Downloaded from www.sjweh.fi on March 29, 2024



Scand J Work Environ Health 2012;38(3):193-208
https://doi.org/10.5271/sjweh.3259
Published online: 01 Nov 2011, Issue date: May 2012

A systematic review of the effectiveness of occupational health and safety training

by Robson LS, Stephenson CM, Schulte PA, Amick III BC, Irvin EL, Eggerth DE, Chan S, Bielecky AR, Wang AM, Heidotting TL, Peters RH, Clarke JA, Cullen K, Rotunda CJ, Grubb PL

Affiliation: Institute for Work & Health, 481 University Ave, Suite 800, Toronto, ON M5G 2E9, Canada. lrobson@iwh.on.ca

Refers to the following texts of the Journal: 1999;25(3):255-263
2003;29(4):288-296

The following article refers to this text: 2015;41(5):478-485

Key terms: education; evaluation; intervention; occupational health; occupational health and safety; occupational health and safety training; OHS; OHS training; prevention; primary prevention; review; training

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/22045515

Additional material

Please note that there is additional material available belonging to this article on the *Scandinavian Journal of Work, Environment & Health* -website.

This work is licensed under a Creative Commons Attribution 4.0 International License.

Print ISSN: 0355-3140 Electronic ISSN: 1795-990X

Review

Scand J Work Environ Health. 2012;38(3):193–208. doi:10.5271/sjweh.3259

A systematic review of the effectiveness of occupational health and safety training

by Lynda S Robson, PhD,[1] Carol M Stephenson, PhD,[2] Paul A Schulte, PhD,[2] Benjamin C Amick III, PhD,[1, 3] Emma L Irvin, BA,[1] Donald E Eggerth, PhD,[2] Stella Chan, MSc,[4] Amber R Bielecky, MSc,[1] Anna M Wang, BScN,[1] Terri L Heidotting, EdD,[5] Robert H Peters, MSc,[6] Judith A Clarke, MA,[1] Kimberley Cullen, MSc,[1] Cathy J Rotunda, EdD,[7] Paula L Grubb, PhD[8]

Robson LS, Stephenson CM, Schulte PA, Amick BC III, Irvin EL, Eggerth DE, Chan S, Bielecky AR, Wang AM, Heidotting TL, Peters RH, Clarke JA, Cullen K, Rotunda CJ, Grubb PL. A systematic review of the effectiveness of occupational health and safety training. *Scand J Work Environ Health.* 2012;38(3):193–208. doi:10.5271/sjweh.3259

Objectives Training is regarded as an important component of occupational health and safety (OHS) programs. This paper primarily addresses whether OHS training has a beneficial effect on workers. The paper also examines whether higher engagement OHS training has a greater effect than lower engagement training.

Methods Ten bibliographic databases were searched for pre-post randomized trial studies published in journals between 1996 and November 2007. Training interventions were included if they were delivered to workers and were concerned with primary prevention of occupational illness or injury. The methodological quality of each relevant study was assessed and data was extracted. The impacts of OHS training in each study were summarized by calculating the standardized mean differences. The strength of the evidence on training's effectiveness was assessed for (i) knowledge, (ii) attitudes and beliefs, (iIi) behaviors, and (iv) health using the US Centers for Disease Control and Prevention's *Guide to Community Preventive Services,* a qualitative evidence synthesis method.

Results Twenty-two studies met the relevance criteria of the review. They involved a variety of study populations, occupational hazards, and types of training. Strong evidence was found for the effectiveness of training on worker OHS behaviors, but insufficient evidence was found of its effectiveness on health (ie, symptoms, injuries, illnesses).

Conclusions The review team recommends that workplaces continue to deliver OHS training to employees because training positively affects worker practices. However, large impacts of training on health cannot be expected based on research evidence.

Key terms education; evaluation; intervention; OHS; OHS training; prevention; primary prevention.

The burden of workplace injuries, illnesses, and fatalities on society is large (1, 2). One common approach to mitigate such adverse outcomes is occupational health and safety (OHS) training. About 15% of the Canadian working population receives OHS training each year (3). Indeed, training is widely regarded as an important component of OHS programs (4–7). However, definitive information on the effectiveness of OHS training is still developing.

OHS training refers to planned efforts to facilitate the learning of OHS-specific competencies (8). Such training typically consists of instruction in hazard recognition and control, safe work practices, proper use of personal protective equipment, and emergency procedures and preventive actions. It may also guide workers on where to find additional information about potential hazards. Finally, OHS training can also empower workers and managers to become more active in making

1 Institute for Work & Health, Toronto, ON, Canada.
2 Education and Information Division, National Institute for Occupational Safety and Health, Cincinnati, OH, USA.
3 School of Public Health, University of Texas, Houston, TX, USA.
4 Cancer Care Ontario, Toronto, ON, Canada.
5 Research Compliance, Cincinnati Veterans Affairs Medical Center, Cincinnati, OH, USA.
6 Office of Mine Safety and Health Research, National Institute for Occupational Safety and Health, Pittsburgh, PA, USA.
7 Education and Information Division, National Institute for Occupational Safety and Health, Morgantown, WV, USA.
8 Division of Applied Research and Technology, National Institute for Occupational Safety and Health, Cincinnati, OH, USA.

Correspondence to: Lynda S Robson, Institute for Work & Health, 481 University Ave, Suite 800, Toronto, ON M5G 2E9, Canada. [E-mail: lrobson@iwh.on.ca]

changes that enhance worksite protection (9). Training interventions sometimes include additional components besides instruction or practice, such as goal-setting, to enhance effectiveness. The distinction between training and education is not universally agreed upon. For some, in contrast to education, training must include a hands-on practice component. For this review, a broad definition of training has been adopted so that training with or without a hands-on practice component is included.

Early attempts to review the OHS training literature were hampered by a lack of evaluative information and there were concerns about its internal validity (10, 11). By the time of the Johnston et al review (12), a substantial number of studies with quasi-experimental designs (13) had accumulated. These studies evidenced that training increases knowledge and targeted OHS behaviors, but the review did not look separately at health-related outcomes (eg, injuries), instead pooling them with true behavior outcomes. A second review by Cohen & Colligan (9) similarly found that the majority of the 80 studies reviewed showed positive effects (not defined) on knowledge and behaviors, rather than mixed or no effects. Of the 80 studies, 42 had quasi-experimental or experimental designs. Injury or illness outcomes were available for about 20 of the studies and they also showed mostly positive results. However, the authors did not feel as confident attributing changes in injury and illness to the training intervention because of threats to the internal validity of the evidence.

When this project was initiated in 2005, no systematic literature review (14) on training effectiveness had yet been published. Further, a preliminary scan indicated that many randomized controlled trials (RCT) had been published in the previous decade. A decision was therefore made to undertake a systematic review of the trial literature published since 1996, the cut-off date in the Cohen & Colligan review (9). There were two primary research questions addressed by the review. The main focus of this paper is concerned with the first question: "does OHS training have a beneficial effect on workers (eg, increase OHS knowledge, improve OHS attitudes, improve OHS behaviors, or protect health)?" In addressing this question, the review team was guided by a conceptual model (figure A in appendix) that drew from existing models (15–18). It depicts training as having an immediate effect on outcomes such as knowledge, attitudes, and behavioral intentions. These outcomes eventually affect behaviors and hazards on the job, which in turn impact outcomes measured in the longer term, such as workplace injuries and illnesses. The model also indicates that these effects are determined by various aspects of the training, trainees, and the workplace environment. Several other systematic reviews reporting on OHS training effectiveness have been published in the period since 2005 (19–23) especially with regards to the prevention of musculoskeletal disorders. The findings of these reviews will be described in the discussion in relation to our findings.

The second question of this review is also reported here, but more briefly since the primary studies needed to address it were scant. The question is "does higher engagement OHS training have a greater beneficial effect on workers than lower engagement training?" This question responded to a systematic review by Burke et al (23) published in 2006. These researchers showed that OHS training had a greater impact when the method of training involved more learner engagement, the theoretical basis for which they elaborated upon elsewhere (24, 25). The researchers operationalized low-engagement training methods as passive, information-based methods, such as lecture or video. High-engagement methods included behavioral modeling, simulation, and hands-on training.

Methods

The methodological steps were the following: conduct literature search, indentify relevant studies, assess the methodological quality of the relevant studies, extract data (evidence) from publications, and synthesize evidence. The methodology is described in detail in a technical report (26) and summarized here.

Literature search

The review team searched ten electronic databases for studies published in English or French during 1996–2007: MEDLINE, EMBASE, PsycINFO, Eric, CCOHS, Dissertation Abstracts, Agricola, Social Science Abstracts, Health and Safety Science Abstracts, and Toxline. The search terms fell into four categories: work-relatedness (5 terms, eg, worker), education/training intervention (15 terms, eg, education, training), OHS outcomes or factors affecting effectiveness (36 terms, eg, accidents, occupational health, safety), and between-group evaluation designs (3 terms, eg, random, comparison). The terms within each category were combined with the Boolean operator *OR* and the categories were combined with the Boolean operator *AND*. The complete set of terms is shown in table A in the appendix. The search also used the following terms to exclude abstracts: health promotion, diet, exercise, smoking, weight loss, and addiction. The electronic search was supplemented by asking several external experts for relevant citations of published or in-press journal articles and by reviewing the reference lists of articles passing this review's relevance assessment screen. The search was first conducted in August 2005 and updated in November 2007.

Relevance assessment

In the first stage of screening for relevance, publication titles and abstracts were reviewed by a single researcher using the criteria shown in table B in the appendix (http://www.sjweh.fi/data_repository.php). In the second and third stages, titles and abstracts, then full publications, respectively, were reviewed by two reviewers independently assessing the following more focused set of criteria: (i) Is the study concerned with the effectiveness of a worker- or workplace-centered OHS training intervention aimed at the primary prevention of workplace injury and/or illness? (ii) Is the study a randomized trial? (iii) Are there pre- and post- measures available for each study group? (iv) Does the study examine a worker, firm, or societal outcome related to OHS training? (v) Is the study published in a scientific peer-reviewed journal? Detailed operationalization of the criteria are available elsewhere (26, pp102–8). Of particular note is that the following types of interventions were excluded from the review: social marketing, secondary prevention, stress management, health promotion, physical fitness, and multi-component interventions when the training component could not be isolated. Any disagreements between reviewers were resolved through consensus and, if required, a third opinion. A joint negative response to any question resulted in the article's exclusion from the review. Multiple articles based on the same study were grouped together for subsequent steps of the review.

Methodological quality assessment

The next step of the review assessed the methodological quality of the relevant studies. Two reviewers independently applied the review's quality assessment instrument to each outcome in a study and met to resolve any disagreements. Following Hayden et al (27), the instrument developed for the project assessed quality in stages; a copy of it is published elsewhere (26, p109). First, sixteen items, based on established instruments (28, 29) adapted to the training literature and refined through pretesting, were used to assess specific biases. The items were concerned with study design, adequacy of randomization method, concealment of intervention allocation, group similarity at baseline, equivalency of any effects of withdrawals across groups, monitoring of intervention implementation, contamination, planned co-interventions, unplanned co-interventions, blinding of outcome assessor, similarity across groups of outcome assessment method, outcome measure validity, outcome measure reliability, appropriateness of statistical testing and procedures, appropriate statistical adjustment for group differences, and intention-to-treat analysis. Second, reviewers were asked to provide summary assessments for each of four domains of potential bias into which the individual items were grouped (ie, comparability of study groups, intervention implementation, outcome assessment, statistical analysis): reviewers were asked whether they were confident (yes=0, partly=1, no=2) that the potential for bias in a particular domain was minimized. The four domain-level assessments were converted to a single limitations score by summing. The limitations scores therefore had a range from 0=no limitations to 8=most limitations. This transformation yielded a metric analogous to that used in the Centers for Disease Control and Prevention's *Guide to Community Preventive Services* (30) (hereafter the Guide), applied here for evidence synthesis (see below). As such, a limitations score of 0–1=good methodological quality; 2–4=fair; and ≥5=limited.

Data extraction and coding

Two reviewers independently performed data extraction and coding using a form developed by the research team (26, p124), with discrepancies resolved using consensus. Items in the form were concerned with the research questions, study design, study population, group characteristics at baseline and follow up, interventions, contamination, co-interventions, outcome measurement methods, results, and statistical analysis. Training interventions were categorized by level of learner engagement, based on the method used in the meta-analysis by Burke et al (23). Training was considered "low engagement" when it only involved the presentation of factual material by an expert source, with no or little interaction (eg, lectures with minimal interaction, videos, computer instruction with no interaction or feedback). Training was considered "medium engagement" when there was a stronger element of interactivity, with or without feedback (eg, lectures with discussion afterwards, computer instruction with interaction, and discussions or problem-solving activities presented in an interactive format). Training was considered "high engagement" when there was an application of concepts from training in a real or simulated environment (eg, behavioral modeling, hands-on training in simulated or actual work environments). Training was also coded according to the category of hazard it addressed (ergonomic, safety, chemical, biological, physical). Outcomes were classified as belonging to one of four categories: (i) knowledge; (ii) attitudes and beliefs (ie, attitudes, beliefs, perceived risk, self-efficacy, behavioral intentions); (iii) behaviors (ie, behaviors, behavior-dependent hazards, behavior-dependent exposures); (iv) health (ie, early symptoms, injury, illness).

When a single study reported on multiple measures in an outcome category of interest, measures were selected from the data extraction forms by the lead author for further synthesis using the following set

of rules. First, measures were automatically excluded if they had not been measured at baseline or in both groups being compared. Second, measures considered most appropriate to the intent of the intervention and evaluation were selected in preference to others. For example, the measure of upper-body musculoskeletal symptoms was used in preference to lower or total body musculoskeletal symptoms when the intervention focus was office ergonomics. The third rule was to favor independent-rater assessments (eg, clinician or external observer) over worker self-reports, when both were available (31–34). If more than one outcome measure remained after this selection procedure was applied, they were all reported in the detailed results tables in the appendix (tables C–G).

The effect of an intervention on an OHS outcome was first summarized in terms of its direction and statistical significance based on the analysis of the authors of the primary study. Effect size was computed to facilitate evidence synthesis. Since the most common type of outcome data in the reviewed studies was continuous, the standardized mean difference (*d*) was selected as the metric. It was computed from post-intervention data by dividing the between-group difference in means by the pooled standard deviation (35). When study results were expressed in a form other than continuous (ie, prevalences, ordinal frequencies, odds ratios, rates), they were transformed to *d* using established methods (35–38). Effect size was computed only after confirming that the two groups were equivalent at baseline with respect to the outcome (ie, the probability that groups were different was P>0.05). Standard errors were calculated using established formulas (35–39).When some of the data required to calculate effect sizes was missing from published study results, the original authors were sent a request for these data, which was repeated once if they did not respond. We ultimately obtained additional data for the analysis from one research group by these means (40).

Evidence synthesis

A qualitative synthesis of evidence was undertaken, using the methods of the Guide (30, 41). This methodology involves constructing a "body of evidence" with regards to an outcome of interest from the results of multiple relevant studies of fair or good methodological quality (defined above). If feasible, the effects within the body of evidence are summarized by their median and interquartile range. Statistical pooling of effects is used when appropriate. The Guide's algorithm assesses the strength of a body of evidence as insufficient, sufficient, or strong based on consideration of five of its aspects: (i) methodological quality of study results, (ii) study design, (iii) quantity of studies, (iv) consistency of effects (regarding their direction), and (v) size of effect. In contrast to some synthesis methods, the statistical significance of individual study results does not play a role in the algorithm's assessment.

In this review, separate bodies of evidence were constructed for each of the four outcome categories (knowledge, attitudes and beliefs, behaviors, and health.) Further subdivision of the results according to population, intervention features and types of outcomes was not pursued, due to a relative scarcity of data. The consensus opinion of the research team was that statistical pooling was inappropriate in this review, due to the heterogeneous nature of the subjects, interventions, and outcomes in the studies identified as relevant to the review's first research question. The algorithm used in this study to determine the strength of a body of evidence (table 1) is a simplification of the Guide's algorithm, which results when there are only RCT in the body of evidence, as in this review. As such, only four aspects of a body of evidence are considered here: (i) methodological quality of study results, (ii) quantity of studies, (iii) consistency of effects (regarding their direction), and (iv) size of effect. The criteria for sufficient and large effect sizes were set by review team members with experience in OHS training intervention research prior to applying the algorithm to the bodies of evidence. Table 1 shows the effect-size criteria used in the application of the algorithm to the evidence addressing research question 1 (ie, evidence from training versus no-training control contrasts). The corresponding criteria used with evidence addressing research question 2 (ie, evidence from lower versus higher engagement training contrasts) were 0.25 times those in table 1. They can be found in table H in the appendix.

Transforming the effect-size data available in the detailed evidence tables (ie, tables C–G in the appendix) into the final bodies of evidence (ie, table 2 and table I in the appendix) involved data exclusion and reduction. First, in keeping with the Guide's synthesis method, data of limited methodological quality were excluded (and those of fair or good quality were retained). Second, to avoid over-representing studies with many reported outcomes, conceptually similar outcomes from the same study were collapsed by reporting only their median. For example, three effects of training on musculoskeletal symptoms in the upper spine were reported in table F (see appendix) for the Greene et al study (34), corresponding to the intensity, frequency, and duration of symptoms. These were summarized in the final evidence synthesis table (table 2) by the median value, 0.27.

Two types of post hoc sensitivity tests were conducted on the evidence synthesis findings concerned with research question one (ie, tables 2 and 3). In one test, instead of allowing multiple, yet conceptually distinct, effects from the same study to contribute to the final body of evidence in a major outcome category, only the median of any multiple effects contributed so

Table 1. Algorithm applied to a body of evidence from training versus control studies to determine its strength. [*d*=standardized mean difference; IQR=interquartile range]

Strength of evidence	Methodological quality [a]	Minimum quantity	Consistency of effects [b]	Effect size criterion [c]
Strong	Good	≥2 studies	IQR (or range) of *d* does not include zero	Sufficient: Knowledge = 1.0 Attitudes & Beliefs = 0.5 Behaviors = 0.4 Health = 0.15
	Good or Fair	≥5 studies	IQR (or range) of *d* does not include zero	Sufficient: Knowledge = 1.0 Attitudes & Beliefs = 0.5 Behaviors = 0.4 Health = 0.15
	Meet methodological quality, quantity and consistency criteria for sufficient but not strong evidence			Large: Knowledge = 1.5 Attitudes & Beliefs = 1.0 Behaviors = 0.8 Health = 0.3
Sufficient	Good	Not applicable	IQR (or range) of *d* does not include zero	Sufficient: Knowledge = 1.0 Attitudes & Beliefs = 0.5 Behaviors = 0.4 Health = 0.15
	Good or Fair	≥3 studies	IQR (or range) of *d* does not include zero	Sufficient: Knowledge = 1.0 Attitudes & Beliefs = 0.5 Behaviors = 0.4 Health = 0.15
Insufficient	The criterion in any one of the four domains not met			

[a] Methodological quality categories for studies: Good (0–1 limitations score), Fair (3–4), and Limited (≥5 or more).
[b] Interquartile range of *d* was determined when there were ≥5 effect sizes in the body of evidence; otherwise the full range was used.
[c] Criteria for sufficient and large effect sizes were defined by the research team. The median effect size of a body of evidence needed to be equal to or greater than the criterion.

Table 2. Final bodies of evidence on training effectiveness from training versus no-training control trials for each of knowledge, attitudes, behaviors and health [*d*=standardized mean difference; FB=feedback; SE=standard error]

Authors; intervention [a] (specific levels of learner engagement [b]; number of sessions)	Methodological quality [c]	Calculated effect sizes							
		Knowledge		Attitudes		Behaviors		Health	
		d	SE	*d*	SE	*d*	SE	*d*	SE
Banco et al (55); box cutter safety (H;1); 12 months	Fair							0.06	0.23
Brisson et al (31); office ergonomics, <40 years (H;2); 6 months	Fair					i) 0.30* ii) 0.33*			
Brisson et al (31); office ergonomics, ≥40 years (H;2); 6 months	Fair					i) 0.18* ii) 0.28*			
Eklöf et al (40), Eklöf & Hagberg (46) office ergonomics, individual FB (H;1); 6 months	Fair					i) 1.09 ii) 0.95	0.51 0.56	-0.13	0.47
Eklöf et al (40), Eklöf & Hagberg (46); office ergonomics, supervisor FB (H;1); 6 months	Fair					i) 1.71 ii) 1.35	0.59 0.50	-1.34	0.53
Eklöf et al (40), Eklöf & Hagberg (46); office ergonomics, group FB (H;1); 6 months	Fair					i) 1.98 ii) 2.36	0.53 0.63	-0.37	0.48
Greene et al (34); office ergonomics (H;2); 2 weeks	Fair	1.45	0.25	i) 0.82 ii) 0.87	0.23 0.23	1.16	0.23	i-iii) -0.12* iv-vi) 0.27*	
Harrington & Walker (45); office ergonomics (L;1); 0 weeks	Fair	3.58	0.46						
Held et al (33); skin care in wet work (H;3); 5 months	Good					0.42*		0.05	0.12
Rasmussen et al (49); farm safety (H;2); 6 months	Good					+ [e]		0.06	0.13
Wright et al (52); Universal Precautions (M;1); 2 months	Fair					1.25	0.28		

[a] More details of interventions are in table 4.
[b] Specific levels of learner engagement contrasted in the study, determined as described in methods section: low (L), medium (M), or high (H).
[c] Refers to the assessed methodological quality of the specific outcome data included in the table: Good, 0-1 methodological limitations score; Fair, 2-4.
[d] *d* values and SE from tables C–F were carried over to table 2 when the methodological quality of the data was fair or good. When a study reported multiple measures of a similar concept, only their median reported here (indicated by *). The standard error for each of the contributing individual measures can be found in tables C–F. A positive value for d indicates the higher engagement training is more effective than the lower engagement training. Numbered bullets preceding individual effect sizes correspond to those used in tables C–F.
[e] Effect size not calculable, but direction is indicated.

Table 3. Determination of the strength of evidence for training's effect on each outcome. [IQR=interquartile range; N=number of studies]

Outcome	Summary of body of evidence in table 2				Assessment of body of evidence using evidence synthesis algorithm [a]			Strength of evidence [b]
	N		IQR	Median of effect sizes	Number of good or fair studies	Consistency	Median of effect sizes	
	Good	Fair						
Knowledge	0	2	1.45–3.58	2.52	Too few	Consistent	Large	Insufficient
Attitudes	0	1	0.82–0.87	0.84	Too few	Consistent	Sufficient	Insufficient
Behaviors	2	4	0.33–1.35	1.09	Enough	Consistent	Large	Strong
Health	2	3	-0.25–0.06	-0.04	Enough	Inconsistent	Less than sufficient	Insufficient

[a] Descriptors indicate the result of assessing a feature of a body of evidence using the evidence synthesis algorithm shown in table 1.
[b] The resulting conclusion about strength of evidence following the assessment of a body of evidence using the algorithm.

that each study was represented only once. In the other sensitivity test, instead of just including only the study results of fair or good methodological quality, as specified by the Guide, we also included the studies of limited methodological quality.

Other post hoc analyses

Two other post hoc analyses were conducted. The first was concerned with the involvement of commercial funding sources in the primary studies. The lead author examined the affiliations and acknowledgement sections of the journal articles included in the review to identify commercial organizations. The potential impacts on the review's conclusions were considered. The second analysis was concerned with the selective reporting of outcomes in the original studies. The first author compared the outcome measures mentioned in the methods sections of articles against the outcomes identified by reviewer pairs in the data extraction step in order to see whether there were likely to have been measured but not reported outcomes.

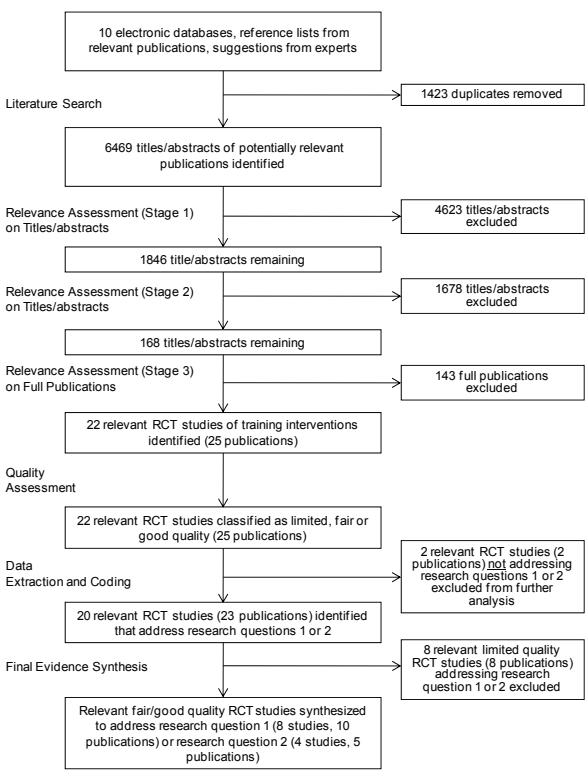


Figure 1. An overview of the review process

Results

Overview of literature search and screening

The electronic search of ten databases yielded 7801 citations; and the manual search of reference lists from relevant articles and experts generated another 91 potentially relevant citations. After removing duplicates, 6469 unique citations remained for relevance screening, from which 22 RCT of OHS training were identified (see figure 1). Their features are summarized in table 4.

Description of eligible studies

The 22 studies most often addressed ergonomic hazards (10 studies), but there were ≥2 studies addressing each of the other 4 hazard categories (safety, chemical, biological, physical). The 2 most frequently studied occupational groups were healthcare and office workers (6 studies each) and the remaining occupations were varied. Usually, the study population was comprised mostly of experienced workers, but in three cases those studied were still trainees (32, 42, 43).

In total, 36 training interventions were included in the 22 trials. Typically, interventions involved multiple methods to deliver the training content. Most common were lectures (20 interventions), printed materials (14), hands-on practice (14), and feedback (12). It should be

Table 4. Summary of study characteristics. [I=intervention; C=control.]

Authors	Hazard category [a] / Training type	Interventions (level of engagement [b]) and randomization method	Occupation/ Workplace/ Country (N=number of participants [c])	Outcomes measured (time of follow-up) [d]
Arnetz & Arnetz (50)	S / Prevention of violence towards health care workers	I: Continuous registration of violent incidents on checklist; structured program for regular discussion of specific violent incidents registered in workplace, over one year. Written guidelines for the feedback/group discussion, based on points summarized on checklist. (M) C: No training control (continuous registration of violent incidents only). 47 workplaces randomized.	Nurses/ Health care workplaces incl emergency depts, geriatric psychiatric and home health care sites/ Sweden (N=1500*)	B (0 wks)
Banco et al (55)	S / Safety training (use of cutters)	I1: Safety training and use of new safety cutters, including instruction and practice. One session; 15 minutes. (H) (Intervention I1 not used in analysis.) I2: Safety training and use of old cutters. One session; 15 minutes. (H) (Intervention I2 used in analysis.) C: No training control (use of old cutters only). 9 similar-sized workplaces randomized.	Supermarket workers/ Super-markets/ USA (N=900*)	H (12 months)
Bohr (53, 54)	E / Office ergonomic training	I1: Participatory education (hands-on demo, problem solving, application to work area). One session; two hours. (H) I2: Traditional education (lecture, informational handout, Q&A session). One session; one hour. (L) C: No training control. Individuals randomized.	Computer users/ Centralized reservation facility in transportation company/ USA (N=154)	B, H (3, 6 and 12 months)
Brisson et al (31)	E / Office ergonomic training	I: Ergonomics training (demonstrations, simulations, discussions, lectures and self-diagnosis on work stations). Two sessions; three hours each, at a two-week interval. (H) C: No training control. 40 work units, stratified by size and nature of work, randomized.	Clerical workers (computer users)/ University/ Canada (N=658*)	B, H (6 months)
Duffy & Hazlett (32)	E / Preventive voice care training	I1: Direct voice care training (vocalization, posture, respiration, release of tension in vocal apparatus, resonance and voice projection). One session; duration not reported. Also received one session of indirect voice care training. (H) I2: Indirect voice care training (information on voice production, factors associated with healthy voice). One session; duration not reported. (L) C: No training control. Individuals randomized.	Teacher trainees/ Schools/ Ireland (N=55)	H (2 months)
Eklöf et al (40), Eklöf & Hagberg (46)	E / Ergonomic and psychosocial work environment intervention	I1: Feedback on individual and on group, directed to individuals; related to normative info; given orally & with printed reports. Discussion. One session; 38 minutes. (H) I2: Same as I1, except feedback is only on the group and is directed to supervisors only. One session; 61 minutes. (H) I3: Same as I1, but feedback is only on the group and is directed to the group. One session; 85 minutes. (H) C: No training control. 36 workgroups, stratified by organization (n=8), randomized.	White collar computer users/ Nine organizations; various sectors/ Sweden (N=396)	B, H (6 months)
Gray et al (63)	E / Lift & transfer training	I: Educational program (demo, videos, lectures, practice sessions, resource team, binder for feedback, manual available & pictograms). Five sessions; four hours per session weekly for five weeks. (H) C: No training control. 2 comparable work units randomized.	Nursing personnel/ Long-term care and rehabilitation hospital/ Canada (N=250*)	K (0 weeks)
Greene et al (34)	E / Office ergonomic training	I: Active ergonomic training intervention (didactic interactions, discussion and problem-based activities). Two sessions; three hours per session. (H) C: No training control, but received intervention at week 4 of the study period. Individuals randomized.	Computer users/ University/ USA (N=87)	K, A, B, H (2 weeks)
Harrington & Walker (45)	E / Home office ergonomics training	I: Computer-based training, screens containing interaction, animation or a colour graphic to keep learner focused. Includes screen-to-screen navigation so learner can move forward, pause, repeat a topic or quit the lesson. Program "combines text, graphics, color illustrations, animation, and sound, to provide a fully interactive media-rich learning environment." One session; 45 minutes. (L) C: No training control. Individuals randomized.	Teleworkers/ Home or telecommuting centres (business, academic, government agency)/ USA (N=102)	K, A (0 weeks)
Harrington & Walker (56)	P / Fire safety training	I1: Computer-based instruction; screens contained narration, interaction, animation or video; some with questions and interactive games. Two sessions; average 30 minutes each. (M) I2: Instructor-led (lectures & printed materials). Two sessions; average 40 minutes each. (L) C: No training control. Individuals, stratified by shift and department, randomized.	All staff/ Life-care community facility/ USA (N=141)	K, A (0 weeks)
Held et al (33)	C / Skin care program	I: Train-the-trainer. Education on skin care (video, instruction, role play, booklet, reinforcement meeting) directed at trainers. Two sessions; four hours each with 14 weeks in between and one meeting with instructors six weeks after last session for reinforcement. (H) C: No training control. 7 workplaces, stratified by size, randomized.	"Wet workers" (nurses, cleaners, kitchen, staff)/ Geriatric care facilities/ Denmark (N=375)	B, H (5 months)

Table 4. Continued

Authors	Hazard category [a] / Training type	Interventions (level of engagement [b]) and randomization method	Occupation/ Workplace/ Country (N=# participants [c])	Outcomes measured (time of follow-up) [d]
Hickman & Geller (44)	S / Safety self-management (mining)	I1 and I2: Safety self-management training, including goal-setting, ways to self-reward for meeting goals, group exercises to demonstrate personal use of self-monitoring form (one two-hour session). Four weeks of self-recording and feedback. I1 had self-recording of intended safety-related work behaviors before beginning of shift. I2 had self-recording of safety-related work behaviors after shift. (H) Individuals, matched on baseline behaviors, randomized.	Miners/ Above-ground quarry/ USA (N=15)	B (0 weeks)
Hong et al (62)	P / Hearing protection training	I1: Tailored with feedback: Computer-based training tailored to worker's hearing test, self-reported hearing protective device (HPD) use, self-efficacy and perceptions. Reinforcement of any HPD use. Practice with HPDs. Handout with hearing test results, individualized information and opportunity to ask questions. One session; 43 minutes. (H) I2: Commercial video with feedback: Computer-delivered video meeting OSHA requirements. Handout with hearing test results, standard information and opportunity to ask questions. One session; 33 minutes. (M) Individuals randomized.	Construction workers, heavy equipment / USA (N=612)	A (0 and 12 months) B (12 months)
Jensen et al (51)	E/ Lifting technique training	I: Transfer technique intervention: train-the-trainer. Two four-hour sessions of mainly classroom education, followed by observation and feedback in work setting. Training aimed to reduce biomechanical risks. (H) C: Control received training in topic of their "choice in matters unrelated to the intervention program." 19 work groups, stratified by ward, randomized.	Health care workers/ Eldercare services/ Denmark (N=210)	B (3, 6 and 9 months) H (15 months)
Löffler et al (42)	C/ Skin care program	I1: Lecture, group problem-solving, practice with individual feedback. Seven sessions over three years; duration of sessions unknown. (M) I2: Informational paper. One time. (L) 14 nursing schools randomized.	Nursing students/ Health care organizations/ Germany (N=521)	H (about 2.5 yrs after 1st session)
Lusk et al (65, 66)	P / Hearing protection training	I1: Tailored: Computer-based training tailored to worker's self-reported practice. Used factual, cognitive approaches; demonstration; directed practice; vicarious experience; persuasion and role-modeling techniques. Presented in interactive format, with feedback. One session; 30 minutes. (H) I2: Non-tailored: As above, but delivered to all participants in a uniform manner. One session; 30 minutes. (H) I3: Control:Video. One session; 30 minutes. (L) In each intervention group, there were four possibilities: Boosters at 30 days; Boosters at 30 & 90 days; Boosters at 90 days; No boosters. Individuals randomized.	Factory workers/ Large automotive factory/ USA (N=2219)	B (6 to 18 months)
Perry & Layde (57)	C / Safe pesticide handling	I1: Education intervention (lecture, slides, presentation by respected area farmer, demonstration and opportunity for hands-on practice). One session; three hours. (H) I2: Standard re-certification meeting for pesticide applicators. One session. (L) Individuals randomized.	Farmers/ Private dairy farms/ USA (N=400)	K, B (6 months)
Rasmussen et al (49)	S / Safety in farming	I: Farm safety check (feedback, written report with recommendations) (1/2 day) and safety course (lecture, meeting with injured farmers, demonstration, discussion of recommendations, action planning) (one day, within one to four weeks after farm safety check). (H) C: No training control 201 farms randomized.	Farmers/ Farms/ Denmark (N=990)	B,H (6 months)
Rizzo et al (47)	E / Preventive office ergonomic training	I1: Instructor-directed: Seminar, video, pamphlets, and concluding discussion in which instructor summarized and responded to individual questions. One session; one hour. (L) I2: Self-directed (videos and pamphlets). One session; 45 minutes. (L) C: No training control. 3 comparable work groups randomized.	Computer users/ Information technology/ USA (N=150*)	K (15 months)
Van Poppel et al (58)	E / Ergonomic training on lifting (back pain prevention)	I: Lifting instruction, including theory and practice, and including instruction in individual work settings. Three sessions (two hours, 1.5 hours, 1.5 hours respectively) over 12 weeks. (H) C: No training control. 36 work groups, stratified by work type (N=6), randomized.	Manual material handlers / Cargo department of airline company/ Netherlands (N=312)	H (6 months)
Wang et al (43)	B / Blood-borne pathogens prevention training	I1: Educational intervention about blood-borne pathogens (lecture, video and printed materials). One session; 60-minute lecture, 20-minute video. (L) I2: Standard education about vaccination only. (L) 2 classes randomly assigned to I1 or I2.	Nursing students/ Hospital/ China (N=106)	K, B (4 months)
Wright et al (52)	B / Universal Precautions training	I: Computer-assisted instruction with problem-solving scenarios and feedback. Number of sessions unknown; self-paced. (M) C: No training control. Individuals randomized.	Registered nurses/ Large teaching hospital/ USA (N=60)	B (8 weeks)

[a] Hazard type: biological (B), chemical (C), ergonomic (E), physical (P), safety (S).

[b] Level of learner engagement: low (L), medium (M), high (H).

[c] Initial number of study participants: the size of the study sample following exclusions on the basis of eligibility, initial inability to contact and initial refusal to participate. * indicates that an estimation was made by the reviewers or an approximation was made by the authors.

[d] Outcome categories: knowledge (K); attitudes and beliefs (A); behaviors (B); health (H).

noted that the number of sessions involved in the training was usually modest. Of the 34 interventions where the number of sessions could be assessed, 23 involved only a single session; 8 involved two sessions; and 1 intervention each involved 3, 5, and 7 sessions, respectively. The length of sessions was also modest. Of the 28 sessions where duration could be assessed, 12 lasted <1 hour, 9 were 1–2 hours, and 7 lasted ≥3 hours.

Of the 22 trials, 16 included a comparison between a study group receiving training and a control group receiving none, thereby addressing research question 1. Three of these trials and four of the remaining six trials included a comparison between a study group receiving lower engagement training and a group receiving higher engagement training, thereby addressing research question 2. The two remaining trials (of the 22 relevant trials) (43, 44) addressed neither of the two research questions because they involved only a comparison of two training interventions with the same level of engagement.

Methodological quality

Table J in the appendix summarizes the assessed methodological quality of all 22 trials. The study sample size varied widely (range 15–2219, median 209), as did the assessed methodological quality (limitations score range 0–8, median 4). Review of the domain-level assessments shows that reviewers often lacked the confidence that the risk of bias was minimized (ie, "partly" or "no" response options selected more often than "yes"). Analysis at the level of each of the quality assessment criteria revealed this arose from inadequate reporting of a variety of study aspects (randomization method, effect of withdrawals on group similarity, intervention implementation, contamination, co-occurring workplace events, statistical adjustments to correct for group differences), lack of blinding of outcome assessors (related to the heavy use of self-report measures in these studies), and lack of consideration of the effect of participant withdrawals on results.

Effects of training on knowledge

As shown in table C in the appendix, data were available from five training versus no-training control trials that examined the effect of training on knowledge. All interventions showed positive, statistically significant results, and the calculated effect sizes were large. Results from only two (34, 45) of the five studies were considered to be good/fair methodological quality (ie, limitations score=0–4) and therefore only these are represented in the final body of evidence in table 2. Both of these studies were concerned with office ergonomics. The study by Greene et al (34) involved two three-hour sessions of didactic presentations, discussion, and problem-based activities, delivered to various computer users in a university. The other was a 45-minute media-rich computer-based training (45) delivered to a variety of teleworkers. The median *d* derived from these two studies (2.52) far exceeded the evidence synthesis algorithm's criterion of sufficient (1.0) or large (1.5) and the range of *d* did not include zero, indicating consistency of the direction of effects. However, since there were only two fair quality studies, application of the algorithm classified the reviewed evidence on the effectiveness of training on knowledge as insufficient (table 3).

Effects of training on attitudes and beliefs

Synthesis results for the evidence on attitudes and beliefs followed the same pattern seen for the evidence on knowledge. Only 3 of the 22 studies examined attitudes and beliefs (table D in the appendix) and the effects in this category ranged from small and negative (and statistically insignificant) to large and positive (and statistically significant). Only the two effect estimates derived from the Greene et al (34) study of office ergonomic training were of sufficient methodological quality to be included in the final body of evidence (table 2) and application of the algorithm therefore classified the evidence on attitudes as insufficient.

Effects of training on behaviors

Ten studies contributed data on behavioral effects, which were typically measured at six months follow-up (table E in the appendix). The effects seen in most studies were positive, with some of these being large and statistically significant (33, 34, 40, 46–48), others with size undetermined but statistically significant (49), and others more modest in size or non-significant (31–33). Two studies yielded small, negative effects (50, 51). One of these (50) was statistically significant, but had poor internal validity, since there had been a major drop in the study sample size, which was distributed differently over the training and control groups. Results from 6 of the 10 studies were of fair/good methodological quality and 5 of them provided 13 effect sizes to a final body of evidence (table 2); they had an interquartile range of 0.33–1.35, indicating consistency in the directions of effects, and a median of 1.09, which surpasses the pre-set criterion for large (0.8). As such, there is strong evidence for the effectiveness of training on behaviors in the workplace (table 3).

The final body of evidence on behaviors is based on three office ergonomics studies (31, 34, 40), one study of dermatitis prevention among those doing "wet work" in geriatric facilities (33), one study of the adoption of precautions against blood-borne diseases by nurses (52), and a study of farm safety (49). One of the office

ergonomics interventions was already described above (34). In a second, a two-session, six-hour, multi-component training intervention (lectures, demonstrations, simulations, and self-diagnosis on work stations) was delivered to university-based visual display unit users, who were primarily clerical workers (31). In the third, a variety of white-collar computer users were randomized at the level of work group to one of three interventions of about an hour in length (40). The interventions differed in terms of the target of feedback (individual, supervisor, group) regarding work group exposure to ergonomic and psychosocial hazards; in addition, the feedback to individuals also included information on their individual exposures. Turning now to the other three studies contributing to the final body of evidence on behaviors (table 2), the Held et al study (33) evaluated a train-the-trainer program, consisting of two 4-hour sessions (separated by 14 weeks) of video, oral instruction, role play, and information, followed by a coaching session 6 weeks later. The farming intervention studied by Rasmussen et al (49) included a half day safety check of the farm with a written report and recommendations; and then 1–4 weeks later, a one-day course with many active components (lecture, meeting with injured farmers, demonstration, discussion of recommendations, and action planning). Finally, the Wright et al study (52) was a self-paced computer-based program of unknown duration. Behaviors were measured in three studies using self-administered questionnaires (33, 40, 49) and in the other three studies using observations (31, 34, 52).

Effects of training on health

The effects of training on health outcomes, drawn from ten studies, are shown in table F in the appendix. They were usually measured at six months otherwise more follow-up and musculoskeletal symptom measures predominated (six studies). The majority of the training interventions involved a hands-on practice component and were therefore classified as high engagement. Only two studies showed statistically significant effects (33, 53, 54), and in the one of these where an effect size could be computed (33), the effect size was very small (0.05). Among the remaining eight studies, which had statistically insignificant effects, the effects were usually small and either positive or negative; the largest positive effect among them was 0.37 (34). Five studies contributed results to the final body of evidence in table 2. However, since the directions of study effects were inconsistent (interquartile range -0.25–0.06) and their size (median -0.04) was below the effect size criterion for sufficient (0.15), evidence was classified as insufficient.

Of the five studies that contributed to the final body of evidence on health effects, four were already described above (and had shown positive effects on behaviors): two of these studies were concerned with office ergonomics and measured musculoskeletal symptoms by self-administered questionnaire after two weeks (34) or six months (40); another was the train-the-trainer study of those doing "wet work" in geriatric facilities, in which a clinical assessment of dermatitis was made after five months (33); and the fourth study (49) provided an estimate of effect based on 6 months of weekly injury reporting by farmers. The fifth study (55) involved a 15-minute hands-on training of grocery store workers in case cutter use, with its effect on injuries measured after one year using administrative statistics.

Consideration of heterogeneity

The reason for inconsistency among the health outcome data was explored as it is suggestive of heterogeneity in the study populations, interventions, or measurement methods. If the reason for the heterogeneity is apparent, separate evidence synthesis statements might be warranted for sub-groupings of the body of evidence (30). The health outcome data in table 2 were considered in light of this. The variation in the type of hazard addressed by the training and the corresponding type of outcome measured was found to be related to outcomes: the negative effects in table 2 were derived from the two ergonomic studies involving self-reported musculoskeletal symptoms, whereas the three studies that addressed safety or chemical hazards had small and positive effects. However, a separate synthesis of the health outcome results for the non-ergonomic studies was not pursued, since the conclusion would have remained that the strength of evidence was insufficient as the median effect size still did not meet the size criterion.

Post hoc sensitivity analyses

The review team explored the robustness of the findings in a sensitivity analysis by allowing limited quality studies to count toward the quantity criteria when applying the evidence synthesis algorithm. This meant that the strength of a body of evidence would be considered sufficient if there were at least three studies (of any quality) with consistent effects of sufficient size. As a result of this reanalysis, the strength of the evidence on training's effectiveness on knowledge and on attitudes and beliefs changed from insufficient to sufficient. This was not unexpected since the conclusion of insufficient for these outcomes in the main analysis was attributable to an inadequate number of studies, not to a lack of consistency or small effect size. On the other hand, the review's findings of strong evidence of training's effectiveness on behaviors in the workplace and insufficient evidence of its effectiveness on health were robust

and remained the same. A second sensitivity analysis explored the effect of allowing each study to contribute only one effect size to the evidence synthesis for a given outcome. This manipulation did not change the rating of the evidence by the algorithm.

Final evidence synthesis findings regarding level of engagement

There were seven trials that contrasted training interventions with differing levels of learner engagement, thereby addressing research question 2. The findings on effects extracted from these studies are shown in table G in the appendix. Only four studies contained outcome data assessed to be of fair or good methodological quality, thereby contributing to the final evidence syntheses (tables I and K in the appendix). No effects on knowledge were determined in these studies and only a single effect size was available for each of the attitudes and health categories (0.13 and 0.60, respectively). Notably, for both of these outcome categories, the size of the single effect was greater than the corresponding minimum size criterion (0.12 and 0.04, respectively); but each final body of evidence overall was considered insufficient in strength because there were too few studies. In contrast, there were a sufficient number of studies that measured behavioral outcomes, but the median of effects on behaviors (0.06) was below the criterion set for that outcome (0.10). The body of evidence on the relative effectiveness of higher versus lower engagement training on behaviors was therefore also considered insufficient in strength.

Examination of funding sources

The potential for bias arising from the nature of the study funding source was examined post hoc by the first author. No commercial sources of funding supported the studies included in the review, but in two studies (45, 56) the lead authors had a commercial interest in the computer-based training interventions being studied. Since only one of these contributed to the final body of evidence on training's effect on knowledge (45), for which an insufficient number of studies were found according to the synthesis algorithm, there was no threat to the review's conclusion about knowledge.

Examination of selective reporting of outcomes

The methods section of each of the 22 relevant articles was reviewed post hoc to see whether some outcomes had been measured but not reported upon in the respective results section, which would be suggestive of selective reporting of outcomes. This situation was applicable to two articles (49, 57), but in both cases the unreported measures would have been grouped here with attitudes outcomes, with one article contributing to table D and the other to table G (both found in the appendix). As such, the results would not have changed the evidence synthesis conclusions, since an insufficient number of studies would have persisted. We found no cases where behavioral outcomes were mentioned in the methods section but not reported in the results.

We also considered whether there may have been outcomes measured, but their collection not reported in the methods section. Of most potential concern to this review would be unreported (non-significant, small) behavioral outcomes, since we had concluded the body of behavioral evidence was strong. There were three studies (32, 55, 58) that measured health outcomes, indicating that they had an adequate timeline in which to measure behaviors yet did not report them. However, only one of the studies (55) made it to the final evidence synthesis stage (table 2), but its non-academic style of reporting and research setting (food retail) render it likely that behaviors were actually not measured.

Discussion

Principal findings

This review found a general lack of high quality randomized trials in the area of OHS training effectiveness that meet the relevance criteria. The modest number of fair or good methodological quality trials available for any outcome of interest (no more than six per outcome), coupled with their heterogeneity in terms of populations, interventions, and outcome measures limited the ability of the review to draw more definitive conclusions. This was particularly the case for the effect of training on knowledge and attitudes and beliefs. For these outcomes, evidence was rated as insufficient due to a lack of studies, although the synthesis algorithm's criteria for consistency and size of effects were met. In contrast, there were a sufficient number of higher quality studies reporting on training's effects on behaviors and health. With regard to behaviors, the review found strong evidence of training's effectiveness. For health, evidence was insufficient to conclude that training was effective because observed effects did not meet the size criterion and were inconsistent in direction. There was also insufficient evidence that higher engagement training was more effective than lower engagement training in improving target worker OHS behaviors, but this was based on only three studies, two of which involved very brief interventions directed toward hearing protection.

Strengths and limitations

There are several notable strengths of this systematic review: (i) a research team with expertise in OHS training and systematic review methodology; (ii) a thorough search of the published literature; (iii) a restriction of relevant study designs to randomized trials, thereby maximizing the internal validity of the synthesized evidence; and (iv) the conduct of sensitivity analyses. There were some limitations too. First, the data were relatively sparse due to the restriction in study design (RCT with pre- and post-measurements), time period (1996–2007), and language (English or French). Data were especially sparse in the final evidence synthesis stage (table 2) since the outcome data with greater risk of bias had been excluded (ie, the study data assessed as having limited methodological quality). This sparseness could limit the robustness of the results. The literature search was terminated in November 2007, so RCT reported since then could affect the conclusions. Second, the bodies of evidence used in the final evidence syntheses (table 2) – though qualitative – grouped together a variety of populations, interventions, and outcome measures. A robust exploration of which aspects of populations, interventions, and outcome measurement methods were prime determinants of outcomes was precluded by the sparseness of data. Until this exploration can be achieved by researchers, our findings could be misleading about certain sub-categories of population, intervention, and outcome. A third limitation is that the evidence synthesis algorithm relies on expert opinion when determining what criteria will be used for classifying the effect sizes of a body of evidence as insufficient, sufficient, or large. This expert judgment is a critical determinant of the conclusions that will be drawn about the strength of evidence. This review's effect size criteria were determined by three senior research team members specialized in training research before they had knowledge of the final synthesis results.

Relation to other research

This review was intended to update the report by Cohen & Colligan (9), which reviewed the literature published in English up to 1996. It was also methodologically enhanced by using systematic review techniques. The results of both reviews are consistent in concluding there are, generally, positive effects of training on knowledge, attitudes, and behaviors. These results are also consistent with the findings of a recent meta-analysis by Burke et al (23) who reviewed quasi-experimental studies published in English between 1971–2003.

On the other hand, this review differs from these two studies (9, 23) with respect to the conclusion drawn about health outcomes (ie, injuries, illnesses, symptoms). Our review found that the health effects were too small (and inconsistent in their direction) to be considered effective. In contrast, the Cohen & Colligan review (9) found mostly positive effects on health outcomes though, perhaps notably, they expressed concern about the internal validity of these effects. Furthermore, their review did not consider the size of effects. Burke et al (23) also reported a positive effect of training on health; however, when results were drawn from the subset of their reviewed studies with stronger research designs (ie, those with a comparison group), the observed effects were small (mean d=0.25) (59), and limited to the subset of training interventions with a high degree of learner engagement. In addition to the above reviews, which considered the effectiveness of training addressing any type of OHS hazard, four recent reviews (19–22), including one meta-analysis of randomized trials (22), have focused on interventions directed at preventing musculoskeletal disorders. None of these reviews found that research evidence supported the effectiveness of training in preventing these disorders.

With regards to examining the role of learner engagement on training effectiveness, this review took a different methodological approach than Burke et al (23). We estimated relative effectiveness directly through trials that compared a lower engagement study arm with a higher one. In contrast, Burke et al (23) first pooled all available study results for low, moderate, and high engagement training, respectively, and then contrasted the mean effect sizes for the three groups. Our approach avoids confounding by factors related to the subjects, intervention features, and methods of outcome measurement to a greater extent; but it resulted in a very sparse data set. The general conclusion of the Burke et al review (23) was that OHS training had a greater impact when the method of training involved more learner engagement. Our findings on this question can be viewed as mixed. On the one hand, the single effect estimates obtained in each of the attitudes and health outcome categories met or exceeded the corresponding effect size criteria, consistent with higher engagement training being more effective than lower engagement. On the other hand, for the one outcome (behaviors) where there was a sufficient number of good/fair quality studies to meet the evidence synthesis criteria of quantity and quality, the median effect size did not meet the size criterion. However, it should be noted that the three studies contributing evidence on behaviors all involved interventions with a single session, which in two studies was less than an hour.

Future research

This study was not able to investigate meaningfully the separate contribution of various features of population,

intervention and outcome to the size of effect, due to a relative sparseness of data. We suggest that future reviews consider including studies with a non-randomized trial design. We found that a sample of 11 otherwise eligible studies with such a design had similar methodological quality as the randomized trials (26, p13).

In terms of primary research in this area, both our review and the one by Burke et al (23) categorized learner engagement post hoc through descriptions available in the reviewed publications. We encourage researchers to continue to develop means of understanding and operationalizing the concept so that it can be intentionally manipulated and measured as a study variable in future training intervention trials. Another worthwhile direction would be to understand the basis for the sizeable effect on health (0.60) observed in the Löffler et al (42) study, which addressed research question 2. This study of nursing trainees and dermatitis prevention contrasted a seven-session medium-engagement training provided over three years with a single provision of an information pamphlet.

With respect to the reporting of future research on OHS training interventions, we suggest that there is room for improvement. Much of the lack of confidence this team had about bias being minimized in the primary studies arose from inadequate reporting in the studies. Use of one of the available guidelines like CONSORT (60) or TREND (61) is recommended, as well as greater use of intention-to-treat analysis. A second issue is apparent in the lack of knowledge and attitudes outcome data available to this review, relative to behaviors and health outcomes data – a surprising finding, given the typically greater ease of collecting knowledge and attitudes data. There were seven studies (33, 40, 49, 51, 58, 65, 66) that collected behavioral or health information pre- and post-intervention by questionnaire. Some of these questionnaires could presumably have measured knowledge and attitudes too. We encourage health researchers to include the measurement of these intervening variables in their research designs, since it can enhance the comprehensiveness and validity of an intervention evaluation (64) and contribute to training theory.

Implications for practice

The authors recommend that workplaces continue to deliver OHS training to employees as a means of addressing OHS risk, since training has been found to positively impact employee work practices in this and other reviews. However, based on the conclusion here and elsewhere of there being either no, uncertain, or low effectiveness of training in preventing illness and injury when delivered as a lone intervention, we strongly suggest that decision-makers consider more than just education and training when addressing a risk in the workplace. Large impacts of training alone cannot be expected based on research evidence.

Acknowledgements

We appreciate the contributions of the following organizations and people who provided methodological expertise, commentary, research assistant services, library services, editorial services, or other organizational support: Laura Blanciforti, Lani Boldt, Michael J Burke, Hee Kyoung Chun, Elaine Cullen, Anita Dubey, Randy Elder, Alyson Folenius, Andrea Furlan, Jill Hayden, Sheilah Hogg-Johnson, Carol Kennedy, Kiera Keown, Quenby Mahood, Lyudmila Mansurova, Cindy Moser, Cameron Mustard, Shanti Raktoe, Dan Shannon, and the IWH Measurement Group. We also appreciate the feedback provided by reviewers and representatives of the following Ontario lay organizations: Canadian Centre for Occupational Health and Safety, Electrical & Utilities Safety Association, Industrial Accident Prevention Association, Ontario Federation of Labour, Ontario Ministry of Labour, Ontario Service Safety Alliance, Workers Health and Safety Centre, and Workplace Safety & Insurance Board (WSIB).

The project was supported in part through a Prevention Reviews pilot initiative funded by the WSIB and directed to the Institute for Work & Health, Toronto, Ontario. Apart from giving feedback as one of many stakeholder lay organizations (see above), the funding agency was not involved in the research or the reporting of it. The findings and conclusions are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

References

1. Leigh J, Macaskill P, Kuosma E, Mandryk J. Global burden of disease and injury due to occupational factors. Epidemiology. 1999;10:626–31. http://dx.doi.org/10.1097/00001648-199909000-00032.

2. Schulte PA. Characterizing the burden of occupational injury and disease. J Occup Environ Med. 2005;47:607–22. http://dx.doi.org/10.1097/01.jom.0000165086.25595.9d.

3. Smith PM, Mustard CA. How many employees receive safety training during their first year of a new job? Inj Prev. 2007;13:37–41. http://dx.doi.org/10.1136/ip.2006.013839.

4. Redinger CF, Levine SP. Development and evaluation of the Michigan Occupational Health and Safety Management System Assessment Instrument: a universal OHSMS performance measurement tool. AIHA Journal. 1998;59:572-81.

5. Canadian Standards Association (CSA). CSA Z1000-06.

Occupational health and safety management. Mississauga, ON: CSA; 2006.

6. British Standards Institute (BSI). Occupational health and safety management systems - Requirements. London: BSI; 2007.
7. American National Standards Institute/American Industrial Hygiene Association (AIHA). American national standard for occupational health and safety management systems. AIHA; 2005.
8. Noe RA. Employee training and development. Boston: McGraw-Hill; 2005.
9. Cohen A, Colligan MJ, Sinclair R, Newman J, Schuler R. Assessing occupational safety and health training: a literature review. Cincinnati, OH: National Institute for Occupational Health and Safety; 1998. DHHS (NIOSH) Publication No. 98–145.
10. Ellis L. A review of research on efforts to promote occupational safety. J Safety Res. 1975;7:180–9.
11. Kroemer KHE. Personnel training for safer material handling. Ergonomics. 1992;35:1119–34. http://dx.doi.org/10.1080/00140139208967386.
12. Johnston JJ, Cattledge GT, Collins JW. The efficacy of training for occupational injury control. Occup Med. 1994;9:147–58.
13. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin; 2002.
14. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med. 1997;126:376–80.
15. Cannon-Bowers JA, Salas E, Tannenbaum SI, Mathieu JE. Toward theoretically based principles of training effectiveness: a model and initial empirical investigation. Mil Psychol. 1995;7:141–64. http://dx.doi.org/10.1207/s15327876mp0703_1.
16. Alvarez K, Salas E, Garofano CM. An integrated model of training evaluation and effectiveness. Hum Resour Dev Rev. 2004;3:385–416. http://dx.doi.org/10.1177/1534484304270820.
17. Kraiger K, Ford JK, Salas E. Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. J Appl Psychol. 1993;78:311–28. http://dx.doi.org/10.1037/0021-9010.78.2.311.
18. Loos GP, Fowler T. A model for research on training effectiveness. Cincinnati, OH: National Institute for Occupational Health and Safety; 1999. DHHS (NIOSH) Publication No. 99–142.
19. Brewer S, Eerd DV, Amick IB, III, Irvin E, Daum KM, Gerr F, et al. Workplace interventions to prevent musculoskeletal and visual symptoms and disorders among computer users: A systematic review. J Occup Rehabil. 2006;16:325–58. http://dx.doi.org/10.1007/s10926-006-9031-6.
20. Brewer S, King E, Amick BC, Delclos G, Spear J, Irvin E, et al. Systematic review of injury/illness prevention and loss control programs. Toronto: Institute for Work & Health; 2007.
21. Kennedy CA, Amick BC, III, Dennerlein JT, Brewer S, Catli S, William R, et al. Systematic review of the role of occupational health and safety interventions in the prevention of upper extremity musculoskeletal symptoms, signs, disorders, injuries, claims and lost time. J Occup Rehab. 2010;20:127–62. http://dx.doi.org/10.1007/s10926-009-9211-2.
22. Martimo KP, Verbeek J, Karppinen J, Furlan AD, Takala EP, Kuijer PP, et al. Effect of training and lifting equipment for preventing back pain in lifting and handling: systematic review. Br Med J. 2008;336:429–31. http://dx.doi.org/10.1136/bmj.39463.418380.BE.
23. Burke MJ, Sarpy SA, Smith-Crowe K, Chan-Serafin S, Salvador RO, Islam G. Relative effectiveness of worker safety and health training methods. AJPH. 2006;96:315–24. http://dx.doi.org/10.2105/AJPH.2004.059840.
24. Burke MJ, Holman D, Birdi K. A walk on the safe side: the implications of learning theory for developing effective safety and health training. In: Hodgkinson GP, Ford JK, editors. International review of industrial and organizational psychology. Chichester,U.K.: Wiley; 2006. p 1–44. http://dx.doi.org/10.1002/9780470696378.ch1.
25. Burke MJ, Scheurer ML, Meredith RJ. A dialogical approach to skill development: the case of safety skills. Hum Resour Manage Rev. 2007;17:235–50. http://dx.doi.org/10.1016/j.hrmr.2007.04.004.
26. Robson L, Stephenson C, Schulte P, Amick B, Chan S, Bielecky A, et al. A systematic review of the effectiveness of training & education for the protection of workers. Toronto: Institute for Work & Health. Cincinnati, OH: National Institute for Occupational Safety and Health; 2010. DHHS (NIOSH) Publication No. 2010–127.
27. Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognosis studies in systematic reviews. Ann Intern Med. 2006;144:427–37.
28. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non-randomised intervention studies. Health Technol Assess. 2003;7:1–173.
29. van Tulder M, Furlan A, Bombardier C, Bouter L. Updated method guidelines for systematic reviews in the cochrane collaboration back review group. Spine. 2003;28:1290–9. http://dx.doi.org/10.1097/01.BRS.0000065484.95996.AF.
30. Briss PA, Fielding J, Hopkins DP, Woolf SH, Hinman AR, Harris JR. Developing an evidence-based guide to community preventive services - methods. Am J Prev Med. 2000;18:35–43. http://dx.doi.org/10.1016/S0749-3797(99)00119-1.
31. Brisson C, Montreuil S, Punnett L. Effects of an ergonomic training program on workers with video display units. Scand J Work Environ Health. 1999;25(3):255–63.
32. Duffy OM, Hazlett DE. The impact of preventive voice care programs for training teachers: a longitudinal study. J Voice. 2004;18:63–70. http://dx.doi.org/10.1016/S0892-1997(03)00088-2.
33. Held E, Mygind K, Wolff C, Gyntelberg F, Agner T. Prevention of work related skin problems: an intervention study in wet work employees. Occup Environ Med. 2002;59:556–61. http://dx.doi.org/10.1136/oem.59.8.556.
34. Greene BL, DeJoy DM, Olejnik S. Effects of an active

ergonomics training program on risk exposure, worker beliefs, and symptoms in computer users. Work. 2005;24:41–52.

35. Lipsey MW, Wilson DB. Practical meta-analysis. Thousand Oaks, CA: Sage Publications; 2000.
36. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med. 2000;19:3127–31. http://dx.doi.org/10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M.
37. Cooper H, Hedges LV. The handbook of research synthesis. New York: Russell Sage Foundation; 1994.
38. Rothman KJ, Greenland S. Modern epidemiology, 2nd ed. Philadelphia: Lippincott-Raven; 1998.
39. Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology, 2nd ed. New York: Oxford; 1996.
40. Eklöf M, Hagberg M, Toomingas A, Tornqvist EW. Feedback of workplace data to individual workers, workgroups or supervisors as a way to stimulate working environment activity: a cluster randomized controlled study. Int Arch Occup Environ Health. 2004;77:505–14. http://dx.doi.org/10.1007/s00420-004-0531-4.
41. Guide to Community Preventive Services [Internet]. Atlanta, GA: The Community Guide Branch, Epidemiology Analysis Program Office, Office of Surveillance, Epidemiology, and Laboratory Services, Centers for Disease Control and Prevention . [updated 2011 Feb 28; cited 2011 Mar 3]. Available from: http://www.thecommunityguide.org/index.html.
42. Löffler H, Bruckner T, Diepgen T, Effendy I. Primary prevention in health care employees: a prospective intervention study with a 3-year training period. Contact Dermatitis. 2006;54:202–9. http://dx.doi.org/10.1111/j.0105-1873.2006.00825.x.
43. Wang H, Fennie K, Guoping H, Burgess J, Williams AB. A training programme for prevention of occupational exposure to bloodborne pathogens: impact on knowledge, behaviour and incidence of needle stick injuries among student nurses in Changsha, People's Republic of China. J Adv Nurs. 2003;41:187–94. http://dx.doi.org/10.1046/j.1365-2648.2003.02519.x.
44. Hickman JS, Geller ES. A safety self-management intervention for mining operations. J Safety Res. 2003;34:299–308. http://dx.doi.org/10.1016/S0022-4375(03)00032-X.
45. Harrington SS, Walker BL. The effects of ergonomics training on the knowledge, attitudes, and practices of teleworkers. J Safety Res. 2004;35:13–22. http://dx.doi.org/10.1016/j.jsr.2003.07.002.
46. Eklöf M, Hagberg M. Are simple feedback interventions involving workplace data associated with better working environment and health? A cluster randomized controlled study among Swedish VDU workers. Appl Ergon. 2006;37:201–10. http://dx.doi.org/10.1016/j.apergo.2005.04.003.
47. Rizzo TH, Pelletier KR, Serxner S, Chikamoto Y. Reducing risk factors for cumulative trauma disorders (CTDs): the impact of preventive ergonomic training on knowledge, intentions, and practices related to computer use. Am J Health Promot. 1997;11:250–3.
48. Wright BJ, Turner JG, Daffin P. Effectiveness of computer-assisted instruction in increasing the rate of universal precautions-related behaviors. Am J Infection Control. 2005;25:426–9. http://dx.doi.org/10.1016/S0196-6553(97)90093-6.
49. Rasmussen K, Carstensen O, Lauritsen JM, Glasscock DJ, Hansen ON, Jensen UF. Prevention of farm injuries in Denmark. Scand J Work Environ Health. 2003;29:288–96.
50. Arnetz JE, Armetz.B.B. Implementation and evaluation of a practical intervention programme for dealing with violence towards health care workers. J Adv Nurs. 2000;31:668–80. http://dx.doi.org/10.1046/j.1365-2648.2000.01322.x.
51. Jensen LD, Gonge H, Jors E, Ryom P, Foldspang A, Christensen M et al. Prevention of low back pain in female eldercare workers: randomized controlled work site trial. Spine. 2006;31:1761–9. http://dx.doi.org/10.1097/01.brs.0000227326.35149.38.
52. Wright BJ, Turner JG, Daffin P. Effectiveness of computer-assisted instruction in increasing the rate of universal precautions--related behaviors. Am J Infect Control. 1997;25:426–29. http://dx.doi.org/10.1016/S0196-6553(97)90093-6.
53. Bohr PC. Efficacy of office ergonomics education. J Occup Rehab. 2000;10:243–55. http://dx.doi.org/10.1023/A:1009464315358.
54. Bohr PC. Office ergonomics education: a comparison of traditional and participatory methods. Work. 2002;19:185–91.
55. Banco L, Lapidus G, Monopoli J, Zavoski R. The Safe Teen Work Project: a study to reduce cutting injuries among young and inexperienced workers. Am J Ind Med. 1997;31:619–22. http://dx.doi.org/10.1002/(SICI)1097-0274(199705)31:5<619::AID-AJIM17>3.0.CO;2-1.
56. Harrington S, Walker BL. A comparison of computer-based and instructor-led training for long-term care staff. Journal of Continuing Education in Nursing. 2002;33:39–45.
57. Perry MJ, Layde PM. Farm pesticides: outcomes of a randomized controlled inervention to reduce risks. Am J Prev Med. 2003;24:310–5. http://dx.doi.org/10.1016/S0749-3797(03)00023-0.
58. van Poppel MNM, Koes BW, van der Ploeg T, Smid T, Bouter LM. Lumbar supports and education for the prevention of low back pain in industry: a randomized controlled trial. JAMA. 1998;279:1789–94. http://dx.doi.org/10.1001/jama.279.22.1789.
59. Cohen J. Statistical power analysis for the behavioral sciences, 1st ed. New York: Academic Press; 1988.
60. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med. 2010;8:834–40. http://dx.doi.org/10.1186/1741-7015-8-18.
61. Des J, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. AJPH. 2004;94:361–6. http://dx.doi.org/10.2105/AJPH.94.3.361.
62. Hong O, Ronis DL, Lusk SL, Lee G.-S. Efficacy of a computer-based hearing test and tailored hearing protection

intervention. Int J Behav Med. 2006;13:304–14. http://dx.doi.org/10.1207/s15327558ijbm1304_5.

63. Gray J, Cass J, Harper DW, O'Hara PA. A controlled evaluation of a lifts and transfer educational program for nurses: a multifaceted educational program can increase nurses' knowledge of the mechanics and procedures underlying safe lifting and transferring. Geriatr Nurs. 1996;17:81–5. http://dx.doi.org/10.1016/S0197-4572(96)80175-3.

64. Robson LS, Shannon HS, Goldenhar LM, Hale AR. Guide to evaluating the effectiveness of strategies for preventing work injuries. How to show whether a safety intervention really works. Cincinnati, OH: National Institute for Occupational Safety and Health; 2001. DHHS (NIOSH) Publication No. 2001–119.

65. Lusk SL, Ronis DL, Kazanis AS, Eakin BL, Hong O, Raymond DM. Effectiveness of a tailored intervention to increase factory workers' use of hearing protection. Nurs Res. 2003;52:289–95 http://dx.doi.org/10.1097/00006199-200309000-00003.

66. Lusk SL, Eakin BL, Kazanis AS, McCullagh MC. Effects of booster interventions on factory workers' use of hearing protection. Nurs Res. 2004;53:53–8. http://dx.doi.org/10.1097/00006199-200401000-00008.

Received for publication: 7 August 2010