A methodological inquiry into the evaluation of smoking cessation programmes

Masahiko Kaneko

Abstract

This paper examines a methodological controversy and aims to show the advantages of introducing an alternative methodological approach, i.e. the ‘scientific realist approach’, into evaluation studies on health education programmes for smoking cessation. The methodological difficulties of the existing standard evaluation model, i.e. the quasi-experimental approach, are examined. This model fails to investigate how the programme setting influences outcomes (context problem) and draws attention away from understanding why programmes work because it adopts a ‘successionist’ logic (causation problem). An alternative methodology, the scientific realist approach, is proposed in order to cope with such problems. This approach adopts the ‘generative’ logic which looks at the matter of causation internally. This logic posits that the working of underlying mechanisms within a more basic level of reality causes relationships between visible events. According to the scientific realist approach, the actual outcomes of smoking cessation programmes follow from the workings of potential mechanisms whose functioning is afforded by contexts conducive to their operation.

Introduction

The evaluation of social programmes has one crucial purpose—‘to measure the effects of a programme against the goals it set out to accomplish as a means of contributing to subsequent decision-making about the program and improving future programming’ [(Weiss, 1972) p. 4].

However, the actual results of evaluation studies have often been less satisfactory, making subsequent decision making somewhat less than rational. This is indeed the case with evaluation studies on health education programmes. Many well-designed, large-scale community-based trials for health promotion have been conducted in recent years. However, ‘although a few had a degree of success, several have ended in disappointment. Generally, size of effects has been meager in relation to the effort expended’ [(Susser, 1995) p. 156]. Health education programmes for smoking cessation are no exception to this trend. The results might lead sceptics to the conclusion that community programmes do not work. What causes such disappointing results?

The initial aim of this paper is to point out that the methodology of existing evaluation studies on smoking cessation holds much of the blame for the failure to address the important practical tasks to be pursued in getting people to stop smoking. Second, this paper aims to introduce an alternative methodology, the ‘scientific realist approach’, which contains the promise that evaluation studies might provide more useful suggestions for improving such programmes.

COMMIT: an example of a quasi-experimental model

Evaluation research has seen itself as the standard bearer for the application of the ‘scientific method’ for measuring the implementation and outcomes of...
policies and programmes. The dominant paradigm of evaluation research in the public health field is the quasi-experimental model which has been adapted from the experimental model in medicine to fit the particular needs of this field. Its basic procedure is known to all and can be summarized as follows.

Take two more or less matched groups. Statistical testing usually requires that samples should be selected through random allocation. If the groups are truly matched through random selection, this can be called ‘real experimentation’. However, the practical and political problems of using the experimental method lead us to use various quasi-experimental methods. One of these methods consists of using a non-equivalent control group in which individuals or groups with similar characteristics are assigned to comparison groups instead of assigning group members through random selection (Cook and Campbell, 1979).

The experimental group receives the programme while the control group does not. The situation in the two groups after programme implementation is compared with the situation before the programme was started. If there is any difference in outcomes between the two groups, it is inferred to be the result of the programme since the experimental and control groups are identical to begin with, and the only difference between them is application of the programme.

Let us now turn to a review of a quasi-experimental evaluation of a particular smoking cessation intervention trial in order to better understand how this experimental-versus-control group comparison model works. The Community Intervention Trial for Smoking Cessation (COMMIT) is a useful case study for our purposes, because this community-based trial contains a technically excellent and statistically sophisticated quasi-experimental evaluation (COMMIT Research Group, 1991). This project, which leading North American researchers in the public health field took part in, was a large-scale programme costing tens of millions of dollars and was 5 years in the making.

In COMMIT, 11 matched pairs of communities were selected with the aim of mirroring the diversity of North American community life. These matched pairs were situated in the western, southwestern, midwestern and eastern US, and southeastern Canada. To investigate the intervention effects in diverse settings, they also varied in terms of population size, race/ethnicity, male/female ratio, educational level, income, etc. Within each pair, the two communities were chosen so as to be geographically separate from each other in order to maintain the independence between them in relation to intervention activities. They were also chosen so as to be matched for socio-demographic factors such as population size, percentage of white people, proportion of females, age distribution, extent of urbanization, estimated smoking prevalence rates and access to community intervention channels (e.g. health care services). One community was randomly assigned to the intervention group and the other served as a comparison.

The intervention was directed at the community as a whole. The community was used as the unit of analysis. The goal of COMMIT was to create a social climate that did not support tobacco use. This project assumed that the best way to change behavior would be to intervene through social structures within a community and that interventions delivered through community channels could reach larger proportions of the smoking population than individual-oriented interventions.

The research group set up four kinds of intervention channels: (1) public education through the media and community-wide events, (2) health care providers, (3) work-sites and other organizations, and (4) cessation resources. Within these channels, 58 specified mandated activities were carried out largely by community volunteers or local staff. The primary target was heavy smokers.

The research group analyzed the results of this trial as follows (COMMIT Research Group, 1995a). The group divided the subjects into a heavy smoker group and a light-to-moderate smoker group, and compared the average quit rate across all 11 intervention and control communities. As shown in the Table I, the aggregate evidence showed that there was only a minor difference between the rate of quitting in the experimental communities (0.180) and the controls (0.187) among the heavy smokers. Among the light-to-moderate smokers, the average
Table I. Numbers (n) of individuals in the cohortsa and fraction (f) of those who met the definition of quitting, with imputation for those unknown in 1993 (MAR analysis) b

<table>
<thead>
<tr>
<th>Pair</th>
<th>Heavy smoker cohort (n = 10 019)</th>
<th>Light-to-moderate smoker cohort (n = 10 328)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention (n = 4976)</td>
<td>Comparison (n = 5043)</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>f</td>
</tr>
<tr>
<td>1</td>
<td>442</td>
<td>0.139</td>
</tr>
<tr>
<td>2</td>
<td>531</td>
<td>0.163</td>
</tr>
<tr>
<td>3</td>
<td>475</td>
<td>0.164</td>
</tr>
<tr>
<td>4</td>
<td>428</td>
<td>0.204</td>
</tr>
<tr>
<td>5</td>
<td>440</td>
<td>0.183</td>
</tr>
<tr>
<td>6</td>
<td>450</td>
<td>0.164</td>
</tr>
<tr>
<td>7</td>
<td>432</td>
<td>0.262</td>
</tr>
<tr>
<td>8</td>
<td>455</td>
<td>0.193</td>
</tr>
<tr>
<td>9</td>
<td>455</td>
<td>0.215</td>
</tr>
<tr>
<td>10</td>
<td>426</td>
<td>0.136</td>
</tr>
<tr>
<td>11</td>
<td>442</td>
<td>0.155</td>
</tr>
<tr>
<td>Community means</td>
<td>0.180</td>
<td>0.187</td>
</tr>
</tbody>
</table>

*a In this table, n represents the sum of those with observed outcome and those with imputed outcome; estimates of quit rates based on such n have greater uncertainty than would occur with the same n if all had been observed.

bMAR = missing at random

*P (one-sided) = 0.068; 90% confidence interval = –0.031, 0.019.

**P (one-sided) = 0.004; 90% confidence interval = 0.014, 0.047.

Source: The COMMIT Research Group [(COMMIT Research Group, 1995a) p. 186].

rate of quitting in the intervention communities was higher than in the controls. Thus, the research group concluded that while COMMIT had a statistically significant, though minor, intervention effect upon the light-to-moderate smokers, it ‘did not significantly affect the primary outcome measure—quit rates among heavy smokers’ [(COMMIT Research Group, 1995a) p. 189]. As Susser notes rather drily, ‘The results are certainly below the expectations of the sponsors’ [(Susser, 1995) p. 157].

**Difficulties of the quasi-experimental model**

The COMMIT design, i.e. the quasi-experimental model (simple experimental-versus-control group comparison), has two difficulties.

**Context problem**

Experiments try to deal with contextual variation by matching and thereby attempting to rule out the impact of social context rather than investigating its actual influence on outcomes. The influences of social context tend to be overlooked because controlling for extraneous influences is seen as a fundamental part of the quasi-experimental model in order to guarantee that it is only the programme that has an effect. The experimental design thus tends to flatten out contextual variation as its very objective. This is well illustrated in the case of COMMIT.

A detailed examination of the data in Table I reveals enormous differences among the 11 matched pairs of communities. For example, the quit rate of heavy smokers in the intervention community in pair 9 was much higher than in the controls, which suggests an effectively working programme. By contrast, pair 3 demonstrates a null result and the data for pair 1 shows that the control group considerably outperformed the experimental group. Similar variations are also found in the light-to-moderate...
smoker cohort. Thus, whilst the net or aggregate change in the total population was close to zero, this masks the obvious marked success of the programme in some communities and its abject failure in others.

This would compel most social scientists to ask, ‘What are the differences in the characteristics of the 11 matched pairs of communities?’ As has been shown, they varied ‘in terms of [population] size, ethnic diversity, degree of urbanization, and the access of smokers to community intervention channels’ [(COMMIT Research Group, 1991) p. 1621]. Numerous studies have revealed that there is a considerable difference in smoking behavior across different demographic groups (such as ethnicity, educational level, occupational status and marital status groups). Access to and interest in community intervention channels probably also vary across these groups, and thus would obviously influence the results of intervention.

In spite of such crucial differences among matched pairs of communities, the research group paid no attention to them. In order to investigate the net intervention effect on changes in smoking behavior, the group sought to ‘separate the effects of the intervention activities from the inherent differences between communities’ (COMMIT Research Group, 1991). In other words, the researchers regarded the differences as deviations from a norm and simply took an average. The group made the mistake of attempting to grasp the population uniformly and as a consequence overlooked the profound differences within it. The obvious question is ‘why?’ Our answer can only be the methodological one, i.e. that experimentation remains method driven, and that the need to exert control and exclude every rival causal influence has led followers of this school instinctively to attempt to remove what should be central to a real explanation from the picture.

It has been demonstrated that smoking prevalences among different social groups have always varied widely. There are, for instance, considerable differences in educational level, which was one of the socio-demographic factors whose extent varied among the 11 pairs of communities. In the US, which was basically the COMMIT domain, educational differences in smoking prevalence have increased. While the smoking rates among college graduates decreased remarkably, from 33.7% in 1966 to 18.4% in 1985, the prevalence among those with less than a high school diploma decreased slightly, from 36.5 to 35.7%, during the same period (USDHHS, 1989). The key hallmark of the recent smoking prevalence has been its ‘segmentation’. Smoking has become a problem of social sub-groups rather than of the whole population. It is more fruitful to consider ‘that what appears to be even distribution across a community is an artifact of averaging and that, e.g., a community’s smoking is more accurately modeled as occurring in pockets of high prevalence’ [(Fisher, 1995) p. 160]. One concludes that it is indispensable for researchers to address such pockets and attempt to find what method is effective for which social sub-group, rather than relying on one almighty, all-purpose method.

Causation problem

The quasi-experimental model is likely to adopt ‘successionist’ logic. This is one of the models that Harré (Harré, 1985) distinguished as the two ‘great metaphysical theories of causation’. Successionist logic looks at the matter of causation externally. It describes constant conjunctions between separate events much like the action of billiard balls. The relation between cause and effect can be observed as one billiard ball collides with another and forces it to move.

Thus, the quasi-experimental model focuses on whether a programme works, thereby drawing attention away from understanding why the programme works. In other words, this model only focuses on the input (intervention) and the output (outcomes). How the two events in the model are related is hidden within a ‘black box’. It relates intervention to outcomes without any necessary obligation to interpolate an explanation. One only says that the programme works if the outcome of the experimental group is better than that of the control one and that the programme does not work if the opposite is the case. As such, the end product is no better than a description.
Evaluation of smoking cessation programmes

In COMMIT, we clearly should be asking, why did the intervention work within some matched pairs? As surveys reveal (Ben-Shlomo et al., 1991; Japanese Ministry of Health and Welfare, 1993), people stop smoking for a host of different reasons. Thus, smokers within the intervention communities in COMMIT may have stopped because they learned more about the adverse effects of smoking through programme materials or because they were persuaded convincingly by their friends, family and colleagues to do so. The act of quitting smoking is quintessentially conscious decision making. Health education programmes for smoking cessation offer people reasons and resources to quit. Such initiatives cannot work until people choose to accept some of those reasons and resources offered. It is not that programmes work but that people choose to make them work.

This notion of ‘programme process’ is treated in the quasi-experimental model as yet another threat to internal validity. Thus, the traditional demands of experimentation influence programme practitioners to make great efforts to formalize and equalize their provision. In COMMIT, 58 mandated activities were specified in the standardization of the intervention. These activities were checked by a ‘process evaluation’ of the trial in the intervention communities that included the detailed monitoring of compliance with the protocol and the dissemination of information found useful in some sites to all communities (Corbett et al., 1991). There was, in short, an attempt to make the intervention activities across 11 intervention communities as similar as possible. However, this is not to say that different people experienced these activities in the same way. The point is not whether key components of a programme are all present and well defined, but what components are considered key by each person. This is not something that can be held constant by process evaluation. What evaluation research should address is what features of smoking cessation programmes actually are persuasive in changing the habits of smokers. The quasi-experimental model with the ‘successionist logic’ is inappropriate as a means of addressing this task. Evaluation research requires an alternative methodology to replace the quasi-experimental orthodoxy.

Scientific realist approach

In these days of the ‘paradigm wars’, many alternative strategies are available to the evaluator. One highly promising alternative goes by the name of ‘scientific realism’, which has come into prominence over the past three decades across different disciplines (Harré, 1985; Pawson, 1989; Sayer, 1992; Bhaskar, 1997). Scientific realism adopts the ‘generative logic’ of causation. Generative logic is the alternative model of causation distinguished by Harré (Harré, 1985). This logic posits that there is a more basic level of reality, i.e. that comprising ‘underlying mechanisms’, rather than just surface (or observable) events. It is the working of these underlying mechanisms that causes the relations between surface events. To borrow an example from Pawson and Tilley (Pawson and Tilley, 1997), the relation between igniting gunpowder (one event) and the explosion (the other event) is caused by the chemical reaction of gunpowder (the underlying mechanism). When the relation between one event and another is explained generatively, there is an attempt to understand how the relationship itself comes about. In other words, generative explanation shows ‘how things work by going beneath their surface (observable) appearance by delving into their inner (hidden) workings’ [(Pawson and Tilley, 1997) p. 65].

Let us consider evaluation research on health education programmes for smoking cessation in terms of the scientific realist methodology. A social programme is essentially conducted to change the status quo. A smoking cessation programme is conducted in order to decrease present levels of smoking prevalence. Thus, the intended outcome of the programme is a change in the smoking ratio in the targeted population.

Mechanisms

A multiplicity of ideas is likely to be involved and plural mechanisms for change are likely to be triggered within any programme. As an example, con-
sider the situation in which a doctor in an antenatal clinic attempts to persuade pregnant smokers to stop smoking. This is one type of health provider channel in COMMIT. The actual intervention itself may consist of a visit to the clinic by the pregnant smoker with her partner, and the doctor may explain the harmful effects of mothers’ smoking on unborn children to both of them and offer some advice on how to stop. Even a simple social transaction like this may generate several mechanisms for change.

First, some smokers may quit by dint of knowing more about the harmful effects. This mechanism, i.e. the ‘medicalization mechanism’, is a standard method for preventing smoking in general. The whole furor about smoking stems from these harmful effects. It is thus probably true to say that all health education programmes for smoking cessation have included this mechanism to a greater or lesser extent. However, it has been demonstrated that there is little difference between smokers and non-smokers in their general knowledge about the dangers of smoking (Baric et al., 1976; Ben-Shlomo et al., 1991). This may mean that the ‘medicalization mechanism’ is not necessarily effective.

Second, some pregnant smokers may stop in response to the demands of their partner who has discovered more about its harmful effects (Mizutani et al., 1992). This mechanism can be referred to as the ‘primary group encouragement mechanism’. According to the findings of a series of studies on self-directed changes in smoking habits (Prochaska and DiClemente, 1983; Prochaska et al., 1991), this mechanism may have the greatest influence early in the period when people are making an active attempt to abstain.

Third, if the doctor also advises substitute methods of stress management or weight control, which are considered to be benefits of smoking for women (Graham, 1976; Klesges et al., 1989; Floyd et al., 1993), some smokers may stop the habit because of the particular pertinence of such factors to them. This mechanism can be called the ‘substitution mechanism’.

Fourth, people often shift their values and behaviors to those of a group that they are in contact with or admire. Thus, pregnant smokers may give up smoking because of the exemplary influence of the non-smokers with whom they mix. One way of implementing this mechanism is for health care providers to involve pregnant smokers with social groups that have a lower prevalence of smoking, i.e. high-educated or health-related occupation groups. This mechanism can be referred to as the ‘role model mechanism’.

The mechanisms for change in this list may well be elicited simultaneously within a simple smoking cessation intervention. The actions of these mechanisms are not mutually exclusive. All social programmes have such multiple aspects. No social initiative is solitary. Therefore, one crucial task of evaluation research is to understand what it is about a programme that makes it work. This means, in realist terminology, that evaluation research should explore underlying mechanisms that cause the relation between programmes and their outcomes.

Context

The above statement brings us to the next key point in the realist explanatory tool-kit. Whilst exploration of underlying mechanisms is paramount, it is also posited that the relationship between underlying mechanisms and their effects is not fixed, but contingent, as Sayer (Sayer, 1992) argues. For example, gunpowder cannot explode if it is damp, or if it is not adequately compacted, or if there is no oxygen present. The successful firing of the mechanism (chemical reaction) depends on the gunpowder being present under the right conditions.

This is also true of health education programmes for smoking cessation. The programmes are always introduced into pre-existing social contexts, i.e. communities with different characteristics of their own.

For instance, the amount of past experience in organizing community-wide event activities may have an influence as a social context (‘experience context’). Next, the existence of competing community forces may differ markedly between communities (‘competing force context’). The presence of a major tobacco company falls within this context. Furthermore, the level of environmental consciousness prompted by local pollution levels...
Evaluation of smoking cessation programmes

("environmental consciousness context") and multiple hardships in a community such as unemployment or poverty ("social pathology context") can also be considered as social contexts.

These contexts may well influence the results of programmes. Thus, in communities that habitually hold many successful community-wide events, this "experience context" may well facilitate mobilization of communities around smoking-cessation activities and affect trial outcome. For instance, the health providers may cooperate with smoking cessation programmes actively, so that the "medicalization mechanism" is effective. On the other hand, in communities where a factory of a major tobacco company is located, family members may very well work in the factory or the factory may act as a role model in these communities. In this case, the "competing force context" may impede the workings of the "primary group encouragement mechanism" or the "role model mechanism".

Thus, programmes sometimes succeed and sometimes fail. In the real social world, which is an open system, it is difficult to control these contexts in advance. Therefore, another important task of evaluation research is to examine the extent to which the social contexts enable or disable operation of the underlying mechanisms.

According to the scientific realist approach, the actual intervention outcomes, i.e., the changes in smoking prevalence, follow from the workings of a range of potential mechanisms whose functioning is made possible by contexts conducive to their operation; in short: Outcome = Mechanism + Context ([Pawson and Tilley, 1997] p. 55–82). The variations in the results of smoking cessation programmes are generated by the presence of various combinations of mechanisms and contexts. Evaluation research should make it the main task to examine the workings of mechanisms and contexts through the actual outcomes.

Research design

Based on the scientific realist approach, in the ideal design of empirical evaluation, before and after data would be collected to provide an overall picture of outcomes as in the quasi-experimental model. However, much more attention should be paid to obtaining data that reveal mechanism-context patterns. Beginning with the query from the scientific realist matrix, "what might work for whom in what circumstances", it is possible to anticipate and make comparisons of the variation in outcome patterns across groups. These groups are not defined by the simple experimental-versus-control-group comparison that the quasi-experimental model is likely to adopt, but are defined by the mechanism-context framework. For example, one might make a comparison in outcome patterns between a community that has the custom of holding many successful community-wide events and a community that does not, or among communities whose knowledge and experience about environmental matters are different. These and dozens of other such configurations will allow us to pinpoint the action of the programme.

It is possible to accumulate programme knowledge by reinterpreting findings within the quasi-experimental model from the standpoint of the scientific realist approach. The remarks and afterthoughts made to account for failure or unexpected results in experimental trials often suggest significant programme mechanisms and contexts. Another research group conducted smoking cessation programmes in different urban areas according to the experimental-versus-control-group comparison (Windsor et al., 1985, 1993; Gielen et al., 1997). In one trial, there was a large difference between the outcomes in the experimental group and the control group among African-American pregnant women in the area with "few social problems". However, in the area with "many social problems" (e.g., unemployment or poverty), there was little difference between them. Although these subjects were similar in respect of ethnicity, which is one of the classic individual attributes for smoking prevalence and quitting, they had different predictors of smoking cessation and reduction during pregnancy simply because they lived in different areas. As one of the reasons explaining the discrepant findings, the research group points out that "the salience and priority given to smoking cessation and support for quit-
ting may be lower among the Baltimore cohort [with social problems]’ [(Gielen et al., 1997) p. 253]. This shows the need to take account of social contexts. In scientific realist terminology, this means that a ‘social pathology context’ will always influence the workings of a ‘medicalization mechanism’ and a ‘primary group encouragement mechanism’.

Thus, a systematic range of comparisons across a series of studies and reinterpretations of existing findings will enable us to promote better understanding of the interplay of mechanisms, context and outcomes in health education.

Conclusion

‘Universalism is not universally applicable to the scientific endeavor’ [(Susser and Susser, 1996b) p. 675, italics are added].

At the conclusion of his review and evaluation of smoking cessation methods, Schwartz (Schwartz, 1987) suggests that there is no single best method, and that it is important to use multiple cessation methods which can deal with different types of people and different uses of cigarettes. There is no single intervention that can be recommended for universal adoption. In order to make a multiple method programme more effective, its components (particular cessation methods) have to focus on the specific needs of targeted smokers. Thus, we should address the question: What is it about smoking cessation programmes that works for whom in what circumstances?

The scientific realist approach seeks specification rather than universality (Pawson and Tilley, 1997). However, specification does not mean the fragmentation of our evaluation knowledge into a thousand-fold different individual reasons for giving up smoking. It means generalization within localized structures. We end up with knowledge of which type of programme works in which set of circumstances.

Acknowledgements

I am grateful to Ray Pawson and the anonymous referees for their valuable comments.

Notes

1. This problem of the ‘black box’ paradigm has recently been also criticized by leading epidemiologists (Pearce, 1996; Susser and Susser, 1996a,b). They argue that the black box paradigm relates exposures to diseases on the individual level without taking account of intervening factors or pathogenesis. These critics say that epidemiology should transform itself from a science that identifies risk factors for disease into one that analyses the systems which generate patterns of diseases in populations.

2. For a more extensive discussion on the scientific realist approach to evaluation research, see Pawson and Tilley (Pawson and Tilley, 1997).

References


Evaluation of smoking cessation programmes


US Department of Health and Human Services (1989) Reducing the Health Consequences of Smoking. Public Health Services, Center for Disease Control, Atlanta, GA.


Received on September 12, 1997; accepted on May 3, 1998