impact on musculoskeletal disorders of changing physical and psychosocial work environment conditions in the automobile industry. *International Journal of Industrial Ergonomics*. 2001; 28:31–45.
- Kernan B, Sheahan C. An investigation into heuristics for alternative worker selection in discrete event simulation. *Journal of Simulation*. 2012; 7:61–67.
- Leider PC, Boschman JS, Frings-Dresen MH, van der Molen HF. Effects of job rotation on musculoskeletal complaints and related work exposures: a systematic literature review. *Ergonomics*. 2015a; 58:18–32. [PubMed: 25267494]
- Leider PC, Boschman JS, Frings-Dresen MH, van der Molen HF. When is job rotation perceived useful and easy to use to prevent work-related musculoskeletal complaints? *Applied ergonomics*. 2015b; 51:205–210. [PubMed: 26154219]
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009; 151:W65–94. [PubMed: 19622512]
- Mathiassen SE. Diversity and variation in biomechanical exposure: what is it, and why would we like to know? *Applied ergonomics*. 2006; 37:419–427. [PubMed: 16764816]
- Michalos G, Makris S, Rentzos L, Chyssolouris G. Dynamic job rotation for workload balancing in human based assembly systems. *CIRP Journal of Manufacturing Science and Technology*. 2010; 2:153–160.
- Moreira MCO, Costa AM. Hybrid heuristics for planning job rotation schedules in assembly lines with heterogeneous workers. *International Journal of Production Economics*. 2013; 141:552–560.
- Padula RS, Pires RS, Alouche SR, Chiavegato LD, Lopes AD, Costa LO. Analysis of reporting of systematic reviews in physical therapy published in Portuguese. *Rev Bras Fisioter*. 2012; 16:381–388.
- Park JK, Jang SH. Association between Upper Extremity Musculoskeletal Disorders and Psychosocial Factors at Work: A Review on the Job DCS Model's Perspective. *Safety and health at work*. 2010; 1:37–42. [PubMed: 22953161]
- Rissen D, Melin B, Sandsjo L, Dohns I, Lundberg U. Psychophysiological stress reactions, trapezius muscle activity and neck and shoulder pain among female cashiers before and after introduction of job rotation. *Work and Stress*. 2002; 16:127–137.

- Roquelaure Y, Mechali S, Dano C, Fanello S, Benetti F, Bureau D, Mariel J, Martin YH, Derriennic F, PenneauFontbonne D. Occupational and personal risk factors for carpal tunnel syndrome in industrial workers. *Scandinavian Journal of Work Environment & Health*. 1997; 23:364–369.
- Sato, TdO, Coury, HJCG. Evaluation of musculoskeletal health outcomes in the context of job rotation and multifunctional jobs. *Applied ergonomics*. 2009; 40:707–712. [PubMed: 18675951]
- Schelvis RM, Oude Hengel KM, Burdorf A, Blatter BM, Strijk JE, van der Beek AJ. Evaluation of occupational health interventions using a randomized controlled trial: challenges and alternative research designs. *Scandinavian journal of work, environment & health*. 2015; 41:491–503.
- Tharmamphornphilas W, Norman BA. A methodology to create robust job rotation schedules. *Annals of Operations Research*. 2007a; 155:339–360.
- Tharmamphornphilas W, Norman BA. A methodology to create robust job rotation schedules. *Annals of Operations Research*. 2007b; 155:339–360.

Appendix 1. Groups of terms used for search strategies

Groups of terms	Search terms	
Population (Work related)	company or companies	operator
	contractor	organization
	department	occupation or occupations
	employee	plant
	employer	retail
	employment	skilled trade
	firm	staff
	factory or factories	supervisor
	industry	team
	institution	work
	labor	work environment
	laborer	workstation
	leadership	worker
	manager	workplace
Intervention (Job rotation)	ergonomics	schedule production
	flexible work	personnel rotation
	inservice training	task allocation
	job organization	task rotation
	job transfer	task performance
	job rotation	task modify
	job design	organizational learning
	multi criteria	work schedule
	multi task	work rotation
	multiskilling	
Outcome or factors then affecting	schedule	
	absenteeism	engagement
	accidents occupational	functional capacity assessment
	accommodate	health protection
	benefit duration	health risk management

Groups of terms	Search terms	
	benefit	healthy workplace strategy
	biomechanics	healthy workplace
	claim	job accommodation
	cost effectiveness analysis	job adaptation
	cumulative trauma disorder	job control
	disability management program	job demand
	early intervention	job performance
	employee satisfaction survey	job satisfaction
	employee satisfaction	job turnover
	job modification	promote recovery
	joint labor management initiative	reassign
	long term disability benefit	recovery
	long-term disability	reduced cost
	lost time	rehab
	lost workday	reintegration
	labor force participation	return on investment
	management of individual	return to work
	organizational policy	stay at work
	musculoskeletal injuries	short-term disability
	musculoskeletal pain	time management
	musculoskeletal disorders	training
	musculoskeletal system	health protection
	musculoskeletal diseases	work capacity
	occupational diseases	work ability
	occupational exposure	work resumption
	occupational health	work intervention
	outcome assessment	work intervention
	performance indicators	work adjustment
	performance management	workers compensation
	periodic medical examination	workload
	personnel management	WRMSDs
	prevention	WRMSD
	productivity ration	
	productivity	

Appendix 2. Quality Assessment of Controlled Intervention Studies

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the study described as randomized, a randomized trial, a randomized clinical trial, or an RCT?			

Criteria	Yes	No	Other (CD, NR, NA)*
2. Was the method of randomization adequate (i.e., use of randomly generated assignment)?			
3. Was the treatment allocation concealed (so that assignments could not be predicted)?			
4. Were study participants and providers blinded to treatment group assignment?			
5. Were the people assessing the outcomes blinded to the participants' group assignments?			
6. Were the groups similar at baseline on important characteristics that could affect outcomes (e.g., demographics, risk factors, co-morbid conditions)?			
7. Was the overall drop-out rate from the study at endpoint 20% or lower of the number allocated to treatment?			
8. Was the differential drop-out rate (between treatment groups) at endpoint 15 percentage points or lower?			
9. Was there high adherence to the intervention protocols for each treatment group?			
10. Were other interventions avoided or similar in the groups (e.g., similar background treatments)?			
11. Were outcomes assessed using valid and reliable measures, implemented consistently across all study participants?			
12. Did the authors report that the sample size was sufficiently large to be able to detect a difference in the main outcome between groups with at least 80% power?			
13. Were outcomes reported or subgroups analyzed prespecified (i.e., identified before analyses were conducted)?			
14. Were all randomized participants analyzed in the group to which they were originally assigned, i.e., did they use an intention-to-treat analysis?			
Quality Rating (Good, Fair, or Poor) (see guidance)			
Rater #1 initials:			
Rater #2 initials:			
Additional Comments (If POOR, please state why):			

* CD, cannot determine; NA, not applicable; NR, not reported

Guidance for Assessing the Quality of Controlled Intervention Studies

The guidance document below is organized by question number from the tool for quality assessment of controlled intervention studies.

Question 1. Described as randomized

Was the study described as randomized? A study does not satisfy quality criteria as randomized simply because the authors call it randomized; however, it is a first step in determining if a study is randomized

Questions 2 and 3. Treatment allocation—two interrelated pieces

Adequate randomization: Randomization is adequate if it occurred according to the play of chance (e.g., computer generated sequence in more recent studies, or random number table in older studies).

Inadequate randomization: Randomization is inadequate if there is a preset plan (e.g., alternation where every other subject is assigned to treatment arm or another method of allocation is used, such as time or day of hospital admission or clinic visit, ZIP Code, phone number, etc.). In fact, this is not randomization at all—it is another method of assignment to groups. If assignment is not by the play of chance, then the answer to this question is no.

There may be some tricky scenarios that will need to be read carefully and considered for the role of chance in assignment. For example, randomization may occur at the site level, where all individuals at a particular site are assigned to receive treatment or no treatment. This scenario is used for group-randomized trials, which can be truly randomized, but often are “quasi-experimental” studies with comparison groups rather than true control groups. (Few, if any, group-randomized trials are anticipated for this evidence review.)

Allocation concealment: This means that one does not know in advance, or cannot guess accurately, to what group the next person eligible for randomization will be assigned. Methods include sequentially numbered opaque sealed envelopes, numbered or coded containers, central randomization by a coordinating center, computer-generated randomization that is not revealed ahead of time, etc.

Questions 4 and 5. Blinding

Blinding means that one does not know to which group—intervention or control—the participant is assigned. It is also sometimes called “masking.” The reviewer assessed whether each of the following was blinded to knowledge of treatment assignment: (1) the person assessing the primary outcome(s) for the study (e.g., taking the measurements such as blood pressure, examining health records for events such as myocardial infarction, reviewing and interpreting test results such as x ray or cardiac catheterization findings); (2) the person receiving the intervention (e.g., the patient or other study participant); and (3) the person providing the intervention (e.g., the physician, nurse, pharmacist, dietitian, or behavioral interventionist).

Generally placebo-controlled medication studies are blinded to patient, provider, and outcome assessors; behavioral, lifestyle, and surgical studies are examples of studies that are frequently blinded only to the outcome assessors because blinding of the persons providing and receiving the interventions is difficult in these situations. Sometimes the individual providing the intervention is the same person performing the outcome assessment. This was noted when it occurred.

Question 6. Similarity of groups at baseline

This question relates to whether the intervention and control groups have similar baseline characteristics on average especially those characteristics that may affect the intervention or outcomes. The point of randomized trials is to create groups that are as similar as possible except for the intervention(s) being studied in order to compare the effects of the interventions between groups. When reviewers abstracted baseline characteristics, they noted when there was a significant difference between groups. Baseline characteristics for intervention groups are usually presented in a table in the article (often Table 1).

Groups can differ at baseline without raising red flags if: (1) the differences would not be expected to have any bearing on the interventions and outcomes; or (2) the differences are not statistically significant. When concerned about baseline difference in groups, reviewers recorded them in the comments section and considered them in their overall determination of the study quality.

Questions 7 and 8. Dropout

“Dropouts” in a clinical trial are individuals for whom there are no end point measurements, often because they dropped out of the study and were lost to followup.

Generally, an acceptable overall dropout rate is considered 20 percent or less of participants who were randomized or allocated into each group. An acceptable differential dropout rate is an absolute difference between groups of 15 percentage points at most (calculated by subtracting the dropout rate of one group minus the dropout rate of the other group). However, these are general rates. Lower overall dropout rates are expected in shorter studies, whereas higher overall dropout rates may be acceptable for studies of longer duration. For example, a 6-month study of weight loss interventions should be expected to have nearly 100 percent followup (almost no dropouts—nearly everybody gets their weight measured regardless of whether or not they actually received the intervention), whereas a 10-year study testing the effects of intensive blood pressure lowering on heart attacks may be acceptable if there is a 20–25 percent dropout rate, especially if the dropout rate between groups was similar. The panels for the NHLBI systematic reviews may set different levels of dropout caps.

Conversely, differential dropout rates are not flexible; there should be a 15 percent cap. If there is a differential dropout rate of 15 percent or higher between arms, then there is a serious potential for bias. This constitutes a fatal flaw, resulting in a poor quality rating for the study.

Question 9. Adherence

Did participants in each treatment group adhere to the protocols for assigned interventions? For example, if Group 1 was assigned to 10 mg/day of Drug A, did most of them take 10 mg/day of Drug A? Another example is a study evaluating the difference between a 30-pound weight loss and a 10-pound weight loss on specific clinical outcomes (e.g., heart attacks), but the 30-pound weight loss group did not achieve its intended weight loss target (e.g., the group only lost 14 pounds on average). A third example is whether a large percentage of participants assigned to one group “crossed over” and got the intervention provided to the other group. A final example is when one group that was assigned to receive a particular drug at a particular dose had a large percentage of participants who did not end up taking the drug or the dose as designed in the protocol.

Question 10. Avoid other interventions

Changes that occur in the study outcomes being assessed should be attributable to the interventions being compared in the study. If study participants receive interventions that are

not part of the study protocol and could affect the outcomes being assessed, and they receive these interventions differentially, then there is cause for concern because these interventions could bias results. The following scenario is another example of how bias can occur. In a study comparing two different dietary interventions on serum cholesterol, one group had a significantly higher percentage of participants taking statin drugs than the other group. In this situation, it would be impossible to know if a difference in outcome was due to the dietary intervention or the drugs.

Question 11. Outcome measures assessment

What tools or methods were used to measure the outcomes in the study? Were the tools and methods accurate and reliable—for example, have they been validated, or are they objective? This is important as it indicates the confidence you can have in the reported outcomes. Perhaps even more important is ascertaining that outcomes were assessed in the same manner within and between groups. One example of differing methods is self-report of dietary salt intake versus urine testing for sodium content (a more reliable and valid assessment method). Another example is using BP measurements taken by practitioners who use their usual methods versus using BP measurements done by individuals trained in a standard approach. Such an approach may include using the same instrument each time and taking an individual's BP multiple times. In each of these cases, the answer to this assessment question would be “no” for the former scenario and “yes” for the latter. In addition, a study in which an intervention group was seen more frequently than the control group, enabling more opportunities to report clinical events, would not be considered reliable and valid.

Question 12. Power calculation

Generally, a study's methods section will address the sample size needed to detect differences in primary outcomes. The current standard is at least 80 percent power to detect a clinically relevant difference in an outcome using a two-sided alpha of 0.05. Often, however, older studies will not report on power.

Question 13. Prespecified outcomes

Investigators should prespecify outcomes reported in a study for hypothesis testing—which is the reason for conducting an RCT. Without prespecified outcomes, the study may be reporting ad hoc analyses, simply looking for differences supporting desired findings. Investigators also should prespecify subgroups being examined. Most RCTs conduct numerous post hoc analyses as a way of exploring findings and generating additional hypotheses. The intent of this question is to give more weight to reports that are not simply exploratory in nature.

Question 14. Intention-to-treat analysis

Intention-to-treat (ITT) means everybody who was randomized is analyzed according to the original group to which they are assigned. This is an extremely important concept because conducting an ITT analysis preserves the whole reason for doing a randomized trial; that is,

to compare groups that differ only in the intervention being tested. When the ITT philosophy is not followed, groups being compared may no longer be the same. In this situation, the study would likely be rated poor. However, if an investigator used another type of analysis that could be viewed as valid, this would be explained in the “other” box on the quality assessment form. Some researchers use a completers analysis (an analysis of only the participants who completed the intervention and the study), which introduces significant potential for bias. Characteristics of participants who do not complete the study are unlikely to be the same as those who do. The likely impact of participants withdrawing from a study treatment must be considered carefully. ITT analysis provides a more conservative (potentially less biased) estimate of effectiveness.

General Guidance for Determining the Overall Quality Rating of Controlled Intervention Studies

The questions on the assessment tool were designed to help reviewers focus on the key concepts for evaluating a study’s internal validity. They are not intended to create a list that is simply tallied up to arrive at a summary judgment of quality.

Internal validity is the extent to which the results (effects) reported in a study can truly be attributed to the intervention being evaluated and not to flaws in the design or conduct of the study—in other words, the ability for the study to make causal conclusions about the effects of the intervention being tested. Such flaws can increase the risk of bias. Critical appraisal involves considering the risk of potential for allocation bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues addressed in the questions above. High risk of bias translates to a rating of poor quality. Low risk of bias translates to a rating of good quality.

Fatal flaws: If a study has a “fatal flaw,” then risk of bias is significant, and the study is of poor quality. Examples of fatal flaws in RCTs include high dropout rates, high differential dropout rates, no ITT analysis or other unsuitable statistical analysis (e.g., completers-only analysis).

Generally, when evaluating a study, one will not see a “fatal flaw;” however, one will find some risk of bias. During training, reviewers were instructed to look for the potential for bias in studies by focusing on the concepts underlying the questions in the tool. For any box checked “no,” reviewers were told to ask: “What is the potential risk of bias that may be introduced by this flaw?” That is, does this factor cause one to doubt the results that were reported in the study?

NHLBI staff provided reviewers with background reading on critical appraisal, while emphasizing that the best approach to use is to think about the questions in the tool in determining the potential for bias in a study. The staff also emphasized that each study has specific nuances; therefore, reviewers should familiarize themselves with the key concepts.

Appendix 3. Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies

Criteria	Yes	No	Other(CD, NR, NA)*
1. Was the research question or objective in this paper clearly stated?			
2. Was the study population clearly specified and defined?			
3. Was the participation rate of eligible persons at least 50%?			
4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?			
5. Was a sample size justification, power description, or variance and effect estimates provided?			
6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured?			
7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?			
8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)?			
9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
10. Was the exposure(s) assessed more than once over time?			
11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants?			
12. Were the outcome assessors blinded to the exposure status of participants?			
13. Was loss to follow-up after baseline 20% or less?			
14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)?			
Quality Rating (Good, Fair, or Poor) (see guidance)			
Rater #1 initials:			
Rater #2 initials:			
Additional Comments (If POOR, please state why):			

* CD, cannot determine; NA, not applicable; NR, not reported.

Guidance for Assessing the Quality of Observational Cohort and Cross-Sectional Studies

The guidance document below is organized by question number from the tool for quality assessment of observational cohort and cross-sectional studies.

Question 1. Research question

Did the authors describe their goal in conducting this research? Is it easy to understand what they were looking to find? This issue is important for any scientific paper of any type. Higher quality scientific research explicitly defines a research question.

Questions 2 and 3. Study population

Did the authors describe the group of people from which the study participants were selected or recruited, using demographics, location, and time period? If you were to conduct this study again, would you know who to recruit, from where, and from what time period? Is the cohort population free of the outcomes of interest at the time they were recruited?

An example would be men over 40 years old with type 2 diabetes who began seeking medical care at Phoenix Good Samaritan Hospital between January 1, 1990 and December 31, 1994. In this example, the population is clearly described as: (1) who (men over 40 years old with type 2 diabetes); (2) where (Phoenix Good Samaritan Hospital); and (3) when (between January 1, 1990 and December 31, 1994). Another example is women ages 34 to 59 years of age in 1980 who were in the nursing profession and had no known coronary disease, stroke, cancer, hypercholesterolemia, or diabetes, and were recruited from the 11 most populous States, with contact information obtained from State nursing boards.

In cohort studies, it is crucial that the population at baseline is free of the outcome of interest. For example, the nurses' population above would be an appropriate group in which to study incident coronary disease. This information is usually found either in descriptions of population recruitment, definitions of variables, or inclusion/exclusion criteria.

You may need to look at prior papers on methods in order to make the assessment for this question. Those papers are usually in the reference list.

If fewer than 50% of eligible persons participated in the study, then there is concern that the study population does not adequately represent the target population. This increases the risk of bias.

Question 4. Groups recruited from the same population and uniform eligibility criteria

Were the inclusion and exclusion criteria developed prior to recruitment or selection of the study population? Were the same underlying criteria used for all of the subjects involved? This issue is related to the description of the study population, above, and you may find the information for both of these questions in the same section of the paper.

Most cohort studies begin with the selection of the cohort; participants in this cohort are then measured or evaluated to determine their exposure status. However, some cohort studies may recruit or select exposed participants in a different time or place than unexposed participants, especially retrospective cohort studies—which is when data are obtained from the past (retrospectively), but the analysis examines exposures prior to outcomes. For example, one research question could be whether diabetic men with clinical depression are at higher risk for cardiovascular disease than those without clinical depression. So, diabetic men with depression might be selected from a mental health clinic, while diabetic men without depression might be selected from an internal medicine or endocrinology clinic. This study recruits groups from different clinic populations, so this example would get a “no.”

However, the women nurses described in the question above were selected based on the same inclusion/exclusion criteria, so that example would get a “yes.”

Question 5. Sample size justification

Did the authors present their reasons for selecting or recruiting the number of people included or analyzed? Do they note or discuss the statistical power of the study? This question is about whether or not the study had enough participants to detect an association if one truly existed.

A paragraph in the methods section of the article may explain the sample size needed to detect a hypothesized difference in outcomes. You may also find a discussion of power in the discussion section (such as the study had 85 percent power to detect a 20 percent increase in the rate of an outcome of interest, with a 2-sided alpha of 0.05). Sometimes estimates of variance and/or estimates of effect size are given, instead of sample size calculations. In any of these cases, the answer would be “yes.”

However, observational cohort studies often do not report anything about power or sample sizes because the analyses are exploratory in nature. In this case, the answer would be “no.” This is not a “fatal flaw.” It just may indicate that attention was not paid to whether the study was sufficiently sized to answer a prespecified question—i.e., it may have been an exploratory, hypothesis-generating study.

Question 6. Exposure assessed prior to outcome measurement

This question is important because, in order to determine whether an exposure causes an outcome, the exposure must come before the outcome.

For some prospective cohort studies, the investigator enrolls the cohort and then determines the exposure status of various members of the cohort (large epidemiological studies like Framingham used this approach). However, for other cohort studies, the cohort is selected based on its exposure status, as in the example above of depressed diabetic men (the exposure being depression). Other examples include a cohort identified by its exposure to fluoridated drinking water and then compared to a cohort living in an area without fluoridated water, or a cohort of military personnel exposed to combat in the Gulf War compared to a cohort of military personnel not deployed in a combat zone.

With either of these types of cohort studies, the cohort is followed forward in time (i.e., prospectively) to assess the outcomes that occurred in the exposed members compared to nonexposed members of the cohort. Therefore, you begin the study in the present by looking at groups that were exposed (or not) to some biological or behavioral factor, intervention, etc., and then you follow them forward in time to examine outcomes. If a cohort study is conducted properly, the answer to this question should be “yes,” since the exposure status of members of the cohort was determined at the beginning of the study before the outcomes occurred.

For retrospective cohort studies, the same principal applies. The difference is that, rather than identifying a cohort in the present and following them forward in time, the investigators go back in time (i.e., retrospectively) and select a cohort based on their exposure status in the past and then follow them forward to assess the outcomes that occurred in the exposed and

nonexposed cohort members. Because in retrospective cohort studies the exposure and outcomes may have already occurred (it depends on how long they follow the cohort), it is important to make sure that the exposure preceded the outcome.

Sometimes cross-sectional studies are conducted (or cross-sectional analyses of cohort-study data), where the exposures and outcomes are measured during the same timeframe. As a result, cross-sectional analyses provide weaker evidence than regular cohort studies regarding a potential causal relationship between exposures and outcomes. For cross-sectional analyses, the answer to Question 6 should be “no.”

Question 7. Sufficient timeframe to see an effect

Did the study allow enough time for a sufficient number of outcomes to occur or be observed, or enough time for an exposure to have a biological effect on an outcome? In the examples given above, if clinical depression has a biological effect on increasing risk for CVD, such an effect may take years. In the other example, if higher dietary sodium increases BP, a short timeframe may be sufficient to assess its association with BP, but a longer timeframe would be needed to examine its association with heart attacks.

The issue of timeframe is important to enable meaningful analysis of the relationships between exposures and outcomes to be conducted. This often requires at least several years, especially when looking at health outcomes, but it depends on the research question and outcomes being examined.

Cross-sectional analyses allow no time to see an effect, since the exposures and outcomes are assessed at the same time, so those would get a “no” response.

Question 8. Different levels of the exposure of interest

If the exposure can be defined as a range (examples: drug dosage, amount of physical activity, amount of sodium consumed), were multiple categories of that exposure assessed? (for example, for drugs: not on the medication, on a low dose, medium dose, high dose; for dietary sodium, higher than average U.S. consumption, lower than recommended consumption, between the two). Sometimes discrete categories of exposure are not used, but instead exposures are measured as continuous variables (for example, mg/day of dietary sodium or BP values).

In any case, studying different levels of exposure (where possible) enables investigators to assess trends or dose-response relationships between exposures and outcomes—e.g., the higher the exposure, the greater the rate of the health outcome. The presence of trends or dose-response relationships lends credibility to the hypothesis of causality between exposure and outcome.

For some exposures, however, this question may not be applicable (e.g., the exposure may be a dichotomous variable like living in a rural setting versus an urban setting, or vaccinated/not vaccinated with a one-time vaccine). If there are only two possible exposures (yes/no), then this question should be given an “NA,” and it should not count negatively towards the quality rating.

Question 9. Exposure measures and assessment

Were the exposure measures defined in detail? Were the tools or methods used to measure exposure accurate and reliable—for example, have they been validated or are they objective? This issue is important as it influences confidence in the reported exposures. When exposures are measured with less accuracy or validity, it is harder to see an association between exposure and outcome even if one exists. Also as important is whether the exposures were assessed in the same manner within groups and between groups; if not, bias may result.

For example, retrospective self-report of dietary salt intake is not as valid and reliable as prospectively using a standardized dietary log plus testing participants' urine for sodium content. Another example is measurement of BP, where there may be quite a difference between usual care, where clinicians measure BP however it is done in their practice setting (which can vary considerably), and use of trained BP assessors using standardized equipment (e.g., the same BP device which has been tested and calibrated) and a standardized protocol (e.g., patient is seated for 5 minutes with feet flat on the floor, BP is taken twice in each arm, and all four measurements are averaged). In each of these cases, the former would get a “no” and the latter a “yes.”

Here is a final example that illustrates the point about why it is important to assess exposures consistently across all groups: If people with higher BP (exposed cohort) are seen by their providers more frequently than those without elevated BP (nonexposed group), it also increases the chances of detecting and documenting changes in health outcomes, including CVD-related events. Therefore, it may lead to the conclusion that higher BP leads to more CVD events. This may be true, but it could also be due to the fact that the subjects with higher BP were seen more often; thus, more CVD-related events were detected and documented simply because they had more encounters with the health care system. Thus, it could bias the results and lead to an erroneous conclusion.

Question 10. Repeated exposure assessment

Was the exposure for each person measured more than once during the course of the study period? Multiple measurements with the same result increase our confidence that the exposure status was correctly classified. Also, multiple measurements enable investigators to look at changes in exposure over time, for example, people who ate high dietary sodium throughout the followup period, compared to those who started out high then reduced their intake, compared to those who ate low sodium throughout. Once again, this may not be applicable in all cases. In many older studies, exposure was measured only at baseline. However, multiple exposure measurements do result in a stronger study design.

Question 11. Outcome measures

Were the outcomes defined in detail? Were the tools or methods for measuring outcomes accurate and reliable—for example, have they been validated or are they objective? This issue is important because it influences confidence in the validity of study results. Also important

is whether the outcomes were assessed in the same manner within groups and between groups.

An example of an outcome measure that is objective, accurate, and reliable is death—the outcome measured with more accuracy than any other. But even with a measure as objective as death, there can be differences in the accuracy and reliability of how death was assessed by the investigators. Did they base it on an autopsy report, death certificate, death registry, or report from a family member? Another example is a study of whether dietary fat intake is related to blood cholesterol level (cholesterol level being the outcome), and the cholesterol level is measured from fasting blood samples that are all sent to the same laboratory. These examples would get a “yes.” An example of a “no” would be self-report by subjects that they had a heart attack, or self-report of how much they weigh (if body weight is the outcome of interest).

Similar to the example in Question 9, results may be biased if one group (e.g., people with high BP) is seen more frequently than another group (people with normal BP) because more frequent encounters with the health care system increases the chances of outcomes being detected and documented.

Question 12. Blinding of outcome assessors

Blinding means that outcome assessors did not know whether the participant was exposed or unexposed. It is also sometimes called “masking.” The objective is to look for evidence in the article that the person(s) assessing the outcome(s) for the study (for example, examining medical records to determine the outcomes that occurred in the exposed and comparison groups) is masked to the exposure status of the participant. Sometimes the person measuring the exposure is the same person conducting the outcome assessment. In this case, the outcome assessor would most likely not be blinded to exposure status because they also took measurements of exposures. If so, make a note of that in the comments section.

As you assess this criterion, think about whether it is likely that the person(s) doing the outcome assessment would know (or be able to figure out) the exposure status of the study participants. If the answer is no, then blinding is adequate. An example of adequate blinding of the outcome assessors is to create a separate committee, whose members were not involved in the care of the patient and had no information about the study participants’ exposure status. The committee would then be provided with copies of participants’ medical records, which had been stripped of any potential exposure information or personally identifiable information. The committee would then review the records for prespecified outcomes according to the study protocol. If blinding was not possible, which is sometimes the case, mark “NA” and explain the potential for bias.

Question 13. Followup rate

Higher overall followup rates are always better than lower followup rates, even though higher rates are expected in shorter studies, whereas lower overall followup rates are often seen in studies of longer duration. Usually, an acceptable overall followup rate is considered 80 percent or more of participants whose exposures were measured at baseline. However,

this is just a general guideline. For example, a 6-month cohort study examining the relationship between dietary sodium intake and BP level may have over 90 percent followup, but a 20-year cohort study examining effects of sodium intake on stroke may have only a 65 percent followup rate.

Question 14. Statistical analyses

Were key potential confounding variables measured and adjusted for, such as by statistical adjustment for baseline differences? Logistic regression or other regression methods are often used to account for the influence of variables not of interest.

This is a key issue in cohort studies, because statistical analyses need to control for potential confounders, in contrast to an RCT, where the randomization process controls for potential confounders. All key factors that may be associated both with the exposure of interest and the outcome—that are not of interest to the research question—should be controlled for in the analyses.

For example, in a study of the relationship between cardiorespiratory fitness and CVD events (heart attacks and strokes), the study should control for age, BP, blood cholesterol, and body weight, because all of these factors are associated both with low fitness and with CVD events. Well-done cohort studies control for multiple potential confounders.

Some general guidance for determining the overall quality rating of observational cohort and cross-sectional studies

The questions on the form are designed to help you focus on the key concepts for evaluating the internal validity of a study. They are not intended to create a list that you simply tally up to arrive at a summary judgment of quality.

Internal validity for cohort studies is the extent to which the results reported in the study can truly be attributed to the exposure being evaluated and not to flaws in the design or conduct of the study—in other words, the ability of the study to draw associative conclusions about the effects of the exposures being studied on outcomes. Any such flaws can increase the risk of bias.

Critical appraisal involves considering the risk of potential for selection bias, information bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues throughout the questions above. High risk of bias translates to a rating of poor quality. Low risk of bias translates to a rating of good quality. (Thus, the greater the risk of bias, the lower the quality rating of the study.)

In addition, the more attention in the study design to issues that can help determine whether there is a causal relationship between the exposure and outcome, the higher quality the study. These include exposures occurring prior to outcomes, evaluation of a dose-response gradient, accuracy of measurement of both exposure and outcome, sufficient timeframe to see an effect, and appropriate control for confounding—all concepts reflected in the tool.

Generally, when you evaluate a study, you will not see a “fatal flaw,” but you will find some risk of bias. By focusing on the concepts underlying the questions in the quality assessment tool, you should ask yourself about the potential for bias in the study you are critically appraising. For any box where you check “no” you should ask, “What is the potential risk of bias resulting from this flaw in study design or execution?” That is, does this factor cause you to doubt the results that are reported in the study or doubt the ability of the study to accurately assess an association between exposure and outcome?

The best approach is to think about the questions in the tool and how each one tells you something about the potential for bias in a study. The more you familiarize yourself with the key concepts, the more comfortable you will be with critical appraisal. Examples of studies rated good, fair, and poor are useful, but each study must be assessed on its own based on the details that are reported and consideration of the concepts for minimizing bias.

Appendix 4. Quality Assessment of Case-Control Studies

Criteria	Yes	No	Other (CD, NR, NA)*
1. Was the research question or objective in this paper clearly stated and appropriate?			
2. Was the study population clearly specified and defined?			
3. Did the authors include a sample size justification?			
4. Were controls selected or recruited from the same or similar population that gave rise to the cases (including the same timeframe)?			
5. Were the definitions, inclusion and exclusion criteria, algorithms or processes used to identify or select cases and controls valid, reliable, and implemented consistently across all study participants?			
6. Were the cases clearly defined and differentiated from controls?			
7. If less than 100 percent of eligible cases and/or controls were selected for the study, were the cases and/or controls randomly selected from those eligible?			
8. Was there use of concurrent controls?			
9. Were the investigators able to confirm that the exposure/risk occurred prior to the development of the condition or event that defined a participant as a case?			
10. Were the measures of exposure/risk clearly defined, valid, reliable, and implemented consistently (including the same time period) across all study participants?			
11. Were the assessors of exposure/risk blinded to the case or control status of participants?			
12. Were key potential confounding variables measured and adjusted statistically in the analyses? If matching was used, did the investigators account for matching during study analysis?			
Quality Rating (Good, Fair, or Poor) (see guidance)			
Rater #1 initials:			
Rater #2 initials:			
Additional Comments (If POOR, please state why):			

* CD, cannot determine; NA, not applicable; NR, not reported

Guidance for Assessing the Quality of Case-Control Studies

The guidance document below is organized by question number from the tool for quality assessment of case-control studies.

Question 1. Research question

Did the authors describe their goal in conducting this research? Is it easy to understand what they were looking to find? This issue is important for any scientific paper of any type. High quality scientific research explicitly defines a research question.

Question 2. Study population

Did the authors describe the group of individuals from which the cases and controls were selected or recruited, while using demographics, location, and time period? If the investigators conducted this study again, would they know exactly who to recruit, from where, and from what time period?

Investigators identify case-control study populations by location, time period, and inclusion criteria for cases (individuals with the disease, condition, or problem) and controls (individuals without the disease, condition, or problem). For example, the population for a study of lung cancer and chemical exposure would be all incident cases of lung cancer diagnosed in patients ages 35 to 79, from January 1, 2003 to December 31, 2008, living in Texas during that entire time period, as well as controls without lung cancer recruited from the same population during the same time period. The population is clearly described as: (1) who (men and women ages 35 to 79 with (cases) and without (controls) incident lung cancer); (2) where (living in Texas); and (3) when (between January 1, 2003 and December 31, 2008).

Other studies may use disease registries or data from cohort studies to identify cases. In these cases, the populations are individuals who live in the area covered by the disease registry or included in a cohort study (i.e., nested case-control or case-cohort). For example, a study of the relationship between vitamin D intake and myocardial infarction might use patients identified via the GRACE registry, a database of heart attack patients.

NHLBI staff encouraged reviewers to examine prior papers on methods (listed in the reference list) to make this assessment, if necessary.

Question 3. Target population and case representation

In order for a study to truly address the research question, the target population—the population from which the study population is drawn and to which study results are believed to apply—should be carefully defined. Some authors may compare characteristics of the study cases to characteristics of cases in the target population, either in text or in a table. When study cases are shown to be representative of cases in the appropriate target population, it increases the likelihood that the study was well-designed per the research question.

However, because these statistics are frequently difficult or impossible to measure, publications should not be penalized if case representation is not shown. For most papers, the response to question 3 will be “NR.” Those subquestions are combined because the answer to the second subquestion—case representation—determines the response to this item. However, it cannot be determined without considering the response to the first subquestion. For example, if the answer to the first subquestion is “yes,” and the second, “CD,” then the response for item 3 is “CD.”

Question 4. Sample size justification

Did the authors discuss their reasons for selecting or recruiting the number of individuals included? Did they discuss the statistical power of the study and provide a sample size calculation to ensure that the study is adequately powered to detect an association (if one exists)? This question does not refer to a description of the manner in which different groups were included or excluded using the inclusion/exclusion criteria (e.g., “Final study size was 1,378 participants after exclusion of 461 patients with missing data” is not considered a sample size justification for the purposes of this question).

An article’s methods section usually contains information on sample size and the size needed to detect differences in exposures and on statistical power.

Question 5. Groups recruited from the same population

To determine whether cases and controls were recruited from the same population, one can ask hypothetically, “If a control was to develop the outcome of interest (the condition that was used to select cases), would that person have been eligible to become a case?” Case-control studies begin with the selection of the cases (those with the outcome of interest, e.g., lung cancer) and controls (those in whom the outcome is absent). Cases and controls are then evaluated and categorized by their exposure status. For the lung cancer example, cases and controls were recruited from hospitals in a given region. One may reasonably assume that controls in the catchment area for the hospitals, or those already in the hospitals for a different reason, would attend those hospitals if they became a case; therefore, the controls are drawn from the same population as the cases. If the controls were recruited or selected from a different region (e.g., a State other than Texas) or time period (e.g., 1991–2000), then the cases and controls were recruited from different populations, and the answer to this question would be “no.”

The following example further explores selection of controls. In a study, eligible cases were men and women, ages 18 to 39, who were diagnosed with atherosclerosis at hospitals in Perth, Australia, between July 1, 2000 and December 31, 2007. Appropriate controls for these cases might be sampled using voter registration information for men and women ages 18 to 39, living in Perth (population-based controls); they also could be sampled from patients without atherosclerosis at the same hospitals (hospital-based controls). As long as the controls are individuals who would have been eligible to be included in the study as cases (if they had been diagnosed with atherosclerosis), then the controls were selected appropriately from the same source population as cases.

In a prospective case-control study, investigators may enroll individuals as cases at the time they are found to have the outcome of interest; the number of cases usually increases as time progresses. At this same time, they may recruit or select controls from the population without the outcome of interest. One way to identify or recruit cases is through a surveillance system. In turn, investigators can select controls from the population covered by that system. This is an example of population-based controls. Investigators also may identify and select cases from a cohort study population and identify controls from outcome-free individuals in the same cohort study. This is known as a nested case-control study.

Question 6. Inclusion and exclusion criteria prespecified and applied uniformly

Were the inclusion and exclusion criteria developed prior to recruitment or selection of the study population? Were the same underlying criteria used for all of the groups involved? To answer this question, reviewers determined if the investigators developed I/E criteria prior to recruitment or selection of the study population and if they used the same underlying criteria for all groups. The investigators should have used the same selection criteria, except for study participants who had the disease or condition, which would be different for cases and controls by definition. Therefore, the investigators use the same age (or age range), gender, race, and other characteristics to select cases and controls. Information on this topic is usually found in a paper's section on the description of the study population.

Question 7. Case and control definitions

For this question, reviewers looked for descriptions of the validity of case and control definitions and processes or tools used to identify study participants as such. Was a specific description of "case" and "control" provided? Is there a discussion of the validity of the case and control definitions and the processes or tools used to identify study participants as such? They determined if the tools or methods were accurate, reliable, and objective. For example, cases might be identified as "adult patients admitted to a VA hospital from January 1, 2000 to December 31, 2009, with an ICD-9 discharge diagnosis code of acute myocardial infarction and at least one of the two confirmatory findings in their medical records: at least 2mm of ST elevation changes in two or more ECG leads and an elevated troponin level. Investigators might also use ICD-9 or CPT codes to identify patients. All cases should be identified using the same methods. Unless the distinction between cases and controls is accurate and reliable, investigators cannot use study results to draw valid conclusions.

Question 8. Random selection of study participants

If a case-control study did not use 100 percent of eligible cases and/or controls (e.g., not all disease-free participants were included as controls), did the authors indicate that random sampling was used to select controls? When it is possible to identify the source population fairly explicitly (e.g., in a nested case-control study, or in a registry-based study), then random sampling of controls is preferred. When investigators used consecutive sampling, which is frequently done for cases in prospective studies, then study participants are not considered randomly selected. In this case, the reviewers would answer "no" to Question 8. However, this would not be considered a fatal flaw.

If investigators included all eligible cases and controls as study participants, then reviewers marked “NA” in the tool. If 100 percent of cases were included (e.g., NA for cases) but only 50 percent of eligible controls, then the response would be “yes” if the controls were randomly selected, and “no” if they were not. If this cannot be determined, the appropriate response is “CD.”

Question 9. Concurrent controls

A concurrent control is a control selected at the time another person became a case, usually on the same day. This means that one or more controls are recruited or selected from the population without the outcome of interest at the time a case is diagnosed. Investigators can use this method in both prospective case-control studies and retrospective case-control studies. For example, in a retrospective study of adenocarcinoma of the colon using data from hospital records, if hospital records indicate that Person A was diagnosed with adenocarcinoma of the colon on June 22, 2002, then investigators would select one or more controls from the population of patients without adenocarcinoma of the colon on that same day. This assumes they conducted the study retrospectively, using data from hospital records. The investigators could have also conducted this study using patient records from a cohort study, in which case it would be a nested case-control study.

Investigators can use concurrent controls in the presence or absence of matching and vice versa. A study that uses matching does not necessarily mean that concurrent controls were used.

Question 10. Exposure assessed prior to outcome measurement

Investigators first determine case or control status (based on presence or absence of outcome of interest), and then assess exposure history of the case or control; therefore, reviewers ascertained that the exposure preceded the outcome. For example, if the investigators used tissue samples to determine exposure, did they collect them from patients prior to their diagnosis? If hospital records were used, did investigators verify that the date a patient was exposed (e.g., received medication for atherosclerosis) occurred prior to the date they became a case (e.g., was diagnosed with type 2 diabetes)? For an association between an exposure and an outcome to be considered causal, the exposure must have occurred prior to the outcome.

Question 11. Exposure measures and assessment

Were the exposure measures defined in detail? Were the tools or methods used to measure exposure accurate and reliable—for example, have they been validated or are they objective? This is important, as it influences confidence in the reported exposures. Equally important is whether the exposures were assessed in the same manner within groups and between groups. This question pertains to bias resulting from exposure misclassification (i.e., exposure ascertainment).

For example, a retrospective self-report of dietary salt intake is not as valid and reliable as prospectively using a standardized dietary log plus testing participants’ urine for sodium

content because participants' retrospective recall of dietary salt intake may be inaccurate and result in misclassification of exposure status. Similarly, BP results from practices that use an established protocol for measuring BP would be considered more valid and reliable than results from practices that did not use standard protocols. A protocol may include using trained BP assessors, standardized equipment (e.g., the same BP device which has been tested and calibrated), and a standardized procedure (e.g., patient is seated for 5 minutes with feet flat on the floor, BP is taken twice in each arm, and all four measurements are averaged).

Question 12. Blinding of exposure assessors

Blinding or masking means that outcome assessors did not know whether participants were exposed or unexposed. To answer this question, reviewers examined articles for evidence that the outcome assessor(s) was masked to the exposure status of the research participants. An outcome assessor, for example, may examine medical records to determine the outcomes that occurred in the exposed and comparison groups. Sometimes the person measuring the exposure is the same person conducting the outcome assessment. In this case, the outcome assessor would most likely not be blinded to exposure status. A reviewer would note such a finding in the comments section of the assessment tool.

One way to ensure good blinding of exposure assessment is to have a separate committee, whose members have no information about the study participants' status as cases or controls, review research participants' records. To help answer the question above, reviewers determined if it was likely that the outcome assessor knew whether the study participant was a case or control. If it was unlikely, then the reviewers marked "no" to Question 12. Outcome assessors who used medical records to assess exposure should not have been directly involved in the study participants' care, since they probably would have known about their patients' conditions. If the medical records contained information on the patient's condition that identified him/her as a case (which is likely), that information would have had to be removed before the exposure assessors reviewed the records.

If blinding was not possible, which sometimes happens, the reviewers marked "NA" in the assessment tool and explained the potential for bias.

Question 13. Statistical analysis

Were key potential confounding variables measured and adjusted for, such as by statistical adjustment for baseline differences? Investigators often use logistic regression or other regression methods to account for the influence of variables not of interest.

This is a key issue in case-controlled studies; statistical analyses need to control for potential confounders, in contrast to RCTs in which the randomization process controls for potential confounders. In the analysis, investigators need to control for all key factors that may be associated with both the exposure of interest and the outcome and are not of interest to the research question.

A study of the relationship between smoking and CVD events illustrates this point. Such a study needs to control for age, gender, and body weight; all are associated with smoking and CVD events. Well-done case-control studies control for multiple potential confounders.

Matching is a technique used to improve study efficiency and control for known confounders. For example, in the study of smoking and CVD events, an investigator might identify cases that have had a heart attack or stroke and then select controls of similar age, gender, and body weight to the cases. For case-control studies, it is important that if matching was performed during the selection or recruitment process, the variables used as matching criteria (e.g., age, gender, race) should be controlled for in the analysis.

General Guidance for Determining the Overall Quality Rating of Case-Controlled Studies

NHLBI designed the questions in the assessment tool to help reviewers focus on the key concepts for evaluating a study's internal validity, not to use as a list from which to add up items to judge a study's quality.

Internal validity for case-control studies is the extent to which the associations between disease and exposure reported in the study can truly be attributed to the exposure being evaluated rather than to flaws in the design or conduct of the study. In other words, what is ability of the study to draw associative conclusions about the effects of the exposures on outcomes? Any such flaws can increase the risk of bias.

In critical appraising a study, the following factors need to be considered: risk of potential for selection bias, information bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues addressed in the questions above. High risk of bias translates to a poor quality rating; low risk of bias translates to a good quality rating. Again, the greater the risk of bias, the lower the quality rating of the study.

In addition, the more attention in the study design to issues that can help determine whether there is a causal relationship between the outcome and the exposure, the higher the quality of the study. These include exposures occurring prior to outcomes, evaluation of a dose-response gradient, accuracy of measurement of both exposure and outcome, sufficient timeframe to see an effect, and appropriate control for confounding—all concepts reflected in the tool.

If a study has a “fatal flaw,” then risk of bias is significant; therefore, the study is deemed to be of poor quality. An example of a fatal flaw in case-control studies is a lack of a consistent standard process used to identify cases and controls.

Generally, when reviewers evaluated a study, they did not see a “fatal flaw,” but instead found some risk of bias. By focusing on the concepts underlying the questions in the quality assessment tool, reviewers examined the potential for bias in the study. For any box checked “no,” reviewers asked, “What is the potential risk of bias resulting from this flaw in study design or execution?” That is, did this factor lead to doubt about the results reported in the

study or the ability of the study to accurately assess an association between exposure and outcome?

By examining questions in the assessment tool, reviewers were best able to assess the potential for bias in a study. Specific rules were not useful, as each study had specific nuances. In addition, being familiar with the key concepts helped reviewers assess the studies. Examples of studies rated good, fair, and poor were useful, yet each study had to be assessed on its own.

Figure 1.
Flowchart of Systematic Review Process

Table 1

Characteristics of the studies included in Systematic Review

Author/Year	Manufacturing Industry	Job characteristics	Country	Study Design			
				RCT	CS	C	CC
Filus (2012)	Home appliances/Assembly line	Repetitive (cycle shorter than 90 seconds)	Brazil		x		
Asensio-Cuesta(2012a)	Automobile parts/Assembly line	High repetitive movements	Spain		x		
Asensio-Cuesta (2012b)	Automobile parts/Assembly line	High repetitive movements	Spain		x		
Guimarães (2012)	Footwear	Task with different complexities	Brazil	x			
Diego-Mas (2009)	Automobile parts/Assembly line	High repetitive movements	Spain		x		
Sato (2009)	Office products	Repetitive movements and Materials handling	Brazil		x		
Dawal (2009)	Automotive	Not reported	Malaysia		x		
Dawal (2007)	Automotive	Not reported	Malaysia		x		
Tharmmaphornphilas (2007)	Manufacturing cells	Lifting tasks	United States		x		
Balogh (2006)	Wooden boards for parquet flooring/Assembly lines	Higher repetitive (67 seconds)	Sweden		x		
Frazer (2003)	Automotive/Assembly line	Repetitive movements (60 seconds)	Canada		x		
Fredriksson (2001)	Automobile/Assembly line	Repetitive movements (cycle time between 15 and 90 seconds)	Sweden	x			
Camahan (2000)	Manufacturing cells	Lifting tasks	United States		x		
Roquelaure (1997)	Television, Shoes and Automobile brakes	High repetitive movements (under 30 seconds)	France				X

RCTs – Randomized Control Trial; CS – Cross Sectional; C - Cohort; CC – Case Control; NR not reported; NA – not applicable..

Table 2

Methodological quality of studies

Author/Year	Quality Assessment Tools														T	QR
	Controlled Intervention Studies [†]															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
Guimarães (2012)	-	-	-	-	NA	+	-	-	+	+	+	-	+	-	5/14	Fair
Fredriksson (2001)	-	-	-	-	-	+	-	+	+	+	+	-	+	-	6/14	Fair
Observational Cohort and Cross-Sectional Studies ^{††}																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14		
Filus (2012)	+	-	-	-	-	-	-	+	-	-	-	-	NA	-	2/14	Poor
Asensio-Cuesta (2012a)	+	-	-	-	-	-	-	+	-	-	+	-	NA	-	3/14	Poor
Asensio-Cuesta(2012b)	+	-	-	-	-	-	-	+	-	-	+	-	NA	-	3/14	Poor
Diego-Mas (2009)	+	-	-	-	-	-	-	+	-	-	+	-	NA	-	3/14	Poor
Sato (2009)	+	+	-	+	-	-	-	+	+	-	+	-	NA	-	6/14	Fair
Dawal (2009)	+	+	+	+	-	-	-	+	+	-	+	-	NA	-	7/14	Fair
Dawal (2007)	+	+	+	+	-	-	-	+	+	-	+	-	NA	-	7/14	Fair
Tharmnaphornphilas (2007)	+	-	+	-	-	-	-	+	-	-	-	-	NA	-	3/14	Poor
Balogh(2006)	+	+	+	+	+	-	-	+	-	-	-	-	NA	-	6/14	Fair
Frazer (2003)	+	-	+	-	-	-	-	+	-	-	+	-	NA	-	4/14	Poor
Camahan (2000)	+	-	+	-	-	-	-	+	-	-	-	-	NA	-	3/14	Poor
Case Control Studies ^{†††}																
	1	2	3	4	5	6	7	8	9	10	11	12				
Roquelaure (1997)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	12/12	Good

+ (Yes), - (No), NR - not reported, NA – not applicable, CD – cannot determine, T- Total punctuation; QR - Quality Rating (67 % or more - Good, 33–66% - Fair, 33% or less - Poor); Quality Tool - [†] Appendix 2 (Criteria: Randomization (1 and 2); Allocation (3, 6 and 13); Blinding (4 and 5); Outcomes (9, 10 and 11); Dropout rate (7,8 and 12); Confounding (14);

^{††} Appendix 3 (Criteria: Setting (1 and 2); Allocation (3, 4 and 5); Outcomes (6,7,8,9,10,11 and 12); Dropout rate (13); Confounding (14);

^{†††} Appendix 4 (Criteria: Setting (1 and 2); Allocation (3, 4, 5 and 6); Randomization (7 and 8); Outcomes (9 and 10); Blinding (11); Confounding (12).

Table 3

Description the studies included in systematic review.

Author/Year	Purpose	Population/Tasks	Outcome	Follow up	Results
Filus (2012)	Identify the fatigue level and perception when job rotation is held every one, two or three hours.	N= 11. Two groups (Group A - N = 5 ; Group B N=6) Range of age: 20 – 30 years Least 6 months experience in the assembly line	Fatigue (Lactate level m/mol) Perception of Fatigue (range 0 – 10 points, 10 is maximum).	Group A. Before and after each job rotation. Group B. Start and end shifts, 3 weeks. All measurements were collected on the last day of each week.	Group A. Baseline - 0.6 (SD 1.0) m/mol 1 hour: 2.0 (SD 0.7) m/mol 2 hour: 2.9 (SD 0.8) m/mol 3 hours: Not evaluated Group B. Baseline – 0.9 (SD 1.1) m/mol 1 hour: 4.0 (SD 3.0) m/mol 2 hour: 2.9 (SD 0.7) m/mol 3 hours: 7.3 (SD 4.9) m/mol There are significant difference between levels of lactic acid from the easy (tape) and difficult (silicon) jobs. Not reported p value.
Asensio-Cuesta (2012a)	Use the genetic algorithm (GA) to evaluate the level of exposure to the repetitive movement of workers and generate job rotation schedules.	N=14/14 workstation Workstation 13 - One worker had vision problems and who was kept in this task. Workstation 14 - worker with musculoskeletal will be allocated in activities with lower-risk level.	Job rotation schedule by exposure levels Right and Left side of the body. OCRA index: lesser than 2.3 (low level of risk), medium level of risk (between 2.3 and 3.5), high level of risk - greater than 3.5. Multitask OCRA index (repetitive task analyze) if necessary. Ergonomic best solution (E)	Immediately	OCra Index : Right side: average –2.9 (SD 0.3) Left side: average - 2.4 (SD 0.2) Best solution – E8 (in 10 attempts) The solution E8: no more than one consecutive rotation to high-risk task and better recovery from cumulative fatigue.
Asensio-Cuesta (2012b)	Generate a job rotation schedules that prevents musculoskeletal disorders.	N=50 (all workers) N=16 (16 jobs)	Ergonomic and Competent Rotation (ECRot solution)	Immediately	The third and fourth columns show how much of the fitness value was due to ergonomic criteria (E1) and how much to the competence criteria (E2). The ergonomic fitness value 45.43. Average competence fitness value 48.81. The solution with the best fitness (93.39) was obtained in generation 5208 of the fifth run. The maximum allowable time at a workstation (tmax), set at 3 h.
Guimarães (2012)	Compare two workers group with two different job rotation sequences	N= 17 (9 workers - 5 assembly and 4 stitching sector) - 8 workers (4 from the assembly and 4 from the stitching sector)	Job satisfaction WMSD (average %) Absenteeism (average %) Turnover (average %) Absence (average %) Rework (average %) Spoilage (average %) Production rate (units/hour)	3.5 years	Job satisfaction - p = 0.01 WMSD (from 7.00 to 1.41%; P<0.05) Absenteeism (from 6.63 to 3.60; P<0.05) Turnover (from 4.26 to 0.85; P<0.05) Absence (from 2.22 to 0.44; P<0.05) Rework (from 5.42 to 0.81; P<0.05) Spoilage (from 2.90 to 0.89; P<0.05) Production rate (from 12.00 to 12.36) p>0.05
Diego-Mas (2009)	Define and obtain the best job rotation solution.	N=18 (18 workstation)	Frequency of movements (0 – 3 points), perform the movement (0–3 points). General, Mental and	Immediately	Workload: Average 24.90 for the of each worker. Break of 10 min between 1 and 2 rotations. 2, 3 to 50 min, and including a

Author/Year	Purpose	Population/Tasks	Outcome	Follow up	Results
Sato (2009)	Compare effects of job rotation schedules in workers no sick leave and those with sick leave.	N=40 (30 Female ; 10 Male), 26 no sick leave (never). 14 sick leave by musculoskeletal symptoms. Age: 34.0 (SD 6.3) years Tenure : 14.48(SD 5.2) years Tasks: (carrying loads, feeding machines).	Communication capacities (necessary and not necessary). Ergonomic best solution (E) Musculoskeletal complaints. Discomfort (VAS- Visual analogic scale- 100 mm in length) Perception exertion - RPE - Borg (6 – 20 points) Ergonomic Workplace Analysis (EWA) – (range 1 to 5 points).	Three moments (beginning, during work and end of the workday).	break of 10 min between rotations. With these new conditions, Workers with and without previous sick leave RPE and VAS (P>0.05). RPE (P<0.05) discriminate progressive workload levels at different shift.
Dawal (2009)	Evaluate the effect of job organization factors on job satisfaction.	N=170 (two companies – Auto 1 and Auto 2). Male Auto 1. 85% of the workers age 26 years or more. Auto 2. 90% of the workers age below 26 years. Age: 26.8 (SD 5.3) years. Work experience 6.5 (SD 4.9) years	Job organization factors (Job rotation, work methods, training problems solving and goal setting) and Job satisfaction (Likert scale: 1–5 points)	Immediately	Job rotation Auto 1. $\alpha = 0.83$ Auto 2. $\alpha = 0.83$ Job satisfaction Auto 1. $\alpha = 0.89$ Auto 2. $\alpha = 0.82$ Job satisfaction & Organization factors. Auto 1. $R^2 = 0.804$ Auto 2. $R^2 = 0.667$. There is significant and positive correlation between job rotation and work method with job satisfaction.
Dawal (2007)	Analyze the relationship between job satisfaction and job organization.	N=170 (two companies – Auto 1 and Auto 2) Auto 1. 85% of the workers age (26 years or more). Auto 2. 90% of the workers age below 26 years). Average age: 26.8 (SD 5.3) years. Male Work experience 6.5 (SD 4.9) years.	Job organization factors (Job rotation, work methods, training problems solving and goal setting) and Job satisfaction (Likert scale: 1–5 points; 1 (very little) through 3 (moderate) to 5 (very much)).	Immediately	Job rotation Auto 1. $\alpha = 0.83$ Auto 2. $\alpha = 0.83$ Job satisfaction Auto 1. $\alpha = 0.89$ Auto 2. $\alpha = 0.82$ Job satisfaction & job rotation Auto 1. $R = 0.647$ (P>0.05) Auto 2. $R = 0.504$ Workers - more than 70% in both companies are satisfied with job rotation.
Tharmmaphornphilas (2007)	Evaluate job rotation schedules to reduce low back injury risk.	N=4 (4 tasks); N=8 (8 tasks) N=12 (12 tasks) 90th - Female 50th - Male and Female 25th - Male	Number of lost days for low back injury risk. Job Severity Index (JSI) - greater than 1.5 is significant injury potential.	Immediately	JSI 50th Male - 1.23 90th Female- 1.60 25th Male - 1.30 50th Female - 0.98
Balogh (2006)	Compare different workloads in different production organization	N=152 (N=25/5 tasks were chosen for the direct technical measurements). Female Manual line - three work tasks. Semi-automated line- two work task. Automated line: four work tasks. Age: 38.5 (range 20–61) years. Work experience 9.8 (range 0.25–38) years.	Nordic Questionnaire (Musculoskeletal symptoms) Physical examination of the neck and upper extremity. Direct measurement: Electromyography (EMG): trapezius, the infraspinatus and the forearm extensor muscles. Posture work and movements Inclinometers: (head, upper back and upper arms). Electrogoniometers by wrists.	Immediately	Postures and movements there were differences (P>0.05) between all three lines (Manual; Semi-automated; Automated). Manual and Semi-automated line - lower muscular activities and little change postures and movements. Automated line – physical workload with very variable and more flexible work.
Frazer (2003)	Determine the effects of job rotation on injury risk of low back pain.	N=1 (Job A and Job B) A 50th percentile male (1.74 m, 76 kg) was to use in the biomechanical model.	Low Back Pain Reporting Index (LBPRI) - (range from 0.0 to 1.0 association probability)	Immediately	Job A=LBPRI score 0.46 Job B=LBPRI score 0.81 Job A + Job B – score of 0.72 (39% increase at Job A and 21% decrease at Job B).

Author/Year	Purpose	Population/Tasks	Outcome	Follow up	Results
Fredriksson (2001)	Incorporate safety criteria in scheduling algorithms for job rotation to reduce the potential for injury.	<p>Intervention Group N=57 (35 Female and 24 Male). Female - 39.1 (SD 8.7) years. Male - 32.5 (SD 9.6) years. The line system (7–8 workers/3–4 workstations). Six-month training</p> <p>Reference Group N=45 (10 Female and 35 Male). Female - 34.8 (SD 9.4) years. Male - 33.5 (SD 9.3) years. All groups - ¾ more than 10 years in work experience. The lineout system (average time 20min – pair of workers).</p>	Time Weighted Average (TWA) – estimate risk scores by the duration of time. Perception exertion - RPE - Borg. Physical workload: Postures and movements - Observation registration (PEOFlex) using a Physiometer (PHYS). Nordic Questionnaire: (musculoskeletal disorders - aches, pain or discomfort). Sick-leave (number of visits in health care center)	1 year and 11 months	<p>Intervention group: RPE- increased from an average of 6.9 to 8.4 points ($P<0.05$). MDS symptoms (0.3 before and 0.4 after). The rate of sick leave 14.3 before and 15.8 days after).</p> <p>Reference group: RPE - from an average of 6.7 to 7.3 points. No statistically significant changes in the physical workload. Low back symptoms (0.02 before and 0.15 after; $P<0.05$). The rate of sick leave: 5.9 days before and 6.8 days after).</p> <p>Low back – OR =2.2 – 7.2 JSI 50th Male - 1.54 (SD - 0.05) 90th Female- 1.53 (SD 0.05) 25th Male - 1.49 (SD 0.07) 50th Female - 1.49 (SD 0.05)</p>
Camahan(2000)	Incorporate safety criteria in scheduling algorithms job rotation to reduce the potential for injury	N=4 (2 Female - 90th and 50th Percentiles; 2 Male – 50th and 25th)	Job Severity Index (JSI). if the values are greater than 1.5 is significant injury potential.	Immediately.	
Roquelaure (1997)	Evaluate the association between occupational factors (job rotation) with carpal tunnel syndrome (CTS) in industrial workers	<p>Case group (Carpal tunnel syndrome) N=65 (55 Female, 10 Male) Age 41.1 (SD 7.5) years old</p> <p>Reference group N=65 (55 Female, 10 Male) Age 41.2 (SD 7.2) years</p>	Medical history about Carpal tunnel syndrome. Occupational exposure (Biomechanics risks) Job rotation (Yes or No)	1 year and 11 months.	<p>No Job rotation (OR 6.3; $P<0.05$) were associated with carpal tunnel syndrome. Exertion of effort - tools or the parts handled (OR 9.01). Two risk factors related to motion repetitiveness, length of the shortest elementary operation of < 10 s. (OR 84, and lack of change in task or breaks for at least 15% of the daily worktime, excluding breaks for meals (OR 6.0; $P<0.05$). Cycle time or <30 s was not associated with MSDs. CTS ($P = 0.36$).</p>

Table 4

Strategies to implement job rotation schedules

Author/Year	Ergonomic Parameters	Protocol and Intervals (Time)	Tools/Schedule	Priorities
Filus (2012)	Biomechanical - Alternations muscular groups. Cognitive: Perception about complexity of the job rotation.	One, two and three hours First week- Each one hour. Second week –Each 2 hours Third week – Each 3 hours	Simple definition Easy, moderate and difficult level.	1 Biomechanics analysis 2 Staff engaged.
Asensio-Cuesta (2012b)	Biomechanical - movements (arm, elbow, neck, shoulder, fingers, trunk and legs). Organizational and Cognitive: standing, climbing, driving vehicles, hearing, writing, using computer (mouse and keyboard). Competence Criteria – company strategies and policies, adaptability and self-control, learning skills, engagement and responsibility, willingness. P and enthusiasm, meticulous and rigorous work (Psychosocial).	Four job rotations per day. 1 st , 2 nd and 3 rd (2 hours) and 4 th (1 hour).	Equation ECRot model - Computational Software Genetic Algorithm (GA). 39 criteria were used to characterize the workers and jobs included in the rotation schedules	1 Know the performance the tasks by the worker. Diversity of tasks: 1 Minimization of fatigue accumulation. 2 Competence and Physical skills demanded by the tasks.
Asensio-Cuesta (2012a)	Biomechanical – Level of risk (low, medium and high). Frequency, awkward postures and movements of the upper limbs, force, stereotypy, lack of postural variation, inadequate recovery periods). OCRA method. Organizational: Parameters of a GA (number of generations, population size, crossover and mutation probabilities).	Number of Job rotation and pause are established by production and working hours.	Equation Computational Software Genetic Algorithm (GA). Organization job rotation schedule considering risk biomechanical factors.	1 Know the performance the tasks by the worker. 2 Biomechanical risk level. 3 Ergonomic staffs (Analysis continues to observe sensitive changes).
Guimarães (2012)	Macroergonomic Work Analysis (MA) – participative approach (company and workers). (1) physical environment, (2) workstation characteristics, (3) task design, (4) task risk, (5) task content, and (6) issues related to the organizational culture (unions and society, effects of government legislation, regulations and trade. Learning Curve (LC) - workers' learning rate and performance.	Each 2 hours (beginning). Time decided by workers (each 1.5 hours) after implementation.	Simple definition Different complexity levels of task: From difficult to easy/ From easy to difficult.	1 Training for all workers (3.5 years). 2 After initial criteria's the workers themselves established the job rotation.
Sato (2009)	Biomechanical – postures adopted (standing, sitting or walking). Cognitive - Level of the concentration (high for sorting products or low for exclusively manual activities. Intensity of the work (heavy or light).	Every hour. Job rotation with sick workers.	Simple definition High physical/low cognitive demands and high cognitive /low physical demands.	1 Biomechanical demands. 2 Cognitive demands.
Dawal (2009)	Psychosocial – job satisfaction Organizational - power decisions, self-regulation and worker autonomy.	Not reported	Not reported	Not reported
(Dawal, 2007)	Psychosocial – job satisfaction	Not reported	Not reported	Not reported

Author/Year	Ergonomic Parameters	Protocol and Intervals (Time)	Tools/Schedule	Priorities
Tharmmaphornphilaas (2007)	Organizational - job-related decisions, self-regulation and worker autonomy Biomechanical – lifting weight, Lifting frequency, Lifting height, Horizontal lifting distance.	Different protocol of work to different scenarios.	Equation Mathematical models deterministic and stochastic problems. Two versions were investigated: Minimizing the maximum number of lost days (StochMinMaxDays) and Minimizing the total number of lost days (StochMinTot-Days).	1 Lifting weight (kg) 2 Lifting rate (lifts/minute) 3 Lifting distance (cm)
Balogh (2006)	Biomechanical - muscular load, work postures and movements.	All operators - Five work tasks per day.	Not reported	Not reported
Frazer (2003)	Biomechanical – anthropometric information (gender, height and weight), Magnitude force hand. Loads lifting limit. Postural demands.	For two different jobs per day.	Equation Computational software - Biomechanical model: estimate the moment of force and reaction forces for the major body joints (low back).	1 Peak L4/L5 shear force and moment cumulated.
Fredriksson (2001)	Psychosocial /Organizational - occupational pride, job demands, stimulation from work, opportunities to influence work, social support at work.	Four times a day.	Not reported	1 Training operators.
Camahan (2000)	Biomechanical – the physical demands of the tasks. The gender of workers. The lifting capacities of the worker. Demands and time spent performing others tasks.	One or two hours – dependents of the characteristics of task.	Equation Computational Software Genetic Algorithm (GA)	1 Lifting weight (kg), Lifting rate (lifts/minute), Lifting distance (cm) 2 Gender and lifting capacity.
Roquelaure (1997)	Biomechanical - Number of motions per time (average of 10 work cycles), Frequency of movements (cycle per time). A checklist based on the RULA method (elbow, forearm, and wrist), frequency of certain activities (pinching, gripping, screwing, pulling, pushing, lifting, turning), nature of the motion, Hand exerted forces. Organizational - number of parts produced, type of tools used, the materials, and workplace.	Not reported	Not reported	1 Two or more workstations to considerer job rotation. 2 Biomechanics factors.