INSTITUTE
FOR WORK & HEALTH
INSTITUT DE RECHERCHE
SUR LE TRAVAIL ET
LA SANTÉ

Interventions in health-care settings to protect musculoskeletal health: a systematic review

Full report

sharing best evidence

About this report:

Authors:
Benjamin C. Amick III, Ph.D.[1,2,3], Jessica M. Tullar, M.P.H.[2], Shelley Brewer, M.Sc., CSP[2], Emma Irvin[1], Quenby Mahood[1], Lisa Pompeii, Ph.D.[2], Anna Wang, R.N.[1], Dwayne Van Eerd, M.Sc.[1], David Gimeno, Ph.D.[4], Bradley Evanoff, M.D., M.P.H.[5]

1 The Institute for Work & Health, Toronto, Canada
2 The University of Texas, School of Public Health, Division of Environmental and Occupational Health Sciences, Southwest Center for Occupational and Environmental Health, Houston, TX, U.S.A.
3 The University of Texas, School of Public Health, Division of Health Promotion and Behavioral Sciences, Houston, TX, U.S.A.
4 International Institute for Society and Health, Department of Epidemiology and Public Health, University College London, London, U.K.
5 Washington University at St. Louis, School of Medicine, St. Louis, MO, U.S.A.

Acknowledgements:
The authors wish to thank: Dan Shannon for obtaining bibliographic information and other materials; Jim Collins for providing references and input; Jane Gibson, Tony Culyer, Anita Dubey, Kiera Keown and Cameron Mustard for their editorial advice; and Shanti Raktoe for administrative support. Jessica Tullar and Shelley Brewer are supported by an Occupational Injury Prevention Training Grant (T42 OH008421) from the U.S. National Institute for Occupational Safety and Health. Ben Amick is supported by St. Luke's Episcopal Health Charities.

If you have a question about this or any of our reports, please contact us at:

Institute for Work & Health
481 University Avenue, 8th Floor
Toronto, Ontario M5G 2E9

E-mail: info@iwh.on.ca

Or you can visit our web site at www.iwh.on.ca

Please cite this report as: Amick B, Tullar J, Brewer S, Mahood Q, Irvin E, Pompeii L, Wang A, Van Eerd D, Gimeno D, Evanoff B. Interventions in health-care settings to protect musculoskeletal health: A systematic review. Toronto: Institute for Work & Health, 2006.

Table of Contents

List of Appendices

List of Tables

List of Figures

Foreword

In recent years, the Institute for Work & Health has been actively engaged in building relationships with Prevention System agencies and organizations in Ontario.

In these encounters, we often hear that potential research users want more evidence about the effectiveness of interventions aimed at protecting workers' health. We are also told that even when research evidence exists, it is often hard to access, difficult to understand and is not always presented in language and formats suitable to non-scientific audiences.

In response to these needs, the Institute for Work & Health has established a dedicated group to conduct systematic reviews of relevant research studies in the area of workplace injury and illness prevention. In instances where there are too few studies to conduct a full Systematic Review we may provide our audiences with a narrative review.

- Our systematic review team monitors developments in the international research literature on workplace health protection and selects timely, relevant topics for evidence review.
- Our scientists then synthesize both established and emerging evidence on each topic through the application of rigorous methods.
- We then present summaries of the research evidence and recommendations following from this evidence in formats which are accessible to non-scientific audiences.

The Institute will consult regularly with workplace parties to identify areas of workplace health protection that might lend themselves to a systematic review of the evidence.

We appreciate the support of the Ontario Workplace Safety & Insurance Board (WSIB) in funding this four-year Prevention Systematic Reviews initiative. As the major funder, the WSIB demonstrates its own commitment to protecting workers' health by supporting consensus-based policy development which incorporates the best available research evidence.

Many members of the Institute's staff participated in conducting this Systematic Review. A number of external reviewers in academic and workplace leadership positions provided valuable comments on earlier versions of the report. On behalf of the Institute, I would like to express gratitude for these contributions.

Dr. Cameron Mustard
President, Institute for Work & Health
December, 2006

1.0 Introduction

Health-care workers are at a high risk of developing musculoskeletal (MSK) symptoms, injuries and disorders, particularly low-back pain. Reported injury rates in health-care workers equal or exceed rates in other industries that are traditionally considered hazardous (1). The total cost of such injuries is unknown, but in 2000, the U.S. Veteran's Administration – one large hospital system – spent over $23 million (US) for job-related injuries related to patient care (2). The prevalence of low-back pain in nursing personnel has been reported at rates between 30 and 60 per cent (3; 4; 5; 6). In 2005, 60 per cent of Canadian nurses said their jobs presented them with high physical demands (7). Low-back pain has been identified as a major reason why nurses leave their profession (4).

Musculoskeletal disorders in health-care workers have been attributed in large part to patient transfer and lifting activities. Biomechanical studies have shown that these activities place high levels of compressive force on low-back structures, far exceeding the lifting limits recommended by the U.S. National Institute of Occupational Safety and Health (NIOSH) (8). Shoulder, knee and other disorders have also been associated with patient lifting and transfer tasks (9). Various interventions have been implemented to reduce back and other MSK disorders among health-care workers. These include worker education programs, physical conditioning or exercise, disability management, organizational policies and use of mechanical lifts or other patient transfer equipment (10).

Because biomechanical exposures are thought to contribute greatly to the high rates of MSK injuries in health-care workers, mechanical patient handling and transfer devices have been a major focus of efforts for prevention. Numerous facilities have instituted "zero-lift" policies banning manual lifting. Nursing organizations have promulgated guidelines recommending the use of mechanical lifting devices. Some U.S. states have enacted legislation encouraging or requiring health-care facilities to have lifting devices available (11). The Government of Ontario has committed over $80 million (CDN) to purchase and install 10,000 new overhead lifts in Ontario health-care institutions.

Recent research suggests MSK injuries in health-care settings may result from non-patient handling activities: patient-related assaults (12), slips, trips and falls. Additionally, MSK injuries also occur from non-patient related health-care jobs or tasks, such as maintenance work. Surprisingly, except for one review of injury prevention for patient lifting (10), no systematic reviews have been conducted on a broad spectrum of interventions to reduce MSK injuries in health-care settings. Stakeholders such as facility managers, occupational health and safety professionals, ergonomic consultants, etc., are thus faced with making decisions without evidence-based reviews.

The systematic review process provides a structured methodology for evaluating the literature and synthesizing evidence regarding prevention strategies (13; 14). Such reviews also identify gaps in the existing literature.

The purpose of this systematic review was to identify studies that evaluated the effects of occupational safety and health interventions on MSK health among health-care workers.

Studies that met our design and quality criteria were evaluated in detail, and data were synthesized from these studies. The review included both primary and secondary prevention studies. Based on our synthesis, we make recommendations about improving work-related MSK health outcomes.

1.1 Organization of the report

Following this introduction, readers will find:

- a description of the methods we used to search for and select relevant studies
- details on quality assessment, data extraction and a best evidence synthesis
- results of the systematic review, including information about the number of studies found, the methodological quality and study characteristics
- results of our synthesis of evidence according to intervention categories
- results of our partial data extraction for studies that did not proceed to evidence synthesis due to insufficient methodological quality
- conclusions
- messages about the current state of the peer-reviewed literature and recommendations for future occupational health and safety (OHS) intervention research and evaluation.

2.0 Materials and methods

Health-care intervention studies were systematically reviewed using a consensus process that was developed by Cochrane (15) and Slavin (13), and adapted by the review team. Additionally, as part of the review, the team incorporated a partial data extraction step following principles proposed by Côté (16).

A review team comprising 10 researchers from the U.S., Canada and the U.K. participated in the process. Reviewers were identified based on their expertise in conducting epidemiologic or intervention studies related to MSK disorders among health-care workers, or their experience in conducting systematic reviews. Review team members had backgrounds in epidemiology, ergonomics, nursing, occupational medicine and safety engineering.

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

- formulate research question and search terms
- identify articles expected in literature search by all review team members
- contact international content experts to identify key articles
- convene stakeholder meetings to review research question, definitions, search terms and inclusion criteria
- conduct literature search and pool articles with those submitted by experts
- conduct Level 1 review to exclude non-relevant studies based on six screening criteria
- conduct Level 2 review to assess methodological quality of remaining relevant articles based on 19 criteria
- conduct Level 3 review to extract data from relevant articles that were identified for evidence synthesis
- complete evidence synthesis
- perform partial data extraction for articles not used in evidence synthesis due to insufficient methodological quality
- convene stakeholder meetings to review evidence synthesis and develop key messages.

The primary research question addressed was: “Do occupational safety and health interventions in health-care settings have an effect on musculoskeletal health status?”

To answer this question, we reviewed primary and secondary prevention intervention studies conducted at worksites.

Primary prevention approaches were considered interventions aimed at preventing healthy employees from developing MSK symptoms, injuries or disorders. Secondary prevention was classified as interventions designed to prevent people with MSK symptoms, clinically recognized disorders or injuries from further morbidity, disability or mortality. To be consistent with other Institute for Work & Health (IWH) reviews, we used the terms primary and secondary, acknowledging that our definition of secondary prevention combines definitions of tertiary prevention (preventing disability onset) and secondary prevention (identifying asymptomatic or pre-clinical employees and getting them to early treatment).

The inclusion of secondary intervention studies was based on the following two assumptions:

- there are too few primary prevention intervention studies to warrant a systematic review
- secondary prevention interventions represent an important source of information on the health improvement of samples representative of worksites.

Because of the small numbers of studies in this area, we did not exclude non-randomized trials that met our methodological quality criteria.

An important goal of this review was to advance thinking on intervention research and on the state of current literature on this topic. Thus, our review was inclusive, not exclusive, which allows us to communicate to both stakeholder and scientific audiences.

Originally this review was intended to answer a question specific to "long-term care and nursing homes," but we expanded the scope to "health-care settings." This occurred because stakeholders encouraged a broader review of all health-care settings, and the review team was concerned there were too few intervention studies in long-term care and nursing home settings. Therefore, we included "acute care" settings (i.e. hospitals). Given the differences that exist between acute care and long-term care (e.g. organization, economics and patient population), we agreed that in answering the basic research question it might be important to categorize conclusions within settings.

Three key definitions were needed prior to performing a literature search. We created specific definitions of the terms "health-care setting or worker," "intervention" and "musculoskeletal health" to determine the breadth of the literature search.

Health-care setting or worker was defined broadly as inpatient care settings or workers in such settings. This included hospitals, assisted living facilities,

long-term care facilities, rehabilitation hospitals, medical centres, emergency departments, tertiary care centres and nursing homes. The definition did not include home health-care workers or stand-alone ambulatory care facilities such as individual doctor's offices, urgent care centres, walk-in clinics, dental offices or physical/occupational therapy rehabilitation centres. In short, in ambulatory care facilities, patients were not part of an inpatient facility or an emergency department. We made these exclusions because our review team agreed that exposures would be significantly different in inpatient versus outpatient facilities. We also excluded laboratory studies, commercial pharmacies, optometry stores, psychologists' offices, chiropractic clinics and stand-alone alternative medicine centres including acupuncture, homeopathic, naturopath and massage clinics.

It would be beneficial to consider workers who only had direct exposure to the intervention. One example would be workers involved in patient lifting activities in studies of lift equipment interventions. However, most studies used denominators of all employees, not just those exposed to the intervention. Therefore, we could not exclude auxiliary staff because some studies did not distinguish which workgroups were included in the description of hospital or nursing home employee. For instance, these descriptions could have included hospital transport staff, food services staff or maintenance staff.

Interventions were defined as any occupational health and safety intervention designed to protect MSK health. We used the traditional hazard control tiers of engineering controls, administrative controls and personal protective equipment use. A broad definition allowed for a more comprehensive literature search to determine what was currently considered an "intervention." We excluded interventions in which the primary outcome was violence reduction, and interventions designed only to meet regulatory requirements (e.g. for respiratory issues, needle-stick injuries and blood-borne pathogens).

Musculoskeletal health included MSK symptoms, disorders or clinical diagnoses. We included workers' compensation and regulatory injury reporting systems, despite the validity and reliability vulnerabilities of these data sources. Hereafter we refer to workers' compensation claims and regulatory injury reports as "administrative outcomes." Administrative outcomes are important to stakeholders because they use these outcomes to evaluate the effectiveness of interventions and to plan programs. We excluded studies where muscle loading was the only outcome because our focus was MSK injury outcomes. The importance of muscle loading research (e.g. EMG measurements and self-reported "perceived exertion") as plausible pathways to MSK injury is raised in the discussion. We also excluded surgeries, cancers and gynecological/pregnancy-related MSK symptoms, disorders and diagnoses.

The review team considered peer-reviewed scientific articles published or in press in the English, Spanish, French and Swedish languages. Language proficiency of team members was the primary reason for exclusions. Book chapters, dissertations and conference proceedings were excluded since it was expected that key findings would be reported in the peer-reviewed literature.

Stakeholder Engagement

Stakeholder groups from the health-care sector were invited to provide feedback on specific aspects of the review. They also received a presentation on the systematic review process. Meetings were held at the Institute for Work & Health (IWH) in Toronto, Canada and at the University of Texas School of Public Health in Houston, U.S. Holding meetings in two places allowed for a broad range of stakeholder perspectives to be captured. Practitioners, health-care managers, health-care employees and policy makers were invited (see Appendix A for a list of stakeholders).

The purpose of the stakeholder meetings was to solicit input from relevant parties related to the following topics:

- the research question
- search terms
- information that stakeholders would want to make decisions
- the quality assessment process to evaluate the literature.

In Toronto, nine stakeholders representing insurance companies, government agencies, occupational health associations and lift equipment manufacturers attended a two-hour meeting. In Houston, there was a 90-minute meeting with three stakeholders in person and seven via teleconference. The stakeholders represented professional associations, hospitals, nursing homes, NIOSH and medical centres.

Stakeholders were supportive and expressed interest in review results. They suggested a number of additional search terms including:

- residents
- home health care
- back injuries
- safe patient lifting
- emergency medical technicians
- injuries other than patient lifting injuries
- disabilities.

The review team assessed all proposed search term additions with IWH library professionals. Terms that duplicated existing items or did not add articles were not included. Importantly, one set of search terms (i.e. lifting)

helped in constructing a valid search.

When asked where they sought information, stakeholders said they turned to websites (e.g. IWH, NIOSH), conference proceedings and professional organization journals. Stakeholders expressed concern that excluding the non-peer-reviewed literature might cause the review team to miss a large part of the literature. The review team agreed that incorporating the non-peer-reviewed literature was beyond the scope of this review, but felt it would be critical to disseminate results through the communication channels identified by stakeholders.

There were three sources of outcome data: administrative MSK health outcomes, such as workers' compensation claim data or regulatory injury reporting records; self-reported MSK outcomes; and clinical outcomes. Both stakeholder groups endorsed clinical health outcomes as the most valuable source. Secondly, stakeholders preferred administrative outcomes, as they used this information in decision-making.

Stakeholders wanted the review to answer the following questions: "What was the most effective intervention?" and "What worked and how much did it cost?" Since other systematic reviews at the IWH are considering the economic evaluation of programs, we focused on the first question.

2.1 Literature search

The literature search is based on the research question and our definitions of health-care setting or worker, intervention and musculoskeletal health. Key terms were identified and combined to search the following databases: MEDLINE, EMBASE, CINAHL, Academic Source Premier, PsychInfo and Business Source Premier.

Search terms were identified for three broad areas: intervention terms, health-care setting terms and MSK health outcome terms (see Table 1). The search categories were chosen to be exclusive within each area. The search strategy combined the three areas using an AND strategy, and combined terms within each category using an OR. An example of a search would be training programs OR lifting equipment AND assisted-care AND cumulative trauma disorders. This would identify an article that had training programs and lift equipment *interventions* in an assisted living *health-care setting* that focused on cumulative trauma disorder as the *MSK health outcome*.

Table 1: Search terms

Search strategy: terms within a row are combined with OR and between rows with AND.

Intervention terms	training programs, orientation programs, lifting, lift devices, mechanical lift devices, zero lift, "no lifts," minimal lifts, maximum lifts, active lifts, passive lifts, hoists, patient transfer, patient assist, material transfer, material handling, manual lift devices, manual assistance, transfer aides, transfer assistance, transfer device, slide board, organizational and policy (administrative) changes, disability management, medical management, participatory ergonomics programs, staffing, shift-work, ergonomic, job redesign, work redesign, equipment redesign, job enlargement, task rotation, work hardening, work place safety, work safety, return to work programs, prevention exercises, strength training, flexibility program, body mechanics, lifting teams, back school, psychosocial work organization, patient handling, resident handling, ceiling lifts, overhead lifts, functional abilities evaluation, functional abilities screening, physical demand analysis, engineering controls, personal protective equipment (PPE), administrative control, antifatigue mats, back belts, non-skid flooring, shoe choice, non-slip soles, slippery floor signs, wet floor signs, umbrella covers, intervention research, intervention studies, interventions NOT: blood-borne pathogens, infection prevention and control, EAP or substance abuse and drug treatment programs, radiation safety courses, hazard(ous) communication, smoking cessation programs, hazwoper, needle-stick, violence prevention (if not in conjunction with something else on the inclusion), cytotoxics, glutaraldehyde formalin, formaldehyde, spill training, lab safety, chemical hygiene, hazardous waste emergency planning, lockout/tagout, energy control, mercury, infection control, ethylene oxide, universal precautions, fire prevention and control, suicide prevention, conflict resolution
Health-care setting terms	assisted-care, assisted living, nursing home, hospital, acute-care, skilled nursing facility, old age facilities, old age homes, residential care facility, long-term care, long term care facility, medical centre, tertiary care centres, direct-care workers, patient techs, CNAs (certified nursing assistants), nurses aides, nursing assistant, personal support workers, patient sitter, transporter, porter, orderlies, attendants, LVNS (licensed vocational nurses), PCA (patient care assistants), nurses, nursing, RN, registered practical nurses, nurse practitioner, nurse clinicians, clinical health nurses, medical aides, nurse anesthetist, physician, surgeon, surgical techs, or techs,

	residents, clerk, intern, radiology tech, diagnostic imaging, ultrasound, OT/PT, rehab therapists, recreation workers/activities workers, hospital workers, health-care workers, health-care aides, restorative care aides, long term care aides, retirement aides, nursing home workers, dietary aides, laundry aides, laundry, food services, housekeeping, maintenance, part-time workers, contract workers, mental health in-patient, emergency department, emergency services, pharmacists, pharmacist aides, pharmacist techs, allied health personnel, paramedic NOT: out-patient, ambulatory, pharmacy, stand-alone ambulatory, stand-alone medicine centres, chiropractic, walk-in clinic, homeopathic, naturopath, massage, psychologist, urgent care clinic, urgent care centre
Musculoskeletal health outcome terms	arm injuries, cumulative trauma disorders, tendonitis, tendinopathy, tenosynovitis, rotator cuff, neck injuries, synovitis, muscle weakness, forearm injuries, wrist injuries, hand injuries, osteoarthritis, "sprains and strains", soft tissue injuries, arthralgia, finger injuries, tendon injuries, bursitis, nerve compression syndromes, myofascial pain syndromes, neuralgia, causalgia, radiculopathy, polyradiculoneuritis, polyneuritis, muscular diseases, carpal tunnel syndrome, shoulder impingement syndrome, thoracic outlet syndrome, tennis elbow, epicondylitis, cervico-brachial neuralgia, ulnar nerve compression syndrome, musculoskeletal diseases, musculoskeletal disorders, repetitive trauma, musculoskeletal system, musculoskeletal injuries, musculoskeletal symptom, RSI, neck pain, back pain, back injuries, degenerative disc disorders, degenerative disc diseases, intervertebral disk displacement, herniated disc, bulging disc, lumbar strain, cervical strain, thoracic strain, upper extremity/AND pain, lower extremity/AND pain, knee injuries, hip injuries, leg injuries, disability NOT: cancer, surgery, pregnancy, gynecological symptoms, gynecological diseases

Before the literature search, the review team identified a list of 22 “must-have” articles to test the sensitivity of our search. An initial search missed two of the 22 articles. Upon investigation, we found that one of these articles was an electronic publication in advance of the print version and was not listed in any database. The second missing article required a different search

term (i.e. lifting) that had not been included in the "intervention" category (Table 1). A second search including the new terms captured all 21 published articles and was therefore considered to have face validity.

The review team also contacted 16 content experts to solicit relevant articles that were not expected to be identified by the search. Six experts responded and four suggested articles. Five of these articles had been accepted for publication. Two experts sent chapters that pertained to the topic. A look through the chapters' reference lists revealed no new references. Only the five articles accepted for publication were moved forward to Level 1 review.

2.2 Level 1 - Selection for relevance

The broad search strategy captured many studies not relevant to our research question. A Level 1 relevance review was designed to identify and exclude these as quickly as possible. Reviewers read only the article title and abstract (when available). Article relevance at Level 1 was based on six criteria, described in Table 2.

If reviewers did not know how to answer a question as part of the screening criteria, they were instructed to mark "unclear" (see Appendix B for Level 1 guide to reviewers). In such cases the article would move forward to the next stage where more information would be available for a decision about inclusion or exclusion. Reviewers entered answers for all levels of the process on Systematic Review Software (17). SRS allowed centralized article tracking and access.

Table 2: Level 1 – Screening questions and the response that leads to exclusion*

LEVEL 1	
1. Did an intervention occur in a health-care setting?	NO
2. Was the reference from a peer-reviewed publication (in-press or accepted for publication)?	NO
3. Was the language English, Spanish, Swedish or French?	NO
4. Did the study only have post-intervention measurements, with no control group?	YES
5. Was individual health data collected?	NO
6. Was outcome measure MSK symptoms/disorders/injury?	NO

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

Because a number of articles passed the abstract review stage with unclear responses, we added an additional step, which became Level 1b. It had the same criteria as Level 1a, but the full article was screened in Level 1b. One team member reviewed each article at Level 1a, while two reviewed each article at Level 1b. At Level 1b, relevant articles were moved forward for

Level 2 review when the two reviewers reached consensus on answers to all six questions.

Since a single reviewer conducted the Level 1a review, there was a possibility for selection bias. Therefore a quality control (QC) check was done with an independent reviewer (QC reviewer).

The QC reviewer assessed a randomly chosen set of 12 studies from each of the seven reviewers who participated in Level 1 review (n=84). Each set included six studies excluded at Level 1a, and six that would continue to subsequent review levels.

QC reviewer responses were entered into SRS software so they could be directly compared to a team member's responses. Of the 84 articles reviewed by the QC reviewer, there were 28 cases in which the QC reviewer disagreed with the original reviewer. In 25 of 28 cases (90 per cent), the QC reviewer excluded the study while the original reviewer included it. Therefore, the original reviewer was more likely to be inclusive than the QC reviewer. We did not consider over-inclusion a problem since the article would be reviewed at the next level for relevance. Furthermore, the QC reviewer was not part of the review process, and missed hearing important decisions and approaches that were not captured in the reviewer guide.

More significant were the three cases in which the reviewer excluded the article and the QC reviewer included it. Upon investigation, we found that two articles were less than three pages long and, therefore, in line with our criteria in the reviewer guide, would have been excluded at Level 1b (see Appendix B). The third article was a regulatory intervention that was not included in our definition of "intervention" and its exclusion was appropriate. Therefore, we consider the Level 1a review process reasonable and not vulnerable to significant selection biases.

Fixed versus open study populations

While conducting the Level 1 review, it became clear that if we excluded studies that did not collect individual health data, as per our screening questions (Table 2), we would exclude relevant studies. Upon inspection, the review team realized that this criterion most often excluded studies in which group injury rates were compared. For example an MSK injury rate for a unit (worksite) was calculated using the number of individual injuries divided by unit exposure time. Individual exposure time was not collected. These injury rate comparison studies used open populations. In such studies, whoever is employed in worksites, departments, wards or units contributed to the injury rate (18). However, the overall group is affected by natural worksite changes including turnover. Therefore workers may leave or begin working during the study either before or after the intervention has been implemented.

We used the term "open population" to refer to the type of study in which details about workers entering and exiting the study were unknown. Often the term "dynamic cohort," rather than open population is used to contrast with a fixed or inception cohort (18). However, we considered open population studies to be different than dynamic cohort studies. In dynamic cohorts, information is known about all individuals entering or exiting the study as well as, typically, the amount of time each person is participating. In open population studies, the information about individuals who are in or out of the study at any one point in time is unknown, as is the time each person has participated in a study. In a fixed cohort, all individuals participating are known at baseline, whereas in a dynamic cohort new individuals can enter the cohort and baseline participants can exit. While this potentially creates multiple levels of study designs, in this paper we recognized only two levels, open and fixed populations. We only found two studies that could be considered dynamic cohorts (Collins 2004; Carrivick 2001).

Some methodological advantages to following the same set of individuals over time are:

- Any changes in MSK health could be observed in a population known to have received the intervention. There is tremendous turnover in health-care settings. If a fixed group of individuals is not followed, new people can enter the sample with unknown MSK health problems. They also may not experience the full intervention (e.g. they may have access to the new lift equipment but not be trained in the equipment's use).
- Following specific individuals enables researchers to observe who continues to participate or chooses to drop out. For example, unhealthy people may drop out of studies, which makes the intervention appear to work better. In fact the population has changed due to a "healthy worker" effect.

Both stakeholder groups expressed concern about restricting the systematic review to studies with individual level health data. The review team reconsidered the criterion and decided that it was important to include a broader range of studies with a wider spectrum of methodological quality. Consequently the group needed to develop a new criterion.

The review team agreed that studies would be considered relevant if the health outcome of "injuries" was collected at an individual level, through workers' compensation claims or regulatory injury reports. However, the review team agreed that studies that **only** used "injury counts" would still be excluded at Level 1. To be included the study needed to calculate an "injury rate." The issue of counts versus rates is discussed in the next section.

Based on the revised criterion, the group reviewed all studies that had been excluded because of lack of individual health data (n=16). In total, 50 per cent were considered relevant and went on to Level 2 review. The other 50 per cent had injury counts, not rates, and were excluded at Level 1.

Injury counts versus injury rates

Injury counts are problematic in systematic reviews for a number of reasons. The review team excluded studies that used injury counts as a health outcome measure because of the uncertainty that arises when comparing studies (see Figure 1). Two sites can have the exact same change in the number of injuries. However, depending on the number of workers who participated in the intervention, the conclusion about the effect of the intervention might be different.

Figure 1: Injury counts versus injury rates

Comparing across sites and studies:

Hospital A 100 employees	Hospital B 1000 Employees
5 fewer injures= 5/100 FTE = .05/FTE	5 fewer injuries= 5/1000 FTE = .005/FTE

Therefore, the rate of decrease in Hospital B is much smaller than Hospital A's, even though the injury count is the same.

2.3 Level 2 - Quality assessment

Articles that passed the Level 1 review were further evaluated for methodological quality during a Level 2 review. The team identified 19 methodological criteria to assess quality, which are shown in Table 3. Study quality is important, because a previous review of health promotion literature found that lower quality studies were more likely to find positive effects (19). Each article was independently reviewed by two team members. To reduce bias, the same two members did not review all of the same articles. Instead, each reviewer was randomly paired with at least two other team members. Reviewer pairs were required to reach consensus on all 19 quality criteria. Team members did not review articles they had consulted on, authored or co-authored.

Because each methodological criterion was not considered equal, the review team assigned weights *a priori* for each criterion. The four-point weighting ranged from "somewhat important" (1 point) to "very important" (4 points) (see Table 3).

Table 3: Level 2 - Quality appraisal questions and weights

Question	Weight
1. Was the research question/objective clearly stated?	2
2. Was a primary hypothesis clearly stated?	1
3. Was the intervention implementation described?	3
4. Was the calendar duration of the intervention documented?	1
5. Was the length of follow-up three months or greater?	2
6. Were concurrent comparison (control) group(s) used?	4
7. Was the intervention allocation randomized?	4
8. Were sample inclusion/exclusion criteria described?	2
9. Was the sampling frame representative of the target population?	2
10. Was the participation rate reported and greater than 40% for employees?	3
11. Did the researchers describe the study participants at baseline by demographics, exposure or outcome?	2
12. Were baseline characteristics presented by group?	3
13. Were differences between those employees who remained in the study and those who dropped out analyzed?	3
14. Did withdrawals affect groups equally?	3
15. Were the effects of the intervention on some exposure parameters documented?	3
16. Was contamination between groups described or documented?	1
17. Were covariates/potential confounders for MSK disorders measured (i.e. gender, age, non-work activities)?	3
18. Was adjustment made for covariates/potential confounders?	2
19. Were statistical methods adequately described?	3

Disagreements between each pair of reviewers were identified and reviewers resolved differences by discussion. In cases where agreement could not be reached, a third reviewer was consulted to ensure consensus was obtained (see Appendix C for the Quality Appraisal (QA) guide to reviewers).

Methodological quality scores for each article were based on a weighted sum of 19 quality criteria. The highest possible weighted score was 47. Each article received a quality ranking score by dividing the weighted score by 47 and multiplying by 100. The quality ranking score was used to group articles into four categories of quality: high (80% to 100%), medium-high (60% to 79%), medium (40% to 59%) and limited (less than 40%). The categories were determined by team consensus with reference to the review methodology literature (13; 15; 20) and a past intervention review (21).

Study quality appraisal and evidence synthesis

Since study quality is one of the cornerstones of evidence synthesis, the review team reflected on the quality appraisal process to ensure that all studies were treated similarly. This was especially important for open population studies, since many of the validity issues driving the QA ranking were grounded in appraisal of fixed population studies. Open population studies could have included novel approaches unanticipated in the Guide for Quality Appraisal (Appendix C). Therefore to ensure consistency between reviewers on open population studies, consensus was reached on a series of issues (see Appendix D for Quality Appraisal Decisions).

Consequently a total of 27 studies changed their QA score. As a result, two studies moved below the 60% medium-high quality ranking cut-point.

2.4 Level 3 – Full data extraction

The quality ranking represents the review team's assessment of the internal, external, construct and statistical conclusion validity of each study (22). Each validity type is important in determining how much weight to give to any one study's reported effects. A lower overall validity reflected greater uncertainty among the review team as to whether the findings were the result of chance or design. Therefore full data extraction and evidence synthesis were only completed on medium-high and high quality studies.

Data were extracted by reading and recording details from each paper. The extracted data were used to build summary tables to inform evidence synthesis and to develop our overall conclusions.

Data extraction was performed independently by two reviewers. Again, reviewer pairs were rotated to reduce bias. Team members did not review articles they had consulted on, authored or co-authored. Differences in data extracted between reviewers were identified and resolved. In cases where agreement could not be reached, a third reviewer was consulted to ensure consensus was obtained.

The team developed standardized data extraction forms based on existing forms and data extraction procedures (23; 24) (see Appendix F for the Data extraction (DE) guide for reviewers).

Reviewer pairs extracted data on: year of study; type of health-care setting; study design; sample characteristics; length of follow-up; intervention characteristics; MSK health outcomes and whether those outcomes were self-reported, administrative or clinically-based; statistical analyses; covariates/confounders; and study findings (see Table 4 for the complete list of data extraction questions). The review team decided to record the effects reported for the longest follow-up period when considering study findings.

During data extraction, reviewers reconsidered the methodological quality rating scores recorded in the Level 2 review. Any quality rating changes that the reviewer identified were proposed to the full team for consensus. Final ratings are documented in the methodological quality appraisal table (Table 7).

Initially, we planned to calculate the effect sizes for each article to evaluate the strength of associations uniformly (25; 26; 27; 28). However, this approach was abandoned early in the process once we realized the amount of heterogeneity in outcome measures and study methods, and the lack of data necessary to calculate effect size in some studies.

Table 4: Full data extraction (FDE) items

1. State the research question/objective.
2. State the primary hypothesis.
3. State additional hypotheses not listed in question #2.
4. Write the last name of the first author and the year of publication.
5. List the jurisdiction where the study was completed.
6. Describe what type of health-care organization(s) that the study was conducted in.
7. List the job titles/classification of the participants in the study.
8. List the inclusion criteria described in the study.
9. List the exclusion criteria described in the study.
10. What is the study design?
11. What type of prevention intervention did the study investigate?
12. Describe all interventions evaluated.
13. Was there confirmation the intervention occurred?
14. How long after the intervention implementation did confirmation occur?
15. What was the duration of the intervention in months/days/hours?
16. Indicate the time period between the baseline measurement and all subsequent follow-up measurements.
17. Describe overall (study) group.
18. Describe the intervention group(s).
19. Describe the referent group(s).
20. When were potential covariates/confounders measured?
21. Select from this list all covariates/confounders that were evaluated for inclusion in the final analysis.
22. Provide a list of covariates/confounding variables that were controlled for in the final test of the intervention's effectiveness.
23. Describe the differences in covariates/confounders for those that participated in the study vs. those that were invited but did not participate, if possible by experimental group.
24. Describe the differences in covariates/confounders for those that participated in the study vs. those that were lost to follow-up, if possible by experimental group.
25. Does the study use “administrative” records to collect measurements of MSK health outcomes?
26. Does the study use self-reported questionnaire records as completed by the employee to collect measurements of MSK health outcomes?
27. Does the study use clinical exams or clinical records as completed by the clinician to collect measurements of MSK health outcomes?
28. Was the population studied “fixed” or “open”?
29. What sources were used to “count” employee injuries?

30. How were employee hours collected?
31. Indicate at what level employee hours were ascertained and/or estimated.
32. Were injury rates calculated?
33. If injury rates were calculated, list the equation(s).
34. Did the study discuss how researchers handled any of the following special issues related to administrative record keeping: temporary or contract employees; employees who floated between units; turnover rate; reinjury to the same employee?
35. Check all body regions where symptoms were ascertained by questionnaire.
36. Describe when follow-up MSK health outcomes (symptoms) were measured.
37. Were MSK symptoms measured at the same time of day or shift?
38. Check all body regions where specific clinical disorders were ascertained by physical examination or laboratory test.
39. Was masking of physical assessment done?
40. Was a standard protocol used for the clinical exams?
41. Please check the types of final analyses done for testing the observed effects of the intervention.
42. Describe for each outcome of interest (MSK) the observed intervention effects.
43. Remark on the findings or enter information that is unique about the study that may not be adequately captured in the other DE questions.

The following general guidelines were used to present findings:

- present findings as the authors did
- when only a global statistical test of the impact of multiple interventions on MSK health effects was conducted, this was presented as "all interventions"
- if a reviewed study did not have a MSK primary outcome but MSK health data was reported, we included the evidence in the synthesis
- when specific MSK health data values were not reported, values were abstracted from figures.

Decision rules were developed to present findings when more than one outcome was used to evaluate the intervention:

- MSK health effects described as improvements were noted as "positive"
- MSK health effects described as deteriorations were noted as "negative"
- MSK health effects described as not significant were noted as "no effect" and the direction of change, if any, was indicated in parentheses
- between-group MSK health effect comparisons (e.g. intervention versus control) were presented where study design allowed
- where no statistical tests were presented for the observed MSK health effects, this was noted in parentheses
- if there was a discrepancy about MSK health effects between statements or tests in the abstract and results, the reviewers relied on the results.

2.4 Evidence synthesis

The high level of heterogeneity required the use of a synthesis approach adapted from Slavin and others (13; 23; 16) known as "best evidence synthesis." This approach considers the article's quality, the quantity of articles using the same prevention strategy and the consistency of the findings (Table 5). "Quality" refers to the methodological strength of the studies as determined in the Level 2 review (quality appraisal ranking). "Quantity" refers to the number of studies that provide evidence on the same intervention category. "Consistency" refers to the similarity of results observed across the studies. A strong level of evidence exists when there are three high-quality studies with convergent effects.

Our evidence synthesis guidelines were adapted from other IWH prevention intervention reviews (21; 29; 23; 30). While the review team first used the evidence synthesis to answer the global question ("Do occupational safety and health interventions in health-care settings have an effect on musculoskeletal health status?"), levels of evidence were also reviewed for intervention categories represented in the literature.

In synthesizing evidence we needed to develop decision rules when a study used more than one outcome to evaluate an intervention. A study with any positive results and no negative results on a single intervention was classified as a *positive effect* study. A study with both positive effects and no effects was also classified as a *positive effect* study (e.g. there was a positive effect on one outcome such as back pain, but no effect on another outcome, such as neck pain). A study with only no effects was classified as a *no effect* study. A study with any negative effects was classified as a *negative effect* study. Synthesis of the reviewed evidence on a particular intervention category was ranked on the following scale: strong evidence; moderate evidence; mixed evidence; insufficient evidence (see Table 5 below). In all cases:

- application of the evidence guidelines for each intervention category relied on review team consensus
- the synthesis conclusions were accepted by all review team members.

Finally, the review team agreed to synthesize evidence when the study reported a statistical test of the intervention's effect on the MSK outcome. Statistical test included a description of a test statistic (e.g. a chi square value) with a p-value, or presentation of confidence intervals. A statistical test was needed to provide statistical confidence that the observed intervention was due to the intervention's effect, and not to chance.

Table 5: Best evidence synthesis guidelines

Level of evidence	Minimum quality	Minimum quantity	Consistency
Strong	High (>80%)	>=3 studies	All high-quality studies converge on the same findings.
Moderate	Medium-high (60-79%)	>=2 studies	Majority of medium-high quality studies converge on the same findings.
Mixed	Medium-high (60-79%)	>=2 studies	Medium-high and better quality studies have inconsistent findings.
Insufficient	None of the above criteria are met.		

2.5 Partial data extraction

Many studies with administrative outcomes were ranked as medium or limited quality. Given the importance of administrative outcomes to stakeholders, the review team felt that describing this part of the intervention literature was essential to guide future research. Therefore, the team conducted a partial data extraction (PDE) on articles that lacked sufficient methodological quality to proceed to evidence synthesis.

Following Côté (2001), the review team decided to extract some key features described in medium and limited quality studies (e.g. setting and design). Given the distinct methodological issues in administrative records, details of the approach used to measure health outcomes and statistical analyses were described. Using the information in the full data extraction (FDE) and PDE tables, we constructed figures to highlight differences between the studies synthesized into evidence and those that were not (see Figures 7, 8, 9, 10 in Results).

Partial data extraction was performed jointly by two reviewers. The reviewers did not review articles they had consulted on, authored or co-authored. Differences in data extracted between reviewers were identified and resolved to reach consensus. In cases where agreement could not be reached, a third reviewer was consulted to ensure consensus was obtained.

The reviewer pair extracted data on: year of study; type of health-care setting; study design; intervention characteristics; MSK outcomes and whether those outcomes were self-reported, administrative or clinically-based; statistical analyses; and study findings (see Table 6 for the complete list of the partial data extraction questions).

During the partial data extraction process, reviewers reconsidered the methodological quality rating scores recorded during the Level 2 review.

Any quality rating changes that the reviewer identified were proposed to the full review team for a consensus decision. All final ratings are documented in the methodological quality appraisal table (Table 7).

Table 6: Partial data extraction (PDE) items

1. Write the last name of the first author and the year of publication.
2. List the jurisdiction where the study was completed.
3. Describe what type of health-care organization(s) the study was conducted in.
4. List the job titles/classification of the participants in the study.
5. List the inclusion criteria described in the study.
6. List the exclusion criteria described in the study.
7. What is the study design?
8. What type of prevention intervention did the study investigate?
9. Describe all interventions evaluated.
10. Does the study use "administrative" records to collect measurements of MSK health outcomes?
11. Does the study use self-reported questionnaire records as completed by the employee to collect measurements of MSK health outcomes?
12. Does the study use clinical exams or clinical records as completed by the clinician to collect measurements of MSK health outcomes?
13. Was the population studied "fixed" or "open"?
14. What sources were used to "count" employee injuries?
15. Indicate at what level employee hours were ascertained and/or estimated.
16. If injury rates were calculated, list the equation(s).
17. Did the study discuss how researchers handled any of the following special issues related to administrative record keeping: temporary or contract employees; employees who floated between units; turnover rate; reinjury to the same employee?
18. Please check the types of final analyses done for testing the observed effects of the intervention.
19. Describe for each outcome of interest (MSK) the observed intervention effects.
20. Remark on the findings or enter information that is unique about the study that may not be adequately captured in the other DE questions.

3.0 Results

3.1 Literature search and selection for relevance

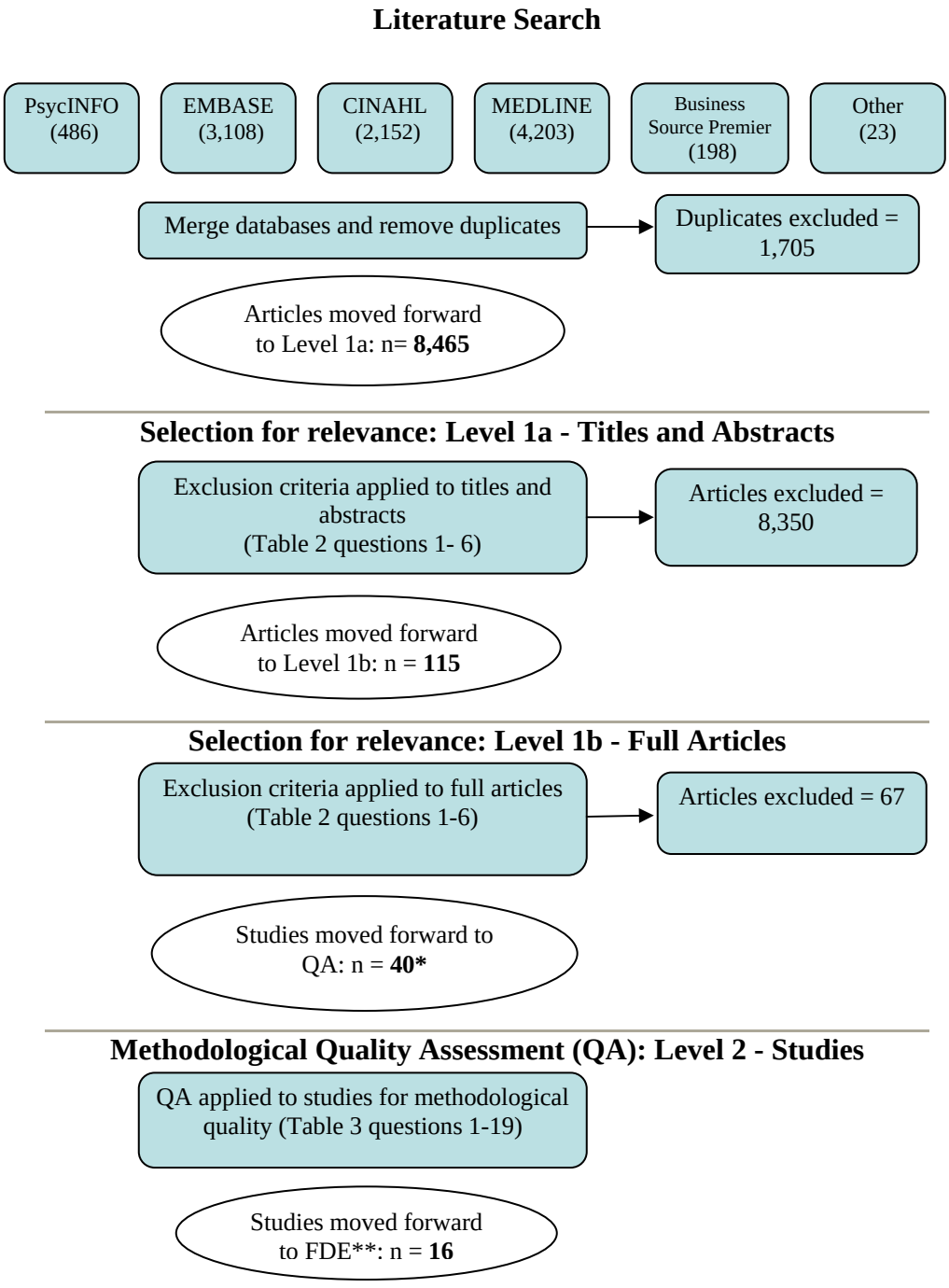
We identified 10,147 articles using the search terms listed in Table 1. After different databases were merged, duplicate articles were removed and articles from content experts were included, we were left with 8,465 articles (Figure 2).

A total of 8,350 articles were excluded during the Level 1 review because they did not meet the inclusion criteria (refer to Table 2 for criteria).

A total of 115 articles proceeded to Level 1b review. Using the exclusion criteria in Table 2, two team members reviewed the articles that had an "unclear" response at Level 1a to the first question, "Was there an intervention in a health-care setting?" This led to the exclusion of 67 additional articles (for more details about articles excluded by Level 1 criteria, see Appendix G). Eight articles were grouped with other articles that described results from the same study.

This yielded 40 studies that proceeded to Level 2 methodological quality appraisal. These 40 studies were each reviewed by two reviewers using the quality appraisal questions in Table 3.

Figure 2: Flowchart of systematic review process



* Eight articles were grouped with other articles that described results from the same study.
- Carrivick 2005 with Carrivick 2002
- Fanello 2002 with Fanello 1999
- Lagerstrom 1997b with Lagerstrom 1997a
- Landstad 2001 with Landstad 2000
- Cooper 1998 and Cooper 1996 with Yassi 1995
- Chhokar 2005 and Spiegel 2002 with Ronald 2002

**Full data extraction

3.2 Methodological quality appraisal

The 40 studies that met our relevance criteria were assessed for methodological quality using 19 criteria (Table 7). The criteria were weighted according to the importance of each item as decided by the entire review team.

The weighted criteria were used to develop a relative quality score for each study. This is described in section 2.3. Studies ranked high and medium-high quality were included in evidence synthesis.

High quality studies

Two studies were of high quality. The high quality studies were quite consistent in their quality scores, since they both met 15 of the 19 criteria (85%). However, neither study stated a hypothesis nor adjusted for covariates in testing for the intervention effect. Only one of the two studies included at least a three-month follow-up time or described contamination between groups.

Medium-high quality studies

We classified 14 studies as medium-high quality (range 60-74%). These studies generally scored well on the following criteria: using concurrent comparison (control) group(s); describing sample inclusion/exclusion criteria; presenting baseline characteristics; and measuring covariates/ confounders. However, few of these studies met the following criteria, as indicated: stating a hypothesis (4/14); describing differences between participants and those lost to follow-up (3/14); indicating contamination between groups (4/14); and adjusting for covariates/confounders (4/14). The medium-high quality studies also did not meet the criteria of randomizing the intervention (5/14) and reporting participation rates >=40% (8/14), while the high quality studies did meet these two criteria.

Medium quality studies

We classified 20 studies as medium quality (range 40-57%). Most of these studies presented baseline characteristics (19/20). Few of these studies met the following criteria, as indicated: randomizing the intervention (2/20); describing differences between participants and those lost to follow-up participants (3/20); indicating contamination between groups (1/20); and adjusting for covariates/confounders (2/20). The medium quality studies often did not include a control group (9/20) or present baseline characteristics by group (7/20) while the medium-high quality studies did.

Limited quality studies

We classified four studies as limited quality (range 17-26%).

Table 7: Methodological quality assessment (QA) (n=40)

Criteria*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**Quality Rank-ing**
Weight	2	1	3	1	2	4	4	2	2	3	2	3	3	3	3	1	3	2	3	
Author, year																				
High quality (H) (two studies)																				
Harma, 1988	1	0	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	85%
Maul, 2005	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	85%
# of studies meeting criteria	2	0	2	2	1	2	2	2	1	2	2	2	2	1	2	1	2	0	2	
Percentage of criteria met	100%	0%	100%	100%	50%	100%	100%	100%	50%	100%	100%	100%	100%	50%	100%	50%	100%	0%	100%	
Medium-high quality (MH) (14 studies)																				
Bru, 1994	1	0	1	0	1	1	1	1	0	1	1	1	0	0	0	1	1	1	1	72%
Carrivick, 2002	1	0	1	0	1	0	0	1	0	1	1	1	0	0	1	1	1	1	1	62%
Collins, 2004	1	1	1	0	1	1	0	1	0	1	1	0	1	0	0	0	1	1	1	64%
Dehlin, 1978	1	0	1	1	0	1	0	1	0	0	1	1	1	0	1	0	1	0	1	62%
Dehlin, 1981	1	1	1	1	0	1	0	1	0	0	1	1	0	0	1	1	1	0	1	60%
Donchin, 1990	1	0	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1	0	1	74%
Gundewall, 1993	1	0	1	1	0	1	1	0	0	0	1	1	1	0	1	0	1	0	0	60%
LeClerc, 1997	1	1	1	0	1	1	0	1	0	1	1	1	0	0	1	0	1	0	1	66%
Li, 2004	1	0	1	1	1	1	0	1	0	1	1	0	1	0	1	0	1	0	1	66%
Linton, 1989	1	0	1	1	1	1	1	1	0	0	1	1	0	0	1	0	1	0	0	62%
Oldervoll, 2001	1	0	1	1	1	1	0	1	0	1	1	1	0	0	1	0	1	0	1	66%
Smedley, 2003	1	0	1	0	1	1	0	1	0	1	1	1	0	0	1	0	1	1	1	68%
Videman, 1989	1	1	1	1	1	1	0	1	0	0	1	1	0	0	1	0	1	0	1	62%
Yassi, 2001	1	0	1	0	1	1	1	1	0	0	1	1	0	0	1	0	0	0	1	60%
# of studies meeting criteria	14	4	14	8	11	13	5	13	0	8	14	12	4	0	12	3	13	4	12	
Percentage of criteria met	100%	29%	100%	57%	79%	93%	36%	93%	0%	57%	100%	86%	29%	0%	86%	21%	93%	29%	86%	

Table 7: Methodological quality assessment (QA) – continued

Criteria*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**Quality Rank-ing**
Weight	2	1	3	1	2	4	4	2	2	3	2	3	3	3	3	1	3	2	3	
Author, year																				
Medium quality studies (M) (20 studies)																				
Alexandre, 2001	1	1	1	1	0	0	1	1	0	0	1	1	0	0	0	0	1	0	1	51%
Best, 1997	1	1	1	0	1	1	1	0	0	0	1	1	0	0	1	0	1	0	0	57%
Carrivick, 2002	1	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	1	1	43%
Davis, 2004	1	1	1	0	1	1	0	1	0	1	1	1	0	0	0	0	0	0	1	53%
Evanoff, 1999	1	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	1	0	1	55%
Evanoff, 2003	1	1	1	0	1	1	0	1	0	0	1	0	0	0	1	0	0	0	1	47%
Fanello, 1999	1	0	1	0	1	0	0	1	0	1	1	1	0	0	1	0	1	0	1	55%
Fujishiro, 2005	1	1	1	0	1	0	0	1	0	0	1	0	0	0	1	0	1	0	1	45%
Garg, 1992	1	0	1	0	1	0	0	0	0	1	1	0	0	0	1	0	1	0	1	45%
Lagerstrom, 1997	1	1	1	1	1	0	0	0	0	0	1	0	0	0	1	0	1	0	1	43%
Landstad, 2001	1	0	1	1	0	1	0	1	0	1	1	1	0	0	0	1	1	0	1	57%
Nassau, 1999	1	1	1	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	1	45%
Nelson, 2005	1	0	1	1	1	0	0	1	0	1	1	0	0	0	1	0	1	1	1	55%
Nevala, 2004	1	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	1	43%
Peterson, 2004	1	1	1	1	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	40%
Ronald, 2002	1	0	1	1	1	0	0	1	0	1	1	0	1	0	0	0	1	0	1	51%
Skargren, 1996	1	0	1	1	0	0	0	1	0	1	1	1	1	0	1	0	1	0	0	53%
Sobaszek, 2001	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	1	0	1	45%
Tiesman, 2003	1	0	1	1	1	0	0	1	0	0	1	0	0	0	1	0	1	0	1	45%
Yassi, 1995	1	0	1	0	1	1	0	1	0	1	1	1	1	0	0	0	0	0	1	57%
# of studies meeting criteria	20	9	20	11	15	9	2	14	0	12	19	7	3	0	12	1	15	2	17	
Percentage of criteria met	100%	45%	100%	55%	75%	45%	10%	70%	0%	60%	95%	35%	15%	0%	60%	5%	75%	10%	85%	

Table 7: Methodological quality assessment (QA) – continued

Criteria*	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**Quality Ranking**
Weight	2	1	3	1	2	4	4	2	2	3	2	3	3	3	3	1	3	2	3	
Author, year																				
Limited quality studies (L) (four studies)																				
Charney, 1997	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	21%
Guthrie, 2004	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	17%
Lynch, 2000	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	1	26%
Ryden, 1988	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	1	26%
# of studies meeting criteria	4	0	3	0	1	1	0	2	0	0	0	0	0	0	2	0	0	0	3	
Percentage of criteria met	100%	0%	75%	0%	25%	25%	0%	50%	0%	0%	0%	0%	0%	0%	50%	0%	0%	0%	75%	

*Refer to Table 3 for the quality assessment criteria.

3.3 Full data extraction results

The 16 studies of high or medium-high quality proceeded to full data extraction. The review team grouped the various interventions described in the studies to create categories of interventions with consensus. Table 8 shows these categories and descriptions used for full data extraction. Additionally, the table includes the study design and the type of injury prevention.

Intervention categories

We organized these 16 studies into 10 different intervention categories. The categories and number of studies undergoing full data extraction (FDE) are listed below. In categories with more than one aspect to the intervention, we connected the intervention characteristics with the "&" symbol. In studies of more than one intervention category (Dehlin 1981, Donchin 1990), the categories are separated by commas:

- multi-component patient handling interventions – 3 studies
- exercise training – 5 studies
- patient handling training – 1 studies
- equipment & equipment training – 1 study
- participatory ergonomics – 1 study
- cognitive behavioural training – 1 study
- exercise training, patient handling training – 1 study
- broad-based MSK injury prevention program – 1 study
- intensive off-site MSK injury prevention program – 1 study
- back school, exercise training – 1 study

Most interventions involved some type of training (13 of the 16 studies). Many intervention categories were examined by only one study (n=8).

Study design

Eight of the 16 studies were non-randomized field trials and seven were randomized field trials. One study was pre- versus post-intervention with a statistical control group.

Type of injury prevention

Nine studies were both primary and secondary prevention trials; therefore, participants were not excluded based on symptoms or disorders. Seven of the secondary prevention studies only included subjects who reported symptoms or were diagnosed with disorders.

Additional information regarding the intervention is contained in Appendix G.

Table 8: Description of interventions from full data extraction studies (n=16) [refer to key at end of table for abbreviations]

Author, year	Description of Intervention	Study design	Prevention type	QA
	Multi-Component Patient Handling*			
Collins, 2004	I_1: zero lift policy; mechanical lifting equipment and repositioning aids; lift equipment training and medical management	Pre-post w/statistical control	Both	MH
Smedley, 2003	I_1: revised manual-handling policy; equipment: sliding sheets, hi/lo baths, hoists and transfer belts; two-day health and safety training C: no policy change; a non-formal ergonomic program (usual)	NR field trial	Both	MH
Yassi, 2001	I_1: "safe-lift" policy; lifting and transfer equipment; three hours of education on back care, patient assessment and handling techniques I_2: "no strenuous lifting" policy; new mechanical patient lifts and transfer equipment on each ward; three hours of education on back care, patient assessment and handling techniques C: no policy changed; one mechanical total body lift available on the ward and access to sliding devices from a central equipment depot on request only; no training provided	R field trial	Both	MH
	Exercise Training			
Dehlin, 1978	I_1: muscle training (exercise) C_1: lectures on geriatric medicine and nursing care C_2: no training or lectures C_3: asymptomatic exposure parameter controls; no training or lectures	NR field trial	Secondary prevention	MH
Gundewall, 1993	I_1: exercise to increase dynamic endurance, isometric strength and functional coordination C: no exercise	R field trial	Both	MH

Table 8: Description of interventions from full data extraction studies - continued

Author, year	Description of Intervention	Study design	Prevention type	QA
	Exercise Training - continued			
Harma, 1988	I_1: physical training (exercise) C: no exercise	R field trial	Both	H
Maul, 2005	I_1: low-back school (three sessions of one hour each) and physical exercises C: low-back school and no exercise	R field trial	Secondary prevention	H
Oldervoll, 2001	I_1: endurance training (promoting aerobic capacity) twice a week for 17 weeks I_2: strength promotion (SP) classes twice/week for 17 weeks C: wait listed controls	NR field trial	Secondary prevention	MH
	Patient Handling Training			
Videman, 1989	I_1: increased practical patient handling training (ergonomic training) C: traditional patient handling training	NR field trial	Both	MH
	Exercise Training, Patient Handling Training			
Dehlin, 1981	I_1: physical fitness training (exercise) 1_2: ergonomic education on lifting technique C: no training or ergonomic education	NR field trial	Secondary prevention	MH
	Back School, Exercise Training			
Donchin, 1990	I_1: calisthenics (exercise) I_2: back school C: wait-listed controls	R field trial	Secondary prevention	MH
	Cognitive Behavioural Training			
Bru, 1994	I_1: cognitive-behavioural training I_2: relaxation training I_3: combined I_1 and I_2 C: wait-listed controls	R field trial	Secondary prevention	MH

Table 8: Description of interventions from full data extraction studies - continued

Author, year	Description of Intervention	Study design	Prevention type	QA
	Intensive Off-Site MIPP			
Linton, 1989	I_1: low-back program at an off-site clinic & hotel (exercise training, ergonomics training, behavioural training on pain management) C: wait-listed controls	R field trial	Secondary prevention	MH
	Equipment & Equipment Training			
Li, 2004	I_1: one portable full-body sling lift, two portable stand-up sling lifts (''E-Z Lift'' and ''E-Z Stand'' by EZ Way Inc, Minneapolis, Minnesota) and friction-reducing sheets (Maxislides) and training sessions in lift usage C: no lifts purchased	NR field trial	Both	MH
	Participatory Ergonomics			
Carrivick, 2001	I_1: participatory ergonomics team of cleaners identified, assessed and recommended controls of manual handling C_1: hospital orderlies not receiving the intervention C_2: cleaners from another hospital not receiving the intervention C_3: hospital and non-hospital cleaners in the state insurance system	NR field trial	Both	MH
	Broad-Based MIPP			
Leclerc, 1997	I_1: training with exercise and ergonomic changes following a site visit by an ergonomist C: usual injury prevention policies	NR field trial	Both	MH

*Multi-Component Patient Handling - an intervention that included three components: policy change, equipment purchase and training on equipment usage & patient handling

MIPP = MSK Injury Prevention Program
I = Intervention group (subscripts indicate number of groups, e.g. I_2)
C = Control group (subscripts indicate number of groups, e.g. C_2)
R field trial = Randomized field trial
NR field trial = Non-randomized field trial
MSK = Musculoskeletal

QA=Quality appraisal
H=high
M=medium-high

Study Description
Table 9 describes factors affecting comparability across studies, and strength of study design. The health-care setting, country and job title all address the similarity within the populations studied. The type of study design, population, participation rate, sample size and follow-up rate address differences that represent threats to internal validity.

Countries of origin
The most common geographic setting was Sweden (n=4). The remaining studies originated in different regions: there were two from the U.S., two from Norway, two from Finland, and one each from U.K., France, Canada, Israel and Australia. One study did not provide the country of origin.

Health-care setting
A variety of settings were represented. Most studies were set in hospitals (14 studies). A nursing home and a nursing school were the other two settings.

Job titles
Many job titles were listed but the primary job titles were nurse, nursing aide, nursing assistant and licensed practical nurses.

Type of study population and participation rate
Four studies had open populations in which participants could enter and leave the study. Eight studies had a fixed population and followed the same participants over time. Four studies collected both administrative and self-reported data and therefore had open populations with administrative data, but they followed a group of participants over time. Participation rates were not provided in almost half of the studies (7/16). When participation rates were provided they were between 44 and 100%.

Sample sizes and numbers lost to follow-up
The sample sizes in the studies tended to be small but varied from 45 to 1,239. Five studies did not report the total sample size and three studies did not report the sample size for intervention or control groups. Details on participants lost to follow-up were missing in almost a third of the studies (n=5). When reported, the numbers lost to follow-up tended to be small, but varied from 0% to 57%.

Additional information on comparability and validity are reported in detail in Appendix H.

Table 9: Study description for full data extraction studies (n=16) [refer to key at end of table for abbreviations]

Author, year	Health-care setting (country)	Job titles	Study design	Population description	Participation rate	Sample size	Loss to follow-up
Multi-Component Patient Handling*							
Collins, 2004	Nursing homes (U.S.)	certified nursing assistants, registered and licensed practical nurses, physical therapists, restorative aides.	Pre-post w/statistical control	Open	100%	NP	NP
Smedley, 2003	Public hospitals (England)	nurses	NR Field Trial	Open	56%	N = 1239 I_1 = NP C = NP	57% of sample
Yassi, 2001	Acute and tertiary care hospital (Canada)	nurses, unit assistants	R field trial	Admin: Open Self-report: Fixed	NP	N = 346 I_1 = 116 I_2 = 127 C = 103	25% of sample
Exercise Training							
Dehlin, 1978	Geriatric/long-term care hospital (Sweden)	nursing aides	NR field trial	Fixed	NP	N = 66 I_1 = 18 C_1 = 14 C_2 = 14 C_3 = 20	8% of sample
Gundewall, 1993	Wards of a geriatric hospital (Sweden)	nurses, nurses' aides	R field trial	Fixed	NP	N = 69 I_1 = NP C = NP	13% of sample
Harma, 1988	University hospital (Finland)	nurses, nursing aides	R field trial	Fixed	79%	N = 119 I_1 =76 C =43	37% of sample
Maul, 2005	University hospital (NP)	all hospital employees	R field trial	Fixed	51%	N = 148 I_1 = 74 C = 74	24% of sample
Oldervoll, 2001	University hospital (Norway)	registered nurses, auxiliary nurses, laboratory staff, administration staff, cleaning department staff.	NR field trial	Open	54%	N = 65 I_1 = 24 I_2 = 22 C = 19	31% of sample

Table 9: Study description for full data extraction studies – continued

Author, year	Health-care setting (country)	Job titles	Study design	Population description	Participation rate	Sample size	Loss to follow-up
Patient Handling Training							
Videman, 1989	Nursing school (Finland)	student nurses	NR field trial	Self-report: Unclear	NP	N =308 I_1 =151 C =157	NP
Exercise Training, Patient Handling Training							
Dehlin, 1981	Geriatric hospital (Sweden)	Nurses' aides	NR field trial	Fixed	NP	N = 45 I_1 = 15 I_2 = 14 C = 16	13% of sample
Back School, Exercise Training							
Donchin, 1990	University hospital (Israel)	clinical, administrative, technical employees	R field trial	Fixed	70%	N = 142 I_1 =46 I_2 =46 C =50	NP
Cognitive Behavioural Training							
Bru, 1994	Hospital (Norway)	physicians, registered nurses, auxiliary nurses, laboratory staff, kitchen staff	R field trial	Fixed	71%	N=119 I_1 =NP I_2 =NP I_3 = NP C = NP	7% at 4 months
Intensive Off-Site MIPP							
Linton, 1989	Hospital (Sweden)	licensed practical nurses, nursing aides	R field trial	Fixed	NP	N = 66 I_1 = 36 C = 30	0%
Equipment & Equipment Training							
Li, 2004	Community hospital (U.S.)	nurses, nursing assistants, patient care attendants	NR field trial	Admin: Open Self-report: Fixed	Admin: NP Self-report: 44%	N = NP I_1 = 61 C = NP	NP

Table 9: Study description for full data extraction studies – continued

Author, year	Health-care setting (country)	Job titles	Study design	Population description	Participation rate	Sample size	Loss to follow-up
Participatory Ergonomics							
Carrivick, 2001	Hospital (Australia)	cleaners, orderlies	NR field trial	Open	100%	N = NP I_1 = 507 C_1 = 279 C_2 = NP C_3 = NP	NP
Broad-Based MIPP							
Leclerc, 1997	Hospital (France)	nurses and auxiliary nurses	NR field trial	Admin: Open Self-report: Fixed	90%	NP	15%+

*Multi-Component Patient Handling - an intervention that included three components: policy change, equipment purchase and training on equipment usage & patient handling

+ 15% - not comparable to other data because includes hospital as well as manufacturing and office settings. For complete details see Leclerc 1997.

MIPP = MSK Injury Prevention Program

I = Intervention group (subscripts indicate number of groups, e.g. I_2)

C = Control group (subscripts indicate number of groups, e.g. C_2)

MSK = Musculoskeletal

NP= Not provided

R field trial = Randomized field trial

NR field trial = Non-randomized field trial

N= Sample size

Design implementation and analysis
Table 10a and 10b describe design implementation and analysis of the study factors that affected generalizability and comparability among studies. We present the inclusion and exclusion criteria as either relating to the "sample" of health-care workers that the study drew from, or the cases of MSK injuries/illnesses that were described and counted (i.e. "case").

Table 10a describes studies with administrative sources as their MSK outcome studies. It includes four medium-high and high quality studies. Table 10b describes studies with self-reported outcomes. Two studies that used both self-reported and administrative outcomes are listed in both Table 10a and 10b.

Information presented in Table 10a
Table 10a presents the level of information on person-hours (the denominator for injury rates), as well as turnover or reinjury rates – characteristics that might confound MSK outcome measurements in an open population. Finally, the type of outcome and statistical method are presented so that comparability between studies can be evaluated.

Inclusion criteria
To refine their populations to capture workers exposed to the intervention, four studies in Table 10a included job titles and setting. Two of the four studies described inclusion criteria for an MSK "case" (Yassi 2001, Li 2004). Since participants were neither included nor excluded based on symptoms or disorders, the interventions did not distinguish between primary and secondary prevention.

Exclusion criteria
Only two studies listed exclusions (Yassi 2001, Li 2004). Yassi (2001) isolated the sample exposed to the intervention by excluding float staff. Li (2004) specified the injury more explicitly by excluding different injuries that were not hypothesized to be influenced by the intervention.

Outcome description
Three studies using administrative outcomes specified the type of injury related to the intervention effect (e.g. patient handling injury related to a patient handling intervention).

Statistical methods
Half of the studies used multivariate statistical methods and the other half used both multivariate and univariate statistical methods.

Administrative outcomes
Two studies collected hours worked at an individual level. Two studies used unit level data. No study reported a turnover rate or adjusted for study

participants being injured more than once. Only one study reported the injury rate calculation.

Table 10a: Factors that affect generalizability and strength of administrative outcomes from full data extraction (n=4) [refer to key at end of table for abbreviations]

Author, year	Inclusion criteria+	Exclusion criteria+	Person hours description+	Turnover (N or %)	Reinjury addressed?	Injury rate calculation	Outcome type	Statistical methods
Multi-Component Patient Handling*								
Collins, 2004	Sample: nursing home "nursing personnel" (Ind)	NP	productive hours worked only for nursing personnel (Ind)	NP -5% present five years and 5% present all six years of exposure time	NP -reported number of employees reporting more than one injury	NP	"Resident handling" injury rates from OSHA logs, employer records and WC; also "Resident Handling" LWD rate and RWD rate from OSHA logs	Poisson regression to calculate relative risks and z-test to compare rates
Yassi, 2001	Sample: nurses and unit assistants on medical, surgical, and rehabilitation hospital wards (Ind) Case: patient lift/transfer incidents (Ind)	Sample: float pool staff (Ind)	"per 200,000" hours (Unit)	NP	NP	NP	"Patient handling" injury rates and claim duration rate from WC	Mantel-Haenszel chi square test and Cox proportional hazard model

Table 10a: Factors that affect generalizability and strength of administrative outcomes for full data extraction studies – continued

Author, year	Inclusion criteria[+]	Exclusion criteria[+]	Person hours description[+]	Turnover (N or %)	Reinjury addressed?	Injury rate calculation	Outcome type	Statistical methods
Equipment & Equipment Training								
Li, 2004	Sample: nurses on medicine/surgery, intensive care, and subacute care (Ind); directly involved in patient handling (Ind) Case: MSK injuries potentially related to lifting, (e.g. shoulder strains, upper and lower back strains, and knee strains) (Ind)	Case: Injuries from bodily fluids or chemicals exposures, slips, falls, and contusions (Ind)	productive hours worked for each division (Unit)	NP	NP	Injuries or lost days per 100 FTEs, where 1 FTE = 2,000 productive hours per year.	"lifting" MSK injury rate and LWD rate from OSHA 200 log and from WC	Poisson regression to calculate adjusted rate ratios
Participatory Ergonomics								
Carrivick, 2001	Sample: study hospital cleaners and orderlies (Ind); comparison hospital cleaners at (Site); all cleaners in the state insurance system (Site)	NP	employee hours from financial records (Ind)	NP	NP	NP	Lost-time injury rate	Poisson regression to calculate adjusted rate ratios

*Multi-Component Patient Handling -an intervention that included three components: policy change, equipment purchase and training on equipment usage & patient handling

[+](Ind) = Individual Level, (Unit) - Unit Level, (Site) = Site Level

MSK = Musculoskeletal

MIPP = MSK Injury Prevention Program

LWD = Lost work days

FTE = Full-time equivalents

RWD = Restricted work days

WC = Workers' compensation

N = Sample size

Information presented in Table 10b
Table 10b describes the inclusion and exclusion criteria for the studies using self-reported MSK outcomes or clinical information to assess the intervention's effects. The self-reported measurement scales used are also presented in more detail here than in Table 11.

Inclusion criteria
Thirteen of the 14 studies that used self-reported MSK questionnaire data listed inclusion criteria. All 13 specified the setting (site or unit description). Two studies included participants based on gender (e.g. Oldervoll 2001); eight described participants by job title (e.g. Smedley 2003); and two included individuals based on age (e.g. Harma 1988). Seven studies described inclusion criteria that would identify them as evaluating a secondary prevention intervention. In other words, the study only included participants with identified MSK symptoms or disorders (e.g. Dehlin 1978).

Exclusion criteria
Only 10 of the 14 studies that used self-reported MSK outcomes described exclusion criteria. Three studies excluded participants based on job titles (e.g. Smedley 2003). Two studies listed exclusions based on intervention adherence (e.g. Dehlin 1978). Five studies excluded participants based on confounding health outcomes (e.g. cardiovascular disease, Maul 2005). No studies excluded a type of MSK health outcome. One study excluded employees who had taken sick leave greater than three months for any reason to create a symptom-free population for a primary intervention study.

Outcome measurement
Eight of the self-reported MSK measurements asked about back or low-back only, while the remainder collected a combination of information on neck, shoulders, knees or ankles in addition.

Statistical methods
Six studies used multivariate methods such as Poisson or Cox Proportional Hazard regression. Five studies used univariate statistical tests such as chi-square or t-test. One study used a combination of univariate and multivariate methods. Two studies did not use statistical tests to evaluate intervention effects (Gundewall 1993, Videman 1989).

Table 10b: Factors that affect generalizability and strength of self-reported outcome from full data extraction (n=14) [refer to key at end of table for abbreviations]

Author, year	Inclusion[+]	Exclusion[+]	Outcome	Statistical Methods
Multi-Component Patient Handling*				
Smedley, 2003	Sample: hospital nurses (Ind)	Sample: midwives and community based nurses (Ind); (after baseline) health-care assistants retroactively excluded from the comparison site due to a change in payroll systems that "recategorized" employees (Ind)	Self-report: low-back pain lasting more than a day in the past month (presence/absence)	Multivariate logistic regression
Yassi, 2001	Sample: medical, surgical and rehabilitation hospital wards(Unit); nurses and unit assistants (Ind) Case: patient lift/transfer incidents (Ind)	Sample: float pool staff (Ind)	Self-report: VAS low-back pain and shoulder pain in past week	ANOVA
Exercise Training				
Dehlin, 1978	Sample: nursing aides with current low-back symptoms (Ind)	Sample: inability to complete the training (Ind)	Self-report: low-back insufficiency measure (frequency, intensity and duration of symptoms; function limitations and medication usage)◦	Mann-Whitney U Test
Gundewall, 1993	NP	NP	Self-report: number of days experiencing back pain and number of days absent from work due to back pain	No statistical test

Table 10b: Factors that affect generalizability and strength of self-reported outcome for full data extraction studies - continued

Author, year	Inclusion[+]	Exclusion[+]	Outcome	Statistical Methods
		Exercise Training - continued		
Harma, 1988	Sample: shift workers for at least 1.5 years (Ind); age 20-49 years (Ind); hospital nurse or nursing aide (Ind)	Sample: unable to complete the exercise intervention (for I_1) or starting personal exercise (for C) (Ind)	Self-report: MSK symptom index (back symptoms, neck, shoulder or hip symptoms; knee or ankle symptoms)	Wilcoxon matched pair test Mann-Whitney U Test
Maul, 2005	Sample: >30 days low-back pain in past 12 months or 8–30 days LBP together with reported disability in daily tasks in past 12 months (second recruitment) (Ind); 20 -55 years old (Ind); ability to read and write German or Italian (Ind)	Sample: cardiovascular or metabolic diseases (angina pectoris, previous heart surgery, cardiac failure, hypertension, diabetes, hyperthyroidism), progressive radicular neurological defects, inflammatory diseases of the spine, previous spinal surgery, other non-rheumatic diseases that may account for LBP (i.e., kidney, gynecological disease), and pregnancy (Ind); participation in regular strength training within the last six months (Ind); planning to change job or working less than 50% time (Ind)	Self-report: pain assessed by Nordic questionnaire of low-back pain; low-back pain intensity (numeric rating scale, pain drawing and medication use); McGill pain characteristics scale; perceived treatment effectiveness	ANOVA
Oldervoll, 2001	Sample: female hospital employee with pain in the neck, shoulders and/or lower back for at least three months during the past year, and also recurring pain during past 30 days (Ind); self-reported reduced work capacity, sick leave or reduced leisure activity due to pain (Ind); agree to participate (Ind)	Sample: radiating pain in arm or leg, pain from neck, shoulders and/or low back (Ind); heart disease, blood pressure above 160/110 mmHg, lung diseases, diabetes mellitus or cancer (Ind)	Self-report: pain index (localization of pain/duration/function limitation) from Nordic questionnaire of MSK pain: lower back, neck and shoulders	ANOVA for pain index and Student's t-tests on individual body areas

Table 10b: Factors that affect generalizability and strength of self-reported outcome for full data extraction studies - continued

Author, year	Inclusion+	Exclusion+	Outcome	Statistical Methods
Patient Handling Training				
Videman, 1989	Sample: nurses who began nursing school between 1981 and 1984 (Ind)	NP	Self-report: back pain severity, disability and onset	No statistical test
Exercise Training, Patient Handling Training				
Dehlin, 1981	Sample: nurses aides working daytime duty with self-report of low-back insufficiency symptoms (Ind)	NP	Self-report: low-back insufficiency measure (frequency, intensity and duration of symptoms; function limitations and medication usage)°	Mann-Whitney U Test
Back School, Exercise Training				
Donchin, 1990	Sample: at least three annual episodes of low-back pain in previous epidemiological study (Ind); current hospital employee (Ind); agree to randomization (Ind)	NP	Self-report: duration, rate, type of low back pain episode	ANOVA
Cognitive Behavioural Training				
Bru, 1994	Sample: medical doctors, registered nurses, auxilary nurses, laboratory staff, kitchen staff (Ind); female hospital workers with self-reported pain in neck, shoulder and/or low back over the past seven days that had caused leave of absence for some period in the past 12 months (Ind); back pain had to be reported for at least two periods over the last six months (Ind)	Sample: staff working in sterile environment (Ind); medical diagnosis such as rheumatoid arthritis, epilepsy, previous surgery to the spine, osteoporosis, breast cancer, fibromyalgia or pregnancy (Ind)	Self-report: pain intensity (five point Likert scale) and pain duration in past 30 days (four point Likert scale) in neck, shoulders and low back	ANOVA

Table 10b: Factors that affect generalizability and strength of self-reported outcome for full data extraction studies - continued

Author, year	Inclusion[+]	Exclusion[+]	Outcome	Statistical Methods
Intensive Off-Site MIPP				
Linton, 1989	Sample: women sick-listed for back pain some time during the previous two-year period (Ind); currently working (Ind); self-reported pain at study entry (Ind)	Sample: any disease that would be a counterindication for the program based on an orthopedic exam (Ind)	Self-report: low-back pain intensity (0-100 VAS) measured with diary at morning, lunch and evening	ANOVA
Equipment & Equipment Training				
Li, 2004	Sample: nurses on medicine/surgery, intensive care, and subacute care (Ind); directly involved in patient handling (Ind) Case: MSK injuries potentially related to lifting, (e.g. shoulder strains, upper and lower back strains and knee strains) (Ind)	Case: Injuries from bodily fluids or chemicals exposures, slips, falls and contusions (Ind)	Self-report: MSK discomfort in neck, shoulders/upper arm, upper back, lower back, forearm, wrist/hand, hips/buttocks, knees, feet/ankles (Likert scale: 1-uncomfortable to 5-comfortable)	Wilcoxon matched pair test
Broad-Based MIPP				
Leclerc, 1997	Sample: actively employed at the hospital (Ind)	Sample: sick leave longer than three months in the previous 12 months irrespective of the cause (Ind); pregnancy (Ind); temporary work contract, retirement in the following 12 months, or duration of employment less than one year (Ind)	Self-report: MSK Nordic questionnaire of pain intensity and duration in low back, upper back, neck or shoulders	T-test

*Multi-Component Patient Handling - an intervention that included three components: policy change, equipment purchase and training on equipment usage & patient handling

[+](Ind) = Individual Level, (Unit) - Unit Level, (Site) = Site Level

°low-back insufficiency -primarily driven by symptoms

MSK = Musculoskeletal

VAS = Visual Analog Scale

MIPP = MSK Injury Prevention Program

3.4 Evidence synthesis

Table 11 presents a summary of the interventions' effects with the outcome measurement and follow-up time. The guidelines from Table 6 are used to determine the level of evidence. Two studies, Gundewall (1993) and Videman (1981), are not included in the evidence synthesis because they did not make statistical comparisons of intervention effects of MSK outcomes.

The findings for each intervention type and each outcome reported are described using the "I" and "C" convention described. The review team did not find any negative or adverse effects of any intervention on MSK outcomes. Therefore, we consistently report positive effects or no effects.

Overall, to answer the basic question driving this review, we found a **moderate level of evidence** for the effect of OHS interventions on MSK health status in health-care settings. This means that a majority of high and medium-high studies found positive effects on MSK outcomes.

Table 11: Intervention effects on MSK health outcomes as reported in evidence synthesis studies (n=14)
[refer to key at end of table for abbreviations]

Author, year[+]	Outcome	Effect (direction of effect°, comparison) on MSK health outcomes °even if non-significant	Length of follow-up	QA
		Multi-Component Patient Handling*		
Collins, 2004	Admin: "Resident handling" injury rates from OSHA logs, employer records and WC; also "Resident Handling" LWD rate and RWD rate from OSHA logs	**Positive** effect for resident handling injury incidence from OSHA (decrease in I_1 vs. C, where C is "all other injuries") **Positive** effect for resident handling WC rate, injury rate from employer records, LWD rate and RWD rate (decrease in post- vs. pre-intervention)	36 months	MH
Smedley, 2003	Self-report: low-back pain	**No** effect on back pain prevalence (increase in post- vs. pre-intervention)	NP	MH
Yassi, 2001	Admin: "Patient handling" injury rates from WC Self-report: low-back pain and shoulder pain	All interventions: **No** effect on injury rates (decrease in I_1, I_2 vs. C at 12 months) "Safe Lifting": **Positive** effect on low-back pain and shoulder pain (decrease in I_1 vs. C at 12 months) "No Lift": **No** effect on low-back pain and shoulder pain (decrease in I_2 vs. C at 12 months) "No Lift": **Positive** effect on low back pain and shoulder pain (decrease in I_2 vs. C at 6 months) "Safe Lifting": **No** effect on low back pain and shoulder pain (decrease in I_1 vs. C at 6 months)	12 months	MH

Table 11: Intervention effects on MSK health outcomes as reported in evidence synthesis studies - continued

Author, year[+]	Outcome	Effect (direction of effect°, comparison) on MSK health outcomes °even if non-significant	Length of follow-up	QA
		Exercise Training		
Dehlin, 1978	Self-report: low-back insufficiency –driven by symptoms	**Positive** effect for low-back insufficiency (decrease in I_1 vs. C_1) **No** effect for low back insufficiency (decrease I_1 vs. C_2)	NP	MH
Harma, 1988	Self-report: MSK symptom index	**Positive** effect for MSK symptom score (decrease in I_1 vs. C)	NP	H
Maul, 2005	Self-report: pain assessed by Nordic questionnaire of low-back pain; low-back pain intensity (numeric rating scale, pain drawing and medication use); McGill pain characteristics scale; perceived treatment effectiveness	**Positive** effect on perception of therapy reducing pain (decrease in I_1 vs. C_1 at 120 months) **No** effect for current pain in comparison with pre-treatment pain (decrease in I_1 vs. C at 120 months) **Positive** effect on pain drawing (decrease in I_1 vs. C at 12 months) **No** effect for numeric rating scale of pain or McGill pain characteristics (decrease in I_1 vs. C at 12 months)	120 months	H
Oldervoll, 2001	Self-report: pain index	**Positive** effect on pain index (I_1 and I_2: decrease in post- vs. pre-intervention at 7 months) All interventions: **Positive** effect on pain index (I_1, I_2 vs. C at 0 months) Endurance Training: decrease in post- vs. pre-intervention Strength Promotion: decrease in post- vs. pre-intervention	7 months	MH
		Exercise Training, Patient Handling Training		
Dehlin, 1981	Self-report: low-back insufficiency –driven by symptoms	All interventions: **No** effect on low-back insufficiency (I_1, I_2 vs. C)	NP	MH

Table 11: Intervention effects on MSK health outcomes as reported in evidence synthesis studies - continued

Author, year[+]	Outcome	Effect (direction of effect°, comparison) on MSK health outcomes °even if non-significant	Length of follow-up	QA
		Back School, Exercise Training		
Donchin, 1990	Self-report: duration, rate and type of low-back pain episode	All interventions: **Positive** effect for low-back pain duration (I_1, I_2, vs. C) Exercise Training: decrease in I_1 vs. C Back School: no difference between I_2 vs. C	12 months	MH
		Cognitive Behavioural Training		
Bru, 1994	Self-report: pain intensity and pain duration in neck, shoulders and low back	All interventions: **Positive** effect (I_1, I_2, I_3, vs. C) on pain intensity; body area affected varied by intervention Cognitive training: decrease neck and shoulders in I_1 vs. all others Relaxation training: decrease low back in I_2 vs. all others All interventions: **No** effect (I_1, I_2, I_3, vs. C) on pain duration.	4 months	MH
		Intensive Off-Site MIPP		
Linton, 1989	Self-report: low back pain intensity	**Positive** effect on low-back pain intensity (decrease in I_1 vs. C)	6 months	MH
		Equipment & Equipment Training		
Li, 2004	Admin: "lifting" MSK injury rate and LWD rate from OSHA 200 log and from WC Self-report: MSK discomfort in: neck, shoulders/upper arm, upper back, lower back, forearm, wrist/hand, hips/buttocks, knees and feet/ankles	**No** effect for "lifting" MSK injury rates, LWD rates from OSHA & WC (decrease in I_1 vs. C) **Positive** effect for MSK discomfort in all areas (decrease in post- vs. pre-intervention)	Admin: 24 months Self-report: 1 month	MH

Table 11: Intervention effects on MSK health outcomes as reported in evidence synthesis studies - continued

Author, year[+]	Outcome	Effect (direction of effect°, comparison) on MSK health outcomes °even if non-significant	Length of follow-up	QA
Participatory Ergonomics				
Carrivick, 2001	Admin: lost-time injury rate	**Positive** effect for lost time injury rate (decrease in post- vs. pre-intervention) **Positive** effect for manual handling lost-time injuries (decrease in post- vs. pre-intervention)	36 months	MH
Broad-Based MIPP				
Leclerc, 1997	Self-report: MSK pain intensity and duration in low back, upper back, neck and shoulders	**Positive** effect for combined spine and shoulder disorder, upper back disorder and lower back disorder scores (decrease I_1 vs. C) **No** effect for neck disorder or shoulder disorder scores (increase I_1 vs. C)	NP	MH

*Multi-Component Patient Handling -an intervention which included three components: policy change, equipment purchase and training on equipment usage & patient handling

+ Gundewall 2003 and Videman 1989 not included in evidence synthesis

MSK = Musculoskeletal
MIPP = MSK Injury Prevention Program
I = Intervention group
C = Control group
VAS = Visual Analog Scale
LWD = Lost work days
RWD = Restricted work day
WC = Workers' compensation
OHSA=Occupational Health and Safety Administration
MH = Medium-high
H= High
QA = Quality appraisal

Evidence by intervention category
Recall that at least three high quality studies are needed for a strong level of evidence, and at least two medium-high quality studies are needed for a moderate level of evidence.

Multi-component patient handling
The team agreed that "multi-component patient handling" interventions included three components: policy change, equipment purchase, and training on equipment usage and patient handling. Three studies of medium-high quality evaluated interventions that included all three components (Collins 2004, Smedley 2003, Yassi 2001). Two showed positive effects (Collins 2004, Yassi 2001). Therefore, we concluded there was **moderate** evidence that a multi-component patient handling intervention had a positive effect on MSK health.

Exercise training
Six studies evaluated exercise training. Two were high quality (Harma 1988, Maul 2005) and four were medium-high quality (Dehlin 1978, Oldervoll 2001, Dehlin 1981, Donchin 1990). Four described "physical fitness" or "calisthenics" while two described exercises that specifically improved strength and/or endurance. Two high quality and four medium-high quality studies showed positive effects on MSK outcomes. We concluded there was **moderate** evidence that exercise training interventions had a positive effect on MSK health.

Patient handling training
One medium-high quality study showed no effect on MSK outcomes for a training intervention on patient handling (Dehlin 1981). With a single study, there is **insufficient** evidence to determine whether these training interventions on their own have an effect on MSK outcomes.

Back school
One medium-high quality study showed no effect on MSK outcomes for a back school intervention (Donchin 1990). With a single study, there is **insufficient** evidence to determine whether back school interventions have an effect on MSK outcomes.

Cognitive behavioural training
One medium-high quality study showed a positive effect on MSK outcomes for a cognitive behavioural training intervention (Bru 1994). With a single study, there is **insufficient** evidence to determine whether cognitive behavioural training has an effect on MSK outcomes.

Intensive off-site MSK injury prevention program
One medium-high quality study showed a positive effect on MSK outcomes for an intensive off-site MSK injury prevention program (Linton 1989).

With a single study, there is **insufficient** evidence to determine whether an intensive off-site MSK injury prevention program affects MSK outcomes.

Participatory ergonomics

One medium-high quality study showed a positive effect on MSK outcomes for a participatory ergonomics intervention (Carrivick 2001). With a single study, there is **insufficient** evidence to determine whether participatory ergonomics training affects MSK outcomes.

Equipment & equipment training

One medium-high quality study showed a positive effect on MSK outcomes for an equipment & equipment training intervention (Li 2004). With a single study, there is **insufficient** evidence to determine whether equipment & equipment training have an effect on MSK outcomes.

Broad-based MSK injury prevention program

One medium-high quality study showed a positive effect on MSK outcomes for a broad-based MSK injury prevention program (Leclerc 1994). With a single study, there is **insufficient** evidence to determine whether broad-based MSK injury prevention programs have an effect on MSK outcomes.

Effects by MSK outcome source

Table 12 presents study findings by outcome type. Interventions on training alone (e.g. exercise training, patient handling training) were primarily evaluated using self-reported MSK outcomes. Most of the interventions that involved equipment used administrative outcomes (i.e. multi-component patient handling, participatory ergonomics, equipment & patient handling training, equipment & equipment training, and ergonomic devices & consultation). Among the studies that used more than one administrative outcome, effects were consistent. All showed either positive or no effects.

Table 12: Effects summary by type of outcome measurement for studies in evidence synthesis (n=14)

Author, Year	Administrative			Self-report	Clinical	QA	Evidence
	Injury	LWD	WC				
Multi-Component Patient Handling							
Collins, 2004	+	+	+			MH	Positive
Smedley, 2003				∅		MH	No Effect
Yassi, 2001	∅			+/∅		MH	Positive
Exercise Training							
Dehlin, 1978				+		MH	Positive
Harma, 1988				+		H	Positive
Maul, 2005				+		H	Positive
Oldervoll, 2001				+		MH	Positive
Dehlin, 1981				∅		MH	No Effect
Donchin, 1990				+		MH	Positive
Patient Handling Training							
Dehlin, 1981				∅		MH	No Effect
Back School							
Donchin, 1990				∅		MH	No Effect
Cognitive Behavioural							
Bru, 1994				+/∅		MH	Positive
Intensive Off-Site MIPP							
Linton, 1989				+		MH	Positive
Participatory Ergonomics							
Carrivick, 2001	+	+	+			MH	Positive
Equipment & Equipment Training							
Li, 2004	∅	∅	∅	+		MH	Positive
Broad-Based MIPP							
Leclerc, 1997				+/∅		MH	Positive

*Possible Values: + (Positive Effect), ∅ (No Effect) or +/∅ (both Positive and No Effect found). Blank spaces indicate the outcome measurement was not used.

LWD= Lost work days
WC= Workers' compensation
QA= Quality appraisal
MIPP= MSK injury prevention program

3.5 Partial data extraction results

The review team did a partial data extraction on studies that were ranked either medium or limited quality. These studies were not included in evidence synthesis. The team grouped the various interventions described in the studies and with consensus, created categories of interventions Table 13 shows the intervention categories and descriptions for full data extraction. Additionally, the table includes the study design and the type of injury prevention.

Intervention categories

We organized 24 studies into 16 different intervention categories. The categories and number of studies that had partial data extraction (PDE) are listed below. In intervention categories with more than one aspect to the intervention, we connect the intervention characteristics with an "&" symbol.

- multi-component patient handling interventions –2 studies
- exercise training –1 study
- patient handling training –3 studies
- lift team – 1 study
- lift team & patient handling & equipment training – 1 study
- equipment & equipment training –2 studies
- participatory ergonomics –2 studies
- back school – 2 studies
- cognitive behavioural training –1 study
- integrated disability management program – 3 studies
- equipment & patient handling training – 1 study
- shower trolley – 1 study
- ergonomic devices & consultation – 1 study
- exercise & patient handling training -1 study
- exercise & patient handling & stress management training – 1 study
- pre-employment screen & return to work policy – 1 study.

Most of the studies involved some type of training (16 of the 24 studies). Many intervention categories were examined by only one study (n=10).

Countries of origin

The majority of the studies reviewed were from the U.S. (n=12). However, the studies originated from many different regions: there were three from Sweden, three from Canada, two from France, two from Australia and one from Brazil.

Health-care setting

Most studies were set in hospitals (14 studies). Only five studies were conducted in a nursing home or long-term care settings. Five studies included both hospital and other settings in the same study.

Study designs and population
The studies mostly used pre- versus post-intervention designs (n=12) with no control group, but they also included two randomized trials, eight non-randomized trials, and two non-randomized cross-over designs.

Most of the studies had "open" populations in which people entered and left during the study period (n=13). The "fixed" population design, in which the same participants are followed over time, was only used in 11 of the 24 studies. As mentioned previously, open population studies always included an administrative outcome.

Type of outcome measurement
Twelve studies used only administrative outcomes to identify MSK injuries. Six studies used only self-reported MSK symptoms and six used self-reports along with either administrative (n=5) or clinical outcomes (n=1).

Direction of effect
Ten studies showed only positive intervention effects on MSK outcomes, while only one study found no effects. Additionally 13 studies reported a combination of positive, negative or no effects. Of the 10 studies with only positive effects, three reported findings with no statistical test. There were 13 studies with control groups, but three of these did not compare the intervention group changes with the control group for any of their significance testing.

Table 13: Study description and effects from partial data extraction (n=24) [refer to key at end of table for abbreviations]

Author, year	Setting (country)	Study design	Description of Intervention	Pop[§]	Outcome (Admin, Self-report, Clinical)	Direction of effect	QA
Multi-Component Patient Handling*							
Nelson, 2005	Hospital and nursing home (U.S.)	Pre-post no control	I_1: "no lift" policy with patient handling criteria and peer safety leaders; patient handling equipment; ergonomic assessment training	Open	Admin: Injury log to create injury rates, LWD rate, modified duty rate & WC costs	**Positive** effect for injury rates and modified duty days (post- vs. pre-intervention) **No** effect for LWD rate (decrease, post- vs. pre-intervention)	M
Ronald, 2002	Extended care hospital (Canada)	Pre-post no control	I_1: "no manual lift" policy; mechanical lifting equipment; equipment training and patient handling training	Open	Admin: MSK injury rates Self-report: recent injury and pain	**No** effect for MSK injury rates (decrease, post- vs. pre-intervention) **Positive** effect for lifting and handling MSK injury rates (post- vs. pre-intervention) **No** effect for repositioning MSK injury rates (increase, post- vs. pre-intervention) **Positive** effect on self-report recent injury and pain (no statistical test)	M
Exercise Training							
Skargren, 1996	Geriatric ward (Sweden)	NR crossover	I_1C and I_2C: exercise program (strength and cardio)	Fixed	Self-report: MSK symptoms	**Positive** effect overall for MSK symptoms (post- vs. pre-intervention)	M

Table 13: Study description and effects from partial data extraction – continued

Author, year	Setting (country)	Study design	Description of Intervention	Pop[§]	Outcome (Admin, Self-report, Clinical)	Direction of effect	QA
Patient Handling Training							
Best, 1997	Nursing homes (3) (Australia)	R field trial	I_1: hospital orientation and training in patient transfers and procedures C: in-house orientation training only	Fixed	Admin: Accident reports for manual handling/patient handling sprains and strains Self-report: back pain	**No** effect for admin injury data (No statistical test) **Positive** effect in 12-month recall of back pain (I_1 vs. C) **No** effect in three-month recall of back pain (decrease, I_1 vs. C)	M
Fanello, 1999	Hospital (France)	NR field trial	I_1: low-back pain prevention program and patient handling instruction C: did not receive program or instruction	Fixed	Self-report: neck pain, shoulder pain (uni- & bilateral), back pain and low back pain	**Positive** effect in back pain remission vs. new case (post- vs. pre-intervention) **No** effect (increase in neck, shoulder; decrease in back and low back pain, post- vs. pre-intervention)	M
Peterson, 2004	Veteran's home (U.S.)	NR field trial	I_1: patient handling training with NAs reinforced by research assistant I_2: train LVNs, NAs, CNAs C: no training	Fixed	Self-report: pain or discomfort levels over parts of the body	**No** effect (increase in upper body, decrease in back, I_1, I_2 vs. C)	M
Exercise & Patient Handling Training							
Alexandre, 2001	Hospital (Brazil)	R field trial	I_1: exercise & training (advice on back conditions and ergonomics) C: Short education session	Fixed	Self-report: back pain frequency and intensity	**Positive** effect in two-month and seven-day recall of cervical pain frequency (post- vs. pre-intervention) **No** effect in two-month or seven-day recall of thoracic or lumbar pain frequency or intensity for all outcomes (decrease, post- vs. pre-intervention)	M

Table 13: Study description and effects from partial data extraction – continued

Author, year	Setting (Country)	Study design	Description of Intervention	Pop[§]	Outcome (Admin, Self-report, Clinical)	Direction of effect	QA
Back School							
Lynch, 2000	Hospital (U.S.)	NR field trial[+]	I_1: back injury training C: no training	Open	Admin: injury counts not rates therefore no effects synthesized Self-report: back pain days/week	**No** effect on back pain days/week (direction not provided, post- vs. pre-intervention) **Positive** effect in back pain days/week (I_1 vs. C, post-only data in low patient transfer departments) **No** effect on back pain days/week (direction not provided, I_1 vs. C, post-only data in high patient transfer departments)	L
Sobaszek, 2001	Hospital (France)	Pre-post no control	I_1: back school	Fixed	Admin: absenteeism for health reasons other than pregnancy Self-report: health-care use and perception of back pain	**Positive** effect for all outcomes –health-care use, absenteeism for health reasons, clinical progression, perception of back pain (post- vs. pre-intervention)	M

Table 13: Study description and effects from partial data extraction – continued

Author, year	Setting (Country)	Study design	Description of Intervention	Pop[§]	Outcome (Admin, Self-report, Clinical)	Direction of effect	QA
Cognitive Behavioural Training							
Landstad, 2001	Hospital (Sweden)	NR field trial	I_1: personnel support (private and group discussions) C: no personnel support	Fixed	Self-report: pain Clinical: MSK disorder diagnoses	**No** effect for self-reported pain (decrease, I_1 vs. C) **Positive** effect for clinical diagnoses after intervention (I_1 vs. C)	M
Exercise & Patient Handling & Stress Management.							
Lagerstrom, 1997	Hospital (Sweden)	Pre-post no control	I_1: education (patient transfer technique, physical fitness and stress management)	Fixed	Self-report: MSK symptoms (neck/ shoulder, elbows, wrists/hands, upper back, low back, hips, knees, ankles/feet)	**Negative** effect for upper back and hip MSK symptoms (post- vs. pre-intervention) **No** effect neck/shoulders, wrists/hands, elbows, low back, knees, ankles/feet, (direction unclear, post- vs. pre-intervention)	M
Lift Team							
Charney, 1997	Hospitals (9) and long-term care facility (1) (U.S.)	Pre-post no control[+]	I_1: lift team and new written policy	Open	Admin: OSHA 200 log injury rates	**Positive** effect in injury rates and LWD (no statistical test presented)	L
Lift Team & Patient Handling & Equipment Training							
Guthrie, 2004	Hospital (U.S.)	Pre-post no control[+]	I_1: lift team, equipment and back school	Open	Admin: injury rates	**Positive** effect in injury rates (no statistical test presented)	L

Table 13: Study description and effects from partial data extraction – continued

Author, year	Setting (Country)	Study design	Description of Intervention	Pop[§]	Outcome (Admin, Self-report, Clinical)	Direction of effect	QA
Participatory Ergonomics Team							
Carrivick, 2002	Hospital (Australia)	Pre-post no control	I_1: participatory ergonomics team	Fixed	Admin: MSK, non-MSK, manual handling (MH) and non-MH cases used to create lost-time injury rate, lost-time duration rate, and claims cost rate	**Positive** effect for MH and MSK injury rate, lost-time duration and claims cost rate and injury rate for non-MSK and non-MH (post- vs. pre-intervention) **No** effect (increase in non-MSK lost-time duration, decrease in lost-time duration, and claims cost rate, post- vs. pre-intervention)	M
Evanoff, 1999	Hospital/ medical centre (U.S.)	NR field trial	I_1: participatory ergonomics team (ergonomic assessment, training, implementation) C: no team	Open	Admin: OSHA 200 log and WC claims used to create injury rates, lost-time rates and lost-time duration Self-report: MSK pain at different sites and pain severity	**Positive** effect on injury rate, lost-time injury rate (I_1 vs. C) **Positive** effect on lost-time duration (post- vs. pre-intervention) **Positive** effect on MSK symptoms (post- vs. pre-intervention)	M
Equipment & Patient Handling Training							
Garg, 1992	Nursing home (U.S.)	Pre-post no control	I_1: equipment & training	Open	Admin: OSHA 200 log injury rates	**Positive** effect on injury rates (no statistical comparison)	M

Table 13: Study description and effects from partial data extraction – continued

Author, year	Setting (Country)	Study design	Intervention description	Pop[§]	Outcome (Admin, Self-report, Clinical)	Direction of effect	QA
Equipment & Equipment Training							
Evanoff, 2003	Hospitals and Long-Term Care (U.S.)	Pre-post no control	I_1: equipment and 2 hour course on equipment use	Open	Admin: OSHA 200 log injury rates	**Positive** effect for injury rates and LWD injury rate (post- vs. pre-intervention) **Positive** effect for total lost days (no statistical test presented)	M
Tiesman, 2003	Long-Term Care Unit in Hospital (U.S.)	Pre-post no control	I_1: equipment (ceiling mounted patient lift system) and training on equipment	Open	Admin: employer incident report and WC used to calculate injury rates	**Positive** effect for LWD rate and RWD rate (post- vs. pre-intervention) **No** effect for injury rate (decrease, post- vs. pre-intervention)	M
Ergonomic Devices & Consultation							
Fujishiro, 2005	Hospitals, Nursing Homes and MR/DD homes (86 sites) (U.S.)	Pre-post no control	I_1: equipment to reduce bending I_2: equipment to eliminate lifting I_3: equipment to reduce carrying I_4: combination of I_1 - I_3 (part of state funded ergonomic intervention)	Open	Admin: OSHA logs to create MSK injury rates	All interventions: **Positive** effect on injury rates (post- vs. pre-intervention) **Positive** effect on injury rates for I_2, I_3, and I_4 individually (post- vs. pre-intervention) **No** effect on injury rates for I_1 individually (post- vs. pre-intervention)	M

Table 13: Study description and effects from partial data extraction – continued

Author, year	Setting (country)	Study design	Description of Intervention	Pop[§]	Outcome (Admin, Self-report, Clinical)	Direction of effect	QA
Shower Trolley							
Nevala, 2004	Hospitals (2), health-care centres (2) and homes for the aged (3) (Finland)	NR crossover	I_1C: electric shower trolley I_2C: traditional shower trolley	Fixed	Self-report: MSK strain in neck-shoulders, arms, back & legs	**Positive** effect for VAS strain (electric trolley required less MSK strain) (I_1 vs. C, post-only data)	M
Integrated Disability Management Program							
Davis, 2004	Hospital (Canada)	NR field trial[+]	I_1: Prevention & Early Active Return to Work Program (PEARS) C: No PEARS	Open	Admin: MSK lost-time injuries, WC claims	**Positive** effect on earlier RTW for nurses & health science professionals (post- vs. pre-intervention) **No** effect on injury rate (decrease, I_1 vs. C)	M
Ryden, 1988	Hospital (U.S.)	Pre-post no control	I_1: back care program training & light duty policy	Open	Admin: Injury rates	**Positive** effect on injury rates (test of trend, post- vs. pre-intervention unclear)	L
Yassi, 1995	Acute and Tertiary Care Hospital (Canada)	NR field trial	I_1: comprehensive rehab program consisting of assessment and treatment by a physiotherapist under direction of a rehab specialized physician C: injured controls told to seek care through their routine care-givers	Fixed	Admin: WC to calculate back injury rate, lost-time injury rate, and lost-time duration	**Positive** effect on: back injury rate and lost-time injury rate (I_1 vs. C) **Positive** effect on lost-time duration and WC (No statistical test)	M

Table 13: Study description and effects from partial data extraction – continued

Author, year	Setting (country)	Study design	Description of Intervention	Pop[§]	Outcome (Admin, Self-report, Clinical)	Direction of effect	QA
Pre-employment Screen & RTW Policy							
Nassau, 1999	Hospital and Medical Centre (U.S.)	NR field trial	I_1: non-standardized pre-work functional screening and RTW policy I_2: job specific pre-work functional screening and RTW policy C: no screen, no RTW policy	Open	Admin: Injury rates, lost-time injury rate, lost-time duration and WC costs	All interventions: **Positive** effect lost-time injury rate, lost-time duration and WC costs (I_1, I_2 vs. C) **No** effect on injury rate (decrease, I_1, I_2 vs. C)	M

*Multi-Component Patient Handling -an intervention which included three components: policy change, equipment purchase and training on equipment usage & patient handling

§ Population –denotes study design having an "open" population where participants can enter and leave vs. "fixed" where the same participants are followed over time

+ Program Evaluation Study	RWD = Restricted Work Day	LWD = Lost work day	I = Intervention group
R field trial = Randomized field trial	MSK = Musculoskeletal	RTW = Return to work	C = Control group
NR field trial = Non-randomized field trial	LVN= Licensed vocational nurse	WC = Workers' compensation	QA= quality appraisal
VAS= Visual analog scale	CNA= Certified nursing assistant	L=Low	M= Medium

Design Implementation and Analysis
Table 14a and 14b describe information collected regarding generalizability and comparability between studies. We present the inclusion and exclusion criteria as either relating to the "sample" of health-care workers that the study drew from, or the cases of MSK injuries/illnesses that were described and counted (i.e. "case").

Table 14a describes studies with administrative sources as their MSK outcome. Table 14 b describes studies with self-reported outcomes.

Information presented in Table 14a
Table 14a presents the level of information on person-hours (the denominator for injury rates), as well as turnover or reinjury – characteristics that might confound MSK outcome measurements in an open population. Finally, the type of outcome and statistical method are presented so that comparability between studies can be evaluated.

Inclusion criteria
All 17 studies in Table 14a listed some inclusion criteria. Five described the MSK cases they included (e.g. "patient handling MSK" Nelson 2005). Three described inclusion criteria that would identify them as a secondary prevention intervention (i.e. the study included participants with identified MSK symptoms or disorders). One study was categorized by its intervention description as a "disability management program" but did not describe whether it included only symptomatic participants.

Exclusion criteria
Only eight studies described exclusion criteria. Five studies excluded specific MSK outcomes to refine their outcome measure. Since participants were neither included nor excluded based on symptoms or disorders, most studies (n=14) do not distinguish between primary and secondary prevention.

Statistical methods
Eight studies used univariate statistical tests such as chi-square or t-test. Three studies used multivariate methods such as Poisson or Survival regression. Two studies calculated a 95% confidence interval and one study used a combination of univariate, multivariate and confidence intervals. Three studies did not use statistical tests to evaluate intervention effects.

Person-hours description
Nine studies collected employee hours as a measure of exposure at the unit level. Five studies did not describe how they collected the hours worked to calculate their injury rate. One study reported obtaining the actual hours worked for study participants, and one study used total hours worked for the site. Finally one study used the hours worked by the unit for the intervention group, and the hours for the worksite for the control group.

Other factors affecting administrative data: turnover and reinjury.
Only two studies reported a turnover rate. Further, only one study reported adjusting for reinjury by adding multiple injury costs related to one individual together.

Injury rate calculation
Ten studies reported an injury rate calculation.

Outcome measurement
Three studies compared back injury rates, while three studies evaluated MSK injuries. The majority of studies did not specify the type of injury, as 10 studies just compared "injuries."

Table 14a: Factors that effect generalizability and quality of administrative data outcomes from partial data extraction (n=17)
[refer to key at end of table for abbreviations]

Author, year	Inclusion criteria+	Exclusion criteria+	Person-hours description	Turnover (N or %)	Reinjury addressed	Injury rate calculation	Outcome type	Statistical methods
Multi-Component Patient Handling*								
Nelson, 2005	Sample: Workers in inpatient hospital units with a high proportion of consenting dependent patients (Site); Case: only patient handling MSK injury (Ind)	Sample: staff not participating in intervention; Case: non-patient handling MSK injury (Ind)	estimate of total # of unit hrs. worked	NP	NP	# of reported injuries/estimate of total # of hours worked on unit (reported per 100 workers per year)	MSK injuries from injury log, LWD/injury, and restricted WD/injury	Poisson regression, Wilcoxon signed-rank test
Ronald, 2002	Sample: Hospital extended care unit workers (Unit)	Case: non-MSK injury (Ind)	per 100,000 worked hours at unit level	NP	multiple listings per individual therefore totals may exceed 100%	number of MSK injuries per 100,000 worked hours	MSK injuries	Poisson regression

Table 14a: Factors that effect generalizability and quality of administrative data outcomes from partial data extraction – continued

Author, year	Inclusion criteria+	Exclusion criteria+	Person-hours description	Turnover (N or %)	Reinjury addressed	Injury rate calculation	Outcome type	Statistical methods
Patient Handling Intervention								
Best, 1997	Sample: nursing home employees (Site), nursing and allied health staff (Ind)	NP	NP	NP	NP	NP	NP	Unmatched chi-square
Back School								
Lynch, 2000	Sample: all hospital employees (Site)	NP	NP	NP	NP	NP	Lost-time or restricted duty injury	Unmatched t-tests
Sobaszek, 2001	Sample: report of chronic low back pain, with or without chronic radiculalgia, for which other symptomatic etiologies have been ruled out (Ind)	Sample: acute back pain; medical treatment not properly followed previously (Ind)	NP	NP	NP	NP	NP	Chi-square test

Table 14a: Factors that effect generalizability and quality of administrative data outcomes from partial data extraction – continued

Author, year	Inclusion criteria+	Exclusion criteria+	Person-hours description	Turnover (N or %)	Reinjury addressed	Injury rate calculation	Outcome type	Statistical methods
Lift Team								
Charney, 1997	Sample: agreed to requirements (Site); Case: lost-time or WC for upper or lower trunk injury (Ind)	Case: neck or shoulder injuries (Ind)	Production hours: actual hours worked by unit	NP	NP	Total back injury X 200,000 /dept production hours (nursing)	Back injury	No statistical tests
Lift Team & Patient Handling & Equipment Training								
Guthrie, 2004	Sample: work in orthopedic or neurology units (Ind)	NP	NP	NP	NP	NP	Injury	No statistical tests

Table 14a: Factors that effect generalizability and quality of administrative data outcomes from partial data extraction – continued

Author, year	Inclusion criteria[+]	Exclusion criteria[+]	Person-hours description	Turnover (N or %)	Reinjury addressed	Injury rate calculation	Outcome type	Statistical methods
Participatory Ergonomics Team								
Carrivick, 2002	Sample: hospital cleaners or orderlies (Ind)	NP	Number of hours worked by individuals (excluding leave or OT)	NA	WC costs and lost-time duration of reinjury added on to the original injury	NP	Injury rate, lost-time rate, lost-time duration and WC rate	Generalized linear mixed models (GLMM)
Evanoff, 1999	Sample: hospital orderlies (Ind)	NP	Productive hours worked for the unit (I_1) and whole hospital (C_1)	65 of 99 original group left	NP	Injuries or cost/ 100 FTE	Injury rate, reportable rate, LT and LWD	Relative risks and 95% confidence intervals
Equipment & Patient Handling Training								
Garg, 1992	Sample: nursing care facility employees on intense patient care units (Unit)	NP	Work-hours calculated for each unit	NP	NP	Number of injuries/ 200,000 work hours	Injury rate, LWD rate, and RWD rate	No statistical tests

Table 14a: Factors that effect generalizability and quality of administrative data outcomes from partial data extraction – continued

Author, year	Inclusion criteria[+]	Exclusion criteria[+]	Person-hours description	Turnover (N or %)	Reinjury addressed	Injury rate calculation	Outcome type	Statistical methods
Equipment & Equipment Training								
Evanoff, 2003	Sample: units with high injury rate (Unit) Case: MSK injury to nursing personnel (Ind)	Case: body substance or chemical exposure, fall or contusion injury (Ind)	Productive hours worked from each unit	NP	NP	Injury rate/100 FTE	Recordable injury rate, LWD rate	Relative risks and 95% confidence intervals
Tiesman, 2003	Sample: nursing staff (Ind) Case: injury while moving, handling or performing activities of daily living for a patient (Ind)	Case: injuries while not moving a patient (Ind)	Hours worked at unit level	NP	NP	Injuries/ hours worked on the unit (reported per 100,000 worked hours)	Recordable injury rate, LWD rate, RWD rate	Z-scores
Ergonomic Devices and Consultation								
Fujishiro, 2005	Case: MSK disorder on intervention units (Unit)	NP	Reported by administrators at unit level	NP	NP	MSK disorder/ employee hours worked X 200,000	MSK disorder rate	Wilcoxon signed rank test of median rate ratio

Table 14a: Factors that effect generalizability and quality of administrative data outcomes from partial data extraction – continued

Author, year	Inclusion criteria+	Exclusion criteria+	Person-hours description	Turnover (N or %)	Reinjury addressed	Injury rate calculation	Outcome type	Statistical methods
Integrated Disability Management Program								
Davis, 2004	Sample: MSK injured hospital employees (Ind)	NP	Person-years of productive hours by job title at the unit level ; took out sick leave and vacation	NP	NP	NP	Lost-time rate, LWD	Poisson & Cox Regression; Kaplan-Meier curves
Ryden, 1988	Sample: hospital employees (Site)	NP	NP	NP	NP	NP	All injury and back injury	Chi-squared test of trend
Yassi, 1995	Sample: sustain a compensable soft-tissue back injury (Ind) registered nurse (Ind);	Sample: Planned departure from worksite (Ind); pregnancy (Ind); previously identified concomitant medical or chiropractic intervention (Ind)	Unclear	NP	NP	# injuries/total hours paid X 100,000 hours paid	Back injury rate and lost-time injury rate	Chi-square test of rates, student t-test

Table 14a: Factors that effect generalizability and quality of administrative data outcomes from partial data extraction – continued

Author, year	Inclusion criteria+	Exclusion criteria+	Person-hours description	Turnover (N or %)	Reinjury addressed	Injury rate calculation	Outcome type	Statistical methods
Pre-employment Screen & RTW Policy								
Nassau, 1999	Sample: hospital employees (Site)	Sample: no back sprain or strain (Ind)	Total hours worked by the worksite	9-12% turnover each year of study	NP	(Injuries per year/total employees screened and unscreened X total hours worked) X 100 FTEs working 40 hours	Recordable rate and LWD rate	Wilcoxon rank sum test

* Multi-Component Patient Handling - an intervention that included three components: policy change, equipment purchase and training on equipment usage & patient handling

+(Ind) = Individual Level, (Unit) - Unit Level, (Site) = Site Level

LWD = Lost work days

RWD = Restricted work day

I = Intervention group

C = Control group

MSK = Musculoskeletal

MIPP = MSK Injury Prevention Program

NP = Not provided

WC = Workers' Compensation

RTW = Return to work

Information presented in Table 14b
Table 14b describes the inclusion and exclusion criteria for the studies using self-reported MSK outcomes or clinical information to assess the intervention's effects.

Inclusion criteria
All 12 studies that used self-reported MSK questionnaire data listed an inclusion criterion describing the setting (site or unit description). None described MSK cases that they included. Two studies included participants based on gender (e.g. Landstad 2001); five included participants by job title (e.g. Ronald 2002); and one included individuals based on age (Alexandre 2001). Two studies described inclusion criteria that would identify them as a secondary prevention intervention (i.e. the study included participants with identified MSK symptoms or disorders, Alexandre 2001 and Sobaszek 2001).

Exclusion criteria
Only five studies using self-reported MSK outcomes described exclusion criteria. Four excluded workers based on confounding health outcomes (e.g. pregnancy, Skargren 1996). Only one excluded a type of MSK health outcome as a "case." Two studies excluded injured or symptomatic employees that had taken sick-leave to create a symptom-free population for a primary intervention study. Most (n=8) were both primary and secondary prevention interventions.

Outcome measurement
Heterogeneity exists even within the type of self-reported MSK outcome measurement. Three studies used self-reports of back pain only, while four studies discuss multiple body regions for pain.

Statistical methods
Seven studies used univariate statistical tests such as chi-square or t-test. Two studies used multivariate methods such as Poisson or Cox Proportional regression. One study calculated a 95% confidence interval and two studies used a combination of univariate, multivariate or confidence intervals. Four studies did not use statistical tests to compare outcomes.

Table 14b: Factors that effect generalizability and quality of self-reported data outcomes from partial data extraction (n=12)
[refer to key at end of table for abbreviations]

Author, year	Inclusion criteria+	Exclusion criteria+	Outcome description	Statistical methods
Multi-Component Patient Handling*				
Ronald, 2002	Sample: hospital extended care unit workers (Unit)	Case: non-MSK injury (Ind)	Self-report: recent injury and pain	Poisson regression
Exercise Training				
Skargren, 1996	Sample: geriatric ward nurses or nurses aides (Ind)	Sample: pregnancy, cardiac problems, exercise asthma, longer period of sick leave due to MSK problems, subjects who knew they would change jobs (Ind)	Self-report: MSK symptoms	Paired t-test
Patient Handling Training				
Best, 1997	Sample: nursing home employees (Site), nursing and allied health staff (Ind)	NP	Self-report: back pain present (Y/N)	Unmatched chi-square test
Fanello, 1999	Sample: non-clerical hospital employee (Ind)	NP	Self-report: neck pain, shoulder pain (uni- & bilateral), back pain and lower back pain	Matched chi-square test
Peterson, 2004	Sample: nursing students (Ind)	NP	Self-reported: pain or discomfort	ANOVA

Table 14b: Factors that effect generalizability and quality of self-reported data outcomes from partial data extraction – continued

Author, year	Inclusion criteria[+]	Exclusion criteria[+]	Outcome description	Statistical methods
Exercise & Patient Handling Training				
Alexandre, 2001	Sample: age < 50 years (Ind); shift-work in selected areas (Ind); self-reporting of back pain (Ind); interest in participating (Ind)	Sample: severe spinal disorder with medical restriction to exercising (Ind) or history of spinal pain (Ind)	Self-reported: pain	McNemar test, Wilcoxon rank sum for matched pairs
Back School				
Lynch, 2000	Sample: all hospital employees (Site)	NP	Self-report: back pain days/week	Unmatched t-tests
Sobaszek, 2001	Sample: report of chronic low-back pain, with or without chronic radiculalgia, for which other symptomatic etiologies have been ruled out (Ind)	Sample: acute back pain; medical treatment not properly followed previously (Ind)	Self-report: Health-care use and perception of back pain	Chi-square test
Cognitive Behavioural Training				
Landstad, 2001	Sample: female (Ind); minor or no existing health problems (Ind)	Sample: taking long periods of sick leave (Ind)	Self-reported: pain Clinical: MSK disorder diagnoses	Wilcoxon matched-pairs signed-rank test, Mann-Whitney paired test, chi-square test, GLM ANOVA
Exercise & Patient Handling & Stress Management				
Lagerstrom, 1997	Sample: county hospital nursing personnel in medical, surgical or geriatric wards (Unit)	NP	Self-reported: MSK symptoms (neck, shoulder, or low back)	Prevalence ratios and 95% CI, Wilcoxon signed-rank test
Participatory Ergonomics Team				
Evanoff, 1999	Sample: hospital orderlies (Ind)	NP	Self-report: presence/absence of MSK pain at different sites and pain severity	Relative risks and 95% CI

Table 14b: Factors that effect generalizability and quality of self-reported data outcomes from partial data extraction – continued

Author, year	Inclusion criteria+	Exclusion criteria+	Outcome description	Statistical methods
		Shower Trolley		
Nevala, 2004	Sample: female nurses (Ind)	NP	Self-report: VAS (MSK strain in neck-shoulders, arms, back, & legs)	Student's t-test

* Multi-Component Patient Handling -an intervention that included three components: policy change, equipment purchase and training on equipment usage & patient handling

+(Ind) = Individual Level, (Unit) - Unit Level, (Site) = Site Level

VAS= Visual analog scale

NP= Not provided

MSK = Musculoskeletal

MIPP = MSK Injury Prevention Program

Effects by outcome source

Table 15 presents effects by outcome type. The studies on exercise training, patient handling training, back school, etc use self-reported MSK outcomes. Studies that included equipment interventions generally used administrative outcomes. Even within the same study using multiple administrative outcomes, mixed findings were reported (e.g. Nelson 2005).

Table 15: Effects summary by type of outcome measurement for partial data extraction (PDE) studies (n=24)

Author, Year	Administrative			Self-report	Clinical	QA
	Injury	LWD	WC			
Multi-Component Patient Handling						
Nelson, 2005	+	∅				M
Ronald, 2002	+					M
Exercise Training						
Skargren, 1996				+		M
Patient Handling Training						
Best, 1997				+/∅		M
Fanello, 1999				∅		M
Peterson, 2004				∅		M
Exercise & Patient Handling Training						
Alexandre, 2001				+		M
Back School						
Lynch, 2000				+/∅		L
Sobaszek, 2001				+		M
Cognitive Behavioural						
Landstad, 2001				∅	+	M
Exercise & Patient Handling & Stress Management.						
Lagerstrom, 1997				∅		M
Lift Team						
Charney, 1997	+	+				L
Lift Team & Patient Handling & Equipment Training						
Guthrie, 2004	+					L
Participatory Ergonomics						
Carrivick, 2002	+	+	+			M
Evanoff, 1999	+	+	+	+		M
Equipment & Patient Handling Training						
Garg, 1992	+					M
Equipment & Equipment Training						
Evanoff, 2003	+	+				M
Tiesman, 2003	∅	+				M
Ergonomic Devices & Consultation						
Fujishiro, 2005	+					M
Shower Trolley						
Nevala, 2004				+		M
Integrated Disability Management. Program						
Davis, 2004				+		M
Ryden, 1998	+					L
Yassi, 1995	+	+				M
Pre-employment Screen & RTW Policy						
Nassau, 1999	∅					M

*Possible Values: + (Positive Effect), ∅ (No Effect) or +/∅ (both Positive and No Effect found). Blank spaces indicate the outcome measurement was not used.

LWD = Lost work days
WC = Workers' compensation
MIPP = MSK injury prevention program
M = medium
L = low

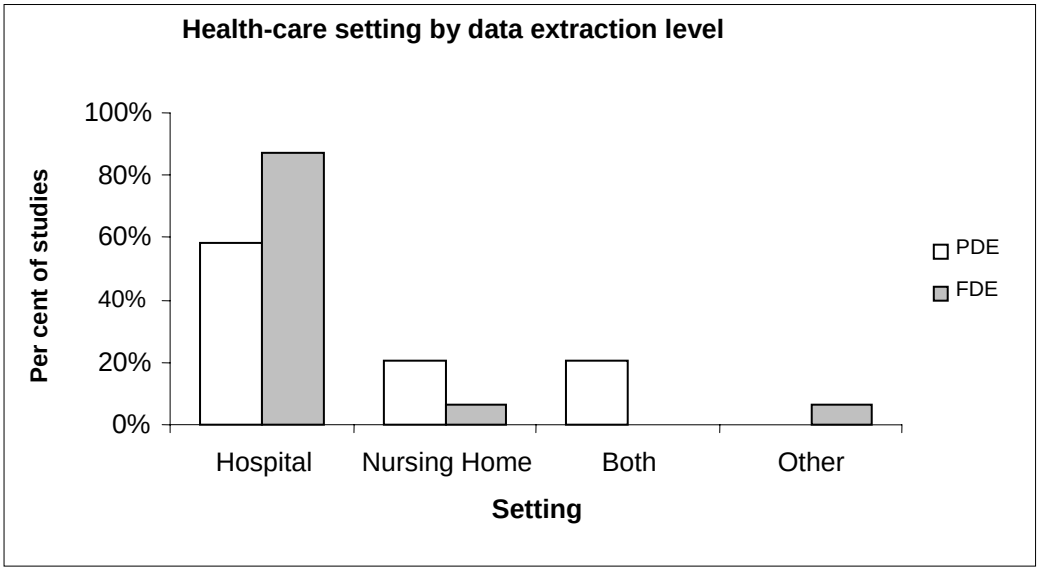
3.6 Comparison of full and partial data extraction (FDE and PDE)

Comparing data extraction levels helps us understand whether certain characteristics are more predominant in higher quality studies. Additionally comparing studies can help direct future research to areas that have been under-examined.

Health-care setting

The original review question included only nursing homes or long-term care settings. Both FDE and PDE had few to no studies based in a nursing home. Most studies were in a hospital only or in multiple settings.

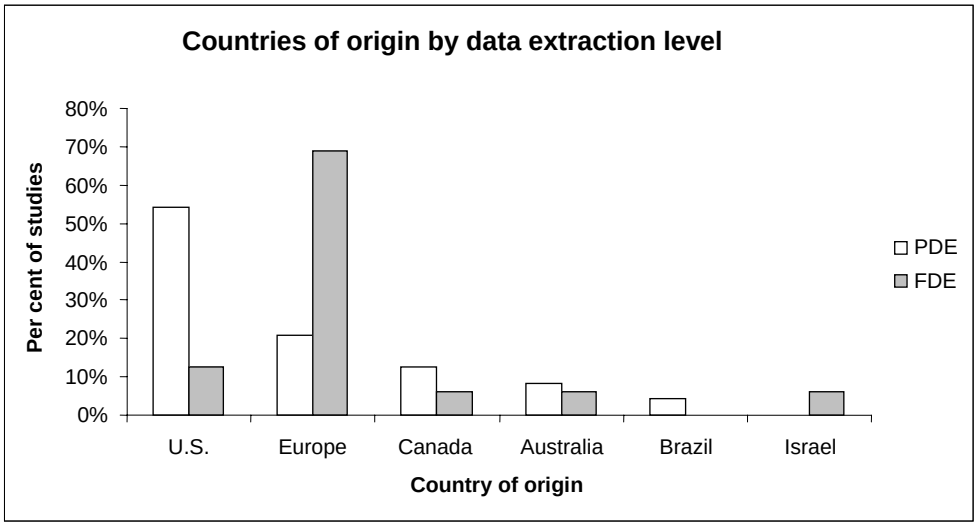
Figure 3: Comparison of health-care setting between studies in PDE and FDE



Geographic setting

Health-care differs across geographic region. Therefore the location where interventions took place was relevant. Interestingly most of the studies in PDE were completed in the U.S., while studies that were included in evidence synthesis took place mostly in Europe.

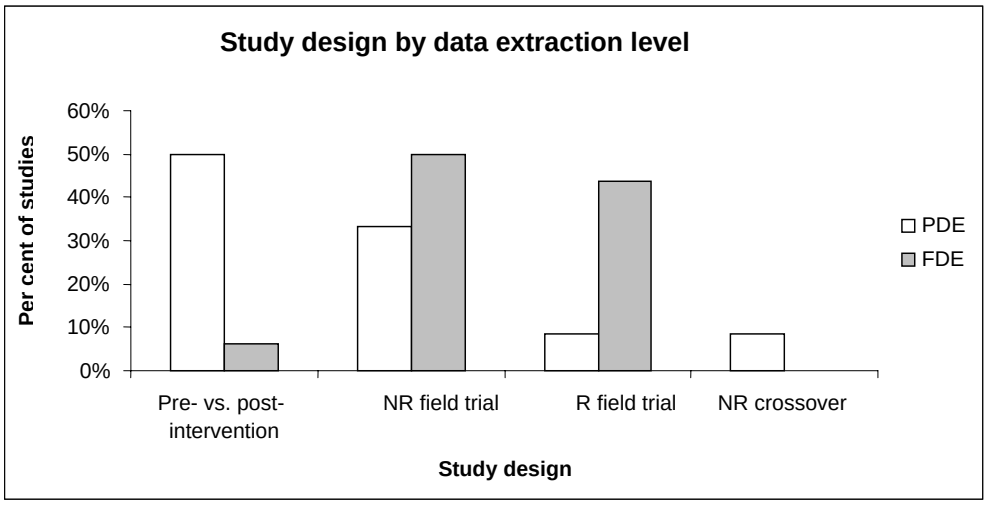
Figure 4: Comparison of geographic setting between studies in PDE and FDE



Study design

Most studies in FDE had control groups, as seen by the fact that more than 90% of the studies were field trials. However, the majority of studies in FDE were not randomized. In PDE, most studies were pre- vs. post-interventions with no control groups, but some studies were randomized (R) and non-randomized (NR) field trials.

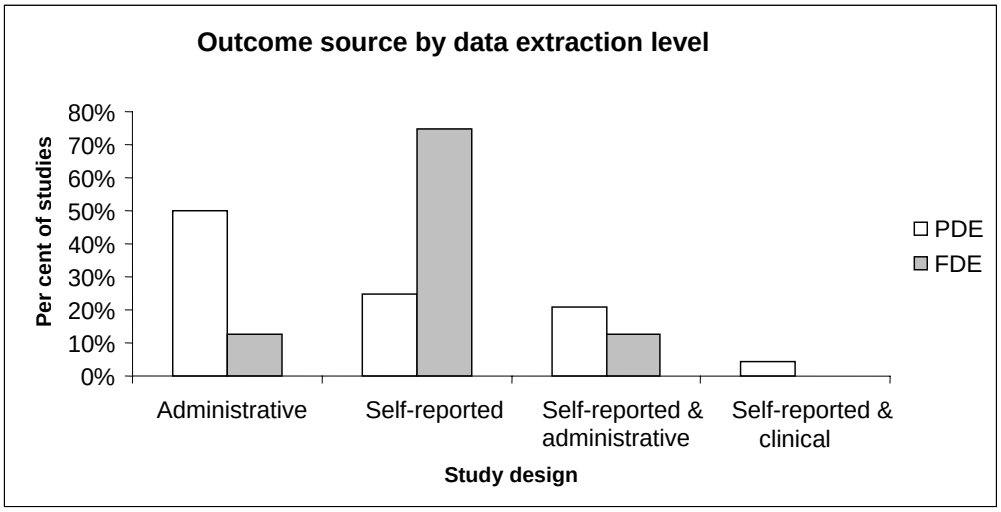
Figure 5: Comparison of study designs used between studies in PDE and FDE



MSK health outcome source

Since the present review is the first to include a wide range of outcome sources, it is important to examine the proportion of studies using each type of outcome. More than 70% of studies in FDE used self-reported questionnaire data while 25% included administrative data sources. A majority of the PDE studies used administrative outcomes. The only study to include clinical measurements was in PDE.

Figure 6: Comparison of MSK outcome source used between studies in PDE and FDE

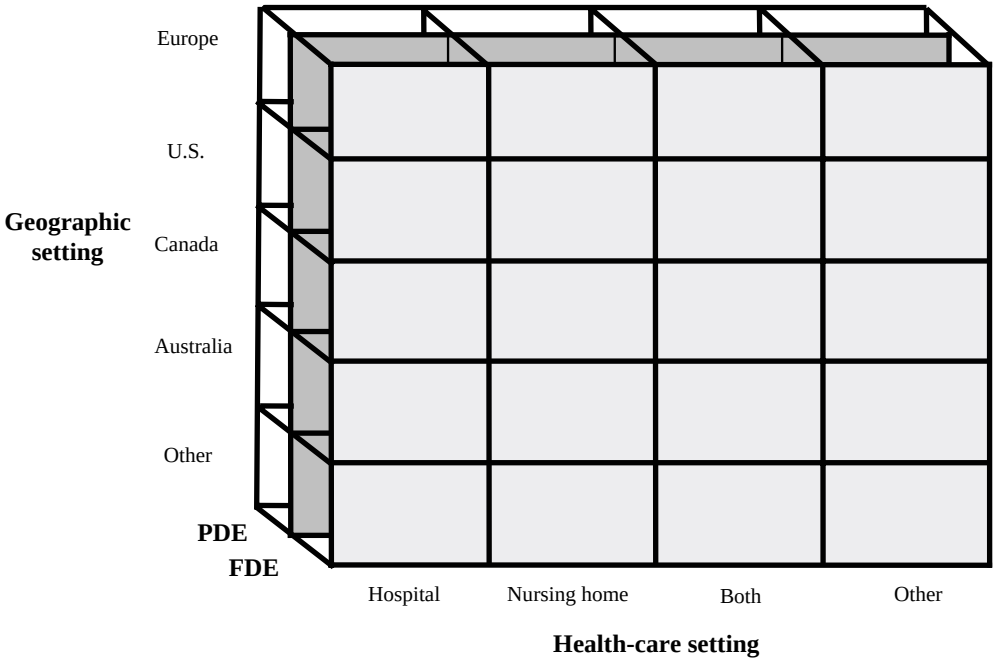


Comparison by more than one characteristic

Setting

Following principles proposed by Côté (2001), we created figures to examine the generalizability of the literature by looking at geographic and health-care setting within partial data extraction (PDE) versus full data extraction (FDE).

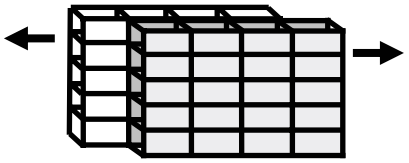
Figure 7: Proposed setting comparison between studies in PDE and FDE



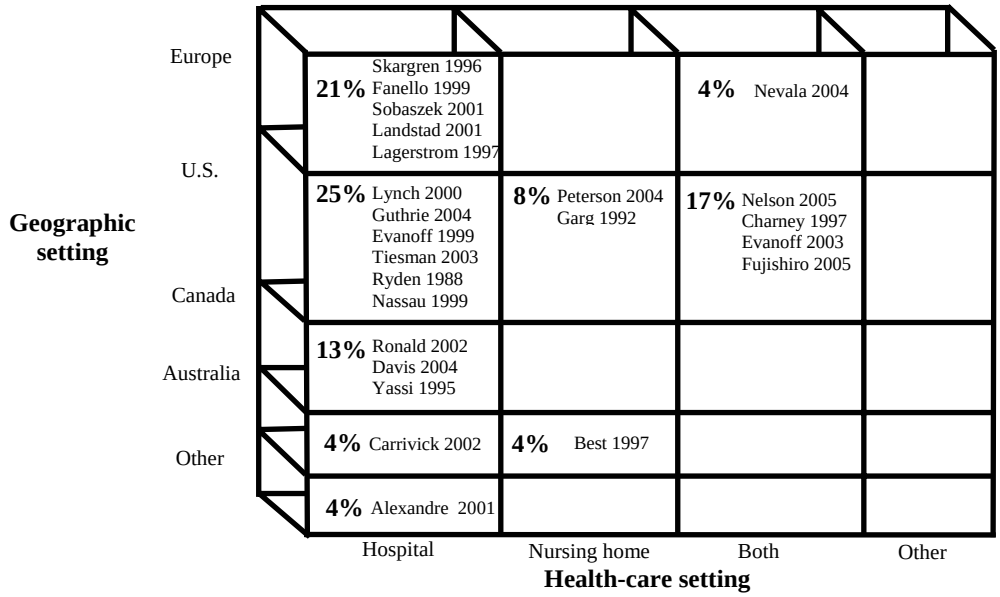
In Figure 8, we split the two levels to see each level more fully and also listed the studies for ease of reference.

Most of the studies (56%) in FDE were in hospitals in Europe, but a significant number of hospital studies in PDE also occurred in Europe (21%). Although only one FDE study was set in a nursing home, more than 20% of the studies in PDE were set in nursing homes in the U.S.

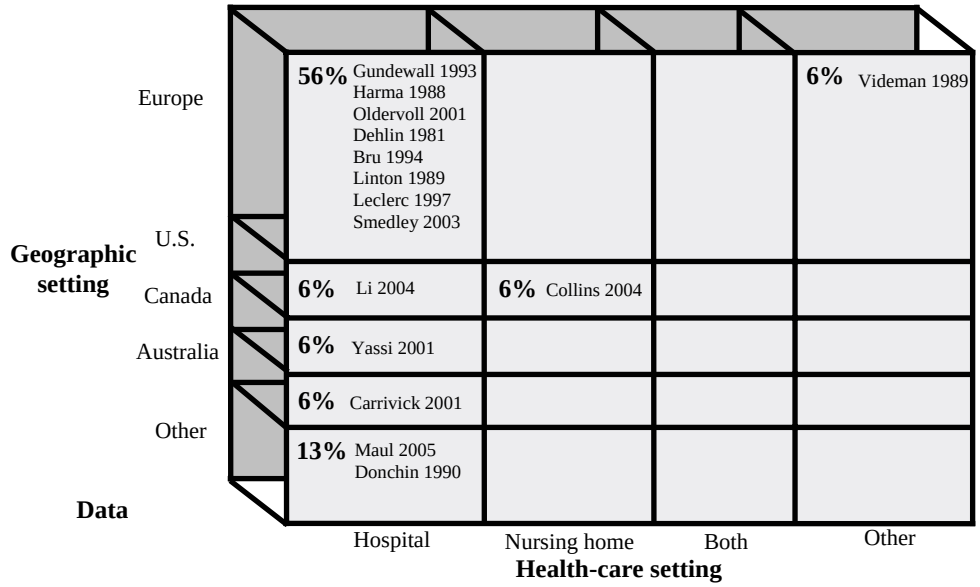
Figure 8: Setting comparison between PDE and FDE studies



Partial Data Extraction



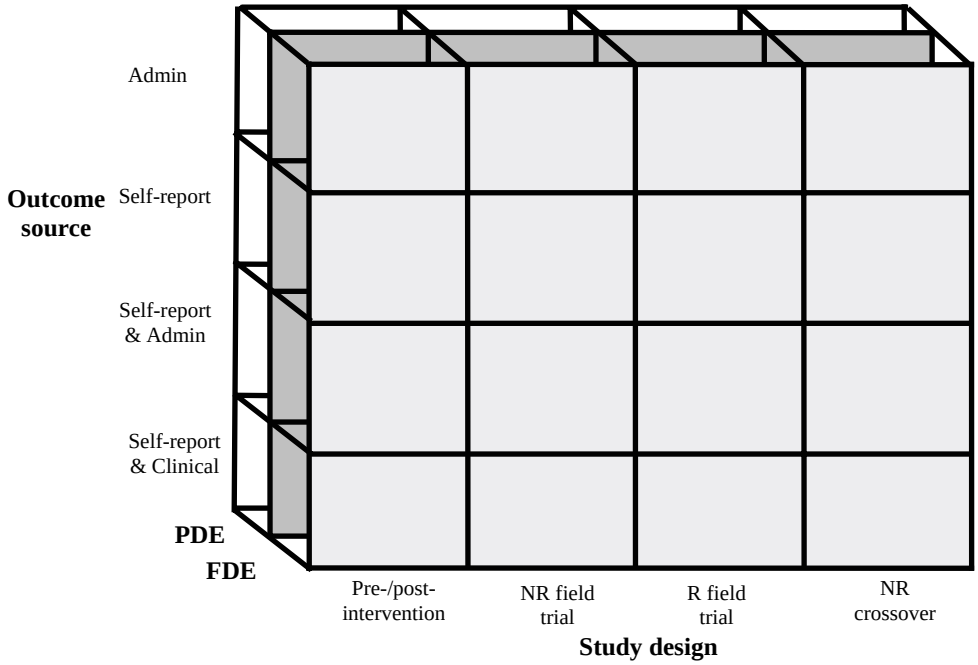
Full Data Extraction



Study design

The following figures were created to examine the study design features in the literature by looking at they type of outcome measurement and study design used between PDE and FDE.

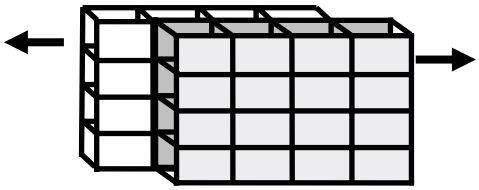
Figure 9: Proposed study design comparison between studies in PDE and FDE



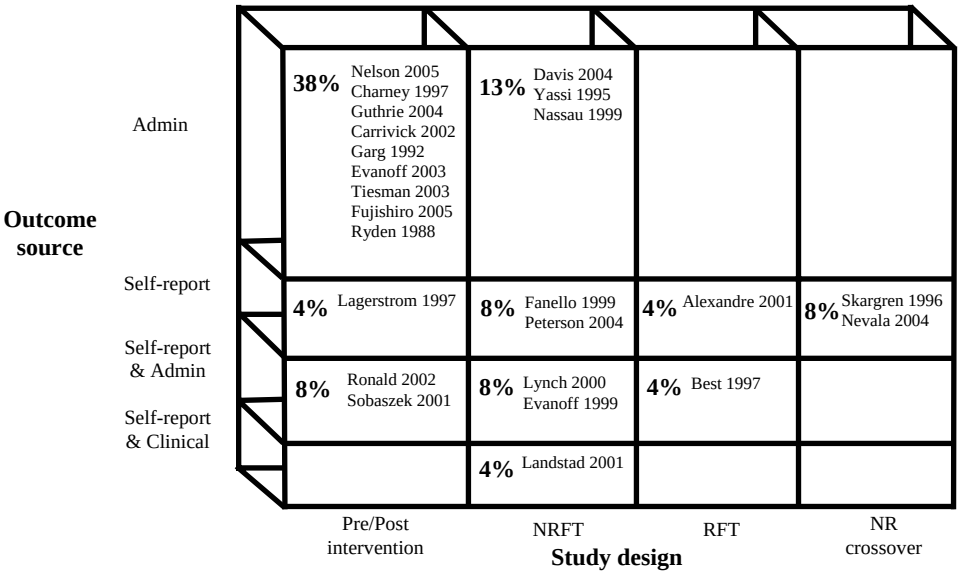
In Figure 10, we split the two levels to see each level more fully and also listed the studies for ease of reference.

In PDE, most of the studies (38%) used administrative data sources and a pre- vs. post-intervention. Most of the studies in FDE used self-reported outcomes but were evenly distributed between randomized and non-randomized field trials.

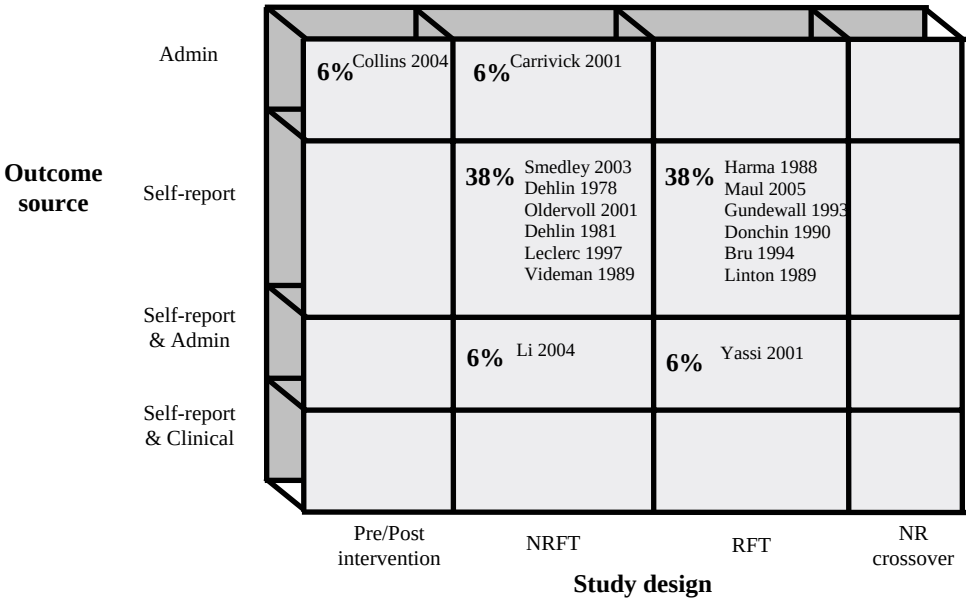
Figure 10: Comparison of study figures between PDE and FDE studies



Partial Data Extraction



Full Data Extraction



4.0 Conclusions

Our systematic review used a standard approach to review the literature, synthesize results and answer the question: "Do occupational safety and health interventions in health-care settings have an effect on MSK health status?"

We found that the literature on this topic was heterogeneous in terms of the types of interventions, study design quality and outcomes measured.

From an initial pool of more than 8,000 articles, we identified 40 relevant studies in which the methodological quality was ranked as high (two studies), medium-high (14 studies), medium (20 studies) and limited (four studies). We did a full data extraction from the 16 medium-high and high quality studies. Since two studies did not include statistical comparisons of MSK outcomes, they were excluded from evidence synthesis, leaving 14 studies. A previous review on patient handling interventions has reported similar quality rankings and findings (10).

Based on our evidence criteria for data synthesis (Table 6), at least three high quality studies with consistent findings were needed to determine the existence of "strong evidence." Therefore with only two high quality studies we had no opportunities to make any statements about a strong level of evidence.

Across all studies we found a **moderate level of evidence** for the effect of OHS interventions on MSK health in health-care settings. This means that a majority of high and medium-high studies found positive effects on MSK outcomes. Additionally, we found no evidence that any injury prevention intervention had a negative or deleterious effect.

Our review included studies in different health-care settings including long-term care organizations and hospitals. As previously mentioned, differences in exposure might exist between these settings due to characteristics including organization, economics and patient population. We were unable to stratify effects between hospitals and long-term care settings because only one study in evidence synthesis was set in nursing homes.

A **moderate level of evidence** was found for a POSITIVE intervention effect on MSK health for:

- **multi-component patient handling interventions**
- **physical exercise interventions.**

There was **insufficient evidence** to determine an effect on MSK outcomes in health-care settings for any of the following interventions on their own,

because only one study per intervention was found: **patient handling training; back school; cognitive behavioural interventions; exercise & patient handling & stress management; intensive off-site MSK injury prevention programs; participatory ergonomics; equipment & patient handling training; equipment & equipment training; broad-based MSK injury prevention program; ergonomic devices & consultation; shower trolleys; integrated disability management; and pre-employment screen & RTW policy.**

In this context, the term "insufficient" refers to the low number of studies, not to the quality of the interventions. In fact, the review team supports more high quality studies of many of these interventions.

All relevant studies had either full or partial data extraction, which enabled a description and comparison across data extraction levels. FDE studies were mainly set in hospitals; mostly originated from Europe; usually used self-reported MSK symptoms; and were mostly non-randomized field trials. In comparison, PDE studies were mostly pre- vs. post-interventions using administrative outcomes in hospitals in the U.S.

4.1 Future work

The review team believes that it is important for further research to be of high methodological quality, to move the evidence base forward and shift the level of evidence from moderate to strong (see Table 3 for quality criteria).

Researchers, funders, employers and organized labour should attend to the effects and study quality reported in Table 11 as one way to gauge the level of interest and investment in further research. Interventions that include equipment such as mechanical lifts and patient transfer aides are of particular importance. The science on the effectiveness of equipment interventions is not strong, yet policies have been developed and many health-care organizations are purchasing the equipment. More high quality studies are needed to better guide policy and practice.

The high and medium-high quality studies reviewed shared certain common threads, regardless of the intervention or outcome. All had concurrent comparison groups or a statistical comparison. Each study was designed to limit threats to internal and external validity.

However, few studies in evidence synthesis used similar MSK outcomes. This challenged the review team when trying to integrate findings, and made it impossible to calculate pooled effect sizes for interventions.

One potential action that funding, policy or research organizations could take would be to convene a conference or series of position papers

advocating standards for injury prevention intervention research in health-care settings.

4.2 Issues in the conduct of systematic reviews

Three issues emerged that warrant further discussion. In raising these issues, we hope to inspire future researchers conducting systematic reviews to solve problems creatively.

Intervention specification

Until there are more intervention studies and more consensus on the types of interventions to be conducted, intervention specification will be an issue in systematic reviews of OHS interventions. There are two issues with intervention specification. The first is our own empirical development of intervention categories. In the category of multi-component interventions, the three interventions had different types of policies, equipment and trainings. The design intent of each in terms of affecting MSK health were not specified, so it is unclear whether the differences between the interventions are as important as the similarities in affecting MSK outcomes. The same case can be made for the physical exercise interventions.

Second, interventions were combined with other interventions preventing the identification of specific effects. A majority (four out of five) of the interventions that included equipment found positive effects on MSK outcomes. However, they could not be grouped together in an intervention category since they included additional intervention characteristics (e.g. patient handling training or policy changes).

Opportunities and challenges using administrative outcomes

Administrative outcomes provide many advantages to researchers. Stakeholders, including policy-makers, rely on these outcomes for regulatory reporting. Outcomes such as workers' compensation claims can be used to estimate the economic burden associated with workplace injuries. Therefore stakeholders support using administrative outcomes to evaluate workplace interventions. Some consider these outcomes more clinically relevant than self-reported low-back pain. Finally, administrative outcomes can be relatively easy to obtain since workplaces already collect the information as part of standard procedures.

Studies that used administrative outcomes in this review often created rates for a unit (or worksite) using unit exposure time (see fixed vs. open study populations in the methods section). The population contributing to these rates changed over time, since employees joined and left the worksite. The review team created the term "open population" to describe this study design and distinguish it from dynamic cohorts or fixed populations. Both of the latter two designs have information on each individual, making it possible to attribute exposure time to each individual.

Open population studies challenged the review team during methodological quality appraisal. Previous systematic reviews have excluded administrative outcomes (e.g. 21) noting the vulnerabilities of reporting systems to a number of biases. Therefore, many quality appraisal criteria used in previous reviews are based on studies using fixed populations, not open populations. Yet, the basic validity issues (i.e. internal, external, statistical, and construct) described in Methods continue to be relevant and some study designs are better suited to address validity issues than others. As studies using administrative outcomes are incorporated into systematic reviews, attention needs to be drawn to the potential unique methodological evaluation challenges. By including a wider spectrum of studies, we had to make some decisions to maintain fair and reasonable methodological criteria.

The review team addressed the following challenges with open populations to improve the systematic review process:

- reporting on who is/is not included in injury rates from workers' compensation, regulatory reports or employee hours (QA question 8)
- identifying worksite consent versus individual consent (QA question 10)
- describing who stops contributing to rates (QA questions 13 and 14)
- addressing the consistency of interventions as population changes (QA question 15).

Each of the criteria that were potentially biased towards fixed population studies were discussed by the review team and our resolutions were presented in detail in Appendix D.

The review team considers that the opportunities far outweigh the challenges of using administrative MSK health outcomes in prevention intervention research. In designing, implementing and analyzing high quality studies using administrative data, the review team recommends researchers and stakeholders consider the following.

First, the team recognizes the many challenges of randomization in studies attempting to reduce biomechanical loads in patient handling, but would still encourage a randomized field experiment where possible. Randomization may need to occur at the work unit or worksite level. Second, the review team agreed that the use of a concurrent comparison group is critical to rule out whether an observed effect is the result of physical, administrative or workforce changes unrelated to the intervention. Third, it is essential to collect socio-demographic information on participants. This is particularly crucial when randomization has not occurred since significant differences may exist between the employees in intervention and control groups. Socio-demographic information can be obtained from self-administered questionnaires, personnel files or other employer data resources. Fourth, when a fixed cohort of employees is not feasible, the review team considers

a dynamic cohort the only reasonable option. To follow a dynamic cohort over time requires information on all individuals entering and exiting the participating work units or worksites. This information is essential to track the duration of individual exposure to the intervention, and work hours used in standard calculations of injury statistics. Finally, a unique challenge in studies using administrative outcomes is the potential reporting biases related to factors not controllable in the intervention (e.g. insurance or employer incentives that take the incentive away for employee injury reporting). A unique opportunity is the ability to clearly state different hypotheses (e.g. injury incidence versus injury duration). Given the tremendous amount of information being gathered at the worksite, it is critical for researchers to have a strong working relationship with organized labour and employers.

As Kristensen (2005) observes, "There may be many good reasons for not performing a randomized controlled trial in an occupational setting. But there are no good reasons for ignoring the problems created by not applying such a design" (28).

Statistical improvements to the literature

Two of the 16 studies that were medium-high and high quality did not statistically test the intervention's effect on MSK outcomes (Gundewall 1993, Videman 1989). When there is tremendous heterogeneity in statistical procedures, it challenges the review team in evidence synthesis. Statistics help tell an important part of the story by showing how certain we are that the results are not due to chance.

Researchers could improve confidence in their findings by ensuring that statistical methods use the strengths in study designs. Several studies in evidence synthesis (Carrivick 2001) and partial data extraction (e.g. Fanello 1999) collected information on control groups, but did not statistically compare differences between intervention and control groups. This comparison allows researchers to rule out alternative hypotheses for an effect, such as a change in injury reporting regulations during the study period.

Another set of design strengths that were underutilized by researchers was information on potential covariates (e.g. socio-demographics or exposure time). Almost all of the FDE studies (15/16) collected some covariates or confounders. However, only four studies adjusted for these when testing the interventions' effects. Statistically adjusting for covariates and confounders is one way to attend to issues that a randomized design more directly addresses (seven FDE studies). While adjusting is not a perfect solution, it does increase confidence in a study's findings.

Finally, researchers could improve studies by not diluting study power, which occurs because they complete multiple statistical tests. Multiple

testing increases the likelihood of finding a false positive effect. Authors should begin by stating a primary testable hypothesis, which fewer than 25% of studies did. Yassi (2001) sets a good example by stating a primary intervention effect on MSK injury rates. Over 50% reported intervention effects for multiple outcomes. Of the seven studies that reported multiple outcomes, four used a global test, which adjusted for multiple testing.

4.3 Strengths of conducting a systematic review

The number of studies published in any given field is more than most practitioners or researchers can easily track or synthesize. This is particularly true in the field of injury prevention in which evidence can be found across many different disciplines. Systematic reviews are useful tools to help researchers, health and safety practitioners, employees, employers, and policy-makers remain current with the evidence.

The systematic review process is designed to be transparent and reproducible. By following an explicit process, systematic reviews aim to eliminate bias in the selection and synthesis of evidence. The goal is to produce an objective appraisal that can help practitioners and researchers resolve uncertainty and inform decision-making. Such uncertainty often occurs when original studies and editorials disagree on the conclusions drawn from the evidence for a particular research question.

Another benefit of a systematic review is that it can help identify gaps in the quantity and quality of studies in a particular area. This information can suggest an agenda for further research and evaluation.

4.4 Limitations of this systematic review

We identified studies by searching the peer-reviewed literature in five electronic databases. We also scanned reference lists from selected studies and references suggested by experts. A broader search of the grey literature, conference proceedings and dissertations might have yielded further relevant evidence. The review team believes that most high quality research will be published in peer-reviewed literature, and thus it is not a substantial limitation to leave out the grey literature.

Because of time constraints, the review team was unable to clarify specific questions with the study authors. The review was limited to articles published in the English, Spanish, Swedish or French languages. Articles in other languages were excluded before their relevance could be assessed. It is possible that these articles might have provided pertinent evidence to answer the study question.

4.5 Strengths of this systematic review

The review team included members with varied backgrounds and specializations (i.e. members had expertise in the systematic review process, ergonomics, MSK health and safety, intervention research and epidemiology). This broad expertise enabled the review team to approach the research question from a number of perspectives.

We also contacted content experts to request potentially relevant published articles or articles in press to ensure that we reviewed as much relevant literature as possible.

The review team used a quality control process to assess the early phase of article exclusion. We also randomly paired reviewers at each phase to improve independent assessment by at least two team members. All decisions were made by team consensus.

The partial data extraction allowed the studies that were not included in evidence synthesis to be characterized, described and compared to higher quality studies. Partial data extraction (PDE) was also an opportunity to do a quality check of reviewers' quality appraisal that might otherwise not have happened without the "extra sets of eyes" conducting the data extraction. At least four reviewers saw each article.

A danger would be for readers to give equal value to the information in the partial data extraction as to the evidence in the full data extraction section. The PDE studies have methodological quality limitations and therefore are not included in our evidence synthesis. The information is only presented to provide stakeholders and researchers with a more complete description of the field.

4.4 Next steps

The current review answers a general question about the effectiveness of occupational health and safety interventions in improving MSK health. The review team believes that the systematic review process should continue to develop in several ways when considering the literature:

- it is important to include non-English articles and grey literature in the process
- review teams should continue to develop transparent, peer-reviewed methods to evaluate studies using administrative outcomes, which have generally not been included in systematic reviews
- if necessary, article authors should be contacted to clarify findings in the published studies
- when possible, studies in which between-group comparisons were not made should be re-analyzed to provide evidence that can be included in data synthesis

- in an effort to produce effect sizes, a full data set should be obtained from researchers.

The information from this review should be used to guide future research in health-care injury prevention interventions, and it alerts stakeholders to the current state of the evidence.

5.0 Messages

Prior to making policy and best practice recommendations, the review team felt there should be a stronger level of evidence. Such recommendations require consistent findings from three high quality studies. Our review did not find this level of evidence. Given that we did find a moderate level of evidence for certain interventions, the review team considered it feasible to recommend several "practices to consider."

The first practice to consider is multi-component patient-handling interventions. The intervention components are:

- worksite policy changes (e.g. zero-lift policies)
- the purchase and implementation of new patient handling equipment
- training on the new equipment and on patient handling.

The positive evidence comes from studies using workers' compensation claims and self-reported MSK symptoms, suggesting this intervention affects a range of MSK health endpoints. The presence of strong biomechanical evidence strengthens our support for making this recommendation. However, the study with the most consistent intervention effects had tremendous turnover in the workforce. Furthermore, because the group of three intervention components is bundled, the team cannot advise stakeholders on whether one specific intervention on its own may be as good as the bundle. A patient handling training intervention had no effect, while a patient handling equipment implementation and equipment training improved self-reported MSK health, but showed no effect on workers' compensation claims, lost work days or MSK injuries.

The second practice to consider is exercise training programs, both aerobic and strength building. A further advantage to physical exercise is that it improves general health and reduces the risk of many chronic diseases.

Many of the exercise programs were conducted over a three- to four-month period requiring each subject to participate at least two times per week. This consistency in the frequency and duration of interventions strengthens our confidence in the recommendation. However, all MSK outcome evidence is from self-reports. Although MSK symptom self-reports have been shown to be valid and reliable (29), injury rates assessed by workers' compensation claims, regulatory reporting systems, or clinician assessment are more widely recognized as practical and convincing evidence to stakeholders.

Given the heterogeneity of types of physical exercise interventions (e.g. cardiovascular versus strength), the review team recommends convening an expert group to define key interventions as a prelude to initiating a series of high quality studies.

An important message to all stakeholders is that the current state of the peer-reviewed literature provides limited high quality evidence to support the MSK health benefits of interventions implemented in hospitals, long-term care facilities and other health-care establishments. Given the significance of the MSK injury problem among health-care professionals and the supporting evidence from biomechanical/ergonomic lab studies, the team considers it important to assess whether one or all components of the multi-component patient handling intervention is required in the context of high methodological quality.

Here are some issues to consider to advance the evidence base:

- Researchers should use concurrent control groups as opposed to study designs with simulated controls, statistical controls or cross-over designs. True concurrent controls contribute results that are more robust.
- Field studies should have adequate sample sizes to reduce the risk of mistakenly concluding an intervention has no effect on MSK health, simply because the sample is too small.
- Fixed cohorts should become the norm in studies using administrative, clinical or self-reported MSK outcome data. Open population studies will never achieve the high quality needed for evidence synthesis due to threats to validity.
- For MSK outcomes we recommend studies follow workers between four and 12 months after the intervention is completed. Studies longer than 12 months may run the risk that workers who participated in the intervention are no longer employed.
- Researchers should present outcomes using standard approaches that are common to the reporting requirements demanded of stakeholders when using workers' compensation, injury records or other regulated injury reporting systems.
- Covariates and confounders should be measured and adjusted for using multivariate statistical models. This is especially true when the researchers are unable to randomize workers into either intervention or control groups.

Given the known problems with MSK injuries among health-care workers, we are frustrated that we are not able to make stronger recommendations. The overwhelming message from our review, which we consider an OHS priority, is that more high quality research must be produced. Well-designed studies, including randomized controlled field trials with adequate sample sizes and appropriate MSK outcome measurement, are sorely needed before policy conclusions regarding specific interventions can be made.

6.0 References

(1) BLS. Incidents of non-fatal occupational injuries and illnesses by industry and selected case types, 2000. 2001; 2004(7/30): 13.

(2) Patient Safety Center of Inquiry. Patient care ergonomics resource guide: safe patient handling and movement. Tampa, FL: Veterans Health Administration and Department of Defense, 2001.

(3) Lagerstrom M, Hansson T, Hagberg M. Work-Related Low Back Problems in Nursing. Scandinavian Journal of Work, Environment and Health. 1998; 24(6): 449-464.

(4) Nelson A, Fragala G, Menzel N. Myths and Facts about Back Injuries in Nursing, American Journal of Nursing. 2003; 103 (2): 32-40.

(5) NRC and IOM Report. Musculoskeletal Disorders and the Workplace. National Research Council and Institute of Medicine. Washington, DC: National Academy Press, 2001.

(6) Videman T. Ojajarvi A, Riihimaki H, Troup J. Low back pain among nurses: a follow-up beginning at entry to the nursing school. Spine. 2005; 30: 2334-2341.

(7) Wilkins K, Shields M. 2005 National Survey of the Work and Health of Nurses. Ottawa, ON: Statistics Canada, 2006.

(8) Garg A. Long-Term Effectiveness of "Zero-Lift Program" in Seven Nursing Homes and One Hospital, 1999: Contract No. U60/CCU512089-02. http://www2.cdc.gov/nioshtic-2/Nioshtic2.htm

(9) National Institute for Occupational Safety and Health (NIOSH). Musculoskeletal disorders and workplace factors: A critical review of epidemiologic evidence for work-related musculoskeletal disorders of the neck, upper extremity, and low back. Washington DC: Department of Human and Health Services (NIOSH); Publication No. 97-141, 1997.

(10) Hignett S. Intervention strategies to reduce musculoskeletal injuries associated with handling patients: a systematic review. Occupational and Environmental Medicine. 2003; 60(9): e6.

(11) Hudson, M. Texas passes first law for safe patient handling in America: landmark legislation protects health care workers and patients from injury related to manual patient lifting. Journal of Long-Term Effects of Medical Implants. 2005; 15(5): 559.

(12) Bensley L, Nelson N, Kaufman J, Silverstein B, Kalat J, et al. Injuries due to assaults on psychiatric hospital employees in Washington State. American Journal of Industrial Medicine. 1997; 31: 92-99.

(13) Slavin RE. Best evidence synthesis: an intelligent alternative to meta-analysis. Journal of Clinical Epidemiology. 1995; 48(1): 9-18.

(14) Reeves S, Koppel I, Barr H, Freeth D, Hammick M. Twelve tips for undertaking a systematic review. Medical Teacher. 2002; 24(4): 358-63.

(15) Cochrane Manual. www.cochrane.org/admin/manual.htm 16 November 2005

(16) Côté P, Cassidy JD, Carroll L, Frank JW, Bombardier C. A systematic review of the prognosis of acute whiplash and a new conceptual framework to synthesize the literature. Spine. 2001; 26(19):E445-E458.

(17) SRS –Systematic Review Software. Trial Stat. www.trialstat.com

(18) Checkoway H, Pearce N, Kriebel D. Research methods in occupational epidemiology. 2nd ed. New York, NY: Oxford University Press, 2004.

(19) Heaney CA, Goetzel RZ. A review of health-related outcomes of multicomponent worksite health promotion programs. American Journal of Health Promotion. 1997; 11: 290–308.

(20) AHRQ Guidelines, http://www.ahrq.gov/

(21) Brewer S, Van Eerd D, Amick BC 3rd, Irvin E, Daum KM, Gerr F, Moore JS, Cullen K, Rempel D. Workplace interventions to prevent musculoskeletal and visual symptoms and disorders among computer users: a systematic review. Journal of Occupational Rehabilitation. 2006; 16 (3): 325-58.

(22) Shadish W, Cook T, Campbell D. Experimental and quasi-experimental designs. Boston, MA: Houghton Mifflin Co, 2001.

(23) Franche R-L, Cullen K, Clarke J, MacEachen E, Frank J, Sinclair S, and the Workplace-based return-to-work intervention literature review group. Workplace-based return-to-work interventions: A systematic review of the quantitative and qualitative literature. Toronto: Institute for Work & Health, 2004.

(24) Smith B. Chronic pain expert advisory panel. Report of the Chronic Pain Expert Advisory Panel. Toronto, ON: Workplace Safety & Insurance Board, 2000.

(25) Cooper H, Hedges LV. The Handbook of Research Synthesis. New York, NY: Russell Sage Foundation, 1994.

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

(27) Tompa E, Trevithick S, McLeod C. A systematic review of the prevention incentives of insurance and regulatory mechanisms for occupational health and safety. Toronto, ON: Institute for Work & Health, 2004.

(28) Kristensen TS. Intervention studies in occupational epidemiology. Occupational and Environmental Medicine. 2005; 62(3):205-210.

(29) Gerr F, Marcus M, Ensor C, Kleinbaum D, Cohen S, Edwards A, Gentry E, Ortiz DJ, Monteilh C. A prospective study of computer users: I. Study design and incidence of musculoskeletal symptoms and disorders. American Journal of Industrial Medicine. 2002; 41: 221-235.

7.0 References used in full data extraction and evidence synthesis

(1) Bru E, Mykletun RJ, Berge WT, Svebak S. Effects of different psychological interventions on neck, shoulder and low back pain in female hospital staff. Psychology & Health. 1994; 9(5): 371-382.

(2) Carrivick PJ, Lee AH, Yau KK. Consultative team to assess manual handling and reduce the risk of occupational injury. Occupational and Environmental Medicine. 2001; 58 (5): 339-344.

(3) Collins JW, Wolf L, Bell J, Evanoff B. An evaluation of a "best practices" musculoskeletal injury prevention program in nursing homes. Injury Prevention. 2004; 10 (4): 206-211.

(4) Dehlin O, Berg S, Andersson GB, Grimby G. Effect of physical training and ergonomic counselling on the psychological perception of work and on the subjective assessment of low-back insufficiency. Scandinavian Journal of Rehabilitation Medicine. 1981; 13(1): 1-9.

(5) Dehlin O, Berg S, Hedenrud B, Andersson G, Grimby G. Muscle training, psychological perception of work and low-back symptoms in nursing aides. The effect of trunk and quadriceps muscle training on the psychological perception of work and on the subjective assessment of low-back insufficiency. A study in a geriatric hospital. Scandinavian Journal of Rehabilitation Medicine. 1978; 10(4): 201-209.

(6) Donchin M, Woolf O, Kaplan L, Floman Y. Secondary prevention of low-back pain. A clinical trial.[see comment]. Spine. 1990; 15(12): 1317-1320.

(7) Gundewall B, Liljeqvist M, Hansson T. Primary prevention of back symptoms and absence from work. A prospective randomized study among hospital employees. Spine. 1993; 18(5): 587-594.

(8) Harma MI, Ilmarinen J, Knauth P, Rutenfranz J, Hanninen O. Physical training intervention in female shift workers: I. The effects of intervention on fitness, fatigue, sleep, and psychosomatic symptoms. Ergonomics. 1988; 31(1): 39-50.

(9) LeClerc A, Landre M, Pietri F, Beaudoin M, David S. Evaluation of interventions for prevention of back, neck, and shoulder disorders in three occupational groups. International Journal of Occupational and Environmental Health. 1997 Jan-Mar; 3 1, 5-12.

(10) Li J, Wolf L, Evanoff B. Use of mechanical patient lifts decreased musculoskeletal symptoms and injuries among health care workers. Injury Prevention. 2004; 10(4): 212-216.

(11) Linton SJ, Bradley LA, Jensen I, Spangfort E. The secondary prevention of low back pain: A controlled study with follow-up. Pain. 1989; 36(2): 197-207.

(12) Maul I, Laubli T, Oliveri M, Kruege H. Long-term effects of supervised physical training in secondary prevention of low back pain.[comment]. European Spine Journal. 2005; 14(6): 599-611.

(13) Oldervoll LM, Ro M, Zwart JA, Svebak S. Comparison of two physical exercise programs for the early intervention of pain in the neck, shoulders and lower back in female hospital staff. Journal of Rehabilitation Medicine. 2001; 33(4): 156-161.

(14) Smedley J, Trevelyan F, Inskip H, Buckle P, Cooper C, Coggon D. Impact of ergonomic intervention on back pain among nurses. Scandinavian Journal of Work, Environment & Health. 2003; 29(2): 117-123.

(15) Videman T, Rauhala H, Asp S, Lindstrom K, Cedercreutz G, Kamppi M, Tola S, Troup JD. Patient-handling skill, back injuries, and back pain. An intervention study in nursing. Spine. 1989; 14(2): 148-156.

(16) Yassi A, Cooper JE, Tate RB, Gerlach S, Muir M, Trottier J, Massey K. A randomized controlled trial to prevent patient lift and transfer injuries of health care workers. Spine. 2001; 26(16): 1739-1746.

8.0 References used in partial data extraction

(1) Best M. An evaluation of manutention training in preventing back strain and resultant injuries in nurses. Safety Science. 1997; 25(1-3): 207-222.

(2) Carrivick PJW, Lee AH, Yau KKW. Effectiveness of a participatory workplace risk assessment team in reducing the risk and severity of musculoskeletal injury. Journal of Occupational Health. 2002; 44(4): 221 – 225.

(3) Secondary Reference: Carrivick PJW, Lee AH, Yau KKW, Stevenson MR. Evaluating the effectiveness of a participatory ergonomics approach in reducing the risk and severity of injuries from manual handling. Ergonomics. 2005; 48(8): 907 – 914.

(4) Charney W. The lift team method for reducing back injuries: a 10 hospital study. American Association of Occupational Health Nurses Journal. 1997; 45(6): 300 – 304.

(5) Davis PM, Badii M, Yassi A. Preventing disability from occupational musculoskeletal injuries in an urban, acute and tertiary care hospital: results from a prevention and early active return-to-work safely program. Journal of Occupational & Environmental Medicine. 2004; 46(12): 1253 – 1262.

(6) Evanoff BA, Bohr PC, Wolf LD. Effects of a participatory ergonomics team among hospital orderlies. American Journal of Industrial Medicine. 1999; 35(4): 358 – 365.

(7) Evanoff B, Wolf L, Aton E, Canos J, Collins J. Reduction in injury rates in nursing personnel through introduction of mechanical lifts in the workplace. American Journal of Independent Medicine. 2003; 44(5): 451 – 457.

(8) Fanello S, Frampas-Chotard V, Roquelaure Y, Jousset N, Delbos V, Jarny J, Penneau-Fontbonne D. Evaluation of an educational low back pain prevention program for hospital employees. Revue du Rhumatisme (English Edition). 1999; 66(12): 711 – 716.

(9) Secondary Reference: Fanello S, Jousset N, Roquelaure Y, Chotard-Frampas V, Delbos V. Evaluation of a training program for the prevention of lower back pain among hospital employees. Nursing & Health Sciences 2002; 4(1-2): 51 – 54.

(10) Fujishiro K, Weaver JL, Heaney CA, Hamrick CA, Marras WS. The effect of ergonomic interventions in healthcare facilities on

musculoskeletal disorders. American Journal of Industrial Medicine, 2005; 48(5): 338 – 347.

(11) Garg A, Owen B. Reducing back stress to nursing personnel: an ergonomic intervention in a nursing home. Ergonomics. 1992; 35(11): 1353 – 1375.

(12) Guthrie PF, Westphal L, Dahlman B, Berg M, Behnam K, Ferrell D. A patient lifting intervention for preventing the work-related injuries of nurses. Work. 2004; 22(2):79 – 88.

(13) Lagerstrom M, Hagberg M. Evaluation of a 3 year education and training program. For nursing personnel at a Swedish hospital. American Association of Occupational Health Nurses Journal. 1997; 45(2): 83 – 92.

(14) Secondary Reference: Lagerstrom M, Josephson M, Pingel B, Tjernstrom G, Hagberg M, Backlund A, Christmansson D, Engkvist I-L, Fager O, Edlund A-C, Hjelm EW, Wikstrom T, Ostman C. Evaluation of the implementation of an education and training programme for nursing personnel at a hospital in Sweden. International Journal of Industrial Ergonomics. 1997; 21(1): 79 – 90.

(15) Landstad BJ, Schuldt K, Ekholm J, Broman L, Bergroth A. Women at work despite ill health: diagnoses and pain before and after personnel support. A prospective study of hospital cleaners/home-help personnel with comparison groups. Journal of Rehabilitation Medicine. 2001; 33(5): 216 – 224.

(16) Secondary Reference: Landstad B, Ekholm J, Schuldt K, Bergroth A, Health-related quality of life in women at work despite ill-health. A prospective, comparative study of hospital cleaners/home-help staff before and after staff support. International Journal of Rehabilitation Research. 2000; 23(2): 91 – 101.

(17) Lynch RM, Freund A. Short-term efficacy of back injury intervention project for patient care providers at one hospital. Aihaj. 2000; 61(2): 290 – 294.

(18) Nassau DW. The effects of prework functional screening on lowering an employer's injury rate, medical costs, and lost work days. Spine. 1999; 24(3): 269 – 274.

(19) Nelson A, Matz M, Chen F, Siddharthan K, Lloyd J, Fragala G. Development and evaluation of a multifaceted ergonomics program to prevent injuries associated with patient handling tasks. International Journal of Nursing Studies. 2005.

(20) Nevala N, Tamminen-Peter L. Ergonomics and usability of an electrically adjustable shower trolley. International Journal of Industrial Ergonomics. 2004; 34(2): 131 – 138.

(21) Owen BD, Keene K, Olson S. An ergonomic approach to reducing back/shoulder stress in hospital nursing personnel: a five year follow up. International Journal of Nursing Studies. 2002; 39(3): 295 – 302.

(22) Peterson EL, McGlothlin JD, Blue CL. The development of an ergonomics training program to identify, evaluate, and control musculoskeletal disorders among nursing assistants at a state-run veterans' home. Journal of Occupational & Environmental Hygiene. 2004; 1(1): D10 - D16.

(23) Ronald LA, Yassi A, Spiegel J, Tate RB, Tait D, Mozel MR. Effectiveness of installing overhead ceiling lifts. Reducing musculoskeletal injuries in an extended care hospital unit. American Association of Occupational Health Nurses Journal. 2002; 50(3): 120 – 127.

(24) Secondary Reference: Chhokar R, Engst C, Miller A, Robinson D, Tate RB, Yassi A. The three-year economic benefits of a ceiling lift intervention aimed to reduce healthcare worker injuries. Applied Ergonomics. 2005; 36(2): 223 – 229.

(25) Secondary Reference: Spiegel J, Yassi A, Ronald LA, Tate RB, Hacking P, Colby T. Implementing a resident lifting system in an extended care hospital: Demonstrating cost-benefit. AAOHN Journal 2002; 50(3): 128 – 134.

(26) Ryden LA, Molgaard CA, Bobbitt SL. Benefits of a back care and light duty health promotion program in a hospital setting. Journal of Community Health. 1988; 13(4): 222 – 230.

(27) Skargren E and Oberg B. Effects of an exercise program on musculoskeletal symptoms and physical capacity among nursing staff. Scandinavian Journal of Medicine & Science in Sports. 1996; 6(2): 122 – 130.

(28) Sobaszek A, Fantoni-Quinton S, Delval M, Rejou P, Mauppin JM, Lefranc D, Thevenon A, Frimat P, Edme JL. Long-term assessment of a sanitary education and lumbar rehabilitation program for health care workers with chronic low back pain at the University Hospital of Lille. Journal of Occupational & Environmental Medicine. 2001; 43(3): 289 – 294.

(29) Tiesman HM, Nelson AL, Charney W, Siddharthan K, Fragala G. Effectiveness of a ceiling-mounted patient lift system in reducing

occupational injuries in long term care. Journal of Healthcare Safety. 2003; 1(1): .34 – 40.

(30) Yassi A, Tate R, Cooper JE, Snow C, Vallentyne S, Khokhar, JB. Early intervention for back-injured nurses at a large Canadian tertiary care hospital: an evaluation of the effectiveness and cost benefits of a two-year pilot project. Occupational Medicine (Oxford). 1995; 45(4): 209 – 214.

(31) Secondary Reference: Cooper JE, Tate RB, Yassi A, Khokhar J. Effect of an early intervention program on the relationship between subjective pain and disability measures in nurses with low back injury. Spine 1996; 21(20): 2329 – 2336.

(32) Secondary Reference: Cooper JE, Tate RB, Yassi A. Components of initial and residual disability after back injury in nurses. Spine 1998; 23(19): 2118 – 2122.

Appendices

Appendix A

List of stakeholders who attended presentations

Toronto (April 2006)

Attendees:

Anne Duffy	Ontario Safety Association for Community Health,
Paulette Sherwood	Ontario Long-Term Care Association
Liz McGroarty	Registered Nurses' Association of Ontario, Occupational Health Nurse
Theresa Crerar	Westpark Health Centre, Ontario
Nancy Kemp	Ontario Nurses' Association
Ilene Stones	Workplace Safety & Insurance Board, Best Practices Division, Ontario
Mary Lou King	Ontario Nurses'Association
Andra Forrest	Arjo, Ontario
Steve Ingram	Waverley Glen, Ontario

Houston (May 2006)

Attendees:

Matt Berkheiser	Safety Program Director, Environmental Health & Safety, MD Anderson, Texas
Jim Collins	Research Scientist, National Institute for Occupational Safety and Health, West Virginia
Chick Deegan	Associate Director, Center of Nursing Leadership, University of Texas at Arlington, Texas
Robert J Emery	Executive Director, Environmental Health & Safety, Univesity of Texas Health Science Center, Texas
Stacey Hubbell	Director of Clinical Operations, Continuing Care Inc., Indiana
Linda Lee	Executive Director and Chief Safety Officer, Environmental Health & Safety, MD Anderson, Texas
Susan Parnell	Instructor of Clinical Nursing, University of Texas School of Public Health, Texas
Robert Salter	Manager, Employee Health Services, Henry Ford Health System, Michigan
Jim Willmann	General Counsel and Director of Governmental Affairs, Texas Nurses Association, Texas

Appendix B

Reviewer guide for Level 1 review

The guide is designed to provide all reviewers with the same information. Each reviewer should become thoroughly familiar with the guide prior to conducting the Level 1 Review. Inter-rater variability should be minimized by each rater's familiarity with the guide. The bolded materials below are included in the table in Memo 1 and in the SRS on-line form.

- ✓ Once an article has made it past the first question, the reviewer must complete all of the fields before submitting the review, no matter the outcome of a criterion.
- ✓ Please click the "submit" button. Otherwise your responses will not be counted.
- ✓ If the reviewer is unclear of the answer from the title and abstract, then mark the "unclear" column and move the paper forward.
- ✓ If the reviewer is marking "unclear," the reviewer must use the text box. This will save time when resolving conflicts.

Q1. Did an intervention occur in a health-care setting (does not include laboratory studies)?
***among health-care employees**
If the research study does not include an intervention then the article is not relevant to the review and therefore should be excluded.

- ❖ Intervention: any occupational safety or health intervention designed to reduce musculoskeletal symptoms, injuries or injury or claim reporting. If it is a review article then an intervention did not occur (tip: look for the word "review" in the MESH terms). Interventions cannot be described in less than 2 pages therefore the "no" option should be selected.
- ❖ Health-care setting: any type of employer/business where health care is provided. Therefore, businesses that do not provide health care should be excluded. Laboratory studies are also excluded here. Our definition of health-care setting does not include stand-alone facilities. The following list was derived to capture what we meant by "stand-alone facilities." Employees in "stand-alone" facilities would have different exposures than those in a hospital, rehab hospital, nursing home, geriatric care center, etc. Therefore interventions implemented within these types of facilities should be excluded.
    - o **out-patient**
    - o **ambulatory**
    - o **pharmacy**
    - o **stand-alone ambulatory care**
    - o **stand-alone medicine centers**
    - o **chiropractic**

- **walk-in clinic**
- **homeopathic**
- **naturopath**
- **massage**
- **psychologist**
- **urgent care clinic**
- **urgent care center**

a) **Yes, in a health-care setting**
b) **No, either not in a health-care setting or an intervention did not occur**
c) **Unclear**

Q2. Is the reference from a peer-reviewed publication (in press or accepted for publication)?
Although some well-designed studies are published in non peer-reviewed publications (e.g. symposia or reports), to simplify the review process those studies that were not peer-reviewed will be excluded. If the reviewer is unclear on whether a journal is peer-reviewed then the group should be queried and a consensus reached for all group members to use. Those studies suggested by content experts that are "in press" or "accepted for publication" will be considered as they have made it through the peer-review process.

a) **Yes**
b) **No**
c) **Unclear**

Q3. Is the language of the article in English, Spanish, Swedish or French?
Articles that are in these languages could be translated by a member of the team or by the Institute for Work & Health. If the article is not in one of these languages then the article will be excluded because we have limited translating resources. Languages other than those accepted will have a cue in SRS to enter the language in the text box next to the "no" response. This way we can track articles in other languages. This action will help answer a question Emma has had in the past. Should many other articles in an excluded language be found, then translation will be investigated.

a) **Yes**
b) **No (Please enter language)**
c) **Unclear**

Q4. Does the study report post-only measures with no control group?
***Note: the answer should be "NO" if there are either pre-intervention measurements or a control group**

A post-only design with no control group has no pre-intervention measures and no control or concurrent comparison group. It is a single group design. Participants could have different pre-intervention characteristics that could account for change. Additionally, secular changes to the workplace could explain observed changes in workplaces. In combination, if a study doesn't include comparisons with pre-intervention measurements as well as no control group then the study can't account for the two biggest threats to validity of workplace research: pre-intervention differences and secular changes. If the study is post-only with no control group it will be excluded.

a) **Yes**
b) **No**
c) **Unclear**

Q5. Does the study include individual health data?
Grouped data (e.g. rates at workplaces) does not allow the researcher to account for heterogeneity of exposure level and covariate levels within groups. Therefore if a study just compares rates between an intervention and control worksite instead of injuries at an individual level, there is missing information that could distort the results. Group only data will be excluded.

a) **Yes**
b) **No**
c) **Unclear**

Q6. Is the outcome a musculoskeletal symptom/disorder/injury (including OSHA log data and workers compensation claims data)?
If the outcome is not a musculoskeletal symptom, disorder or injury then it is not relevant to this review and therefore will be excluded. Musculoskeletal data abstracted from OSHA or other injury-reporting forms and workers' compensation claims databases will be included since this is the type of information stakeholders often use.

a) **Yes**
b) **No**
c) **Unclear**

Appendix C

Quality appraisal guide for reviewers

The guide is designed to provide all reviewers with the same information. Each reviewer should become thoroughly familiar the guide prior to conducting a quality assessment review. Inter-rater variability should be minimized by each rater's familiarity with the guide. The bolded materials below are included in the SRS on-line form.

Questions 1 & 2 are designed to remove articles that could not be removed in Level 1 review due to lack of information. The reviewer is asked to apply the same criteria used in Level 1 review as an initial screen of the article.

If the reviewer answers "Yes" to question 1 then only questions 1 & 2 must be answered and the reviewer can submit.

Q1. Should the paper have been excluded at Level 1?
The reviewer is first asked to determine if the paper should be excluded because it is not an intervention study. The reviewer must consider all 6 exclusion criteria.

a) **Yes**
b) **Unclear**
c) **No**

Q2. If the answer to Number 1 above is "yes" then why? (Check all that apply).
So that the team can effectively summarize the state of the literature, the reviewer is asked to describe the exclusion criteria applied above in question 3.

a) **There is not an intervention or it's not in a health-care setting.**
b) **The article is not from a peer-reviewed journal.**
c) **The language of the article is not English, French, Spanish or Swedish.**
d) **There is either no control group or no pre-intervention measurements.**
e) **The study does not report individual health data.**
f) **The outcome is not musculoskeletal symptoms/disorders.**

If the reviewer answers "Yes" to question 1 then only questions 1–2 must be answered and the reviewer can submit.

Q3. Was the research question/objective clearly stated?
A clear, explicit statement of objectives should be included in the study. This could be stated as an objective or as a research question. If the aim of the study is not clearly stated then results are likely of limited value.

a) **Yes**
b) **Unclear**
c) **No**

Q4. Was a primary hypothesis clearly stated?
A clearly stated research question/objective does not mean a clearly stated primary hypothesis has been stated. Hypotheses usually begin with: "We hypothesize…"; "We expect…"; or "We predict…" and explains that a change in X leads to a change in Y.

A well-designed intervention will have one or two clearly stated and testable primary hypotheses. There are many outcomes that can be considered and stated as secondary or post-hoc hypotheses, but a well-designed study is typically powered with a single outcome. This allows for the alpha region in the statistical test to be devoted to the single hypothesis test.

a) **Yes**
b) **Unclear**
c) **No**

Q5. Was the intervention implementation described?
Describing the intervention includes: the setting of the intervention, i.e. where it was carried out, and specifically what was changed and how. These are important aspects to document. Furthermore, if training was part of the intervention, how was the training done in a consistent way across subjects? If placebos were used, was their implementation described.

Inadequate description of the intervention strategy and implementation makes it impossible to reproduce the intervention in another population.

a) **Yes**
All or most aspects of the intervention are clearly described.
b) **Unclear**
There is not enough information provided, the intervention implementation process is not clearly described.
c) **No**
The intervention process is not described.

Q6. Was the calendar duration of the intervention documented? (Calendar duration is the time it took to implement the intervention.)
The calendar duration refers to the number of months or years over which the intervention took place. The calendar duration could be documented in a self-evident way (e.g. the beginning and end dates of an intervention as distinct from the follow-up period post-intervention implementation).

The duration of the intervention is important to document. Interventions of short duration (i.e. a couple of days or weeks) could have insufficient intensity to have a significant effect on the proposed mechanisms of change. Conversely, interventions that take too long (i.e. 5 yrs) may also hinder the evaluation of the intervention's impact as many other changes are likely to occur in the organization. Workplaces are dynamic environments and many other changes may have taken place during a long period of follow-up, other than the intervention itself, which can confound the results.

a) **Yes**
b) **Unclear**
c) **No**

Q7. Was the length of follow-up 3 months or greater?
If the length of follow-up post intervention was not at least three months (90 days) then there is little likelihood of observing substantively important changes in the musculoskeletal outcomes. Negative intervention studies with short follow-ups are likely to find no significant effects due to lack of a long-enough follow-up period to allow group differences to emerge. Thus, in synthesizing data from a range of intervention studies, differences could be due to differences in length of follow-up. Therefore, the 3 month period is considered a basic standard and we expect most studies to have a longer follow-up period.

a) **Yes**
Follow-up ≥ 90 days
b) **Unclear**
c) **No**
Follow-up ≤ 90 days

Q8. Were concurrent comparison (control) groups(s) used?
Inadequate comparison groups or not utilizing referents at all creates validity problems, which may undermine the conclusions drawn from a study. Therefore, it is important for a study to include a concurrent comparison group. A comparison group can receive a placebo; and thus be considered a comparison. While a control group typically does not receive any treatment. By 'concurrent' it is expected the information on the control or comparison group is collected at the same time as the treatment group. Considering the importance of having a comparison group to document and account for the potential effects of unexpected secular changes in workplaces, having a

closely analogous referent group, with similar work experiences is a methodological strength.

a) Yes
At least one comparison group or control group was used against which intervention's effect were evaluated.
b) Unclear
c) No
No concurrent comparison group or control group were used in this study.

Q9. Was an intervention allocation randomized?
A randomized allocation strategy is part of a strong research design. Randomization of intervention conditions is typically preferred because it avoids systematic confounding by known and unknown factors. If the group membership (intervention vs. non-intervention) was not randomly assigned (or treatment not randomly allocated) then the study must address potential group differences in analysis. IF THERE IS ONLY ONE GROUP THEN CHECK NOT APPLICABLE.

Inadequate description of the intervention randomization allocation strategy makes it impossible to reproduce the intervention in another population. The allocation strategy should be clearly stated in the study to allow for interventions to be reproducible by others. If the researchers state they employed a random allocation but it is unclear how this was done and thus not easily replicable the reviewer should endorse 'unclear.'

a) Yes
Intervention was randomly assigned and assignment was well described.
b) Unclear
c) No
Groups were not randomly assigned or assignment was not well described.
d) Not applicable (No control group)

Q10. Were sample inclusion/exclusion criteria described?
In every study some potential participants are excluded because their participation could bias the findings. If there is no information on sample inclusion or exclusion criteria, then the generalizability of the conclusions may be challenged. Finally, with different sample inclusion/exclusion criteria (e.g. including those with the outcome vs. those without) synthesis of the literature may be difficult.

a) Yes
b) Unclear
c) No

Q11. Was the sampling frame representative of the target population?
The sampling frame is the list or other physical representation of sampling units in the accessible part of the target population. The answer would be "yes" if the author explained how the characteristics of a population compare to characteristics of known populations of health-care workers on key demographics (e.g. age or gender). This question addresses external validity as a methodological strength.

Here we list a couple of examples of sampling frame and target population.

- Half the nursing population in an intervention is male; therefore the sampling frame was not representative of the target population, nurses. The results of that intervention might not be applicable to the general population (mostly female).
- Your target population might be small businesses in medical equipment distribution in State X. To sample those businesses, you may use a list of businesses from the Medical Equipment Distribution Business Association as the sampling frame. This question is concerned with how well that list represents the target population. Ideally, we want the population similar to the population we are comparing it to (e.g. the sample).

  a) Yes
  Sampling frame was representative of the target population.
  b) Unclear
  c) No
  Sampling frame was not described or was not representative of the target population.

Q12. Was participation rate reported and >40% for employees?
The reviewer is being asked to determine if a participation rate was reported and the level of participation. Both the participation rate and the level of participation must be reported as > 40% to answer the question "yes." By participation rate we mean those who were asked to sign inform consent and those who agreed and are participating. A participation rate reflects the potential selection problems introduced when moving from a sample to group of participating workers. This is a single value reported prior to intervention. Participation rates may be calculated from information in the article including tables.

If participation rates were not reported then it is impossible to draw any conclusions about the validity of the study and thus study conclusions since we know nothing about those who participated and those who chose not to participate. We have set as a lower bound a 40% participation rate. The group asserts that any rate lower than 40% makes the study of limited validity. The greater rate of participation (or recruitment) reduces non-response bias.

a) Yes
Participation rate was reported and ≥ 40% for employees/workers.
b) Unclear
An unclear response can only be endorsed if some information is presented and thus researchers are trying to report rates, but do not provide the exact information we requested.
c) No (NP/<40%)
The participation rate was either not reported (NR) or was less than 40% (<40%).

Q13. Did the researchers describe the study participants at baseline by demographics, exposure or outcome?
Please indicate if baseline characteristics are described for the entire study sample at baseline. To answer "yes" participants may be described by job-related factors, individual characteristics, or factors related to exposures and outcomes (for example baseline pain levels across groups).

Describing the baseline characteristics allows for a comparison of the population studied with other research and thus contributes to evidence synthesis. Different populations studied that find similar findings contributes to the external validity of the intervention.

a) Yes
b) Unclear
c) No

Q14. Were baseline characteristics presented by group?
If there are groups then the researchers should present characteristics by group (usually in a table). IF THERE IS ONLY ONE GROUP THEN CHECK NOT APPLICABLE. This allows the reviewer to compare groups based on those characteristics. If a study has multiple arms the researchers should have a table showing that there are no differences between groups. If randomization was used then the table should show that randomization worked.

If there are no major significant differences between the groups on baseline characteristics, exposures or outcomes, one can be confident that differential selection bias was minimal and that the differences attributed to the intervention are not likely affected by these differences.

a) Yes
b) Unclear
c) No
d) Not applicable (No control group)

Q15. Were differences between those employees who remained in the study and those who dropped out analyzed?
Loss to follow-up can be a significant problem especially if it is differential. Comparisons should be made between those who dropped out and those remaining to determine if significant differences exist that could affect the validity of the study. Selection bias can result if certain subjects are systematically more likely to be lost to follow-up than others.

a) **Yes**
b) **Unclear**
c) **No**

Q16. Did withdrawals affect groups equally?
Withdrawals are those study units that are not observed in later stages of the study because they become inaccessible or ineligible. "Yes" requires similar withdrawal rates (e.g. within 5 percentage points). If withdrawals are larger than 10% in one or more groups, "yes" requires that the study show that potential confounders are similarly distributed in the comparison groups. Consideration of confounders is done in two steps, in Question 19 and 20. IF THERE IS ONLY ONE GROUP THEN CHECK NOT APPLICABLE.

Differential attrition of subjects poses a major threat to internal validity. When there are no statistical differences between participants who stay and those who leave, one can be more confident that attrition bias did not occur.

a) **Yes**
Withdrawal rates were similar across groups.
b) **Unclear**
c) **No**
Withdrawal rates were different across groups.
d) **Not applicable (No control group)**

Q17. Were the effects of the intervention on some exposure parameters documented?
This question addresses the mechanisms of change for how the intervention will reduce MSD risk in individuals. In intervention research these are often termed process outcomes. Do the researchers report process outcomes? Process outcomes would not be the primary outcome but would support the hypothesis that the intervention changed health outcomes through the hypothesized pathway between the intervention and changes in health outcomes. An example of this could be the number of lifts done by a machine (therefore replacing manual lifts); muscle loading changes; or did behaviours change because of training.

Most importantly, if the process outcomes don't reflect the hypothesized changes then observed health effects may be due to other factors and not the intervention.

a) Yes
Exposures were assessed.
b) Unclear
c) No
Exposure measurements were not measured or discussed.

Q18. Was contamination between groups described or documented?
Contamination can occur when the interventions assigned to participants in one group are picked-up and adopted by some or all members of the other groups. This can be documented many ways including whether lifts could be moved to "control" floors. IF THERE IS ONLY ONE GROUP THEN CHECK NOT APPLICABLE.

Contamination can introduce bias in the results if comparison groups, for example, have been exposed to some of the interventions intended for the study group, unbeknownst to the researchers. This is an issue particularly when a study uses controls from the same workplace as the intervention group.

a) Yes
Contamination between groups was described or documented.
b) Unclear
c) No
Contamination between groups was not described or documented.
d) Not applicable (No control group)

Q19. Were covariates/potential confounders for musculoskeletal disorders measured (i.e., gender, age, non-work activities)?
Ascertainment of covariates and potential confounders is important to allow the researcher to rule out plausible alternative explanations for observed health differences. Physical risk factors for musculoskeletal disorders include: force, repetition, static loading, time spent in awkward postures, etc. Psychosocial and organizational risk factors can include: social support, job satisfaction, control over one's job, etc.

Musculoskeletal disorders are multi-dimensional in origin and thus there are many covariates and known confounders.

a) Yes
b) Unclear
c) No (Not Measured)

Q20. Was adjustment made for covariates/potential confounders?
Without appropriate multivariate adjustment the conclusions may not be valid.

a) **Yes**
Statistical method used to adjust for confounders is explained and appropriately conducted.
b) **Unclear**
c) **No**
d) **Not Applicable (Not Measured)**

Q21. Were statistical methods adequately described?
The reviewer must use his or her knowledge of statistics to comment on the analysis.

Furthermore, there are often site or group differences that may be important to consider that bias results. For example, in ergonomic interventions the differences between supervisors in supporting the intervention could influence the intervention's success. If there are differences then the group differences must be accounted for in analysis.

a) **Yes**
Statistical methods are described sufficiently, and the methods used were appropriate and properly applied.
b) **Unclear**
c) **No**

Q22. Are there any other potential primary studies listed in this reference list which should be retrieved for consideration (if yes, please include reference ID or author/year/publication, etc.)?
It is important to look in the reference section of relevant studies because usually other studies that may be of potential use for this review are cited, which could have been missed in our search strategy.

a) **Yes**
b) **No**

Q23. Should article proceed to data extraction?
Our goal is to give the reviewer the opportunity to move articles forward to data extraction even if the study had not met many quality criteria.

a) **Yes; because it has met enough of the quality criteria**
b) **Yes; even though it has not met many of the quality criteria (please justify in comment)**
c) **No**

Appendix D

Quality appraisal decisions

Novel Approach: Were concurrent comparison (control) groups used?

Open population studies would construct an analytic comparison group to rule out changes in the workplace occurring naturally over time. This was done in two ways: 1) statistically adjusting for changes in injuries at the worksite that the intervention should not influence (e.g. a lift equipment intervention should affect patient-handling MSK injuries but should not affect slips and trips) or 2) statistically adjusting for changes in MSK injuries in job titles that would not receive intervention benefits (e.g. clerical workers are not likely to benefit from a ceiling mounted lift assist in a patient's room, while nurses will benefit). Although open population studies did not follow a fixed set of individuals who received and did not receive the intervention over time, they did compare people or injuries that were exposed and unexposed to the intervention. Therefore, the group agreed to give the study credit for having a control group.

On the other hand, some fixed cohort studies collected data on a comparison group yet they never statistically evaluated the intervention's effects by adjusting for MSK health effects in the control group. Therefore, the control group was never used to rule out natural workplace changes. From a methodological quality perspective, the observed effects on MSK health are therefore the result of a pre-/post-intervention comparison in the intervention group only. If a study collected information on a control group but did not present statistical comparisons between groups, the review team agreed the study should not receive credit for having a control group.

Novel approach: Were differences analyzed between those employees who remained in the study and those who dropped out?

Open population studies that do not follow the same people over time often do not present data on workers who left the study. However, an open population study could evaluate whether the population changed based on key descriptors of the worksite population (e.g. age, job tenure, gender). This would increase confidence that the intervention had an effect on the change in MSK injury rates, and was not the result of a change in the working population to include more healthy workers, for instance. If a study conducted this evaluation, then the review team agreed to give the study credit for analyzing drop-outs.

Novel approach: Were covariates/potential confounders for MSK disorders measured (i.e. gender, age, non-work activities)?

Many studies collected information on what are generally considered covariates for MSK disorders (e.g. mental health or muscle strength). However, sometimes the researchers treated these variables as outcomes, not as covariates to be statistically controlled in the test of the intervention effect (often process outcomes like muscle strength in exercise interventions). Therefore, the review team decided the study should not get credit for measuring covariates if they were presented as outcomes.

Novel Approach: Was contamination between groups described or documented?

Studies often chose another work unit or another job group as the comparison group because they would not be exposed to the intervention. Some team members felt an explicit evaluation of the potential for contamination was required. For example, the study should report if employees changed jobs from an intervention unit to a control unit. However, the team agreed that choosing comparison groups explicitly because they would not be exposed to the intervention was a deliberate design strategy to prevent contamination between individuals. The study therefore received credit in the QA for describing contamination.

Novel Approach: Was the participation rate reported and greater than 40% for employees?

Open population studies obtain worksite consent instead of individual consent. Therefore authors do not report an individual participation rates. The group agreed that the concept of participation rate was important to rule out selection bias. Selection bias occurs in a self-selected sample, when the observed effects of an intervention were due to the selected sample, and do not represent the population the sample was drawn from. In open populations the sampling frame is technically the worksite, job title, department, ward or unit. The review team considered studies that had complete information on the sampling frame to have 100% participation (i.e. they included injuries for the employee sample, and the actual number of hours worked for included employees). In such cases the authors had addressed the notion of sample representativeness and indirectly self-selection bias. However, if researchers estimated the hours worked then it was unclear as to what kind of participation they may have been able to capture in their estimate. The team agreed to give open population studies credit for participation rates when injuries and actual hours were reported.

Omission: Were sample inclusion/exclusion criteria described?

The team agreed that health outcome inclusion or exclusion criteria – for example, if a study excluded non-MSK injuries – were an inclusion/exclusion criteria and therefore the study got credit in the QA.

Omission: Was an adjustment made for covariates/potential confounders?

In the quality appraisal stage, reviewers gave credit for adjusting for covariates. However, during data extraction, reviewers could not extract a test of an MSK outcome that was adjusted. The group agreed that the study had to adjust for covariates in the statistical test of the MSK outcome to receive credit for adjustment.

Another proposed change was to give credit to studies that matched intervention and control participants on potentially confounding variables at the beginning of the study. However, the study had to maintain the match in the final statistical analyses. Otherwise, if participants dropped out, then the study design characteristic no longer served to adjust for confounders.

The questions of the methodological quality appraisal were designed to determine whether studies addressed vulnerabilities in their design. Specifically, vulnerabilities would mean that alternative hypotheses (related to internal and statistical conclusion validity) could explain the observed effects, or the study was of limited generalizability (related to construct and external validity). We present the quality appraisal questions by each validity type in Table 8 (Shadish 2001).

Therefore, higher quality studies (high and medium-high) address more threats to internal, statistical conclusion, external and construct validity than medium and limited quality studies, leading us to have higher confidence in the validity of the observed effects of the intervention on MSK health.

Appendix E

Quality appraisal (QA) questions by validity type and QA ranking

QA question by validity type	Percentage of studies meeting criterion within QA ranking			
	H	MH	M	L
Internal validity				
3. Was the intervention implementation described?	100%	100%	100%	75%
5. Was the length of follow-up three months or greater?	50%	79%	75%	25%
6. Were concurrent comparison (control) group(s) used?	100%	93%	45%	25%
7. Was the intervention allocation randomized?	100%	36%	10%	0%
11. Were baseline characteristics of study participants presented?	100%	100%	95%	0%
12. Were baseline characteristics presented by group?	100%	86%	35%	0%
13. Were differences between those participants who remained in the study and those who dropped out analyzed?	100%	29%	15%	0%
14. Did withdrawals affect groups equally?	50%	0%	0%	0%
16. Was contamination between groups described or documented?	50%	21%	5%	0%
Percentage of internal validity items met by >70% studies	67%	56%	33%	11%
Statistical conclusion validity				
1. Was the research question/objective clearly stated?	100%	100%	100%	100%
4. Was the calendar duration of the intervention documented?	100%	57%	55%	0%
15. Were the effects of the intervention on some exposure parameters documented?	100%	86%	60%	50%
17. Were covariates/potential confounders for musculoskeletal disorders measured (e.g. gender, age or non-work activities)?	100%	93%	75%	0%
18. Was multivariate adjustment made for covariates/potential confounders?	0%	29%	10%	0%
19. Were the statistical methods adequately described?	100%	86%	85%	75%
Percentage of statistical conclusion validity items met by >70% studies	83%	67%	50%	33%
External validity				

QA question by validity type	Percentage of studies meeting criterion within QA ranking			
	H	MH	M	L
8. Were both sample inclusion/exclusion criteria described? (If No, indicate whether inclusion criteria, exclusion criteria or both were not described in the comment box.)	100%	93%	70%	50%
9. Was the sampling frame representative of the target population?	50%	0%	0%	0%
10. Was the participation rate reported and >40% for participants? (If No, indicate whether participation rate was not reported or was <40%.)	100%	57%	60%	0%
Percentage of external validity items met by >70% studies	67%	33%	0%	0%
Construct validity				
2. Was a primary hypothesis clearly stated?	0%	29%	45%	0%

H=high
MH=medium-high
M=medium
L=low

Appendix F

Guide to the data extraction form for reviewers

This guide must be read before beginning the data extraction. Print this guide (on a colour printer if possible) and have it available to refer to while doing the data extraction. Please extract the data from the articles you review by completing the form on SRS and entering text in the provided areas. Please read the questions carefully, especially the instructions in italics which provide details on how to enter the data. Bolded text provides some additional instructions that will help to ensure that the answers from different reviewers are consistent. Red text provides examples to illustrate specific responses.

All of the questions in the SRS form should have an answer. If an article lacks the information necessary to answer a particular question then the reviewer should enter "**NP**" (not provided) in the text box. It is important that all questions have answers because we will not know if an article did not have the information or a reviewer forgot to enter it if we allow blank answers. Remember, do not extrapolate just provide the information that is presented in the article. You may need to get information out of tables or figures (e.g., to calculate participation rates).

Study Design and Setting:

1. State the research question(s)/objective(s). Please use the exact wording from the article. If more than one objective; then list all objectives. Be clear to only include *the objectives tested* not broader objectives described.

2. State the primary hypothesis. Please use the exact wording from the article or enter "NP". A clearly stated research question/objective does not mean a clearly stated primary hypothesis has been stated. Hypotheses usually begin with: "We hypothesize…"; "We expect…"; or "We predict…" and explains that a change in X leads to a change in Y. If the authors list a series of hypotheses but do not declare which is primary then enter all hypotheses stated in question 3.

3. State additional hypotheses not listed in question #2 (list all and number; type "NP" if not applicable).
Additional hypotheses are hypotheses that do not use the primary health outcome and may include process hypotheses that examine the effect of the intervention on an intermediate outcome. Please use the exact wording from the article or enter "NP".

4. Write the last name of the first author and the year of publication (Author's last name, yyyy). Give the first author's last name and the year (4 digits) the article was published.

5. List the jurisdiction where the study was completed. Provide information regarding the country, region, province, city, etc. where the study was carried out - enter "NP" where information is not available.

Country
Province
Region (e.g. Mid-western USA)
State
City

6. Describe what type of healthcare organization(s) that the study was conducted in. Please use the language from the article to describe succinctly. Describe the organization and the unit as it is part of the setting. For example, the organization may be a hospital but the units are only surgical units in the hospital.

7. List the job titles/classification of the participants that participated in the study. Provide the level of detail given in the study or enter "NP".

8. List the inclusion criteria described in the study. (Please list inclusion criteria clearly) Enter a *numbered* list (see below) of how the study selected their site, unit, or individuals for inclusion. For studies that use "administrative" data to track musculoskeletal outcomes, their inclusion of employees or units could be found in the description of outcome measures. Please also summarize the level for inclusion criteria using the notation "S", "U", or "I". We use an example for administrative data because the inclusion criteria are found in unexpected places.

E.g.
1. Intervention units selected based on previous injury rate (U)
2. Back injuries defined as upper or lower trunk injury resulting in either lost time or health care expenses (I)
Make sure to use hard returns so the numbers are left justified.

9. List the exclusion criteria described in the study. (Please list exclusion criteria clearly)
Enter a *numbered* list (see below) of how the study selected their site, unit, or individuals for exclusion. This could be found in the setting description or in their outcome description. Especially studies that use "administrative" data as musculoskeletal outcomes their exclusion of employees or units could be found in the description of outcome measures. Please also summarize the level for exclusion criteria using the notation "S", "U", or "I".
List any exclusion for types of injuries or employee title excluded in abstraction from the injury record?

E.g.
1. Neck or shoulder injuries (I).
2. Employees in the float pool (U)

10. What is the study design? **(Choose only one)**. Please describe any unique characteristics verbatim about the study design in the comment boxes beside the choice you make. "*Trial*" indicates that the study had an intervention and control group. Often in "administrative" data an explicit control group is not described in the design but may be described in the analysis section.

Caution: Do not describe the intervention in great detail. It will be described in Q12.
***Use notation (I_1 –Intervention #1, I_2 –Intervention #2, C_1 Control Group #1, C_2 Control Group #2, I_1C –crossover with intervention first, I_2C –crossover with intervention second)**.

Randomized Field Trial
Non-randomized Field Trial
Randomized Cross-Over Design
Non-randomized Cross-Over Design
Pre-post Design with NO control
Other

Randomized Field Trial -a field study where the intervention assignment is randomized.
R O X O
O O
Non-randomized Field Trial –a field study where the intervention assignment is not randomized.
O X O
O O
Randomized Cross-Over Design: –a field study where two groups receive the intervention at different times and group assignment is randomized.
R O X O O
O O X O
Non-randomized Cross-Over Design –a field study where two groups receive the intervention at different times and group assignment is not randomized.
O X O O
O O X O
Pre-post Design with NO control –a field study with one group observed before and after receiving an intervention.
O X O

11. What type of prevention did the study investigate? (choose only one). Indicate whether the study evaluated a primary or secondary prevention/intervention. The classical definition of primary prevention is an intervention aimed at preventing healthy people from progressing on to symptom or disorder. The classical definition for tertiary prevention is defined as intervention aiming to prevent people with clinically recognized disorders from further morbidity and mortality. Although these definitions are accepted in public health literature to be comparable to other IWH reviews, we will use the terms primary and secondary (instead of tertiary) for those definitions. Should any studies be found with the classical definition of secondary prevention (an intervention aiming to identify asymptomatic or pre-clinical cases and get them to early treatment –classically surveillance studies) the reviewer should flag the article and notify Jessica.

To determine what the authors "aimed" to do reviewers must only answer based on what was reported by the authors. Therefore any studies where clinical diagnoses or symptoms (as part of a case definition) were used to identify and include participants with disorders will be classified as secondary prevention. If a study excluded employees with clinical diagnoses or symptoms to create a cohort of individuals free from symptoms this would be considered a primary prevention. If no such exclusions were made, then the authors will be assumed to have intended to prevent both "asymptomatic" employees from developing symptom or disorder and "symptomatic" individuals from further morbidity and mortality, therefore will be classified as both. If you choose other please provide details.

Primary prevention
Secondary prevention
Both
Other

Intervention Characteristics:

12. Describe all interventions evaluated.
If control received intervention please describe.
E.g.: I_1 - exercise ("training to improve physical fitness"); I_2 -ergonomics training "to improve lifting technique"; C_1 -no exercise and no "training"
***Organize your description of interventions according to I_1, I_2, C, I_1C, and I_2C**

13. Was there confirmation the intervention occurred? (check all that apply) Provide details in the comment box to support your response.
E.g.: "exercise" could be confirmed either by self-report of exercise logs, attendance in classes, or questionnaire report of exercises done.

Direct measurement by equipment
Observation
Self-report
None

14. How long after the intervention implementation did confirmation occur? Monitoring of attendance would be confirmation "during" the intervention. A questionnaire of self-reported exercise one month after the intervention would be 1 month. Place "NP" in text box if confirmation of the intervention is not available in article.

15. What was the duration of the intervention in months/days/hours? (Note this is not the follow-up time but the actual duration of the intervention implementation). Indicate in months if possible, if not in weeks, days etc. or enter "NP".
***Use notation (I_1, I_2, I_1C, and I_2C) for different intervention groups**.

Eg. Baseline data collected on May 1st, 2000. Intervention implemented June 1st, 2000 continues until June 1st, 2001. Follow-up data collected on May 1st 2002. Note this information may be presented in a number of ways (tables, figures, timelines etc). In this example the duration of intervention is ***I_1 = 12 months.***
For "administrative" data it is best to establish what the intervention period is first (e.g., lifts were installed between April 2002 to July 2002).

16. Indicate the time period between the baseline measurement and all subsequent follow up measurements. Use months to indicate the length of follow up, for example, questionnaires were administered at 6, 12, and 18 months. Indicate in months if possible, if not in weeks, days etc. or enter "NP". Please make sure that you describe all intervention groups and all referent groups using the same group notation throughout the data extraction forms.

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

Often in administrative data there are not multiple time points of outcome data collection. Instead there are time periods over which data are collected. For "administrative" data, it is best to establish what the intervention period is first. Then establish the baseline data period for outcome measurements. This period may be a month, 6 months, or years before the intervention. State the full time-period for which baseline outcome data was collected (e.g., "data was collected 3 years prior to lifts installation" answer: April 1998 to April 2002). Finally, establish the follow-up period (e.g., "We compared to 3 years after the lifts were completed installation" answer: July 2002 to July 2005).

Study Group Questions:

17. Describe overall (study) group. If the design is one group that is pre vs. post with no control then only answer Q.18. –Provide answer(s) for each category. Type "NP" in all comment boxes where information is not available.

Sample Size
Age (mean, SD, range)
% female
Loss to follow up (N)

18. Describe the Intervention group(s). Provide answer(s) for each category - enter "NP" in all comment boxes where information is not available. If design is cross-over then answer for I_1C only.

***Use notation (I_1, I_2, and I_1C)**

Sample size	*Eg: I_1 =, I_2=, … (or I_1C=, I_2C=, …)*
Age (mean, SD, range)	*Eg: I_1 =, I_2=, … (or I_1C=, I_2C=, …)*
% female	*Eg: I_1 =, I_2=, … (or I_1C=, I_2C=, …)*
Loss to follow up (N)	*Eg: I_1 =, I_2=, …(or I_1C=, I_2C=, …)*

19. Describe the Referent group. Provide answer(s) for each category - enter "NP" in all comment boxes where information is not available. If design is cross-over then answer for I_2C only.

***Use notation (C, I_1C, and I_2C)**.

Sample size	*Eg: C_1 , C_2, …(or I_1C=, I_2C=, …)*
Age (mean, SD, range)	*Eg: C_1 , C_2, …(or I_1C=, I_2C=, …)*
% female	*Eg: C_1 , C_2, …(or I_1C=, I_2C=, …)*
Loss to follow up (N)	*Eg: C_1 , C_2, …(or I_1C=, I_2C=, …)*

Not applicable (No control group)

Covariate Questions:

20. When were potential covariates/confounders measured? (check all that apply) If covariates were measured any time prior to intervention this will be counted as baseline. *We do not consider pre-intervention measures of the MSK outcome (i.e., dependant variable) to be a covariate.

E.g., for administrative data a study describes demographics of employees at the time of intervention but uses baseline MSK measurements for three years prior to the intervention = "baseline near intervention implementation".

Baseline at time of outcome MSK measurement
Baseline near intervention implementation
Follow up
Unsure (please describe)
Not applicable (Not measured)

21. Select from the list all covariates/confounders that were evaluated for inclusion in the final analysis. **(check all that apply)** Please give details for each response. Provide details and names of variables if you select other.

*We do not consider pre-intervention measures of the MSK outcome (i.e. dependant variable) to be a covariate.

No covariates measured
Physical/biomechanical work conditions (e.g. force, repetition, or static loading)
Psychosocial/cognitive work conditions (include social support here)
Organizational environment (e.g. specific policies, practices, or safety climate)
Equipment adjustment
Medical conditions (diseases & disorders)
Mental & physical health status
Legal
Family environment
Demographics (include income here)
Work experience
Non-work activities
Other

22. Provide a list of covariates/confounding variables that were controlled for in the final test of the intervention effectiveness. Enter "none" in text box if no covariates controlled for. Covariates include gender, age, non-work activities, education etc. Physical risk factors for musculoskeletal disorders include: force, repetition, static loading, time spent in awkward postures, etc. Psychosocial and organizational risk factors can include: social support, job satisfaction, control over one's job, etc. If many variables are considered, three may be entered in broad categories (e.g. demographic (5), medical (3), etc.)

*We do not consider pre-intervention measures of the MSK outcome (i.e., dependant variable) to be a covariate.

23. Describe the differences in covariates/confounders for those that participated in the study vs. those that were invited but did not participate (if possible by experimental group). If authors determined that these differences were not significant, please describe. Enter "NP" in text box if the information is not available. If non-participants cannot be identified from participants because it is an open population (worksite or work unit based) study then the answer is "NA".

24. Describe the differences in covariates/confounders for those that participated in the study vs. those that were lost to follow-up (if possible by experimental group). If authors determined that these differences were not significant please describe this. Enter "NP" in text box if the information is not available. If non-participants cannot be identified from participants typically because it is an open population (worksite or work unit based) study then the answer is "NA".

Outcome Questions: SRS will drop certain questions depending on the answers to the following 3 outcome questions.

25. Does the study use "administrative" records to collect measurements of MSK health outcomes?
By administrative records we mean regulatory required employer record keeping data (e.g. OSHA logs), voluntary employer record keeping data (e.g. incident reports), or insurance record keeping systems (e.g. worker's comp). Voluntary employer record keeping systems are any record keeping systems that either regulatory agencies or insurance agencies do not require.

Describe succinctly the type of administrative record.

Yes
No

26. Does the study use self-report questionnaire records to collect measurements of MSK health outcomes?
We are only including musculoskeletal symptoms and not function or disability questions. Describe succinctly the nature of the musculoskeletal questionnaire used.

E.g. symptom frequency, VAS pain scale, or intensity.

Yes
No

27. Does the study use clinical exams or clinical records as completed by the clinician to collect measurements of MSK health outcomes?
Describe succinctly the protocol or type of clinical exam.

Yes
No

28. Was the population studied "fixed" or "open"? (check all that apply)
A "fixed" population is one where the population is fixed at some time and the same participants are followed over time. An open population is where individuals can come in and out of the study. In a worksite population, the intervention happens at some point and different individuals can contribute information before and after the intervention (new hires).

Fixed population
Open population
Unclear

"Administrative" Record Questions

29. What sources were used to "count" employee injuries? (check all that apply)

Regulatory required employer record keeping data (e.g., OSHA logs)
Voluntary employer record keeping data (e.g., incident reports)
Insurance record keeping systems (e.g., workers' compensation claims data)

30. How were employee hours collected? (check one only)

Many studies calculate injury rates for a unit or an organization. A critical piece to the calculation is the method of collecting employee hours. Estimations of employee hours by calculating from the number of employees are very different from getting actual employee billed hours from human resources.

If unclear, please describe what the study has done. Jessica will be reviewing all unclears.

Estimation of employee hours worked from an estimated of number of employees
Estimation of employee hours worked from an actual number of employees
Actual employee hours from a specific number of employees
Employee hours not collected
Unclear (please describe)

31. Indicate at what level employee hours were ascertained and/or estimated.

Individual
Unit
Site

32. Were injury rates calculated?

Yes
No

33. If injury rates were calculated, list the equation(s). Please define the numerator and denominator using the author's language explicitly. If the equation is not explicitly explained, type "NP".

34. Did the study discuss how they handled any of the following special issues related to administrative record keeping? (check all that apply and describe in comment box)

Temporary employees, contract employees, or floating employees between units
Turnover rate
Reinjuries to the same employee

Questionnaire Questions

35. Check all body regions where symptoms were ascertained by questionnaire. (check all that apply) Provide details in the comment box to support your response. We are only including musculoskeletal symptoms and not function or disability questions. If unclear and you do not feel the information fits into one of these categories please call Jessica (713-385-5811).

Hand/wrist/elbow (HWE)
Neck/shoulder (NS)
Upper back (UB)
Lower back (LB)
Legs/knees/feet (LKF)
Not attributed to a body part (NAB)

36. Describe when follow-up musculoskeletal health outcomes (symptoms) were measured. (check all that apply) Give details if you select "other". If there is more than one MSK outcome identified please use the notation above for each outcome in the comment box beside your measurement choice.

A single time point
Multiple time points assessed and then averaged
Other

37. Were musculoskeletal (MSK) symptoms measured at the same time of day or shift? (check only one) Indicate the consistency of symptom measurement by checking the appropriate response. If there is more than one MSK outcome identified please use the notation above for each outcome in the comment box beside your measurement choice.

Yes, measured at a consistent time of shift (put time in comment box)
Yes, measured at a consistent time of day (put time in comment box)
No, not measured at a consistent time of day/shift
Unclear or unknown time of day

Clinical Exam Questions:

38. Check all body regions where specific clinical disorders were ascertained by physical assessment or laboratory test. (check all that apply) Provide details in the comment box to support your response. If unclear and do not feel information fits into one of these categories please call Jessica (713-385-5811).

Hand/wrist/elbow (HWE)
Neck/shoulder (NS)
Upper back (UB)
Lower back (LB)
Legs/knees/feet (LKF)
Not attributed to a body part (NAB)

39. Was masking of physical assessment done? Provide details in the comment box to support your response. This question is asking if the clinician was blinded to the intervention group.

Yes
No
Unclear
Not Applicable

40. Was a standard protocol used for the clinical exams?

Yes (list protocol name)
No
Unclear (describe)

Statistical Analysis Questions:

41. Please check the types of final analyses done for testing the observed effects of the intervention. (provide details about the analyses in the comment box) You should select the one that represents the final test not the preliminary analyses. Provide details in the comment box to support your response. Give details if you select "other". If unclear and do not feel information fits into one of these categories please call Jessica (713-385-5811).

ANOVA (ANCOVA)
MANOVA (MANCOVA)
Linear/Logistic Regression
Multilevel Regression (linear or logistic)
Survival Regression
Poisson Regression
Percentage of change
Nonparametric tests
Nonparametric Matched Test
Nonparametric Unmatched Test
Other Parametric Matched Test
Other Parametric Unmatched Test
No Statistical Test

42. Describe for each outcome of interest (MSK) the observed intervention effects. (Be brief and concise i.e., enter "effect size", "risk ratio", "rate differences, "mean differences" etc, the actual number and associated outcome). If there is more than one outcome of interest, please number and identify each using the same names you used in Questions 25 and 27. For administrative data, multiple types of information might be reported. For self-reported and clinical data, please report by body part. PLEASE use notation HWE, NS, UB, LB, LKF, NAB, or O)

***Organize your description of interventions according to I_1, I_2, C, I_1C, and I_2C**

43. Remark on the findings or enter information that is unique about the study that may not be adequately captured in the other DE questions. Be clear and concise.

Housekeeping questions:

44. Check the names of both DE reviewers for this study.
BA, SB, BE, DG, LP, JT, AW

45. Is this the consensus – final - version of the DE form? Please select "no" until consensus has been completed.

Yes
No

Appendix G

Exclusions at Level 1a and 1b

Review phase	Exclusion criteria						Total
	Intervention in health-care setting	From peer-reviewed publication	Language: English, Spanish, Swedish or French	Post only with no control group	Individual health data	Outcome musculoskeletal symptoms/disorders/injuries	
Level 1a	8256	17	34	17	11	15	8350
Level 1b	45			4	10	6	66
Total Excluded	8301	17	34	21	21	21	**8416**

Total exclusions: Level 1a + Level 1b (8350 + 66 = 8416)
8465 - 8416 = 49 articles that we reviewed at QA and DE phases