Scand J Work Environ Health 1995;21(5):321-324

<https://doi.org/10.5271/sjweh.45>

Issue date: Oct 1995

Evaluation of research -- a difficult but necessary task

by [Rantanen J](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/8571087

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Evaluation of research – a difficult but necessary task

The key principle of research is to study and describe nature, the environment, events, and phenomena in a way which is as accurate and reliable as possible and thus to help us understand ourselves and our physical and social environment better. This principle implicitly calls for the evaluation of research programs, their results, research institutions, and research groups or individual researchers in all disciplines. Evaluation is as important for the development of research in occupational health as in any other sector of research.

The scientific community started systematic research evaluations with the help of peer reviews already 300 years ago, and such evaluations focused originally on the scientific technical quality and productivity of individual scientists and scientific results by using what is called “internal criteria” of the scientific community as a reference. The internal control of the scientific community has thereafter remained the most central evaluation system. Several trends within the scientific community currently stress the growing importance of the evaluative approach. First, the increase in the numbers of researchers and growing competition for both positions and grants make comparative evaluations of individual researchers and research groups and programs more actual than before. To a less degree, detected violations in good standards of conduct in research have also stimulated interest to carry out very specific evaluations within the scientific community with the purpose of defending trust in the scientific community.

On the other hand, the high number of published research reports makes it necessary for any reader to evaluate what is worth reading and what can be put aside without spending time nonproductively for qualitatively or substantially less relevant information. This seems to be a common practice since recent studies show that, depending on the discipline, 37–98% of the articles published in scientific journals were not cited at all in a four-year follow-up period. Traditionally the scientific community has mainly focused on the evaluation of the scientific quality of research results and less on other aspects of evaluation. More comprehensive evaluations are, however, increasingly being initiated also by the scientific community.

In the past few years evaluation has also been called for by publicly controlled authorities and private financiers, and this demand is growing steadily because of the increasing interest of decision makers and tax payers to control the use of financial allocations, particularly in times of constricting economies. This new interest has clearly been demonstrated for example by recent legislation (the GPRA Act, 1993) that requires public institutions in the United States, including those operating in research, to plan and document their performance in terms of quantity, quality, and economic efficiency. Governments of the Nordic countries have decided that virtually all publicly funded research institutions should be periodically subjected to evaluation carried out by external (preferably international) evaluators.

Systematic evaluation theories were first developed in industry, where early activities started by expanding the traditional quality control approach to a multistep process evaluation. The principles of process evaluation were gradually transferred to the service sector, including health systems. The systematic evaluations of health programs were initiated in the late 1960s and early 1970s. Health program evaluation is now taken as a constitutional part of health policy in many countries, and numerous evaluation practices and schemes have been developed. A logical consequence of such development has been that health research also be considered in evaluations. About 40 years ago societal interest in big scientific projects evolved, and systematic research evaluations were started by

combining societal "external criteria" with "internal scientific criteria" in evaluations. The "Hindsight" project carried out in the United States was a classical example of such an evaluation. Since then evaluation has been expanded to cover almost all of the larger research programs enjoying public financial support.

The external pressures for evaluation are mainly caused by the interest of financiers to ensure the justification of their financial investments and, in certain cases, by society's interest to ensure the reliability of research carried out for common safety (eg, occupational health) or used as a basis for making critical societal decisions.

Much development in methods for research evaluation is still needed, and several unanswered questions still prevail. We have a universal problem in finding adequate measures and indicators for productivity, scientific-technical quality, and effective research. Neutral universal standards and references for quality and performance are difficult to find, and thus evaluations often remain mere cross-sectional descriptive analyses of the research program or institution concerned.

Only a few evaluation programs have focused on the third important aspect of evaluation — the learning effect. In optimal cases the well-made evaluation exercise helps us learn from both failures and successes, and it helps us determine, for example, how the most productive research innovations are generated. Such an evaluation is very seldom in the interest of financiers or research administrators, but the scientific community itself may benefit more from it than from "actuarial" evaluations.

Bibliometric methods, such as science citation indices and impact factors, have been increasingly utilized for quantitative evaluation, but several weaknesses have also been identified in their use. The most common criticism tells us that the bibliometric methods are more a measure of visibility than of the real scientific value of a research report. On the other hand, self-citing, which has been often mentioned as a source of bias in bibliometric methods, does not seem to affect the results of analyses substantially.

Most evaluation exercises focus on documented performance in the past, while the evaluation should primarily answer questions dealing with foreseeable performance in the future. Such an approach is particularly important when decisions on long-term funding or institutional developments are being made. The assumption of the predictive power of performance demonstrated in the past is, by experience, deemed correct, but such prediction may fail if, for example, the conditions of operation or the problem area is substantially changed.

Very few objective measures are available for assessing the innovativeness of research proposals or research results. Similarly the evaluation methods for measuring the ultimate impact of research on practice are less developed. The lack of relevant standards and references to which the actual outputs of research can be compared is one of the obstacles of evaluation, particularly in the discipline of occupational health, for which several factors other than research data are needed to determine the impact on practice.

In most cases the program or research unit to be evaluated is used as its own control, and this procedure, at best, gives information on trends but not on the actual "goodness" of the activity.

Who should evaluate is a question which is often asked in connection with practical evaluation situations. The critical issue is who is competent enough to have sufficiently deep insight into the scientific substance but is also located far enough from the object of the evaluation to remain neutral. In practice the authorities and financiers who require the evaluations often do not even know enough about the field in question to be able to appoint persons who could evaluate highly specific programs, and thus the researchers or institutions themselves are asked to recommend evaluators. This situation implies a risk of bias in selection. There are, however, some precautions to counteract such selection biases. First, a group of evaluators is likely to be more objective than a one-man evaluation committee, and, second, the more prominent and well-known the recruited scientists are, the more likely it is that they will not

risk their reputation by being anything but as objective as possible. The third factor which is important for ensuring neutrality and independence for the evaluation is, as in several other aspects of research, the widest possible publicity.

The role of self-evaluation as an element of the programming and planning of research activities and as a means of assessing the achievement of objectives has been widely discussed in the past 10 years. Self-evaluation should be encouraged more for each operator and institution involved with research because it may also be the most important learning process. Self-evaluation input and output, if documented properly, may also facilitate external evaluations by collecting the necessary data. Standardized techniques, criteria, and references for self-evaluation need to be developed. Self-evaluation can, however, never substitute for an external evaluation.

Client's and customer's opinions of research results are crucial for any evaluation, and they should, at least in the case of occupational health research, always be considered. Again, to ensure objectivity, wide consultations should be carried out with clients by using standardized questionnaires to acquire information which can be systematically analyzed later. In occupational health research not only occupational health practitioners, employers and workers' representatives should be interviewed, but also the neighboring research fields and academia, including trainers and educators. In interpreting the results of such customer surveys, again a critical approach is needed because several types of competitive, beneficiary, or controversial relations can appear, and they may be effectively hidden in general statements on the value of research or its poor or good applicability in practice. Thus critical judgment of such client opinions by evaluators is also necessary.

If carried out properly, relevant criteria are used and neutral objective judgments are exercised, the evaluation of a research program or institution may be of utmost value for the further development of relevant and high-quality research products. On the other hand, poor, superficial evaluations, no matter what the outcome, may have a hazardous impact on the morale of researchers and on the real development of institutions and programs, as well as on the credibility of evaluation results in general.

Evaluation also comprises several other ethical issues. For example, how much subjective judgment is acceptable on the part of the evaluators in drawing conclusions on the basis of presented data? In addition, the different capacity of research groups or individual scientists in presenting and marketing their programs and results may affect the outcome of the evaluation. All who have ever been active in research are well aware that scientific competence and marketing skills are not always parallel, although both are important. The balance should be found by evaluators as they weigh the substantive content of research and the presentation skills of researchers.

Several other ethical aspects are connected with the evaluation activities. In addition to the critical issue of the neutrality and objectivity of evaluators the consequences of evaluation results contain an ethical element. The output of evaluation may be used for making decisions on extensive practical or policy actions. Continuation or discontinuation of a research program or even a research institution may be decided on the basis of evaluation results. The responsibility of evaluators in such a situation is great, and, whatever is presented, it should be based on deep understanding of the object of evaluation and on the potential consequences of the evaluative statements. Therefore, it is reasonable to expect that the data on which the evaluative conclusions and recommendations are based be freely available for all in the spirit of transparency — particularly for those who are being evaluated. The objects of evaluation should also be given an opportunity to comment on both the draft conclusions and, if requested, the final recommendations in order to contribute their views to the discussion on actions needed as a result of the evaluation. This is a reasonable demand in current participatory societies.

To facilitate the development of competent, objective, and development-oriented evaluations, the theory of evaluation should be developed further; appropriate paradigm, criteria, and indicators for the performance and quality of research should be produced; and the need for appropriate references and

standards should be met. International harmonization of evaluation practices and the exchange of evaluation data would probably improve the quality of evaluations.

New information technologies may also change the possibilities and needs of evaluation. A research result can already be published through a computerized network in a few seconds worldwide, and the opinions and judgments from numerous experts can be collected throughout the world through the Internet even today. New needs for evaluation practices may appear, and new standards may also be needed for reviewing the results before publication. On the other hand, the pace and productivity of research innovations may be enhanced to levels unimaginable 10 years ago, and publication delays may be gradually minimized. Such developments will also affect evaluation strategies. It is no doubt that both internal and external evaluations will continue to be an essential part of research processes even in future information societies. The best way to cope with the growing demands of evaluation is to plan research activities by keeping the forthcoming evaluation in mind and then implementing the plan as competently and effectively as possible — which in fact is the actual objective of the evaluation approach.

Literature

1. Irvine J, Martin BR. Foresight in science: picking the winners. London, Dover: Frances Pinter, 1984:166 p.
2. Irvine J. Evaluating applied research. London, Dover: Frances Pinter, 1989:103 p.
3. Mervis J. Asking science to measure up. *Science* 1995;267:20—2.
4. Alberts B, Shine K. Scientists and the integrity of research. *Science* 1994;266:1660—1.
5. Nichols RW. Gearing up for innovation. *The Sciences* 1995;January/February:4.
6. Matthews R. Storming the barricades. *New Scientist* 17 1995;June:38—40.
7. Davidson K. Liberté, égalité, interneté. *New Scientist* 27 1995;May:38—42.
8. Donabedian A. The quality of medical care. *Science* 1978;200:856—64.
9. Bilich F. Science and technology planning and policy. Amsterdam: Elsevier Science Publishers, 1989:303 p.

Jorma Rantanen, Director General
Finnish Institute of Occupational Health
Topeliuksenkatu 41 a A
00250 Helsinki
Finland