La Medicina del Lavoro Med Lav 2009; *100*, 4: 247-257

Systematic reviews of workplace injury interventions: What are we missing?

Hester J. Lipscomb, Lisa A. Pompeii*, D.J. Myers, Ashley L. Schoenfisch, J.M. Dement
Division of Occupational and Environmental Medicine, Department of Community and Family Medicine, Duke University Medical Center, Durham
* Division of Epidemiology, The University of Texas, School of Public Health, Houston, Texas

Key words
Occupational injury; observational study; evidence-based prevention

Summary
Background: *There are pitfalls associated with applying a biomedical model with its emphasis on experimental designs to the evaluation of workplace injury interventions.* **Objectives:** *Evaluation over enough time is essential in occupational safety when interventions are expected to have a latent effect as well as to assess sustained effects. Controlled trials are not well-suited to this task and are not even possible in circumstances where a policy change, such as legislative action, affects a population of workers simultaneously. Social context influences occupational injury interventions, their evaluation and the wider generalization of findings but is lost in the pooling of data for meta-analyses. Some of these issues can be addressed through recognition of the contribution of diverse observational methodologies in intervention evaluation, improvement and maintenance of robust surveillance systems, and inclusion of qualitative methodologies not typically embraced by epidemiologists or medical researchers.* **Methods:** *Through consideration of an evaluation of a legislative effort to prevent falls from height in construction, we demonstrate lack of flexibility in current methods used for evaluating time series analyses in systematic reviews of occupational injury intervention effectiveness.* **Discussion and conclusions:** *These include the manner in which downward change in slope is assessed and the call to demonstrate a significant initial downward change in level. We illustrate essential contextual detail regarding this intervention that is lost in the pooling of data from multiple studies into a combined measure of effect. This reduction of occupational injury intervention evaluation to one of pure statistical significance is ill-conceived, irresponsible, and should be stopped.*

Riassunto
«Revisioni sistematiche di interventi di prevenzione di infortuni sul lavoro: cosa stiamo perdendo?». *L'applicazione di modelli biomedici, con la loro enfasi sul disegno sperimentale, alla valutazione dell'efficacia degli interventi di prevenzione degli infortuni sul lavoro nasconde delle insidie. Nel campo della sicurezza del lavoro è essenziale che la valutazione sia eseguita su un arco di tempo sufficientemente lungo, quando ci si aspetta che gli effetti degli interventi compaiono dopo una certa latenza oppure quando si valutino effetti persistenti nel tempo. I trial non sono adatti a questo scopo e non è nemmeno possibile eseguirli nel caso di cambiamenti di politica generale, come nel caso di nuove leggi, che interessano contemporaneamente un'intera categoria di lavoratori Il contesto sociale in-*

Pervenuto il 12.1.2009 - Accettato il 11.3.2009
Corrispondenza: Hester J. Lipscomb, Ph. D., Division of Occupational and Environmental Medicine, Box 3834, Duke University Medical Center, Durham, N.C. 27710, United States of America - Tel. 919 684-8175 - Fax 919 286-1620
E-mail: hester.lipscomb@duke.edu

fluenza gli interventi di prevenzione degli infortuni, la valutazione della loro efficacia e la generalizzazione dei risultati, ma questa influenza viene perduta quando si mettono insieme i dati per procedere ad una metanalisi. Ad alcuni di questi problemi si può ovviare tramite: il riconoscimento del contributo di diverse metodologie di osservazione nella valutazione dell'efficacia degli interventi; il miglioramento e il mantenimento di sistemi di sorveglianza affidabili, l'inclusione di metodi qualitativi non utilizzati normalmente da epidemiologi o ricercatori. Partendo dal caso della valutazione dell'efficacia di un intervento legislativo per la prevenzione delle cadute dall'alto in edilizia, viene dimostrata la mancanza di flessibilità dei metodi attualmente utilizzati per valutare i risultati delle serie temporali nelle revisioni sistematiche riguardanti l'efficacia sul campo di interventi di prevenzione, compreso il modo con il quale si valuta il cambiamento di pendenza della curva ed il fatto che sia richiesta la dimostrazione di un decremento iniziale significativo. Viene illustrato il dettaglio contestuale essenziale di questo intervento legislativo, che viene perduto nel mettere insieme i dati di più studi al fine di ottenere una misura di effetto combinata. Ridurre in tal modo la valutazione degli interventi di prevenzione degli infortuni alla pura significatività statistica è mal concepito, irresponsabile e dovrebbe essere abbandonato.

Background

There is a growing call for the practice of evidence-based clinical medicine. Over a decade ago, Sackett et al., in an editorial in the British Medical Journal, described evidence-based medicine as "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients " (43). However, recommendations derived from population-based research do not necessarily provide the best recommendations for any given individual (36, 37). Appropriately, these authors went on to clarify the essential integration of clinical expertise and evidence from systematic research and, in the case of patient care, individual patient's preferences and choice. But in making decisions about therapies, they caution about the use of non-experimental approaches.

The Cochrane Collaboration was developed in 1993 with the goal of improving healthcare decision-making globally through systematic reviews of the effects of healthcare interventions. Fairly recently, the Collaboration has formally extended to occupational health and safety through efforts coordinated by the Finish Institute of Occupational Health (www.cohf.fi) (4). In the light of often scarce financial resources for occupational safety (21, 44), thoughtful evaluation is imperative in the occupational safety research setting. Recommendations for evidence-based practice are certainly relevant to public health decision making just as they are in clinical practice. However, the methods used to establish evidence for public health practice, including occupational safety, must be appropriate to the task and consider the relevant context.

Critical appraisal of a systematic review and meta-analysis

A recent systematic review and meta-analysis concluded there was no evidence that legislation is effective in preventing injuries in the construction industry (19). The meta-analyses included three studies from the US of legislative interventions (8, 23, 46) involving two Federal Occupational Safety and Health Administration (OSHA) regulations (8, 46) and a state level standard change (23). The respective outcomes and populations assessed in each of these studies were deaths from trenching in the U.S., deaths from falls in the U.S., and falls from height among union carpenters in Washington State. Our comments focus on the Washington State analyses and expand upon a letter to the editor of the American Journal of Preventive Medicine where the systematic review was published (22).

The review was limited to (Randomized Control Trials) RCTs, cluster RCTS, controlled before-after studies, or interrupted time series studies (19). The three studies addressing legislative actions were considered time series. Inclusion criteria for

time series analyses were developed by the Effective Practice and Organization of Care (EPOC) review group (10); the intervention must have occurred at a clearly defined time, and there had to be at least 3 data points before and after the intervention. Quality of these studies was rated as adequate or not in six domains:

1) protection from secular changes; the intervention is independent of other changes;

2) reliable statistical inference enabled based on sufficient data points for statistical inference and formal test for trend;

3) the intervention was unlikely to affect data collection;

4) blinded assessment of outcome;

5) completeness of the data set;

6) reliable primary outcome method.

None of the three studies of policy interventions received an adequate score on the first criteria. In fact, if they had, it would have been surprising. It is hard to imagine any situation in which a broad policy change designed to influence industry-wide worker safety could occur independently of other changes.

None of the studies were considered to have reliable statistical inference based on sufficient data points and formal test for trend based on methodologies accepted under EPOC criteria. The number of data points is quite arbitrary in the analyses of administrative data involved in all three of these studies. Investigators can cut the available data more finely to reach the proscribed number of data points, but finer cutting results in less stable measures. In the evaluation of the Washington State Fall Standard, Lipscomb et al., had data on work-related injuries and hours of union work for a cohort of over 16,000 carpenters for a ten-year period including pre-intervention data for two years prior to the standard change and eight years of post-intervention follow-up (23). Of note, injury rates calculated at 6-month intervals in the Washington standard evaluation (23) were condensed to yearly estimates in the review article (see Table 2 in Lehtola et al, 2008) (19).

The Washington State Fall Standard followed a record number of twenty-two fall deaths in Washington in 1988. Considerable publicity followed these events and the promulgation of the standard, but preceded the date the standard became effective in 1991. Given the nature of the data available and this background, the use of autoregressive integrated moving average (ARIMA) models was less than ideal; the pre-intervention trend would have been established in a period of time in which injury rates might very well have been influenced by the publicity surrounding events that led to the intervention.

Poisson regression was chosen as the multivariate analytical tool to allow assessment of injury rate ratios before and after the intervention through progressive lagging of time windows to assess potential latent effects. The models were adjusted for known risk factors for falls among this cohort (age, gender, time in the union). Based on the decreasing overall injury rates in the construction trade, the magnitude of decline of non-fall injuries among this cohort of workers was also controlled for in the analyses. The expectation that the Washington standard would not be effective immediately was clear, and the emphasis of the lagged analyses was exploration of latency. Similar use of Poisson regression has been reported in evaluating changes in injury rates over time following an intervention (17). As in the assessment of dose response, the issue of importance was felt to be the pattern, or lack thereof, rather than the level of statistical significance.

The most substantial decline in fall rates was seen between 3 and 3.5 years after the standard went into effect. The interpretation by Lehtola et al., (19) was a lack of any sustained intervention effect based on a lack of a significant initial downward change in slope or level. In this interpretation neither the circumstances leading to enactment of the legislation or reasonable latency are considered. Knowledge of when interventions designed to influence complex work practices become effective and when they lapse is essential to understanding and improving worker safety. In a workplace injury intervention effectiveness study, where worker reporting (self-report, report to workers' compensation, etc.) captures the outcome of interest, basing a determination of success or failure upon an initial change in level also fails to recognize potential effects of an intervention on reporting.

The review assessment was based on overall injury rates from falls from height despite the more pronounced pattern for more severe cases that resulted in the loss of time from work (which occurs after the 3rd lost day under Washington State compensation). These injuries were more likely to be representative of falls from 6 feet or greater that would be influenced by the standard. The review did not acknowledge a decline in the number of days injured workers lost or the associated workers' compensation costs, including a reduction in medical costs (adjusted for age and gender) in a period when non-fall related medical costs increased.

An acknowledged limitation of the evaluation was the lack of information about compliance and/or enforcement (23). The work of Nelson et al., addressing injury rates among inspected contractors compared to a control group of employers who were not inspected, supported specific deterrence and suggested that a greater decline in injury rates might be expected if more employers were inspected (29). These data were not considered in the review because of concerns about completeness of the data.

HOW DID WE GET HERE?

In the 1960's when Sir Austin Bradford Hill outlined what have come to be known as criteria for causal inference (11), he first listed strength of association. It is not surprising that epidemiologists spend considerable efforts on improving effect measures. Basic tenets of our population-based science call for reduction of misclassification through careful measurement of clearly defined exposures and outcomes, and in recent years the science has become even more enamored with analytical techniques designed to precisely define strength of association. Systematic reviews of literature and meta-analyses, in which data from multiple studies are pooled to improve the precision of analyses, are among these tools. These approaches are now used fairly commonly in situations in which causality is being assessed, as well as when the effects of interventions are being evaluated. In so doing, often strict inclusion criteria are applied.

Choice of study design

Randomized design

While viewed as the 'gold standard' in evaluating effectiveness of clinical treatments, randomized controlled trials (RCTs) also have limitations (3). Some of these limitations have particular relevance when considering the application of RCTs to occupational injury research. Randomized controlled studies should not be considered superior to other study designs without careful consideration of issues surrounding randomization, including the validity of the process for randomizing and selecting the sample to be randomized. The goal of randomization is to have similar control and experimental groups at the initiation of the trial and reduce bias related to the unmeasured differences between groups (18). The blind acceptance of RCTs as the superior study design to assess occupational interventions fails to recognize the very complex and highly dynamic nature of workplaces that actually challenge the validity of randomization and pose threats to the generalizability, or external validity, of the study.

Acute injury occurs because of energy transfer and the proximal cause is usually easy to identify. Rarely is defining the temporality of exposure and injury an issue, and we are not typically concerned about long latency periods. However, occupational injury can, and often does, result from a complex mix of factors that include personal behaviors, as well as tool equipment/design, work norms, environments and relationships among workers and between workers and their supervisors or managers. Injury may also be associated with formal and informal policies that influence the work people do and the conditions in which they work, including their exposures and the availability and acceptance of control measures. All these factors have the potential to influence workplace interventions.

Work environments which contribute to the risk and pattern of occupational injury represent another level of context that is nearly always the result of additional factors. These include demographic patterns and shifts, technology, government regulations and policies, degree of unionization in broad-

er society as well as in the workplace itself, unemployment rates and changes in labor and consumer markets across the globe.

Depending on the design of the experiment, the number of participants or work units to be included, and the prevalence of factors an investigator is hoping to equalize across groups, randomization may not be efficient; matching cases and controls based on designated criteria may come closer to the goal of having similar groups to study. In other situations randomization is not feasible. For example, policy interventions, particularly those initiated at a governmental level, are intended to impact all affected workers simultaneously.

Effectiveness of medical interventions, or at least desired intermediate effectiveness measures, can often be assessed in relatively short time periods. Interventions designed to prevent occupational injury typically require substantial changes in complex environments. While not an issue in assigning causality of acute injuries, consideration of latency may be critical in understanding effects of occupational interventions. Some workplace interventions would need to be in place for several years to gain a clear understanding of effectiveness. In the case of relatively rare events where very large samples would be needed to ensure appropriate statistical power to detect meaningful changes, feasibility, and even usefulness, of tracking large randomized occupational cohorts over substantial periods of time within dynamic workplaces becomes questionable.

Observational studies

"(syn: nonexperimental study) Epidemiologic study in situations where nature is allowed to take its course; changes of differences in one characteristic are studied in relation to changes or differences in other(s), without intervention of the investigator." [A Dictionary of Epidemiology (JM Last, 1988).]

Observational studies have both strengths and limitations, regardless of design; these have been, and continue to be, enumerated through multiple sources (5, 13, 32, 39, 40). In describing sources of error and their frequency, Rothman and Greenland point out that to varying degrees, nearly every observational study will have nearly every type of error (39, 41); randomized controlled trials have most of them as well. A basic concern is bias in results caused by systematic error, whether from study design, information acquisition, participant recruitment/retention, or uncontrolled confounding.

In our focus on what observational data may lack, we can lose sight of the strengths and contributions observational studies have made. "Observational data on human disease and mortality are not intrinsically frail. On the contrary, they are our most crucial source of information on the patterns, causes and trends of disease and death in human beings in their natural habitat – human society. For that, experiments are next to useless" (Michael Coleman, 2007) (6). The emphasis on RCTs as the superior study design for determining an intervention's effectiveness can be seen in the occupational health literature as investigators apologize for failure to have used a trial, even in situations where it was not feasible or even possible, rather than emphasizing the strengths in their observational approaches and use of robust surveillance systems (25, 21).

When choosing a study design, the researcher needs to consider the relative merits of that design, always seeking the most suitable one for the question. But design is always influenced by feasibility. Furthermore, it is expected in observational science that more than one study will be required to establish strong evidence which comes through the synthesis of findings (5). Bradford Hill's consideration of consistency, coherence, and analogy all speak to this important concept (11).

Context

Oftentimes intervention studies fail to adequately address context that may be influential in determining immediate, as well as more long-term, effects of workplace interventions. There are many situations where the investigator serves solely as an evaluator rather than a designer of the intervention. In such cases it is important to understand what led to the recognition of the need for an intervention, its subsequent design, and its eventual

implementation; promoters of and barriers to implementation, adoption and compliance; and factors that could influence measures related to the intervention or the outcome(s) of interest. Changes in workplace environments over time rarely can be predicted and cannot be avoided or controlled by the investigator. The assessment of whether the changes that occur are substantive enough to affect outcomes of interest is, in large part, a qualitative one. Conceptual models of when the intervention realistically should have an effect must also consider particular outcome(s) being measured.

Determinants of health, including working conditions, are influenced by many things that cannot be measured at the individual level such as patterns of social inequality, spoken and unspoken cultural norms, as well as movement of capital to other countries, laws limiting tort cases, and industries' efforts to break unions (7, 9, 14-16, 24, 28, 30, 35, 38, 45). As we combine findings across studies, we further reduce complex data into a series of point estimates that can be used for meta-analyses. Systematic reviews do not have to be limited to RCTs to lose important information. In the synthesis and summarization of material from complex observational or experimental studies detail is always lost.

BACK TO OUR ROOTS AND BEYOND

"…it is…useful to take stock at regular intervals of who we are, where we have come from, and what has happened to our luggage."

Neil Pearce, International Journal of Epidemiology, 2007

As epidemiologists we view ourselves as quantitative scientists. We make assumptions that data are "superior" if they are quantitatively measured, objectively collected, there are a lot of them, they were collected with a probability sample or better yet from a randomized trial. These approaches implicitly ignore context and foster the belief that the world is made up of only individuals. However, we are not always measuring what we think we are, and our analyses are not as pristine as sometimes portrayed. Numerous decisions are made about how we will handle our data and modeling, and interpretation of findings is not an entirely objective process even in studies that meet the strict criteria for inclusion in these reviews.

In attempts to make our work more accessible, and hopefully more useful, we have established criteria for including findings in systematic reviews that are not unreasonable; but they are not flexible. The criteria established by EPOC, while useful perhaps for healthcare decision making, fail to adequately address the challenges faced in the evaluation of workplace interventions and possible alternatives including qualitative methods and quasi-experimental designs (49).

Unfortunately, as epidemiology has become more of a biomedical discipline, it may be failing as the basic science of public health. Scholars of epidemiology (14-16, 27, 28, 33, 45) have emphasized the importance of social, economic, environmental, and cultural determinants of population health. These macro-level factors have great potential to influence effectiveness of interventions designed to prevent injuries in highly dynamic workplaces, and they should likely be considered more in study design. "Knowing 'what works' does not necessarily translate into implementing injury prevention in the real world" (12). The efficacy of interventions observed in randomized and controlled settings may not indicate effectiveness in everyday work settings.

It is important that we continue to explore ways to improve the accuracy and precision of our measurements of outcome, exposure and effects as we seek to refine measures of association. But let us not forget that Hill listed seven other considerations besides strength of association in making causal inference (consistency, specificity, temporality, biological gradient, biologic plausibility, coherence, and analogy) as well as experimental evidence. He considered no single criterion sufficient to establish causality, and, of note, statistical significance was not included in his list. In fact, he comments that formal statistical tests cannot answer questions of causality, "they remind us of the effects that the play of chance can create, and they will instruct us in the likely magnitude of those effects. Beyond that they contribute nothing to the 'proof'

of our hypotheses" (11). Similarly, it seems unlikely that occupational injury intervention evaluation can be reduced to an issue of statistical inference. The precision sought by combining data results in a loss of breadth and richness which is necessary in understanding findings beyond the level of a probability value.

This deviation from sole dependence on statistical measures does not mean that statistical methodologies are not important; in fact, the growing interest in applications, such as multi-level modeling, that allow a more detailed exploration of contextual variables (9) is one useful example. However, multi-level modeling still has all the trappings of other quantitative work, and perhaps misleads us into thinking we are adequately addressing context as we create more complex coefficients. To evaluate broader forces, broader designs are needed; consideration should be given to the use of more historical and comparative designs that provide insight on broader levels. This approach is, after all, where epidemiology started: including the context, culture, and socioeconomic conditions that so strongly influence health (31, 32). In 1997, in his treatise on the failure of academic epidemiology, Shy described the biomedical fallacy, specifically "the error in inferring that risk factors for diseases in individuals can be summed to understand the causes of disease in populations, or that the health of a population can be explained entirely in terms of characteristics of individuals" (45). While concerns have been raised about ecologic fallacy, Shy further pointed out that the biomedical fallacy is likely to have greater consequences for public health.

Hill's 40 year-old recommendations still provide important guidance for analytical thought, synthesis and judgment not just in the interpretation of causal relationships but also as we assess the utility of workplace interventions. We should remind ourselves that the actual process of applying these considerations is a qualitative one that involves searching for the best we can define at the moment. Sometimes external validity can be enhanced with methods beyond traditional epidemiology that get ignored in systematic reviews with strict inclusion criteria (1, 34, 26, 44). It is not always feasible to conduct RCTs and at times they are not even appropriate (2), and criteria established for time series analyses, if applied too rigidly, may fail to adequately consider context. Surveillance systems that are robust, in both accuracy and size, can provide an important foundation and infrastructure for more robust evaluations of policy interventions in particular. Inclusion of contextual variables could markedly improve these systems.

Discussion

In this presentation, we highlighted relative merits, limitations, and challenges when applying the same evidence-based approach being called for in clinical care to the evaluation of interventions designed to prevent injury in the workplace. We framed our work around the premise that occupational injury epidemiology should be an applied science with the goals of understanding factors and circumstances that contribute to injury risk in populations, development of interventions to prevent injuries and their sequelae, and the evaluation of the effectiveness of those interventions. We need to understand why interventions work or not and under what conditions.

Systematic reviews typically exclude the bulk of reviewed research for failure to meet inclusion criteria. The minimization of complexities (reductionism) might be the ideal in a clinical trial where investigators are interested in eliminating noise by controlling for "everything," even though this in itself is a counterfactual ideal. In contrast to such reductionism is holism, where study complexity is embraced. Holism seems more ideal for intervention studies in a workplace setting where the researcher cannot control, or sometimes fully delineate, complexities.

Short-term intervention evaluations (including randomized trials) may not allow adequate time for differences in outcomes to be realistically assessed; these studies should place more emphasis on intermediate measures that indicate adoption. Long-term evaluations will be influenced by many factors other than the intervention that can be difficult to measure and control for, especially in evaluation of

industry-wide interventions such as legislative changes. But this is the real, highly dynamic world in which workers function. Is this not what effectiveness is about – determining if something works in the real world situation? Effectiveness evaluation is about more than statistical inference.

The systematic review we described in this presentation (19) included work of our own in which we assessed changes in falls from height among a large cohort of carpenters after a policy change in the US (23). Randomization was not possible to evaluate this state-wide policy change, and our interest was in long-term rather than immediate effects. Substantive declines in the rate of falls from height, as well as reduced paid lost days and compensation costs for falls from height were noted after considering the temporal trends of non-fall injuries among the same cohort. In our report, we recognized limitations to recommended practices based on the circumstances and strove for an appropriate and reasonably robust analysis given the nature of the existing data and the contextual circumstances surrounding the promulgation of the standard. In the systematic review, the original data were taken out of context and markedly condensed to allow meta-analyses with data from two other studies. The review concluded that there is no evidence that legislation is effective in prevention of injuries in construction, an ill-conceived and perhaps dangerous conclusion based on preconceived constraints as to how data of this nature should be analyzed.

Our assessment can certainly be interpreted as nothing more than defense of our own work. We are not saying this work was exemplary, and many limitations were acknowledged in the manuscript. However, we believe it is more robust than reflected through the transfer and reanalysis of 10 crude de-contextualized data points. We focused our comments on our own work because these are data with which we are quite familiar. However, we suspect there are other investigators whose work has also been unreasonably condensed, with perhaps misleading conclusions, through the process of systematic review and meta-analysis.

Black (3) has pointed out the naïve understanding held by the research community as to how policy is made and our unrealistic expectations for what our work can achieve. We echo his caution in accepting the notion of evidence-based policy without critical appraisal. He emphasizes that researchers "need to understand that there are many sources of evidence, that sensible decisions may not reflect scientific rationality, and that context is all important…" (p. 227)

Unfortunately, the current approach to the systematic review of occupational injury interventions fails to mention potential contributions of qualitative methodologies (e.g., interviews, focus groups, ethnographies, case studies, comparative methods) that provide other sources of data to include in syntheses including important contextual detail that may help us understand why an intervention is effective in one population and perhaps not in another, the meaning of interventions to workers, and information that may be particularly effective in influencing policy.

This discourse undoubtedly leads to the question of whether it is the role of the injury epidemiologist to influence occupational safety policy. If the researcher views herself as a public health practitioner, it is not just her role, it is her responsibility. If she views herself only as a researcher, perhaps it is not, but then what is the point of the work? One could also argue that ethically it is still her responsibility to make sure the work is interpreted correctly, particularly by those who may seek to use her work to influence policy change or lack thereof. In either case, science should not be confused with measurement; measurement is a tool, science is understanding (31).

Conclusions

Systematic reviews of occupational injury interventions can provide useful and concise information about a body of prior work including an important critical appraisal. They provide a way through which the occupational safety community can learn from the collective experiences of others, and this can, among other things, lead to improvement in our subsequent work. Meta-analyses can provide more precise estimates of effect, but we

must recognize that the precision lacks depth, and efforts to improve our understanding of occupational safety interventions should not be focused solely on refinement of effect measures.

What are the realistic expectations of systematic reviews for occupational injury interventions? Systematic reviews and meta-analyses are certainly ways to improve communication among researchers, healthcare-providers and in the case of occupational safety, public health practitioners. Through the process of conducting these reviews, volumes of information are assembled and condensed (43) that can make the information more accessible. However, in so doing they risk reducing multiple pieces of work – sometimes quite complex – into a short appraisal. Consequently, the already reductionist quality of much of our research is magnified; this may contribute to a "paralysis by analysis" (Andrew Watterson, 2007) (48).

In addition to methodological aspects, the criteria for including studies in a meta-analysis of occupational safety interventions must include some consideration of the differences in study settings. Close attention must be paid to the contexts of the studies being pooled since a variety of notable differences between study populations and settings may essentially create misclassification of interventions and outcomes making it more difficult to detect effects and possibly outweighing or invalidating any justifications for combining them. We do a disservice to science and to workers by reducing workplace intervention evaluation to one of pure statistical inference. Intervention effectiveness evaluation for public policy cannot be reduced to simple quantitative measures; meta-analyses of policy interventions are unlikely to add value to these questions and should be stopped.

The importance of observational studies as well as qualitative methods, traditionally not embraced by epidemiologists, should not be forgotten. As we focus efforts on workplace interventions and the assessment of their effectiveness, the importance of the development and maintenance of surveillance systems that support observational methods should not be neglected. Study design alone does not provide a reasonable threshold to assess the utility of an intervention evaluation, and investigators do bear responsibility for providing adequate detail on the intervention and the surrounding context to allow a fair determination of evidence of effectiveness (42, 47). The context upon which design and analysis decisions were based should not be ignored.

There is a propensity, and perhaps a desire, to view study results in a deterministic manner. Systematic reviews foster this tendency with both merits and liabilities. Regarding the latter, we miss potentially important complexities to the context from which a study's findings were birthed, perhaps at the cost of health and safety of workers. Systematic reviews are not a panacea; they are not *the* answer to understanding occupational injury interventions. Rather, they are one tool that can help us build upon the works of each other. If we forget this, they can also keep us from, perhaps less than perfect, but needed action.

"All scientific work is incomplete – whether it be observational or experimental. All scientific work is liable to be upset or modified by advancing knowledge. That does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand."

Sir Austin Bradford Hill

"All the fruits of scientific work, in epidemiology or other disciplines, are at best only tentative formulations of a descriptive nature, even when the work itself is carried out without mistakes. The tentativeness of our knowledge does not prevent practical applications, but it should keep us skeptical and critical, not only of everyone else's work but of our own as well."

Rothman and Greenland

No potential conflict of interest relevant to this article was reported

References

1. Baril-Gingras G, Bellemare M, Brun J: The contribution of qualitative analyses of occupational health and safety interventions: an example through a study of external advisory interventions. Safety Science 2006; *44*: 851-74

2. BLACK N: Why we need observational studies to evaluate effectiveness of health care. Br Med J 1996; *312*: 1215-1218
3. BLACK N: Evidence-based policy: proceed with care. Br Med J 2001; *323*: 275-279
4. The Cochrane Collaboration. Cochrane Occupational Health Field www.cofi.fi
5. CHECKOWAY H, PEARCE N, KRIEBEL D: Selecting appropriate study designs to address specific research questions in occupational epidemiology. Occup Environ Med 2007; *64*: 633-638
6. COLEMAN MP: Commentary: Is epidemiology really dead anyway? A look back at Kenneth Rothman's 'The rise and fall of epidemiology, 1950-2000 AD'. Int J Epid 2007; *36*: 719-723
7. COOK KE: Using critical ethnography to explore issues in health promotion. Qual Health Res 2005; *15*: 129-138
8. Derr J, Forst L, Chen HY, Conroy L. Fatal falls in the US construction industry, 1990 to 1999. J Occup Environ Med 2001; *43*: 853-860
9. DIEZ ROUZ AV: The study of group-level factors in epidemiology: rethinking variables, study designs, and analytical approaches. Epid Rev 2004; *26*: 104-111
10. EFFECTIVE PRACTICE AND ORGANIZATION OF CARE (EPOC): Criteria for systematic reviews. www.epoc.cochrane.org/Files/Website/Reviewer%20Resources/inttime.pdf (accessed 18 Sept 2008)
11. HILL BA: The environment and disease: association or causation? Proceedings of the Royal Society of Medicine 1965: 295-300
12. KENDRICK D: Editorial: on systematic reviews. Inj Prev 2007 doi: 10.1136/ip.2007.016501
13. KLEINBAUM DG, KUPPER L, MORGENSTERN H (eds): *Epidemiologic research: principles and quantitative methods*. New York, NY: Van Nostrand Reinhold, 1982
14. KRIEGER N: The ostrich, the albatross, and public health: an ecosocial perspective – or why an explicit focus on health consequence of discrimination and deprivation is vital for good science and public health practice. Public Health Rep 2001; *116*: 419-423
15. KRIEGER N: Theories for social epidemiology in the 21st century: an ecosocial perspective. Int J Epidemiol 2001; *30*: 668-677
16. KRIEGER N: Why epidemiologists cannot afford to ignore poverty. Epidemiology 2007; *18*: 658-663
17. KUHN L, DAVIDSON LL, DURKIN MS: Use of Poisson regression and time series analyses for detecting changes over time in rates of child injury following a prevention program. Am J Epid 1994; *140*: 943-955
18. LAST JM: *A dictionary of epidemiology*. Oxford NY: Univ press 1988: 91
19. LEHTOLA MM, VAN DER MOLEN HF, LAPPALAINEN J, et al: The effectiveness of interventions for preventing injuries in the construction industry, a systematic review. Amer J Prev Med 2008; *35*: 77-85
21. LIPSCOMB HJ: On the importance of observational methods for evaluation of interventions to prevent occupational injuries. Comment on Mancini G, Baldasseroni A, Laffi G, et al: Prevention of work-related eye injuries: long-term assessment of the effectiveness of a multi-component intervention among metal workers. Occup Environ Med 2005; *62*: 819-820
22. LIPSCOMB HJ, DEMENT JM: A counterview on data quality and the systematic review process for occupational injury interventions: Are we missing the forest for the trees? Amer J Prev Med 2009 (in press)
23. LIPSCOMB HJ, LI L, DEMENT JM: Work-related falls among carpenters in Washington State before and after the Vertical Fall Arrest Standard. Amer J Indus Med 2003; *44*: 157-165
24. LIPSCOMB HJ, LOOMIS DP, MCDONALD MA, et al: A conceptual model of work and health disparities in the United States. Int J Health Serv 2006; *36*: 25-50
25. MANCINI G, BALDASSERONI A, LAFFI G, et al: Prevention of work-related eye injuries: long-term assessment of the effectiveness of a multi-component intervention among metal workers. Occup Environ Med 2005; *62*: 830-835
26. MORSE JM: The power of the anecdote (Editorial). Qual Health Res 2006; *16*: 1019-1020
27. MUNTANER C: Commentary: Social capital, social class, and the slow progress of psychosocial epidemiology. Int J Epid 2004; *33*: 674-680
28. MUNTANER C, LYNCH JW, HILLEMEIR M, et al: Economic inequality, work-class power, social capital, and cause-specific mortality in wealthy countries. Int J Health Serv 2002; *32*: 629-656
29. NELSON N, KAUFMAN J, KALAT J, SILVERSTEIN B: Falls in construction: injury rates for OSHA-inspected employers before and after citation for violating the Washington State Fall Protection Standard. Am J Indus Med 1997; *31*: 296-302
30. O'CAMPO P, EATON WW, MUNTANER C: Labor market experience, work organization, gender inequalties and health status: results from a prospective analysis of US employed women. Soc Sci Med 2004; *58*: 585-594
31. PEARCE N: Epidemiology as a population science. Int J Epidemiol 1999; *28*: S1015-1018
32. PEARCE N: Commentary: The rise and rise of corporate epidemiology and the narrowing of epidemiology's vision. Int J Epid 2007; *36*: 713-717
33. PEARCE N, CHECKOWAY H, KRIEBEL D: Bias in occupational epidemiology studies. Occup Environ Med 2007; *64*: 562-568
34. PORTER JDH: Epidemiological reflections of contribution of anthropology to public health policy and practice. J Biosoc Sci online pub 2005: 1-12

35. QUINN MM, SEMBAJWE G, STODDARD AM, et al: Social disparities in the burden of occupational exposures: results of a cross-sectional study. Amer J Indus Med 2007; *50*: 861-875
36. ROCKHILL B: Commentary: The message is rarely simple: the J-Curve and beyond. Int J Epid 2004; *34*: 44-45
37. ROCKHILL B, SPEIGELMAN D, BYRNE C, et al: Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. J Natl Cancer Institute 2001; *93*: 358-366
38. ROSE G: Sick individuals and sick populations. Int J Epid 1985; *14*: 32-38
39. ROTHMAN KJ, GREENLAND S: Causation and causal inference in epidemiology. Amer J Public Health 2005; *95*: S144-S150
40. ROTHMAN KJ, GREENLAND S (eds): Modern Epidemiology, Pennsylvania, 1998
41. ROTHMAN KJ, GREENLAND S: Causation and causal inference. In Rothman KJ, Greenland S (eds): *Modern Epidemiology*. Chap 2. Philadelphia, Pennsylvania: Lipincott-Raven Publishers, 1998: 7-28
42. RYCHETNIK L, FROMMER M, HAWE P, SHIELL A: Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002; *56*: 119-127
43. SACKETT DA, ROSENBERG WMC, GRAY JAM, et al: Evidence based medicine; what it is and what it isn't. (editorial) Br Med J 1996; *312*: 71-72
44. SHANNON HS, ROBSON LS, GUASTELLO SJ: Methodological criteria for evaluating occupational safety intervention research. Safety Sci 1999; *31*: 161-179
45. SHY CM: The failure of academic epidemiology: witness for the prosecution. Am J Epid 1997; *145*: 479-484. Reply to comment Am J Epid *145*: 487
46. SURUDA A, WHITAKER B, BLOSWICK D, et al: Impact of the OSHA trench and excavation standard on fatal injury in the construction industry. J Occup Environ Med 2002; *44*: 902-905
47. VANDENBROUCKE JP, VON ELM E, ALTMAN DG, et al: Strengthening the reporting of observational studies in epidemiology (STROBE): Explanation and elaboration. Epidemilogy 2007; *18*: 805-835
48. WATTERSON A: Global construction health and safety - what works, what does not, and why? Int J Occup Environ Health 2007; *13*: 1-4
49. ZWERLING C, DALTROY L, FINE LJ, et al: Design and conduct of occupational injury intervention studies: a review of evaluation strategies. Amer J Indus Med 1997; *32*: 164-179

ACKNOWLEDGEMENTS: *Stefano Mattioli and Alberto Baldasseroni are acknowledged and thanked for providing the opportunity to present these ideas and the impetus to write about them. We thank Bradley Evanoff and David Wegman for thoughtful comments on an earlier version of the manuscript*