<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Scandinavian Journal of Public Health*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Järvholm, B., Bohlin, I. (2014)

Evidence-based evaluation of information: the centrality and limitations of systematic reviews.

Scandinavian Journal of Public Health, 42(13 suppl): 3-10

<http://dx.doi.org/10.1177/1403494813516713>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-86280>

Evidence based evaluation of information – The centrality and limitations of systematic reviews

Bengt Järvholm, MD PhD, Department of Public Health and Clinical Medicine, Umeå University

Ingemar Bohlin, PhD, Department of Sociology and Work Science, University of Gothenburg

Correspondence to:

Bengt Järvholm

Dept. Public Health and Clinical Medicine

Umeå University

901 87 Umeå

Sweden

E-mail: Bengt.jarvholm@envmed.umu.se

Tel +46 90 785 2241

Cell phone: +46 70 619 2241

Fax: +46 90 785 2456

Abstract

This introductory paper considers the value and limitations of the methodology of systematic reviews especially according to the evidence-based movement. It explains some terms and organizations producing systematic reviews. It also discusses controversies. The first concerns the criteria by which the quality of individual studies is assessed, the second the possible effects of the affiliation of some reviewers and the third the value of formalization of procedure, i.e. the tension between formal tools and professional judgments. The article contrasts the evidence-based formalism with other formalisms as those by IPCC and IARC. It discusses systematic reviews in social science where interventions are complex, difficult to blind and depending on context. Systematic reviews in working life research are often focusing on prevention. The formal evidence-based process may devalue or disregard findings from mechanistic and observational studies. Hence such reviews falsely conclude that existing knowledge about the risk of the factor is limited or non-existent.

Key words: Social work, working life research, IARC, IPCC

Introduction

The concept of Evidence-Based Medicine was introduced in the early 1990s, as a systematic approach to the evaluation of medical interventions (1, 2). From the outset, the focus of this concept has been on health care interventions, while little attention has been paid to other aspects of medicine, notably mechanisms and causality of disease. As a typical example of questions to which the decision support tools of EBM provide answers, one might mention the efficiency of a certain drug or behavioral therapy for smoking cessation. Experiments, controlled randomized trials in particular, are given priority over observational studies; if several trials are available, only randomized trials are included in the evaluation of the efficiency of drugs and therapies. A recent systematic review found that high intensity behavioral intervention was the most efficient therapy in hospital patients and that the addition of pharmaceutical drugs improved cessation rates (3). Another systematic review, in which pharmaceutical drugs were compared, found that the drugs were more efficient than inert controls and that one drug was better than the other ones evaluated (4). Each review included a meta-analysis; i.e., data from all studies were combined statistically. One review included 50 trials, the other 146 trials. However, the question being evaluated was somewhat different; one review was restricted to patients in hospitals and included behavioral therapy, while the other compared pharmaceutical drugs only. Both are published in respected peer-reviewed journals.

On the other hand, the issue of whether tobacco smoking causes lung cancer has been resolved by observational and mechanistic studies, as randomized trials on humans, for obvious reasons, are infeasible. For instance, the conclusion that environmental tobacco smoke causes lung cancer is based on the observation that non-smoking spouses of smokers have a higher risk of lung cancer than non-smoking spouses of non-smokers, as well as on the finding that side stream smoke contains high concentrations of chemicals that are considered as carcinogens. The acknowledged fact that smoking causes lung cancer in smokers is mainly based on epidemiological studies, whose findings, however, were not accepted without an intense scientific discussion. On one side of the debate were those who considered that the evidence was sufficient to demonstrate a causal link, when combined with some understanding of the mechanisms involved. On the other side of the dispute, further studies were deemed necessary as the data available were considered inconclusive, and possible benefits of smoking were discussed (5-7). Among those rejecting claims for a causal link between smoking and lung cancer was the famous statistician R.A. Fisher; it has been suggested that his unwillingness to accept causality was, to some extent, due to his serving as a consultant for the tobacco industry, as well as to his being a smoker himself (8).

This article considers the value and limitations of the methodology of systematic reviews. As indicated already, systematic reviews are a key resource in evidence-based medicine. Indeed, as the concept of evidence-based medicine has been extended into evidence-based practice and policy more broadly, this manner of synthesizing research findings is now equally instrumental to efforts at improving decision-making and professional practice in many areas beyond medicine and health care. While the centrality of the methodology of systematic reviews is obvious, then, the wide use to which it is put has engendered controversies over the reliability and accuracy of its results. Three themes of these controversies will be focused in this article. The first theme concerns the criteria by which the quality of individual studies is assessed in meta-analyses and systematic reviews. More specifically, we will address some of the limitations of randomized controlled trials, and the relative value of experimental and observational studies. The second theme concerns the possible effects of

some reviewers' affiliations with the corporate sector on conclusions reached in systematic reviews. Our third general theme concerns the value of formalization of procedures, and the ubiquitous tension between formal tools and professional judgments. Many of the examples we will use are Swedish. Our overall aim is to summarize key points in the critique of systematic reviews and shed new light on the limitations of this methodology.

This introductory paper is organized as follows. In the first section, the terms systematic review, meta-analysis and practice guidelines are explained. A section on agencies and organizations producing systematic reviews is followed by a discussion of the quality of reviews. Then some criticisms and sources of biases are described. The next section is devoted to other types of formalization, using the procedures of the IPCC and IARC as examples. Finally, the use of evidence based procedures in social work and working life research is discussed.

Terminology

The methodology of systematic reviews, in which the literature relevant to a given clinical issue is sampled, evaluated and synthesized, is at the heart of evidence-based medicine. The methodology is strongly formalized in some aspects, e.g. the bibliometrics. Other aspects are formalized to a lesser extent, e.g. the criteria used to appraise the quality of the studies identified and to determine which ones should be retained in the analysis. The evaluation of the quality of randomized trials can be reasonably well formalized but it is much of more of a judgment done by the evaluators in observational studies, e.g. the quality of control of confounders. The systematic review is also strongly depending on the topic of the review, i.e. the question the review should ask (compare the above mentioned two reviews of tobacco cessation interventions). A review should have one or several questions it tries to answer. A common formalism is to structure the PICO for the question, i.e. the population, intervention, comparison and outcome (9).

Systematic reviews complying with the methodology of the Cochrane Collaboration are currently regarded as the gold standard in evaluations of efficiency of pharmaceutical drugs. However, some aspects of pharmaceutical drugs, such as long-term adverse effects, are usually studied by observational methods, e.g. the negative effects of Cox-2 non-steroidal anti-inflammatory drugs on the cardiovascular system (10). The decreased risk of selection bias is an advantage of the randomized trial compared to observational studies (11). With few exceptions, reviews of pharmaceutical drugs are based on meta-analyses of several randomized trials. Meta-analysis is a method for statistically estimating the effect from several studies which are typically weighted according to their size.

Even if a systematic review favors a certain intervention it may not be put into clinical practice, e.g. it may be too costly, have too high risks of side effects, or there may not be sufficient infrastructure to use it. The concept of evidence-based decision-making is used for the process of combining findings from a systematic review with judgments from patients, clinicians, politicians and other stakeholders. Though routinely based on results from systematic reviews, practice guidelines are often developed through a process designed to take judgments from various stakeholders into account. Practice guidelines typically reflect judgments by patients, practitioners and even politicians. Though systematic reviews are sometimes assumed to be just summaries of the scientific literature, the role of judgments in the process of producing reviews is often acknowledged.

The success story of the randomized trial and subsequently of meta-analysis and systematic reviews, in evaluating the effects of pharmaceutical drugs, has resulted in similar approaches in evaluating other medical interventions, like rehabilitation. However, there are other obstacles, such as difficulties in describing the intervention; e.g. a multidisciplinary intervention is sometimes used as a means of encouraging employees on sick-leave to return to work. However, a complex rehabilitation may be difficult to describe. A systematic review studying multidisciplinary intervention as a method to increase the return to work in persons with low back pain just gives the reader a vague idea about the intervention per se, while other aspects, like randomization, are more thoroughly described (12). Return to work may also depend on contextual factors such as insurance systems and laws regulating employment and responsibilities for employers. A review of interventions for increasing return to work in persons with musculoskeletal disorders found that the effectiveness of the intervention depended on “health care settings, the disciplines involved, the role of compensation systems, and the selection of patients” (13). Thus, the findings from a randomized trial may be hard to generalize if the intervention strongly depends on contextual factors.

Performing randomized trials on humans is sometimes impossible. Even non-randomized experiments on humans may well be infeasible; consider the acceptability of experiments set up to investigate likely causes of many diseases, abuse or criminal behavior. Many interventions, moreover, such as welfare benefits, compensation during unemployment and insurance amounts, cannot be randomized for ethical reasons. Where randomized trials or even non-randomized experiments are infeasible, one must rely on observational studies. Several guidelines for reporting and evaluating studies on humans have been published; a few are outlined in Table 1.

Production of systematic reviews

Numerous agencies and organizations have been established to produce systematic reviews for evidence-based decisions. The internationally best-known and most respected organization for systematic reviews is the Cochrane Collaboration. This is an international network with both a paid staff and volunteers (www.cochrane.org). It stresses that it is independent and democratically governed. The Swedish Council on Health Technology Assessment (SBU, www.sbu.se) is a governmental authority which has been producing systematic reviews in a since 1987. The Swedish National Board of Social Health and Welfare (www.socialstyrelsen.se) is a governmental agency that develops national clinical guidelines and states that it offers “support for those who make decisions concerning the allocation of resources within Health and Medical Care and Social Services”. The National Institute for Health and Care Excellence (NICE; www.nice.org.uk) is a British independent organization supported by the government that develops guidelines in health and social care. It stresses its independence even though it is sponsored by the government and receives its topic for a guideline from the Department of Health. The formalized methodology of systematic reviews has been adopted beyond health care, too. For instance, the Campbell Collaboration was founded as a sister organization of the Cochrane Collaboration (www.campbellcollaboration.org). It is a network of voluntary researchers who produce systematic reviews on social interventions within education, crime and justice, social welfare, and international development.

National agencies sometimes commission reviews from researchers. The topic is specified by the agency and the methodology decided by the researchers after discussions with the agency; e.g. the

Swedish Working Life Authority has commissioned reviews in different areas of interest from their perspectives. Reviews can also be initiated and produced by the researchers themselves. Organizations that produce reviews may have a staff that oversee the reviewing process and perform some parts of it, e.g. the bibliometrics search.

Quality

Systematic reviews are often contrasted to narrative reviews, i.e., reviews which are not conducted in a formalized manner. Systematic reviews tend to be evaluated in accordance to the degree to which their production follows a formalized schedule, whereas narrative reviews are often evaluated according to the seniority of the authors and the status of the journal in which they are published. While certain guidelines exist for evaluating the quality of systematic reviews, guidelines for assessing narrative reviews are scarce. The quality of a review depends on its purpose; if the purpose is to understand the mechanisms responsible for hypertension, a narrative review may be the first choice, while a systematic review conducted in accordance with the methodology of the Cochrane Collaboration might be the primary choice where the purpose is to compare the efficiency of hypertensive drugs.

Systematic and narrative reviews alike are, in most cases, reviewed before being published. The review of a systematic review is an essential part of its quality. Scientific journals have a peer-review system while organizations undertaking systematic reviews have their own reviewing systems. In science the possibility of criticizing and discussing findings and conclusion is essential, and systematic reviews that are published in English have an advantage. The reader of a systematic review has limited possibilities (in terms of time and knowledge) to make their own evaluation of the quality of a systematic review. Reviews published by Cochrane Collaboration and major journals are regarded as reliable due to their rigorous reviewing system.

Critique and bias

The formalized manner in which information is evaluated in the evidence movement has generated considerable critique. One objection concerns the priority given to randomized trials. Especially in areas where randomized trials are impossible due to ethical or other reasons, the focus on the randomized trial may devalue other kinds of knowledge. However objective the formalized methodology of systematic reviews may appear, it inevitably involves judgments, concerning, for instance, what studies to include in a review. A recent Swedish review on the study of causes for musculoskeletal disorders of the arm, shoulder and neck excluded cross-sectional studies from the evaluation by principle (14). This was criticized by the reviewers of the systematic review as there were few longitudinal studies and, for obvious reasons, no randomized trials (15). The scarcity of information that highly restrictive inclusion criteria produce may encourage the conclusion that some factors that have been regarded as harmful, e.g. repetitive bending in the wrist, should not be considered harmful. A decision to include additional studies in this review would likely have produced a different conclusion.

Evaluations based on inclusion criteria that are difficult to satisfy routinely conclude that the evidence available is insufficient. If factors that are potentially harmful are evaluated in such a manner, this puts the burden of proof on those who are exposed to the factor. The strategy of producing doubt and calling for more knowledge before preventive measures are taken is well known from the tobacco industry and the asbestos industry. Corporate interests were able to postpone action by persistently calling evidence indicating causation in doubt, while more evidence was slowly amassed (16). The same strategy has successfully been adopted by climate skeptics.

The best-known organizations conducting systematic reviews depend on the voluntary participation of scientists. However, evaluations are also produced by private organizations and authorities whose reviewing processes are less transparent. Scientific evidence is a prestigious word, and those who support systematic reviews may have a strong interest in a certain conclusion. A recent Swedish review of privatization of welfare production reached the conclusion that it could not be demonstrated that the private service providers were more efficient than public ones (17). Having been commissioned by an organization with many representatives from the private sector, the review caused an intensive debate. The director of the organization was forced to resign, and so did the head of research.

The formalized character of the procedure is no guarantee of reliability. So, for instance, it has been found that studies sponsored by pharmaceutical companies that have not yielded the desired results have more rarely been published than studies with positive findings. A meta-analysis will therefore offer more favorable results than it would if all studies had been included (18). Indeed, publication bias, a long-standing problem which has often been discussed, is due not only to corporate actors suppressing negative data; there is evidence that authors and journal editors, too, prefer publishing positive data (19-21). Furthermore, authors may just publish favorable findings and not all the findings. To decrease the risk for publication bias through selective reporting the authors should specify important characteristics of a trial at the start. In 2004 the International Committee of Medical Journals Editors stated that only studies that have registered the protocol of their trial before the start would be eligible for publication in their eleven journals (22). They argued that too few studies had been registered on a voluntary basis, and that it was unethical not to publish less favorable results, too. Recently, the editors of a number of leading physiotherapy journals, having found that the proportion of prospectively registered protocols of intervention studies of physical therapy remains small, have taken a similar decision (23). Several registers are available, where key information about the original protocol must be offered. WHO has a website listing registers that contain relevant contents (who.int/ictrp/network/primary/en/index.html).

It has repeatedly been demonstrated that reviews carried out by people with a possible conflict of interest tend to reach conclusions favoring the organizations with which they are associated. A review of 106 reviews of passive smoking and harmful effects on health found that of 39 reviews which concluded that passive smoking was not harmful, 29 had authors who were affiliated with the tobacco industry. A regression analysis found that the only predictor for concluding that passive smoking causes no harm was that the author was linked to the tobacco industry (odds ratio= 88.4 , 95% confidence interval 16.4-476.5) (24). Similarly, the affiliation of authors to pharmaceutical industries has been shown to bias the conclusion of reviews. Hence most journals and organizations require that authors and reviewers disclose any affiliation and conflict of interest (25).

Other types of formalization

A distinction is sometimes made between science proper and “science for policy” (see, for instance, (26)). While consensus formation in normal scientific practice tends to be a very slow process, potential risks and threats which are not well understood may need to be addressed by decision makers within a few years and in some cases even within months or weeks. The traditional mechanism used to collect the best available information under such circumstances is expert committees. National governments and international organizations alike establish advisory bodies whenever the need arises. For instance, the UK government set up a series of expert committees when BSE had been identified in British cattle in the late 1980s (27) and the World Health Organization appointed an emergency committee within days of discovering the H1N1 pandemic in 2009. It has been customary for expert panels to work in a highly informal manner, and in most cases the criteria by which their members are selected have not been made public. In an age where transparency and accountability is increasingly demanded from decision makers in almost any area, this makes the recommendations offered by expert committees vulnerable to criticism.

Formalization of processes and procedures is the key means used to strengthen the legitimacy of mechanisms by which scientific findings are summarized for policy purposes. Though evidence-based medicine originated in a professional rather than a policy context, the entire movement of evidence-based practice and policy may be taken as instantiating this feature of efforts to produce science for policy. Evidence-based methods are not widely adopted in the sciences, however, though meta-analysis is an established method in ecology (28) and many systematic reviews are now being conducted in conservation (29). There are other ways to formalize processes through which scientific evidence is assessed for policy purposes besides the methodology of meta-analysis and systematic reviews. The Intergovernmental Panel on Climate Change (IPCC) and the International Agency for Research on Cancer (IARC) represent other types of formalization.

The IPCC, which was founded in 1988, is widely regarded as the most reliable source of information about climate change and its effects on the environment and on society. Every five or six years, the IPCC publishes a three-volume report assessing and synthesizing the available evidence. In each assessment cycle, two or three hundred scientists are involved as Lead Authors in the teams writing the chapters of each of the three Working Groups, and many more as reviewers. In the first assessment cycle, the report of which was published in 1990, there was little by way of formal procedures, and each Working Group organized its work differently. From the early 1990s on, however, the procedures to be followed in the assessments have been set out in ever more detailed fashion in a series of guidance documents (30, 31). For all these efforts, circumstances casting serious doubt on the reliability of IPCC assessments were revealed in 2009 and 2010 (32). Claims about Himalayan glaciers deriving from the grey literature were reproduced uncritically, and email correspondence among top scientists involved in the IPCC, which was somehow stolen and leaked to the media, was taken to exhibit manipulative behavior.

The response of the IPCC was to commission the InterAcademy Council, an independent expert body, to conduct a review of its procedures. The InterAcademy Council (2010) offered numerous recommendations, all intended to strengthen existing rules of procedure. One recommendation was to increase the transparency of the selection of Lead Authors through a formal set of well-defined

criteria. Another recommendation concerned the management of the peer review of chapter drafts, Each chapter is subject to several rounds of review, first by scientific peers, then by government representatives, too, and the review of each chapter is overseen by at least two Review Editors. The InterAcademy Council recommended that the Review Editors take a more active role by summarizing the comments received and categorizing them under different headings, in order for the entire process to become more targeted and effective. The IPCC has endorsed these and other recommendations of the InterAcademy Council, and further changes have been made in the rules of procedure. Clearly, the trend towards increased formalization continues.

In 2000, another type of guidance for IPCC author teams was introduced, offering recommendations on the manner in which uncertainty be determined and described. A set of tools and procedures were proposed, allowing authors teams to define the nature of the data on which their estimates were based and to indicate their level of confidence in the evidence. IPCC authors are encouraged, that is, to be explicit about the extent to which their conclusions are based on well-established evidence and subjective judgments, respectively (33). While the tools and procedures recommended have been revised several times (Mastrandrea et al. (2010) is the current version), the general aim of this form of guidance remains that of promoting consistency and transparency. This type of formalization, too, was assessed by the InterAcademy Council. Endorsing the general aim of IPCC uncertainty guidance, the InterAcademy Council report suggested ways of improving the tools as well as of increasing compliance with the procedures.

The International Agency for Research on Cancer (IARC), which is affiliated with the World Health Organization, has developed another set of procedures for identifying, assessing and synthesizing evidence in order to determine whether a given agent may cause cancer or not. Unlike the procedures of EBM, those of IARC, as well as those of the IPCC, include all relevant information. In the case of IARC, this means knowledge about mechanisms, experiments on animals, cells, chemical experiment etc.

IARC provides critical reviews of the scientific literature in order to determine whether an agent, e.g. a chemical substance or mixture, is causing cancer in humans. The classification is mainly qualitative and states the evidence in five groups (carcinogenic for humans (1) , probably carcinogenic to humans (2a), possibly carcinogenic to humans (2b), not classifiable as to its carcinogenicity to humans (3), probably not carcinogenic to humans (4)) (34).

IARC was established in 1965 and already around 1970 it started with the evaluation of carcinogenicity. The first publication appeared in 1972 and today more than 100 monographs have been published. In the beginning carcinogenicity was evaluated by data from humans and animal experiments, but it now includes mechanistic and other relevant data, too.

Agents for reviews are selected where there is human exposure and a suspicion of carcinogenicity. IARC has a staff that works with the evaluations. They select the experts that perform the evaluation which is called the “working group”. Working group members are selected on the basis of knowledge and experience and absence of real or apparent conflicts of interests. In the selection process consideration is also given to “demographic diversity and balance of scientific findings and views” (34). Specialists with some conflict of interest can be invited, but are not allowed to participate in the evaluation. There might also be representatives from national agencies or

observers from stakeholders who follow the process, but they, too, are not allowed to influence the evaluation.

Working groups are divided into subcommittees for human, animal, mechanistic and exposure data, each developing a draft and summary. Each study has to be evaluated, and there are certain guidelines for the joint classifications in the subcommittees.

The total evaluation in the working group should end in a consensus among the members, but there may be a poll. Besides a classification into the five groups, the final evaluation includes a “rationale” which explains the classification. There are also guidelines for the overall evaluation, e.g. to be classified in group 1 there should, with some carefully specified exceptions, be sufficient human data of carcinogenicity. The evaluation is presented in a monograph where each study is described and there is a more extensive discussion about classifications. Human data are discussed, as well as results from experimental studies of animals and mechanistic data.

Unlike IPCC assessments, IARC evaluations do not include external reviews. Transparency of the selection of working members is poor, but there has been relatively little critique of IARC evaluation procedures. This may be because the judgments are presented and the criteria for the classifications are clear and possible to follow and criticize.

Systematic reviews in social work and working life research

The papers in this issue are based on contributions to a conference organized by the Swedish Research Council for Health, Work Life and Welfare (FORTE). The aim was to highlight strengths and weaknesses in the methods used for ensuring evidence-based practice, with a focus on systematic reviews in social work and working life research. In this special issue, advocates as well as critics of evidence-based methods are represented.

Interventions in social care are often difficult to randomize, and they are almost always impossible to blind. Interventions tend to be complex, hard to describe and measure, and may depend on personal characteristics of the practitioners involved. A major difficulty concerns contextual factors which are of importance for the performance and outcome of the intervention. E.g. a social intervention which has been demonstrated to be efficient in the US may not be efficient in a European country with a very different social context. External validity is an estimate of the chance of generalizing the findings of a study to other contexts, while internal validity concerns the quality of the study in the context where it was performed. Formalized evidence-based appraisal tools do not include procedures in which external validity is assessed; they focus exclusively on internal validity (35). Procedures allowing transparent evaluation of the external validity of findings would be a welcome development in evidence-based evaluation (36).

Working life research often focuses on prevention, e.g. reducing possible harmful physical factors such as vibration, heavy physical load, repetitive movements etc. Randomized trials are rarely relevant in such situations, and the formal evidence-based process de facto devaluates or disregards findings from mechanistic and observational studies. Hence many reviews falsely conclude that existing knowledge about the risk of the factor is limited or non-existent. From a preventive

standpoint the likely result is that no action is taken. If the reverse question had been asked, “Is it safe to be exposed to vibration, heavy physical load, repetitive movements?”, the conclusion would have been that there is insufficient knowledge to say that it is safe. Thus, the underlying question asked is crucial, and stakeholders may use the results from systematic reviews in quite different ways.

A project for compiling systematic reviews within occupational safety and health has recently started in PEROSH, a collaboration between national institutes of occupational safety and health in Europe (<http://www.perosh.eu/research-projects/perosh-projects/occupational-safety-and-health-evidence-clearinghouse/>). It brings together the results from systematic reviews including not only intervention studies but also reviews of etiology, prognosis, diagnosis and prevalence. The methodology includes predefined steps as e.g. defining a specific and clear question, identifying all relevant records in at least two electronic databases and assessing the study quality. Knowledge is used for a variety of purposes. A person with a disease that he believes is caused by occupational factors is likely to take a quite different view than that of an insurance company or an employer. A trial in court may compare the information available with the law. Scientists with different backgrounds or representing competing research schools habitually interpret the same data differently, yet scientific debates in which opposing views are represented are necessary for the progress of knowledge.

In a famous speech in the 1960s, Bradford Hill criticized the formal way of using statistical testing for conclusions about causality (37). His points have sometimes been used in a more formalized way to prove causality, but the underlying idea in his speech was that formalized procedures have to be discussed and must not be used without careful reflection. He criticized the use of statistical significance as a major criterion for establishing causality. Clearly, the formalized way of making a systematic review has its strengths and limitations. We hope that this special issue will stimulate further discussion. The real long-term risk of the evidence movement, in our view, is that its methods will be separated from research practice and regarded, instead, as administrative procedures which produce evidence that is not open for discussion.

Table 1. Some guidelines for reporting studies on humans.

Topic	Guideline	Reference
Randomized controlled trials of pharmaceutical drugs	CONSORT	(38)
Randomized controlled trials of non-pharmaceutical treatments	CONSORT	(39)
Systematic reviews and meta-analysis on intervention trials	PRISMA	(40)
Systematic reviews and meta-analysis on intervention trials	GRADE	(40)
Systematic reviews and meta-analysis on observational studies	MOOSE	(42)
Non-randomized evaluations of behavioral and public health Interventions	TREND	(43)
Qualitative studies	COREQ	(44)

References

1. Guyatt G. Evidence-Based Medicine. *Ann Intern Med.* 1991;114(Suppl 2):A16.
2. Guyatt G. Evidence-Based Medicine – A new approach to teaching the practice of medicine. *Jama* 1992;268(17):2420-5.
3. Rigotti NA, Clair C, Munafo MR, Stead LF. Interventions for smoking cessation in hospitalised patients. *Cochrane Database of Systematic Reviews.* 2012(5).
4. Mills EJ, Wu P, Lockhart I, Thorlund K, Puhan M, Ebbert JO. Comparisons of high-dose and combination nicotine replacement therapy, varenicline, and bupropion for smoking cessation: A systematic review and multiple treatment meta-analysis. *Ann Med.* 2012;44(6):588-97.
5. Eysenck HJ. Were we really wrong? *Am J Epidemiol.* 1991;133(5):429-33; discussion 34-6.
6. Greenland S. Re: "Those who were wrong". *Am J Epidemiol.* 1990;132(3):585-6.
7. Greenland S. Science versus advocacy: the challenge of Dr. Feinstein. *Epidemiology.* 1991;2(1):64-72.
8. Stolley PD. When genius errs –Fisher,R.A: and the lung-cancer controversy. WHEN GENIUS ERRS - FISHER,R.A. AND THE LUNG-CANCER CONTROVERSY. *Am J Epidemiol.* 1991;133(5):416-25.
9. Schünemann HJ, Oxman AD, Fretheim A. Improving the use of research evidence in guideline development: 6. Determining which outcomes are important. *Health Res Policy Syst.* 2006;4:18.
10. Ray WA, Stein CM, Daugherty JR, Hall K, Arbogast PG, Griffin MR. COX-2 selective non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease. *Lancet.* 2002;360(9339):1071-3.
11. Egger M, Smith GD, Altman DG. Systematic reviews in healthcare : meta analysis in context. 2nd ed. ed. London: BMJ; 2000.
12. Norlund A, Ropponen A, Alexanderson K. Multidisciplinary interventions: review of studies of return to work after rehabilitation for low back pain. *J Rehabil Med.* 2009;41(3):115-21.
13. Staal JB, Rainville J, Fritz J, van Mechelen W, Pransky G. Physical exercise interventions to improve disability and return to work in low back pain: current insights and opportunities for improvement. *J Occup Rehabil.* 2005;15(4):491-505.
14. Edling C, SBU. Arbetets betydelse för uppkomst av besvär och sjukdomar : nacken och övre rörelseapparaten : en systematisk litteraturöversikt. Stockholm: Statens beredning för medicinsk utvärdering (SBU); 2012.
15. Järvholm B, Andersen JH, Veiersted B. [On diseases in the neck and upper musculoskeletal system: Too low scientific quality in the SBU report]. *Lakartidningen.* 2012;109(25):1235.
16. Oreskes N, Conway EM. Merchants of doubt : how a handful of scientists obscured the truth on issues from tobacco smoke to global warming. 1st U.S. ed. New York: Bloomsbury Press; 2010.
17. Hartman L (ed.). Konkurensens konsekvenser : vad händer med svensk välfärd? 2. uppl. Stockholm: SNS förlag; 2011.
18. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One.* 2008;3(8):e3081.

19. Begg C, Berlin J. Publication bias and dissemination of research. *J Natl Cancer Inst.* 1989;81:107-15.
20. Easterbrook P, Berlin J, Gopalan R, Matthews D. Publication bias in clinical research. *Lancet.* 1991;337(8746):867-72.
21. Dickersin K, MIN Y, Warren K, Mosteller F. Publication bias – The problem that won't go away. Doing More Good Than Harm: the Evaluation of Health Care Interventions. *Ann NY Acad Sciences.* 1993;703:135-48.
22. De Angelis C, Drazen JM, Frizelle FA, Haug C, Hoey J, Horton R, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *N Engl J Med.* 2004;351(12):1250-1.
23. Costa LO, Lin CW, Grossi DB, Mancini MC, Swisher AK, Cook C, et al. Clinical trial registration in physical therapy journals: Recommendations from the International Society of Physiotherapy Journal Editors. *Phys Ther.* 2013;93(1):6-10.
24. Barnes DE, Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. *Jama.* 1998;279(19):1566-70.
25. Lundh A, Sismondo S, Lexchin J, Busuioac OA, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev.* 2012;12:MR000033.
26. Funtowicz SO, Ravetz JR. Uncertainty and quality in science for policy. Dordrecht ; Boston: Kluwer Academic Publishers; 1990.
27. Beck M, Asenova D, Dickson G. Public administration, science, and risk assessment: A case study of the UK bovine spongiform encephalopathy crisis. *Public Administration Review.* 2005;65(4):396-408.
28. Gurevitch J, Hedges L. Meta-Analysis: Combining the Results of Independent Experiments. In: Scheiner J, Gurevitch J, editors. Design and Analysis of Ecological Experiments. 2nd ed. Oxford: Oxford University Press; 2001.
29. Pullin A, Knight T. Doing more good than harm - Building an evidence-base for conservation and environmental management. *Biological Conservation.* 2009;142(5):931-4.
30. Agrawala S. Structural and process history of the intergovernmental panel on climate change. *Climatic Change.* 1998;39(4):621-42.
31. Miller C. The Design and Management of International Scientific Assessments: Lessons from the Climate Regime. In: Farrell I, Jäger J, editors. Assessments of Regional and Global Environmental Risks: Designing Processes for the Effective Use of Science in Decisionmaking. Washington DC: Resources for the Future; 2006. p. 187-205.
32. Beck S. Between Tribalism and Trust: The IPCC Under the "Public Microscope". *Nature and Culture.* 2012;7(2):151-73.
33. Moss R, Schneider S. Uncertainties in the IPCC TAR: Recommendations to Lead Authors for More Consistent Assessment and Reporting. In: Pachauri R, Taniguchi T, Tanaka K, editors. Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC. Geneva: World Meteorological Organization; 2000. p. 33-51.

34. IARC. Chemical agents and related occupations. IARC Monogr Eval Carcinog Risks Hum. 2012;100(Pt F):9-562.
35. Cartwright N, Hardie J. Evidence-based policy : a practical guide to doing it better. Oxford ; New York: Oxford University Press; 2012.
36. Lavis JN, Oxman AD, Souza M, Lewin S, Gruen RL, Fretheim A. SUPPORT Tools for evidence-informed health Policymaking (STP) 9: Assessing the applicability of the findings of a systematic review. Health ResPolicy Sys 2009; 7(Suppl 1) :S9.
37. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295-300.
38. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726-32.
39. Boutron I, Moher D, Altman DG, Schulz KF, Ravaud P, Group C. Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Ann Intern Med*. 2008;148(4):295-309.
40. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
41. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490.
42. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283(15):2008-12.
43. Des Jarlais DC, Lyles C, Crepaz N, Group T. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health*. 2004;94(3):361-6.
44. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19(6):349-57.