Journal of Occupational Health Psychology
2008, Vol. 13, No. 1, 1–9

1076-8998/08/$12.00 DOI: 10.1037/1076-8998.13.1.1

Intervention Effectiveness Evaluation Criteria: Promoting Competitions and Raising the Bar

Ted Scharf
National Institute for Occupational Safety and Health, Cincinnati, Ohio

Larry Chapman
University of Wisconsin—Madison

Jim Collins
National Institute for Occupational Safety and Health, Morgantown, West Virginia

Julia Limanowski
National Institute for Occupational Safety and Health, Cincinnati, Ohio

Cathy Heaney
Stanford University

Linda M. Goldenhar
University of Cincinnati College of Medicine

The Intervention Evaluation Competition at the Work, Stress, and Health conference in Miami (March 2006) highlighted the importance of intervention evaluation studies that promote safety and health at work. A retitled, "Best Practices Evaluation Competition," has been included in the March, 2008, Work, Stress, and Health conference, in Washington, DC. This brief note describes the development of the criteria used to evaluate the manuscripts. The criteria are discussed with respect to (a) improving the science of evaluation methodology, (b) promoting the highest ethical standards in intervention evaluation, and (c) using the current criteria as a starting point for continuing to raise the bar for evaluation methodology. The policy implications of the evaluation criteria are discussed as well.

Keywords: intervention evaluation criteria, occupational safety and health interventions, conference competitions

---

Ted Scharf and Julia Limanowski, Division of Applied Research and Technology, NIOSH, Cincinnati, Ohio; Larry Chapman, University of Wisconsin—Madison; Jim Collins, Division of Safety Research, NIOSH, Morgantown, West Virginia; Cathy Heaney, Stanford University, Palo Alto, California; and Linda Goldenhar, University of Cincinnati College of Medicine, Cincinnati, Ohio.

This paper reports the evaluation criteria used in two competitions (2006 and 2008) from the perspective of the team conducting the review. However, these competitions are possible only because researchers in occupational health psychology have been willing to submit their work to this detailed review and competitive evaluation. We thank the authors of these manuscripts for their willingness to participate in the competition and to meet the early deadlines required by the conference schedule. In particular, we thank Anna-Liisa Elo (Finnish Institute of Occupational Health, Helsinki, and the University of Jyväskylä, Finland), who provided comments on an earlier draft of this manuscript from the point of view of a competition participant.

These competitions would not be possible without a dedicated team of reviewers who have been willing to conduct extremely thorough and very specific evaluations in a relatively short period. We acknowledge and thank (in alphabetical order, for both 2006 and 2008): David Ballard (APA, Washington, DC); Ann Brockhaus (ORC Worldwide, Washington, DC); Ann Carruth (Southeastern Louisiana University, Hammond); Elizabeth Dalsey and Don Eggerth (NIOSH); Anna-Liisa Elo (Finnish Institute of Occupational Health, Helsinki, and the University of Jyväskylä, Finland); Jane Ferrie (University College London); Jim Grosch (NIOSH); Cheryl Haslam (Loughborough University, Leicestershire, United Kingdom); Ted Hitchcock, DeLon Hull, Janice Huy, Lynn Jenkins, Jessica Keel, and Katherine Klein (NIOSH); Autumn Krauss (Kronos Inc., Portland, Oregon); Paul Landsbergis (Mount Sinai School of Medicine, New York); Leslie MacDonald, Launa Mallett, and Paul Middendorf (NIOSH); Ivan Most (Strategic Occupational Health Management, Old Orchard Beach, Maine); Larry Murphy, Akinori Nakata, Jeannie Nigam, and Rene Pana-Cryan (NIOSH); Ross Pastel (Fort Belvoir, Virginia); Kellie Pierson, Susan Reutman, Rashaun Roberts, Anita Schill, and Ray Sinclair (NIOSH); Robert Sinclair (Portland State University, Oregon); Carol Merry Stephenson (NIOSH); Mina Westman (Tel Aviv University, Israel); and Trudi Zimmerman (NIOSH).

Finally, we extend our sincere appreciation and thanks to Michael Colligan and Carol Merry Stephenson (NIOSH). These reviewers provided extremely helpful suggestions and observations on an earlier draft of this paper. We also thank Wes Baker, Chrissy Craddock, John Howard, Maggie Ivory, Jessica Keel, Anita Schill, Jamie Snyder, Sid Soderholm, and the entire conference planning crew at the APA and NIOSH for their tremendous assistance with this award, including their innovative and creative solutions for the many logistical problems associated with including this award in the conference.

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health or the American Psychological Association.

Correspondence concerning this article should be addressed to Ted Scharf, NIOSH, C-24, 4676 Columbia Parkway, Cincinnati, OH 45226. E-mail: tscharf@cdc.gov

The past few years have seen a dramatic increase in the number of journal articles addressing occupational research that attempts to promote safer and healthier work environments for individual businesses and throughout entire industries (Routsalainen et al., 2006). This recent increase in intervention research, including training interventions, is extremely important in that—by moving research to practice—the results of research can promote safer ways to work for our nation's workforce.

At the same time, intervention research needs to be evaluated with sufficient rigor to determine whether a given intervention is actually effective at reducing hazards or injuries, or improving health. Thus, although there have been publications describing intervention studies in occupational safety and health, until recently there has been a paucity of publications that include a scientifically rigorous methodology in *intervention effectiveness evaluation* (Goldenhar & Schulte, 1996; Burke et al., 2006). In addition, a study of the papers presented at the Sixth World Conference on Injury Prevention and Control, in Montreal, in May 2002, found an absence of evaluation research (Smith & Shannon, 2003). Fortunately, in the past few years the amount of intervention evaluation research has increased (Sinclair & the NORA Intervention Effectiveness Research Team, 2004; NORA Intervention Effectiveness Research Team, 2004). Furthermore, new materials have been developed and are available to assist researchers and practitioners (Robson, Shannon, Goldenhar, & Hale, 2001; NORA Intervention Effectiveness Research Team, 2004; NORA Organization of Work Team, 2002). A review of occupational health interventions published in 2000 and 2001 showed a marked improvement in the proportion of studies with truly experimental evaluations and acceptable methodologic quality (i.e., 56% in Routsalainen et al., 2006: vs. 6% in Goldenhar & Schulte, 1996). However, there is still a large imbalance in the total number of occupational intervention evaluation studies published compared with the proportion of intervention evaluation studies in other medical and public health fields (Routsalainen et al., 2006). Most recently, a review of 90 interventions in occupational stress that included some evaluation component emphasizes the importance of moving such interventions toward a primary prevention focus (LaMontagne et al., 2007). Finally, a team of researchers from the Institute for Work and Health (Toronto) and the National Institute for Occupational Safety and Health (NIOSH; Education and Information Division, Cincinnati) is preparing an updated review of the literature (publication anticipated Spring, 2008, C. M. Stephenson, personal communication). The goal of this new review is to examine the recent literature (1996–2006) to assess what is known about the factors that make training effective. This review will identify factors that influence the learning process so that employers, trainers, and developers of training can optimize resources applied to training. This review is restricted to randomized trials published in scientific, peer-reviewed journals.

To bring attention to intervention effectiveness research in occupational health psychology, an "Intervention Evaluation Competition Awards" session was prominently featured at the Work, Stress, and Health Conference (March 1–4, 2006) in Miami, Florida, sponsored by the American Psychological Association (APA) and NIOSH (available at: http://www.apa.org/pi/work/wsh2006.html). This award has been renewed as the "Best Practices Evaluation Competition" at the 2008 Work, Stress, and Health Conference in Washington, DC, (March 5–8, available at: http://www.apa.org/pi/work/wsh.html). These two competitions continue the nascent "tradition" of two prior competitions, sponsored by the Intervention Effectiveness Research team—a part of the National Occupational Research Agenda (NORA), coordinated by NIOSH. The Intervention Effectiveness Research (IER) team was one of the 21 teams established by NORA-I, 1996 to 2006 (details regarding the entire NORA effort are available at: http://www.cdc.gov/niosh/nora/default.html).

The purpose of this brief note is to report the review criteria developed by the competition reviewers and to promote these criteria as (a) a present-day standard for intervention evaluation studies in occupational safety and health, and (b) as a realistic starting point for improving intervention evaluation methodology.

Background

Promoting high-quality intervention evaluation studies was always an important part of the NORA IER team agenda. In particular, the NORA IER team wanted to encourage graduate students to become interested in interventions in occupational safety and health. The question at issue was finding a workable venue to implement such promotions. Dr. Larry Chapman first proposed the idea of evaluating posters and presentations at NIOSH-sponsored conferences in 1998. Dr. Chapman's proposal is similar to the study reported by Smith and Shannon (2003) from the 6th World Congress on Injury Prevention and

Control. The logistics required to complete such an evaluation during a 2-day conference led to a proposal in late 1999 to sponsor a preconference competition for the 2000 National Occupational Injury Research Symposium (NOIRS, available at: http://www.cdc.gov/niosh/noirs/menu.html). This competition proposed to adapt the evaluation criteria for research papers reported by Heacock, Koehoorn, and Tan (1997), for ergonomics. Unfortunately, the decision to announce the competition was made late in the conference planning and there was not time for adequate marketing. Too few manuscripts were submitted, and the competition was canceled.

The competition idea was revived for the 2003 NOIRS meeting (available at: http://www.cdc.gov/niosh/noirs/noirsmain2003.html). The scoring criteria were somewhat simplified in comparison to Heacock et al. (1997): (a) description of the need for a safety intervention (15%), (b) description of the safety intervention (25%), (c) explanation of study design and methods (40%), and (d) implications of the evaluation results for the field of occupational injury prevention (20%). Next, the APA, NIOSH, and the NORA IER team cosponsored an intervention evaluation competition at the 2006 Work, Stress, and Health conference. Finally, the Society for Occupational Health Psychology has joined APA and NIOSH in sponsoring the 2008 Best Practices evaluation competition. This brief note focuses on the criteria used to evaluate the best practices and other intervention evaluation papers submitted to the 2006 and 2008 Work, Stress, and Health conferences.

2006 and 2008 Competition Criteria

The criteria from the NOIRS 2003 competition were revised based on Shadish, Cook, and Campbell (2002) and on experience gained during the 2003 review. General criteria for the review were released with the competition announcement (available at: http://www.apa.org/pi/work/bestpractices.html); however, the precise review criteria were not finalized until after the volunteer reviewers had the opportunity to recommend changes and improvements. Based on a 100-point scale, the scores represent a somewhat complicated weighting scheme or priority system. Thus, the final version of the review form (see Table 1) can be taken as one contribution to a standard in intervention evaluation.

Lessons Learned

Below are the key questions to apply to the criteria in Table 1:

1. Are the criteria sufficiently comprehensive to represent the best practices in the science of intervention evaluation methodology?

2. Are the criteria sufficiently discriminating to set a high bar for achievement while recognizing partially successful methodology?

3. Is the scoring system (weighting scheme) appropriate to set high quality standards for intervention evaluations conducted in occupational safety and health?

The first two questions are best answered empirically through an accumulation of new evaluation studies that challenge the present criteria to the breaking point. Future manuscript competitions with more and varied evaluation topics will probe the present criteria to identify where there may be: (a) insufficient specificity (i.e., a lack of sensitivity in the scoring), (b) unchallenging criteria (e.g., every paper scores at the ceiling), or (c) an outright gap (an important criterion is missing).

The third question addresses the scoring system as an explicit set of values applied to the current criteria. In this respect, the scoring system represents not only the best science but also the highest ethical standards currently agreed upon. Thus, the primary value in publishing the current criteria in Table 1 is quite explicitly to encourage: (a) the modification, improvement, or replacement of the current list with better criteria, (b) the development of a more appropriate weighting scheme, and (c) overall improvements in evaluation science.

Policy Implications in Selected Evaluation Criteria

A "p" in Table 1, column 1, notes criteria which promote a particular policy in evaluation methodology. The criteria with policy implications sum to 40% of the total possible score (not counting the request for null results in Section IV). To be sure, one might argue that simply publishing the best science is also making policy. The evaluation criteria clearly expect good quality science. The key point reflected in these criteria is that the most appropriate methodology to use with intact workgroups is often somewhat different from the traditional methodology of experimental science (see also, LaMontagne et al., 2007).

For example, Section II.C. encourages the research

Table 1
Review Criteria for the Best Practices Evaluation Competition Awarded at the Work, Stress, and Health Conference, Washington, DC, March 2008

Item No.	**Evaluation Criteria** (with descriptions[a] and with total possible points[b]) Note: References to "occupational health psychology" represent the topics relevant to the Work, Stress, and Health Conference. Replace this phrase with "occupational safety and health" or a similar phrase for other applications of these criteria.	Scores[b] Subcategory Elements	Subcategories	Main Categories
I. *p*[c]	**Description of the need for the workplace best practice(s):** Problem, background data, and societal, organizational, and individual impact relevant to occupational health psychology. Severity of the problem is not a scoring criterion; this prevents a best practice to eliminate fatalities from displacing a good evaluation of a best practice addressing workplace wellness or health promotion. Clear description of the problem(s) and potential consequences scored here (I.A.+ I. B).			
I.A.	**Health and safety problems addressed:** The problem(s) may reside at the individual, organizational, and/or societal level. For example, high workload may contribute to stress; or the target problem may be substance abuse by individual workers. Solutions/best practices are scored in II, below.		2	**4**
I.B.	**Injury and illness consequences for workers:** The consequences experienced by the workers with respect to the problem(s) identified in I.A.		2	
II.	**Description of the workplace best practice(s)** (II. A + II. B + II. C + II. D).			**24**
II.A.	**Basic descriptive information regarding the best practice(s):** This category provides points for a comprehensive and clear description of the best practice(s) including any specific, testable objectives or research hypotheses, regardless of the quality of the practice itself (II.A.1 + II.A.2).		8	
II.A.1.	**Specific, testable study objectives or research hypotheses:** quality and comprehensiveness of the description of the specific testable objectives or research hypotheses.[d]	2		
II.A.2.	**Overall quality of the description of the best practice(s)**	6		
II.B. *p*[c]	**Organizational levels included in the best practice(s):** This encourages multi-level practices with respect to an organization. Best practices that focus on the organization of the work at more than one level of management will be credited more highly than practices focused on a single level of organization. Thus, for example, an EAP program that treats worker complaints will be rated more highly when it works with management to make changes in the work processes that contributed to the employees' problems (II.B.1 + II.B.2 + II.B.3 + II.B.4 + II.B.5 + II.B.6.).		6	
II.B.1.	**Individual worker**	1		
II.B.2.	**Work crews or groups of workers**	1		
II.B.3.	**Offices, businesses, or union locals**	1		
II.B.4.	**Entire corporations, or groups of small businesses or unions**	1		
II.B.5.	**Corporate or governmental policies and regulations**	1		
II.B.6.	**Interaction of multiple organizational levels of the practices**	1		

Table 1
(continued)

Item No.	Evaluation Criteria (with descriptions[a] and with total possible points[b]) Note: References to "occupational health psychology" represent the topics relevant to the Work, Stress, and Health Conference. Replace this phrase with "occupational safety and health" or a similar phrase for other applications of these criteria.	Scores[b] Subcategory Elements	Subcategories	Main Categories
II.C. *p*[c]	**Process of development of the best practice(s):** Participatory solutions are specifically credited. Some papers may be review studies or other observational evaluations in which the researchers are reporting data for which they did not actually develop a best practice (II.C.1 + II.C.2).		6	
II.C.1.	**Participation of workers and management (to the extent possible):** Are workers and/or management involved in designing and/or maintaining one or more aspects of the practice(s)? What is the degree of participation?	3		
II.C.2.	**Participation of the research/evaluation team in developing the practice(s):** Review studies will not receive credit for developing a best practice.	3		
II.D. *p*[c]	**Innovative and creative approaches or noteworthy methods of addressing a problem:** Is it possible that this program, intervention, or research-to-practice could make a difference in the workplace? No matter how poor (or good) the research methodology, innovation and creativity receive credit here.		4	
III.	**Best practices implementation and evaluation study design and methods: the plan for implementing the practice(s) and for the evaluation of those practices.** Assessment of the study design carries the most weight. Thus, even if the study results are not the very best, researchers still receive credit for a comprehensive and thorough design (III. A + III. B + III. C + III. D + III. E).			**42**
III.A.	**Overall quality of the research design:** Research designs can have a great deal of variety; thus the "other" subcategory receives as many points as all the other subcategories combined (III.A.1 + III.A.2 + III.A.3 + III.A.4 + III.A.5).		16	
III.A.1. *p*[c]	**Use of comparison groups:** Randomized controlled trials (RCTs) describe just one type of comparison group that might be used in these studies. Comparison groups should be selected to reflect a realistic representation of the work environment(s) and target population.	2		
III.A.2.	**Pre-, post-, delayed post-testing, etc., as warranted**	2		
III.A.3.	**Internal validity strengths & weaknesses (of the design):** Aspects of the scientific method. (Construct validity is considered in IV.A.3.)	2		
III.A.4.	**External validity strengths & weaknesses (of the design):** Relevance to the workplace. (Construct validity is considered in IV.A.3.)	2		
III.A.5.	**Other aspects of the research design.**	8		

Table 1
(continued)

Item No.	**Evaluation Criteria** (with descriptions[a] and with total possible points[b]) Note: References to "occupational health psychology" represent the topics relevant to the Work, Stress, and Health Conference. Replace this phrase with "occupational safety and health" or a similar phrase for other applications of these criteria.	Scores[b] Subcategory Elements	Subcategories	Main Categories
III.B. *p*[c]	**Ethical considerations:** Best practices in occupational health psychology that treat an individual worker's problems without addressing the causes of the problems in the workplace may fall into the trap of "blaming the victim/worker." This is just one example of ethical considerations that may be raised with respect to a particular practice and its evaluation.		6	
III.C.	**Population, sample, recruiting participants, assignment to groups:** What is the quality and appropriateness of the targeted sample with respect to the population of interest? How will participants be recruited and assigned to different practice groups for the evaluation study (III.C.1 + III.C.2)?		8	
III.C.1.	**Power analysis or a consideration of power:** Power and other issues in hypothesis testing relevant to the design, including reliability of the measures and likely effect sizes, necessary sample size(s), effects of attrition, and other topics relevant to the population, sample, and inference. [d]	2		
III.C.2.	**Overall description of population, sample, recruiting, and assigning to groups.**	6		
III.D.	**Measurements (both quantitative and qualitative):** Are the researchers using a variety of measures, preferably employing multiple methods of assessment? (III.D.1 + III.D.2 + III.D.3).		8	
III.D.1.	**Variety of constructs measured:** As appropriate to the practice(s), multitrait, multimethod assessments are credited. Are the constructs relevant and appropriate to the goals of the intervention?	4		
III.D.2. *p*[c]	**Measurement of process/efficacy of the best practice(s):** Are there process/efficacy measures of the practice(s)?	2		
III.D.3. *p*[c]	**Measurement of impact/effectiveness of the best practice(s):** Are there outcome/effectiveness measures of the practice(s)?	2		
III E *p*[c]	**Plan to permit process/efficacy evaluation results to modify implementation of the best practice(s):** Will the implementation of the practice(s) be adjusted to reflect intermediate/process evaluation results?			
IV.	**Best practices implementation and evaluation study results and analysis:** Clear reporting of results is emphasized, but the most points are given for showing how the evaluation results support the overall findings of the study. Important note: Negative results, null results, and significant results may all be reported (policy[c]). The results are evaluated for clarity, comprehensiveness, and degree of support for the overall findings. In short, significant results are not a prerequisite for this competition (IV. A + IV. B).		4	**20**

Table 1
(continued)

Item No.	**Evaluation Criteria** (with descriptions[a] and with total possible points[b]) Note: References to "occupational health psychology" represent the topics relevant to the Work, Stress, and Health Conference. Replace this phrase with "occupational safety and health" or a similar phrase for other applications of these criteria.	Scores[b] Subcategory Elements	Subcategories	Main Categories
IV.A.	**Reporting the results (both quantitative and qualitative) (IV.A.1 + IV.A.2 + IV.A.3 + IV.A.4).**		12	
IV.A.1. *p*[c]	**Descriptive data clearly reported**	2		
IV.A.2.	**Analytic data clearly reported**	2		
IV.A.3.	**Report of results that address validity and reliability:** Are relevant questions of validity and reliability addressed clearly and to the degree warranted by the study? Note reports for individual items as well as assessment of construct validity and scale reliability.	4		
IV.A.4.	**Overall quality and comprehensiveness of the report of results:** Is the report of the results complete and clearly described? Are important questions about the study missing?	4		
IV.B.	**Quality of the results with respect to the overall findings and conclusions:** Are the results clearly tied back to the overall purpose, findings, and conclusions regarding the best practice(s)?		8	
V.	**Implications of the evaluation results for occupational health psychology**: How well do the researchers relate their best practice(s) evaluation back to their selected topic(s) in occupational health psychology? (V. A + V. B + V. C + V. D + V.E.)			**10**
V.A.	**Placing this study into the broader context of occupational health psychology**		2	
V.B.	**Limitations of the study:** Are the methodological and other limitations of this study considered in the discussion? How comprehensively have those limitations been addressed in the entire paper? It is not necessary that there be a clearly labeled section titled, "limitations." Score proposed solutions to the identified limitations in V.C. and V.D., below.[d]		2	
V.C. *p*[c]	**Recommendations for improvements to the best practice(s):** Reflecting back to the statement of the need for the best practice(s) (I.).		2	
V.D.	**Recommendations for improved evaluation studies of best practice(s) (new or old)**		2	
V.E.	**New research questions**		2	
TOTAL POSSIBLE SCORE (I + II + III + IV + V):				**100**

Note. Additional prepublication criteria not addressed above, and NOT scored:
1) Mastery of the relevant literature, 2) Theoretical/conceptual framework, and 3) Appropriateness of topic for JOHP/APA.
[a]All of the criteria are considered in the review of papers for the competition, however it is neither anticipated nor required that any single paper address every criterion.
[b]Points in sub-category elements sum to produce sub-category totals; points in sub-categories sum to produce main category totals.
[c]Policy implications for best practices and other interventions included in this criterion or set of criteria.
[d]New criterion; added for the 2008 competition.

team to involve both workers and managers in a joint process of developing the practice or intervention. Similarly, Section III.E. asks if there exists some sort of recursive process to allow input from prior measures of intervention efficacy (III.D.2.) to be used to modify future implementations of the study intervention (III.A.). Classically trained experimental scientists may ask: (a) How do scientists engage their study subjects (II.C.), or (b) how do scientists measure a changing intervention (III.E.), all without introducing bias into the evaluation? Nevertheless, there are many possible solutions—from using key informants, to developing a comprehensive approach to participatory action research (e.g., Brydon-Miller & Tolman, 1997). The development of a simple nursery tool by ergonomists at the University of California, Davis (featured in Baron, Estill, Steege, & Lalich, 2001), is a model for good participation from workers and managers without compromising the scientific rigor of the evaluation methodology (Janowitz et al., 1998). Furthermore, involving partners and stakeholders in planning, developing, executing, revising, and evaluating interventions is an explicit component of the NIOSH program, "Research to Practice, r2p" (available at: http://www.cdc.gov/niosh/r2p/).

The terminology "comparison group" (III.A.1.) was selected explicitly in contrast to "control group." Experimentally trained scientists often do not appreciate that there are limits to the appropriateness of randomized controlled studies, especially with intact workgroups. Indeed, in many situations, appropriate comparison treatments can yield better and more realistic results than the somewhat artificial, albeit classic "placebo control group" (e.g., Mohr, 1995; Victora et al., 2004). Note: this is not to say that randomized controlled trials with placebo control groups cannot be effectively included in carefully designed studies; randomized controlled trials have a very important place in evaluation methodology. The point is that randomized controlled trials constitute one important subset contained within the superset of all possible comparison group designs. This issue is part of a much larger discussion that goes well beyond the evaluation criteria discussed here (e.g., Rosenstock & Thacker, 2000; and the response: NORA Intervention Effectiveness Research Team, 2001).

The scientific method contains no inherent morality or ethics. It is up to every scientist to conduct research that follows the highest ethical standards. Unfortunately, most of the ethical standards to which we all adhere in our research derive from ethical failures, not successes (e.g., Angell, 1997; Lurie & Wolfe, 1997). Nevertheless, the ethical application of social science evaluation methods to the workplace—first pioneered in elementary education—represents a big challenge for all of us (Shadish et al., 2002). More than 20 years ago, Prof. Don Campbell recognized this challenge when he advocated "contagious cross-validation" and "competitive replication" as two important strategies among many to address the evaluation of entire programs (Campbell, 1984). One of the many points in Campbell's essay is that no one study will be without methodological flaws. However, when we examine a collection of similar studies from different communities and workplaces, it is possible to do high-quality science in the applied arena—such as in evaluating best practices and other interventions in occupational health psychology, as well as in other areas of occupational safety and health.

Conclusion

The current evaluation criteria are presented at this time because of the 2006 and 2008 Work, Stress, and Health conference competitions. However, if these criteria are to achieve any true usefulness, it will be as a starting point from which evaluators will continue to "raise the bar" by improving the methodological rigor of such studies and by promoting safety and health at work.

References

Angell, M. (1997). The ethics of clinical research in the Third World. *New England Journal of Medicine, 337,* 847–849.

Baron, S., Estill, C. F., Steege, A., & Lalich, N., (Eds.) (2001). *Simple solutions: Ergonomics for farm workers.* Cincinnati, OH: US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2001–111.

Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative effectiveness of worker safety and health training methods. *American Journal of Public Health, 96,* 315–324.

Brydon-Miller, M., & Tolman, D. L. (Eds.) (1997). Transforming psychology: Interpretive and participatory research methods. *Journal of Social Issues, 53,* 597–827.

Campbell, D. T. (1984). Can we be scientific in applied social science? In R. F. Conner, D. G. Altman, & C. Jackson, (Eds.), *Evaluation studies: Review Annual* (v. 9, pp. 26–48). Beverly Hills, CA: Sage.

Goldenhar, L. M., & Schulte, P. A. (1996). Methodological issues for intervention research in occupational health

and safety. *American Journal of Industrial Medicine, 29,* 289–294.

Heacock, H., Koehoorn, M., & Tan, J. (1997). Applying epidemiological principles to ergonomics: a checklist for incorporating sound design and interpretation of studies. *Applied Ergonomics, 28,* 165–172.

Janowitz, I., Meyers, J. M., Tejeda, D. G., Miles, J. A., Duraj, V., Faucett, J., et al. (1998). Reducing risk factors for the development of work-related musculoskeletal problems in nursery work. *Applied Occupational & Environmental Hygiene,* 13, 9–14. (Now published as the *Journal of Applied Occupational & Environmental Hygiene.*)

LaMontagne, A. D., Keegel, T., Louie, A. M., Ostry, A., & Landsbergis, P. A. (2007). A systematic review of the job-stress intervention evaluation literature, 1990-2005. *International Journal of Occupational and Environmental Health, 13,* 268–280.

Lurie, P., & Wolfe, S. M. (1997). Unethical trials of interventions to reduce perinatal transmission of the human immunodeficiency virus in developing countries. *New England Journal of Medicine, 337,* 853–856.

Mohr, L. B. (1995). *Impact analysis for program evaluation* (2nd ed.). Thousand Oaks, CA: Sage.

NORA Intervention Effectiveness Research Team. (2001). May 2000 Supplement on preventing occupational injuries [Letter to the Editor]. *American Journal of Preventive Medicine, 20,* 308–309.

NORA Intervention Effectiveness Research Team. (2004). *Does it really work? How to evaluate safety and health changes in the workplace.* Cincinnati, OH: US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2004–135.

NORA Organization of Work Team. (2002). *The changing organization of work and the safety and health of working people: Knowledge gaps and research directions.* Cincinnati, OH: US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2002–116.

Robson, L. S., Shannon, H. S., Goldenhar, L. M., & Hale, A. R. (2001). *Guide to evaluating the effectiveness of strategies for preventing work injuries: How to show whether a safety intervention really works.* Cincinnati, OH: US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2001–119.

Rosenstock, L., & Thacker, S. B. (2000). Toward a safe workplace: the role of systematic reviews. *American Journal of Preventive Medicine, 18*(4, Suppl. 1), 4–5.

Routsalainen, J. H., Verbeek, J. H., Salmi, J. A., Jauhiainen, M., Laamanen, I., Pasternack, I., et al. (2006). Evidence on the effectiveness of occupational health interventions. *American Journal of Industrial Medicine, 49,* 865–872.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

Sinclair, R., & the NORA Intervention Effectiveness Research Team. (2004). The National Occupational Safety and Health Intervention Effectiveness Research Agenda at Nine: A look back at the science of looking back after taking action. Unpublished manuscript.

Smith, C. A., & Shannon, H. S. (2003). How much science is there in injury prevention and control? *Injury Prevention, 9,* 89–90.

Victora, C. G., Habicht, J. P., & Bryce, J. (2004). Evidence-based public health: moving beyond randomized trials. *American Journal of Public Health, 94,* 400–405.

Received October 31, 2007
Revision received October 31, 2007
Accepted November 1, 2007 ■