productivity, human performance, and employee productivity and performance. The search strings for the other databases (MEDLINE, CINAHL, PsycINFO, and the Campbell Collaboration) are provided in Appendix B. We entered retrieved titles/abstracts into an EndNote® database, except for the EBSCO and Campbell Collaboration databases, where only included studies were entered.

Selection Processes

The searches resulted in a total of 23,179 citations. The lead investigator for each of the working condition categories reviewed the titles/abstracts for the citations that fell into their working condition category. Table 1 lists the eligibility criteria that were applied during the title/abstract review. For the non-healthcare industries, studies which met the criteria in Table 1 were considered for inclusion if they reported any measure of work results or productivity.

Table 1. Inclusion/exclusion criteria for judging titles/abstracts and full text papers

Code #	Justification for codes
	Inclusions
1	IN: For this category of working conditions
2	IN: For *another* category of working conditions
	Indicate which one:
	A. Physical Environment
	B. Workflow Design
	C. Workforce Staffing
	D. Organizational Factors
	E. Personal/Social
3	IN: Good review or background article
	Exclusions
4	OUT: Does not address any key question
5	OUT: Does not report original data
6	OUT: Wrong population (animal study, etc.)
7	OUT: <specific to topic> Write justification below

guidelines for extraction of information into the data abstraction tables, the investigative team developed data abstraction guidelines (Appendix D).

Assessment of Study Quality

Our system for rating the quality of individual studies was based on previously published methods. Several approaches to evaluating quality were examined. We reviewed concepts from the U.S. Preventive Services Task Force, the *Guide to Community Preventive Services*, the Oxford Centre for Evidence-based Medicine, and AHRQ Evidence Report/Technology Assessment Number 47 *Systems to Rate the Strength of Scientific Evidence*.[31-34]

The approach that was best suited to the types of studies included in this report is based on the system used by the Task Force on Community Preventive Services.[32] This approach differs from that used in a previous evidence report on patient safety practices.[14] Each study was assessed for quality using two ratings. The first, suitability of study design, was a three-tier approach (greatest, moderate, and least) relating the strength of the study design to threats of internal validity. We developed a modified version of design suitability for our topic. We rated studies greatest, moderate, or least based on comparison group status and measures of other factors affecting outcomes (Table 2). The second measure of quality was the quality of study execution. Here, six areas of threats to validity have been described,[32] with the ratings of good, fair, or limited corresponding to 0-1, 2-4, or 5 or more limitations. We used a similar approach and rated study execution as good, fair, or poor based on internal and external validity. Internal validity was assessed by considering such factors as comparability of groups, differential loss to followup, measurement/instrumentation issues, maturation/pre-testing effects, and whether there was a clear description of interventions. External validity was assessed by considering such factors as a sensitized or pre-tested population, specialized/atypical population, selection biases (non-random subject selection), reactive effects of experimental settings, and multiple interventions.

Table 2. Design suitability

Greatest	Concurrent comparison groups and sufficient measures for other factors affecting outcomes
Moderate	Non-concurrent comparison groups or insufficient measures for other factors affecting outcomes
Least	Non-concurrent comparison or no comparison groups and insufficient measures for other factors affecting outcomes

The overall strength of the evidence per topic area was assessed based on the criteria outlined by the Task Force on Community Preventive Services.[32] The quality and quantity of the studies and size and consistency of the results were used to grade the overall strength of the

evidence. Where evidence was available, the body of evidence for a topic area was rated as strong, sufficient, or insufficient according to the parameters outlined in Table 3.

Table 3. Assessment of strength of evidence	
Strong	At least two studies having greatest design suitability, good execution, and consistent results; or at least five studies having greatest design suitability, good or fair execution, and consistent results
Sufficient	At least one study with greatest design suitability and good execution; or at least three studies having moderate or better design suitability, fair or better execution, and consistent results
Insufficient	Too few studies to meet definition of sufficient evidence; or inconsistent results among multiple studies having some design or execution flaws

Methodologic Limitations

In this report we have adapted the methods of the systematic review[35] to collect, evaluate, and synthesize the best available evidence that addresses the key questions. However, these methods were not developed for the domain of inquiry in which we are applying them, and this has implications for the conclusions that can be reached and the degree of certainty with which they can be stated. The methods and assumptions of the systematic review of health interventions, as practiced by organizations such as the AHRQ Evidence-based Practice Centers and the Cochrane Collaboration[36] are not entirely applicable to a broad and diverse domain of inquiry such as that addressed in this report. Differences include those relating to a) underlying research traditions and assumptions; b) the search for relevant literature; c) selection of evidence for inclusion; d) appraisal of the relative validity and generalizability of studies; and e) combination or synthesis of evidence.

For healthcare interventions, the research tradition and assumptions of clinical epidemiology serve as the foundation both for original studies and syntheses of evidence. The underlying conditions, the interventions used to identify or treat them, and the outcomes of interest are generally well defined, with established and agreed upon means of identifying patients for inclusion and measuring outcomes of interest. The research designs used to investigate these interventions and conditions are well established and agreed upon, with defined limitations and biases and a familiar hierarchy of levels of evidence.[33] As a result, studies of a particular intervention for a particular condition can usually be compared, results of selected studies can be combined, and the result often be expressed in terms of a single numeric estimate of effect, with an appropriate and precise estimate of precision.

In contrast, the evidence for this report is drawn from both healthcare and non-healthcare fields and includes diverse domains of inquiry, with different research traditions and assumptions, including cognitive science, sociology, industrial and human factors engineering, and others. In many cases there are no uniform definitions of underlying conditions, interventions, or outcome states, and no single commonly accepted means of identifying or