Evidence for success in health promotion: suggestions for improvement

G. Macdonald, C. Veen1 and K. Tones2

Abstract

This paper argues that health promotion needs to develop an approach to evaluation and effectiveness that values qualitative methodologies. It posits the idea that qualitative research could learn from the experience of quantitative researchers and promote more useful ways of measuring effectiveness by the use of intermediate and indirect indicators. It refers to a European-wide project designed to gather information on the effectiveness of health promotion interventions. This project discovered that there was a need for an instrument that allowed qualitative intervention methodologies to be assessed in the same way as quantitative methods.

Introduction

Health Promotion is, by definition, concerned with improving the health status of populations. Health promotion research is primarily concerned with the assessment of both the results of interventions (that is the outcomes) and the relative effectiveness of the means used to achieve these results. If it does not attempt to do both it may be laying itself open to scientific criticism and political scepticism.

However, it is a mistake to limit effectiveness studies to interventions solely concerned with changes in population health status. As many have argued (Kickbusch, 1986; Whitehead, 1991; Macdonald and Bunton, 1992; Tones and Tilford, 1994), health promotion must be equally concerned with issues to do with equity, healthy public policy, community involvement, health services access and social well-being. Evidence that demonstrates positive change in these areas should also contribute to health promotion effectiveness studies.

In this article, we argue that although health promotion has a history of research and evaluation principally concerned with effectiveness, it has been poor in reviewing its approach to evaluation. This is now of critical importance since practitioners, policy makers and fundholders increasingly want to concentrate resources on programme areas and approaches shown to be effective. Health promotion and its precursor health education have a respectable history of evaluation research but, at the present time, it finds itself part of a general movement of 'evidence-based' health care. A European-wide project, described here, is an attempt to keep health promotion in the forefront of evaluative, evidence-based health care interventions. After a general discussion and reminder of some key methodological issues—including a debate about the most appropriate choice of indicators for health promotion research—a description and critique of the project is offered.

The focus on health outcomes

In recent years there has been a growing and discernible interest in the area of health outcomes following an intervention. This interest has not been confined to clinicians who, understandably,
may be concerned with quality and outcome of treatment but has also been extended to planners and policy makers within and outside the NHS in the UK and indeed elsewhere (Pollitt and Harrison, 1992). There has been a number of reasons for this. Firstly, it is apparent that many clinical interventions, along with medical and other health care practice, are not based on demonstrable need or on evidence (Cochrane, 1972; Eddy and Billings, 1988). Clearly there is a pressing need to introduce rigorous appraisal systems in a world in which ever more sophisticated and expensive medical techniques compete for finite resources. Such systems would allow policy makers the facility of allocating resources not only on the basis of effectiveness but also, in a much broader sense, allow for expenditure to be more closely related to need and based on evidence. On the other hand, the current popularity of appraisal systems may well conceal a cost-containment imperative (Fraber and Costain, 1992). Indeed, in the USA, the UK, the Netherlands, Denmark and elsewhere it might be argued that the popularity of the concept of health gain and the attention paid to health promotion is due in no small part to the need to contain spiralling health costs rather than to a desire to achieve the more altruistic goal of improving people’s health! (Department of Health and Human Services, 1980; Welsh Health Planning Forum, 1989; Department of Health, 1991; Ministry of Welfare, Health and Cultural Affairs, 1992; Ministry of Health, 1992).

In the UK the current pre-occupation with outcome-driven approaches to health care has given rise to strategies that emphasize the primacy of prevention, health promotion and health gain through target setting. Although such an approach may have its critics (Macdonald, 1993) this concern with outcomes and the setting of health status targets does provide a clearer and sharper direction for planners and practitioners alike. It also parallels a growing movement within medical care which seeks to make progress in measuring and using outcomes (Long, 1994).

Apart from cost-containment, two other factors have given impetus to the move towards outcome and effectiveness measures within the UK. Firstly, the focus on quality assurance within health services through multi-professional clinical audit and the introduction of performance indicators gave rise to the need to have appropriate and rigorous outcome measures. Secondly, the purchaser–provider split and the introduction of an internal market within the UK NHS, following the White Paper Working for Patients, made explicit the issue of effectiveness; the need for purchasers to base commissioning of service provision on community needs assessment, the need to build quality assurance into contracts, and the need to develop monitoring mechanisms and targets or health care outcomes—all became imperative. Purchasers need this information in order to make reasoned decisions on where to place contracts to produce the most effective health care interventions. In short, it was argued, purchasers need evidence in order to make reasoned decisions about where to place contracts in order to produce the most effective health care interventions.

Evidence-based health care

For the reasons outlined above, the demand for information on effectiveness and efficiency spurred the establishment of a number of centres and projects designed to measure effectiveness and/or to determine the extent to which existing initiatives are based on evidence. These centres tend to adopt a hierarchical approach and they tend to operate on an essentially logical–positivist paradigm and base their assessments on what has traditionally been regarded as ‘hard’ ‘scientific’ evidence. Such a philosophy has a number of limitations. For instance, financial costs may not receive due attention. Clearly, resources can only be efficiently allocated if, through some form of economic appraisal, competing interventions are costed in addition to the relative magnitude of the morbidity, mortality and geography of disease. Health economics has, therefore, a major contribution to make to effectiveness studies (Drummond and Maynard, 1993). More important, perhaps, are
Evidence for success in health promotion

1. Intervention studies
   1a. Randomized controlled trial (RCT)
   1b. Controlled trial (non-randomized)
   1c. Quasi-experimental design (QED)

2. Observational studies
   2a. Cohort (prospective) study
   2b. Case-control (resective) study
   2c. 'Before' and 'after' studies (no controls)
   2d. Descriptive studies: reports on clinical or practitioner experience—surveys

Fig. 1. A research evidence hierarchy. Adopted from Sheldon et al. (1993).

those limitations which actually derive from the pursuit of 'hard evidence'—as we will note later. For now it seems reasonable to assert that current systematic reviews of published literature and meta analyses of the data are based on a hierarchy of evidence (Sheldon et al., 1993)—with or without the cost-effectiveness element. This hierarchical approach is shown in Figure 1. As Figure 1 demonstrates, the randomized controlled trial (RCT)—or its equivalent in educational and behavioural sciences, the true experimental design—is considered the best or 'gold standard' for quantitative evaluative research for well established reasons—more particularly, because, with a large enough sample population it can eliminate bias and confounding variables. However, this is only the case if the RCT is well designed and carried out. A well conducted quasi-experimental design (i.e. a true experimental design without the benefit of randomization to experimental and control groups) or even a sound observational study may provide better 'evidence' of effect than a poor RCT. Any review of published research studies should, therefore, include an assessment of the quality of the design and not just an analysis of the results (Fowkes and Fulton, 1991).

Health promotion research design

Health promotion, like other aspects of health care delivery, should also be subject to evaluation. Indeed some argue that as a relatively new discipline, health promotion has had to prove itself from its earliest beginning—and has suffered as a result since it has often been obliged to use unrealistic criteria to evaluate outcome (Naidoo and Wills, 1994).

However, evaluation research in health promotion is not only concerned with outcomes, which may arguably be the case with other forms of evidence-based health care (as discussed above), but must also seek to gain insight into the processes involved in programme implementation and the social and environmental context in which they take place (Tones and Tilford, 1994). Qualitative research methodology has, therefore, much to offer in 'illuminating' both outcome and process. For instance, only illuminative research can provide detailed insights into why a given programme has or has not achieved its objectives. It can provide answers into how outcomes were achieved rather than whether they were achieved. Indeed, it could legitimately be argued that the next step for qualitative research is to develop its own hierarchy for quality of method and produce its own 'gold standard' (Macdonald, 1996).

The observations made above do, of course, reflect the truism that different stakeholders demand different things from the research process. The academic researcher whose discipline is based on the logical-positivist tradition will have different needs from the planner or the practitioner. Nutbeam et al. (1990) have illustrated these competing requirements in the development of a form of bipolar sliding scale (Figure 2).

It would normally be accepted that the purpose of the RCT/Experimental Design is to establish a casual relationship between interventions and outcome by comparing two or more groups or cases. In health promotion, it is rarely possible or desirable to use such a design. Apart from the need for illumination, which was mentioned above, there are practical problems. For instance, Lipsey et al. (1985) identified the following.

1. There are numerous practical difficulties inherent in matching a good experimental design to practical programme circumstances. Because of the naturalistic and complex setting...
of health promotion and health education and
the inclusion of multifaceted programmes,
measurement problems arise (i.e. within
living and dynamic communities), which
include considerable opportunity for variation
or influences on the actual implementation.
(2) Conventional evaluators may not be very well
trained to do methodologically exacting
research under field conditions.

Health promotion researchers have, in fact, tended
to rely on the ‘weaker’ quasi-experimental design
and, as a consequence, may get the worst of both
worlds. On the one hand, their results will always
be subjected to challenge by those wedded to the
RCT gold standard; on the other hand, their
research results may contain insufficient informa-
tion to be of value to practitioners. It is therefore
often suggested that experimental designs and,
*a fortiori*, quasi-experimental designs be restricted
to circumstances in which information on cause–
effect relationships are of major concern and,
in the case of quasi-experimental studies, where
appropriate techniques are employed to minimize
the effects of lack of randomization. Furthermore,
since even the simplest health promotion interven-
tion is more complex than the administration of a
pill or even a surgical procedure, some degree of
illumination is usually needed. Accordingly, we
need approaches which study programme develop-
ment and process; we need qualitative research,
formative evaluation, naturalistic observation; we
need research into service delivery and surveys of
need; we need a variety of alternative methods for
answering the broad range of important questions
generated by health promotion and health education
programmes—and we need triangulation and a
good theory base to make sense of the results
created by pluralistic methodology!

We cannot, of course, ignore the political dimen-
sion of the research process and policy makers
figure prominently among research stakeholders—
as we noted from Figure 2. However, the different
needs of policy makers illustrate nicely the often
conflicting requirements for ‘evidence’ of success
with the need for research to provide sufficient
illumination for practitioners to repeat successful
interventions and modify or reject those which have
not been effective. Traditional research designs
provide hard evidence of success or failure but
due to the stringency of their requirements, they
tend to offer few examples of *unequivocally*
effective programmes. Moreover, where effective
programmes are identified ‘external validity’ is often
lacking because of the dearth of illuminative detail.

Part of the problem has to do with a lack of
process evaluation; part of the problem centres
on an unsophisticated and often inappropriate use
of indicators of success.

The importance of intermediate and
indirect indicators

We argued above that health promotion often has
different goals from those exposed by traditional
preventive medical approaches and these typically
include such measures as equity, the attainment of
‘healthy public policy’ and social well-being. We
would further assert that a major principle underlin-
ing the evaluation of health promotion initiatives
is that traditional epidemiological indicators may

![Fig. 2. Health promotion evidence: different needs.](image-url)
Evidence for success in health promotion

be inappropriate and even behavioural outcomes may not be the most useful way of appraising effectiveness and efficiency.

The above assertion might seem a little surprising. After all, if health promotion has a useful contribution to make to achieving preventative outcomes, is it not reasonable to assess success in terms of reducing mortality and morbidity? There are two main reasons for not doing so. First, there is often some degree of uncertainty associated with the links between particular behaviours and specific disease outcomes—at any rate on a population basis. It would, therefore, be unreasonable to claim that health promotion has been ineffective even if it has not lead to a reduction in disease provided that it has influenced lifestyle. For instance, given the uncertainty associated with postulated links between saturated fat consumption and breast cancer, it would be unwise to measure the success of dietary interventions in terms of decline in breast cancer mortality. Of course, the reasons for running a lifestyle change programme which did not in fact have a preventative outcome might legitimately be questioned—but, clearly, the kind of dietary intervention mentioned above might be expected to have a pay-off in respect of other cancers, disorders of the digestive system and cardiovascular diseases. Even then, given the fact that cardinal risk factors for most diseases may fail to explain a substantial amount of the variance in mortality and morbidity, a one-to-one relationship between behaviour change and 'medical' outcome would not be expected.

We could also go further and argue that even lifestyle change or other behavioural measures may not be appropriate indicators of success of health promotion. While in some instances it would be reasonable to expect to see a change in behaviour after a given educational intervention (e.g. a school-based smoking programme might, if properly constructed, delay the onset of smoking in school pupils), in other cases, behavioural outcomes would not be feasible. One such instance is provided by programmes where a particular education input could not be expected to influence a given outcome for many years. A successful cancer education programme which effectively taught young children about the nature of cancers in such a way as to contribute to a reduction in cancerophobia could not be expected to have an impact on, say, a woman's utilization of mammography screening until the girls in question had reached middle age!

An even more important reason for using intermediate indicators derives from the fact that individuals' decisions about health or illness are multiply determined. As we noted above, a health education programme is not at all like the administration of a drug. Even a relatively straightforward face-to-face consultation between nurse and patient is much more complex and sophisticated. Health decisions are not the result of a simple linear progression from provision of knowledge through attitude change to the adoption of healthy practices. Accordingly one single input—simple or complex—is unlikely to achieve a desired output. Consider the case of one of the more simple educational interventions—a short mass media 'commercial' seeking to influence dietary practices. At the very least, the following chain of events would have to take place. The target audience would first need to correctly interpret it and understand it; they would need to believe it and assess its personal relevance; they would typically consider the possible costs and benefits involved in undertaking the changes in diet recommended by the programme, and then assess the likelihood of their having the expertise and stamina needed to make those changes. If these several steps have been successfully negotiated and if the audience has not been antagonized or made anxious by the message, a positive attitude to adopting a healthy diet might result. Of course, before that attitude could be translated into practice, the client group would doubtless need a variety of skills, additional knowledge and, of course, they must be able to find the health products and be able to afford them. Let us simplify the above process into the following sequence: awareness—knowledge—beliefs—attitudes—support—action. If we now assign hypothetical probabilities of each stage being satisfactorily negotiated, the likelihood of success in
terms of behavioural change might be depicted as follows:

<table>
<thead>
<tr>
<th>Awareness</th>
<th>Knowledge</th>
<th>Beliefs/Attitudes</th>
<th>Support</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>30%</td>
<td>85%</td>
<td>31%</td>
<td>40%</td>
<td>3%</td>
</tr>
</tbody>
</table>

This formula is based on 30% of the audience having become aware of the message, 85% of that particular group understanding it, 31% of the 85% believing it and being committed to the recommended action but only 40% having an appropriately supportive environmental and social circumstances. The result (quite creditable if employing the criteria of commercial advertising)—a 3% behaviour change. We should, perhaps, note that the figures presented above are not completely hypothetical. The percentages for awareness, knowledge and beliefs/attitudes are based on the average figures expected for well pre-tested cancer education materials (Romano, 1984). The assumption that 40% of those exposed to the media programme will receive the necessary social, environmental and educational support they need is hypothetical—and probably somewhat optimistic!

In other words, in order to improve effectiveness, the mass media input would need to be complemented by a variety of different educational inter-
Evidence for success in health promotion

ventions. These would be designed to achieve different learning outcomes and be delivered over a period of time in a variety of contexts and settings. Ideally, the educational programme would be supplemented by appropriate healthy public policy. Accordingly, it makes sense to assess each of these inputs separately. For instance, understanding of the nutritional information on which dietary decisions are based might best be provided in schools—and assessed there. Empowerment-related beliefs and attitudes could also be supplied in the personal and social education programme within the school and/or in community development work. Supportive policy might be created by providing healthy options in the works canteen and food labelling in the supermarket. Evaluation would, therefore, centre on the development and implementation of these policies. Taking into account our knowledge of the potential of mass media to influence behaviour, evaluation of this aspect of the total programme might best be limited to concept testing, pre-testing and the measurement of audience penetration, and the intelligibility and acceptability of the message to clients.

On some occasions, even intermediate indicators may not be necessary for valid appraisal of effectiveness of inputs. For instance, we may need to assess the efficiency of a communication skills course provided for practice nurses. Assuming that the course is sound, the trainees will become more skilled in educating their clients. Their new skill will, in turn, ultimately be reflected in more effective work with patients. In the long term they and the course they have completed may contribute to clients’ behaviour change—and may even have some small impact on the prevalence of disease!

Clearly, important decisions have to be made about which particular intermediate and indirect indicators to select. These decisions can only be based on good quality theories—derived for research. Figure 3 seeks to provide a schematic representation which exemplifies the complex inter-relationships of factor existing in the ‘black box’ which frequently conceals what happens between ‘input’ and ‘output’.

These complicated and multifaceted aspects of process evaluation will need to be built into any attempt to develop a suitable effectiveness measurement tool.

The need for Illumination

We noted above the complex interior of the ‘black box’. If we are to gain insight into the reasons why given initiatives succeed or fail, we must penetrate the mysteries of the ‘black box’. In other words, we need ‘illumination’—description in great detail of what is happening in programmes so that we can make reasoned judgements about which particular features have been effective. For this reason, we need qualitative techniques providing ‘thick, rich description’. This does not, of course, mean abandoning the quest for validity—in fact qualitative work can provide very good checks on the validity of conclusions drawn about effectiveness by using ‘triangulation’. In short, qualitative methods can provide excellent ‘judicial’ evidence on which to make decisions by combining and comparing data from various sources. This kind of evidence often has more practical significance than statistically unimpeachable data based on rigorous true experimental designs.

A European review of studies of health promotion effectiveness

Whilst health promotion researchers and practitioners have been concerned with designing appropriate evaluation and effectiveness studies, very little effort has been spent on either assessing the quality of the research design or assessing effectiveness through literature reviews. Much published material on effectiveness (Nutbeam, 1993) has focused on outcome rather than research design. Even, where reviews have combined assessment of methodology with effectiveness, the implications for health promotion interventions and their impact on health status is unclear (Nuffield Institute for Health, 1993) or limited (Garcia et al., 1994).

Recently a major innovative project was designed to develop and implement an instrument
for assessing or analysing effectiveness studies in health promotion and health education (Veen et al., 1994). The project, funded by the European Union and coordinated by the International Union for Health Promotion and Health Education (IUHPE) had two principal aims. The first was to gather information on the effectiveness of health promotion and health education interventions, the quality of the evaluation research and the outcomes; the second was to improve the accessibility of information for researchers and practitioners of health promotion and health education effectiveness.

Sixteen researchers across Europe were asked to review published health promotion and health education interventions and outcomes in 16 topic areas. A review instrument, in the form of a detailed questionnaire, was developed, piloted, revised and distributed to the researchers for completion. The questionnaire included questions on the intervention itself, the evaluation element, summative evaluation outcomes, formative evaluation outcomes, recommendations (from the paper), and the reviewers own assessment of the quality of the paper and intervention using a prescribed checklist.

The reviewer was asked to concentrate on a minimum of 10 and a maximum of 15 effectiveness studies, using a 13-point hierarchical criterion schedule supplied by IUHPE. These 13 criteria included interventions which attempted to change knowledge, attitudes and behaviour. They made use of case-control groups with pre- and post-testing measurements; triangulation of evaluation methods; formative and summative evaluation; and they looked particularly for interventions which were European.

In tandem with development of the review questionnaire a computerised database (EffectBase) was designed and used to store details of the effectiveness studies systematically.

Seventeen publications were produced as a result of the exercise including an evaluation of the project following a meeting in March 1995 (Veen and Macdonald, 1995)—and these are appended to this article.

### Problems

The project was a first major attempt to systematically review health promotion and health education intervention effectiveness, at the international level, although national reviews are now taking place in England. The project did generate a number of issues which are worth highlighting.

1. The review instrument (the questionnaire) was not appropriate for certain types of evaluation research, like multicomponent studies. It included questions on the resources and evaluation as well and was therefore multi-levelled.
2. It was useful for single intervention studies and publications, but of limited value for systematic reviews and meta analysis.
3. The questionnaire concentrated on health education and less on broad health promotion interventions concerned with policy and environmental change.
4. The reviewers found it hard to find up to 15 European papers in some topic areas or ones that completed with the selection criteria, such as quasi-experimental or case-control studies.
5. Most effectiveness studies were primarily concerned with knowledge and attitude change in small groups (often less than 30 participants) and less with behaviour and broader social change issues.
6. The reviewers made a plea for more qualitative, process-orientated evaluation that would complement and add to our understanding of health promotion programme planning, and intervention impact, rather than solely rely on outcome measures.

The project piloted an instrument that could be used to gather information on effectiveness studies in health promotion. The fact that it encountered difficulties in collecting quality data in both outcome orientated evaluation and more so with process evaluation illustrates the problem researchers and programme planners have in health
promotion. The IUHPE is currently addressing part of this issue through the development of minimum standards for quality assurance for health promotion interventions.

A challenge facing academics, researchers and practitioners in health promotion is how to differentiate between good and poor qualitative research. Is it possible to develop a kind of hierarchy in qualitative research design in the same way as the evidence based health care researchers have done? Is it desirable and appropriate? What is more certain is that we need a measurement tool for reviewers of research that goes beyond the IUHPE questionnaire, and provides useful information on qualitative research, process and outcome in the same was as the measurement tools for quantitative evaluative research. Only by doing this can we contribute to effectiveness studies.

Clearly, we are not arguing for a rejection of quantitative methods nor do we deny the importance of RCT’s. We do, however, need to remedy the existing imbalance between quantitative and qualitative. As Partlett and Hamilton noted in an influential book entitled Beyond the Numbers Game, someone who relies on ‘outcome only’ evaluation is:

...rather like a critic who reviews a production on the basis of script and applause meter readings, having missed the performance.

In short, we need to evaluate our approach to evaluation: we need more useful evidence of success.

References


Evidence for success in health promotion
Appendix: list of publications available as result of the project 'Improvement of the effectiveness of health education and health promotion'


Un Outil d'Analyse des Études d'Efficacité de la Promotion de la Santé et de l'Éducation pour la Santé: Mise en point, Utilisation et Recommandations. Order no. INT 7094.


Hosman, C. M. H. and Veitman, N. E. Prevention in Mental Health: A Review of the Effectiveness of Health Education and Health Promotion. Order no. INT 8194.


Siâma, K. Tobacco Control: A Review of the Effectiveness of Health Education and Health Promotion. Order no. INT 7794.


Vuori, I. M. Exercise Promotion: A Review of the Effectiveness of Health Education and Health Promotion. Order no. INT 8694.

Weston, R. Cancer Prevention: A Review of the Effectiveness of Health Education and Health Promotion. Order no. INT 8394.

For further details see the News section of this issue.