What is the "golden standard" for assessing population-based interventions?—problems of dilution bias

L Lindholm, M Rosén

Abstract

Objectives—To identify different types of dilution bias in population-based interventions and to suggest measures for handling these methodological problems.

Design—Literature review plus analysis of data from a population-based intervention against cardiovascular disease in a Swedish municipality.

Main results—The effects of an intervention on mortality and morbidity were much more diluted by non-intervening factors, dissemination to areas outside the intervention area, social diffusion, population mobility and time than by using intermediate outcome measures.

Conclusions—Theoretically, changes in scientifically well documented risk factors, for example, intermediate outcome measures, should be preferred to using morbidity or mortality as outcome measures.

(*J Epidemiol Community Health* 2000;**54**:617–622)

Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden
L Lindholm
M Rosén

Centre for Epidemiology, National Board of Health and Welfare, Stockholm, Sweden
M Rosén

Correspondence to:
Dr Rosén, Centre for Epidemiology, National Board of Health and Welfare, S-106 30 Stockholm, Sweden (mans.rosen@sos.se)

Accepted for publication 12 January 2000

Any decision to intervene must be based on evidence of the risks, costs and effects of an intervention. The final decision must, however, be based on a subjective judgement of whether the intervention is worthwhile. Ambiguity concerning risks, costs or effects, or methodological problems in interpreting the results, makes the decision even harder. A methodological problem such as the selection of end points in evaluations of community-based and population-based coronary heart disease interventions is one such problem that has been discussed in other contexts.[1–4] We define community-based interventions as "interventions in which the unit of allocation to receive a preventive regimen is an entire community".[5] Population-based intervention is where a general but defined population is the target group.

The methodological problems of assessing community-based interventions are multifold. One main problem is the impossibility of randomising people into intervention groups and control groups: we cannot force people to move from one area to another or the control group not to take action. Furthermore, one advantage of a community-based intervention is that it may also create changes in the surrounding environment. This positive effect, on the other hand, poses a dilemma for the evaluator.

The critics of community-based primary interventions in general and cardiovascular disease (CVD) or coronary heart disease (CHD) interventions in particular have focused mainly on the lack of statistically significant net effects on CVD or total mortality.[6–8] McCormick and Skrabanek illustrate this view "Overall, age-adjusted mortality must remain the final arbiter of benefit because it removes any biases from the ascription of the cause of death...".[7] However, even though focusing on the ultimate goal of the intervention is appealing, a more thorough analysis will show that the obvious is not so obvious after all.

In this article, we try to show some methodological limitations of an approach with morbidity and mortality as outcome measures. We then discuss suggestions that might improve evaluations in the future.

A model of dilution

There is a long chain from intervention measures to changes in risk factors to changes in morbidity and mortality. This whole process usually takes many years, and more factors, both risk and protective, are gradually introduced that will influence the changes in subsequent stages. For each new factor introduced and each period passed, there is a risk that the effect of the intervention will be diluted. The following example may illustrate the dilution process (fig 1).

In any trial a limited number of all available measures is selected. Many community-based interventions against CHD have used measures aiming at the three risk factors smoking, high blood pressure and increased serum cholesterol. At the start of these programmes and because of resource constraints, the project group has been obliged to choose among a limited number of measures against these three risk factors. Still, smoking habits may be affected by other societal changes in the community, for example, an increase or decrease in tobacco prices may have a larger impact than the intervention measures.

Not only is the incidence of CHD affected by these three well known risk factors: more than 200 other potential risk factors for CHD have been pointed out,[9] each of which may be influenced by different events in the community. The effects of an intervention will then be still more diluted. CHD mortality is a consequence not only of CHD morbidity; medical treatment may also change the outcome of the disease. In addition, the quality of care may differ over time or between intervention area and control area. Lastly, total mortality is affected by still many more social factors.

Interventions	Riskfactors	Morbidity	Mortality

```
A              Smoking         CHD           CHD         Total
Smoking

B              Cholesterol              Medical
Smoking                                  care

C              BP
Diet

D              ?                        Cancer
Diet

?              ?                        Accidents
.....
                                        ......
```
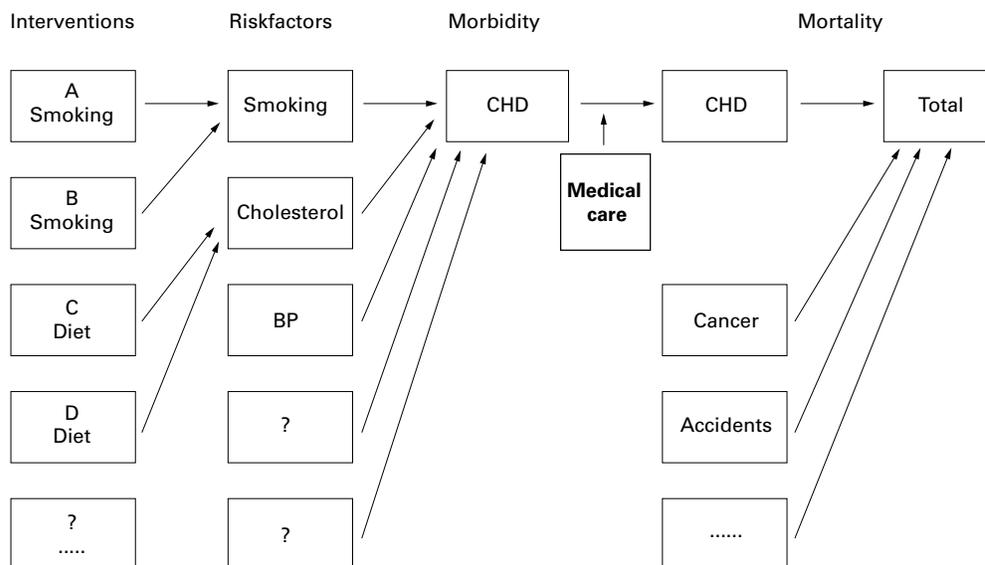
Figure 1   A model of exposure dilution biases—from intervention to mortality outcome.

Dilution bias

The trial design usually compares the intervention area with the reference area, assuming that the first population is exposed to the intervention and the second is not exposed. Yet the assumption that all the people in the intervention area are exposed and none in the reference area is, as pointed out above, unrealistic.

Misclassifications regarding exposure always cause an underestimation and a dilution in the association, and reduce the chance of showing significant differences between intervention area and reference area.

We define dilution biases as biases diluting the "real effect" of an intervention. The dilution model outlined above leads us to introduce six types of dilution biases that affect outcome measures (table 1).

The discussion will imply that it is much more difficult to find effects and associations between intervention measures and end results (changes in morbidity or mortality) than between intervention measures and changes in risk factors or other intermediate outcome measures. We offer examples supporting our reasoning based on theoretical and empirical cases of community-based interventions in general and on a Swedish community programme, the Norsjö project,[10–12] in particular.

Bias attributable to changes in non-intervening factors

The intervention measures chosen will hopefully reduce morbidity and mortality of the community more than would be the case without intervention. This does not necessarily mean that there will be a total reduction of morbidity or mortality. Other non-intervention factors may intervene in the opposite direction in the intervention area or forcefully reduce risks also in the control areas. This creates a potential bias in the evaluation process as some cases illustrate.

Though not a community-based study, the MRFIT (Multiple Risk Factor Intervention Trial) may serve as an example of this kind of bias. The number of CHD deaths in the MRFIT control group was much lower than expected because of the large reductions in smoking and other risk factors in that group, making it more difficult to find significant net effects after 7.5 years of follow up.[13] As all Western societies have had a dramatic decrease in CHD mortality, this has been a general problem for all interventions during the past few decades, for example, the North Karelia project, the Stanford Five-City Project and the Minnesota project.

Table 1   Six types of dilution bias in population-based interventions

Bias attributable to:	Definition
Exclusion of causes and consequences	
1 Changes in non-intervening factors	Non-intervening factors in the community influence morbidity and mortality and distort the effect of an intervention. This is because of the non-randomised nature of community interventions.
2 Single disease measurement	Most evaluations focus on a single disease measure although many behavioural lifestyle changes affect the risks of several diseases.
Exposure dilution	
3 Population mobility	The fact that people move from the intervention area to the control area and vice versa will create a dilution bias, causing the effects to be underestimated in the intervention area and overestimated in the control area.
4 Dissemination effects to other areas	Successful interventions are adopted by others, an effect omitted from the outcome analysis.
Mis-specification of follow up time	
5 Social diffusion to following generations	The adult population exposed to the intervention influences the lifestyles of following generations, an effect usually omitted from the outcome analysis.
6 Time lag	The effect of a risk factor reduction will have a lag time and be distributed during a long follow up time, which creates a dilution effect in the evaluation.

In Sweden, a population intervention in the municipality of Norsjö focused on dietary changes and succeeded between 1985 and 1990 in reducing the level of serum cholesterol in the population by nearly 20% more than in the reference areas.[10–12] The results on CHD morbidity and mortality may, however, have been eliminated by a negative social development in the municipality during the 1990s. The unemployment rate, for example, rose much more in Norsjö than in the surrounding areas in the early 1990s.

Single disease measurement bias
Most interventions focus on few outcome measures for a single disease, for example, CVD, cancer, diabetes or accidents. The fact is, however, that interventions focusing on behavioural lifestyle changes will most probably affect the risks of several diseases. Smoking is the ultimate example, affecting the risk of probably more than 20 diseases. Interventions aiming at changing dietary habits or physical exercise will not only affect CHD, they may also influence the risk of stroke, diabetes, certain cancer sites and osteoporosis. Results from the North Karelia project indicate that the intervention first has an effect on CHD, followed 15 years later by a reduction of lung cancer risks.[14]

Population mobility bias
Population mobility presents another methodological problem not usually considered in the evaluations. As some people move from one area to another, part of the population exposed to the intervention will not be included in the follow up on mortality. Also, some of the non-exposed will move to, and die in, the intervention area. This will create a dilution bias, causing the effects to be underestimated in the intervention area and overestimated in the control area (fig 2).

Of all those aged 25–64 years living in Norsjö in 1985, the starting point of the programme, 121 persons died between 1990 and 1995. Of those, 23% were living outside the municipality at the time of death and were

> KEY POINTS
> - Potential mortality effects of an intervention are diluted by the impacts of non-intervening factors, population mobility and time.
> - The effect of an intervention might have been underestimated by about 50% because of population mobility.
> - Changes in intervened risk factors may be more correct assessment measures than morbidity and mortality.

therefore not included in the analysis. In that case, nearly one quarter of those exposed to the intervention were not included in the outcome analysis. Another problem was that some people had moved into the municipality during a later phase of the intervention and consequently received weaker exposure. Still, they were completely included in the outcome analysis of mortality. In Norsjö, 8% of those who died during the period 1990–95 were not living in the municipality at the start of the project in 1985.

The problem of evaluating population mobility is not negligible. We assume the net risk reduction of the intervention to be 20% (a relative risk of 0.8 in the intervention and 1.0 in the control group) and that one quarter of the exposed have moved outside the intervention area and one quarter of the unexposed have moved into the area. With these assumptions, 75% of those exposed with a risk of 0.8 are living in the intervention area while 25% in the intervention area have a risk of 1.0. This gives an average measured risk reduction of 15% in the intervention area ($0.75 \times 0.8 + 0.25 \times 1.0 = 0.85$). The equivalent measured risk for the control area will be 0.95 ($0.75 \times 1.0 + 0.25 \times 0.8$). The measured net reduction will then be 10.5% (0.85/0.95) instead of an actual 20%.

Dissemination effects to other areas
Successful intervention measures create diffusion. The successful results from the five year follow up in North Karelia invited others to
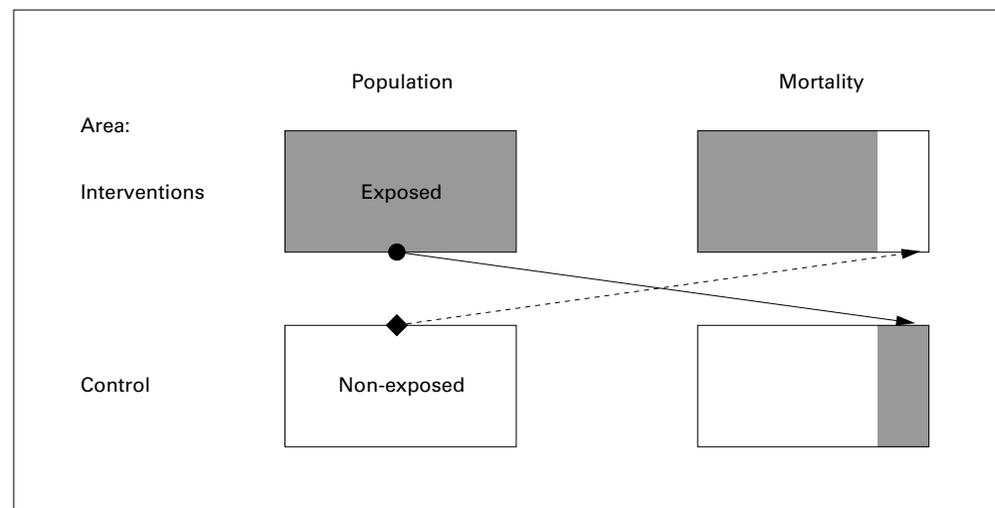

Figure 2   Schematic presentation of dilution effects attributable to population mobility.

imitate the project, also in the reference area.[15][16] The Heart Beat Wales initiative was also rapidly noted and implemented in the reference area.[17] The Stanford Five-City project certainly inspired other US projects ... and so on. Such dissemination effects are difficult to measure and therefore not included in the evaluations.

Returning to the Norsjö project, a labelling system on healthy foods with high fibre and low fat content was introduced. The labelling system, later called the "green keyhole" was then spread over the whole of Sweden with the authority of the National Food Administration.[18] In 1995, a telephone interview survey was conducted among a random sample of 500 Swedish citizens aged 20–64 years. Of 393 persons contacted, 282 responded (response rate 80% women and 61% men). Nearly 85% reported having seen the symbol in the shops where they bought their food. Twenty per cent knew the exact meaning of the message—that is, low fat and high fibre content. Fifteen per cent thought it was low fat or high fibre content while 40% could specify only "healthy food". Thus nearly 80% of all Swedes could use the symbol in the intended way, while about 20% did not get the real message.

Another question was how far the responders actually chose food labelled with the green keyhole. Thirty eight per cent said they always (6%) or often (32%) did so. While we do not know the impact of this measure on dietary changes or health, in this case even a very marginal effect would have a substantial effect on the evaluation of the Norsjö project. The diffusion effects on the whole Swedish population, 8.5 million inhabitants, should be related to the target population of just under 6000 Norsjö inhabitants.

Social diffusion bias to the next generation

A problem closely related to changes in non-intervening factors is that of social diffusion from the intervention. As people, our lifestyles are affected by that of our social environment, for example, children, spouse/cohabitant, friends and colleagues. Messages are also spread by mass media and other sources from one community to another. One study indicated that for 10 people who stop smoking, in the long run two more stop or do not start.[19] The risk of smoking is about twofold for schoolchildren where at least one parent is a smoker and ninefold if the parents permit the child to smoke.[20] The direct effects of social diffusion may not create any biases, but the long range effects, for example the

influence on coming generations, will be excluded from the outcome analysis.

Time lag dilution bias

As far as prevention is concerned, two cases must be distinguished. One is when a risk factor is eliminated (people quit smoking) as a result of preventive measures, and the other is when a risk factor is never developed (young people do not start smoking) as a consequence of an intervention. The time lag in the second case is very long (perhaps two to three decades). In the first case, the lag is much shorter (one to five years).

Using risk factor data from the Norsjö project,[10–12] a Framingham risk equation[21] and data on the time lag between a reduction in cholesterol and a following reduction in CHD mortality,[22] expected mortality for Norsjö during different time periods was estimated (table 2).

The odds ratio increased as the follow up time lengthened—that is, from 1.09 to 1.16 (table 2). The effects of a risk factor reduction will not have their full impact until some years after the intervention has started. This is supported by the fact that the odds ratio increased when the first years were omitted from the analysis, to 1.24. Accordingly, a "wrong" specification of the follow up time causes a serious dilution in the strength of the association, perhaps making it impossible to show a significant reduction in "real" mortality data.

Results from the MRFIT study and the North Karelia project further support the need to specify a long and appropriate time for follow up. The MRFIT study showed a non-significant reduction in CHD mortality of 7% and also a slight non-significant increase in total mortality after 7.5 years of follow up. After 10.5 years of follow up the net effects had increased, both on CHD mortality (−10.6%) and on total mortality (−7.7%).[23] In the North Karelia project, mortality outcomes were more marked in the second five years of the project than in the first five.[15][16] As already pointed out, it is noteworthy that the incidence of male lung cancer in North Karelia, which had been consistently higher than the reference area, had decreased 15 years after the project started below the incidence figures for the reference area, a significant 20% beneficial effect.[14]

Discussion

Randomised controlled studies with "hard" end points such as morbidity and mortality reduction are the ultimate evaluation methods and the "golden standard" for most assessments. For community-based primary interventions, however, this is not an appropriate option. We have argued above, theoretically and with empirical evidence, that the appropriate evaluation method in population interventions should be to measure effects closely related to the intervention, for example, risk factors that have earlier been shown to be causal. The restriction to causal risk factors is important. There must be well documented evidence and scientific consensus that there is a

Table 2 Variations in the odds ratios of fatal and non-fatal CHD events in Norsjö because of the length and location of the follow up period. Estimations based on risk factor reductions in the Norsjö project and the Framingham risk equations

Period of follow up	Prevented cases	Exposure	Diseased	Not diseased	Odds ratio
Years 1–10	16	Exposed	200	2400	
		Non-exposed	184	2416	1.09
Years 1–15	38	Exposed	300	2300	
		Non-exposed	262	2338	1.16
Years 6–15	36	Exposed	200	2400	
		Non-exposed	164	2436	1.24

causal link between the risk factor and the disease that is the subject of the intervention.

Our arguments for choosing intermediate outcome measures are primarily based on the fact that dilution bias creates great problems in an evaluation where morbidity and mortality are the ultimate outcome measures. Problems of dilution bias increase with time as factors other than the intervention measures are continuously introduced.

Another reason to choose intermediate outcome measures such as risk factor changes is that, given the larger number of events, it is easier to arrive at a reliable conclusion. If many people change their risk factor profiles, only some of them will avoid a non-fatal CHD and of those only a fraction will avoid a fatal CHD. Consequently, we are more likely to find significant net effects on risk factors than on morbidity and more likely to find significant effects on morbidity than mortality, a pattern also seen in all the community trials. Even though the results from many studies have been modest or shown minimal effects, they have been more positive when measuring risk factors than when measuring morbidity or mortality.[15] [16] [24–27]

The MRFIT results showed that the net effects on CHD morbidity were, as expected from out hypothesis, much more pronounced than the net effects on CHD mortality. For example, the incidences of angina pectoris, intermittent claudication, congestive heart failure and peripheral arterial disease sank by 21%, 12%, 88% and 16% respectively, while CHD mortality sank by a non-significant 7%.[28]

The foregoing example supports the theory of dilution bias. If mortality is the outcome measure, our estimations of the effects of the Norsjö project, as well as the empirical evidence presented from the North Karelia and MRFIT projects, all show the need for a very long follow up with a subsequent increase in exposure dilution bias. The fact that large intervention trials have shown limited effects on mortality indicates that for smaller trials, with larger random fluctuations, there is no other option than measuring intermediate outcome measures.

This article has focused on different kinds of dilution bias in population-based interventions. There is no simple way to handle these methodological problems and it is merely a question of limiting the magnitude of the problems. The single disease measurement bias can, of course, be handled by choosing several diseases in the outcome analysis. Population mobility can also be handled by thorough follow up of all persons in the population who have been exposed and non-exposed. This is not an easy task in many countries, but it is an option in Sweden and some other countries with a nearly complete follow up of population mobility. For most of the six biases presented, however, we see no other alternative than to measure intermediate outcome variables.

This review shows that the end results of community-based primary interventions are diluted over time and by other social factors. For several reasons, evaluations of these population interventions are trickier than evaluations of clinical procedures where randomised controlled trials are possible. On the other hand, the potential health impact of a population strategy is much greater than for most other available interventions among patients or high risk groups.[29] The danger is that even beneficial action will not be taken because of uncertainty about the evaluation. The expected utility of a planned intervention could still be very high for community-based interventions and thereby support a decision to intervene.

Expected utility is a function of the probability that an intervention has the anticipated effect and the potential effect. Even where the scientific evidence for the effect of intervention A is weaker than for another measure B, the expected utility of A could still be greater. This holds for events where the weight of the potential effects outweighs the probability of that event. Postulate that the evidence for a treatment intervention having an effect is very strong (a subjective probability of 0.95, say) and that the effect would yield a benefit of 200 years of life saved (YLS). The expected utility is then 190 (200 × 0.95). A population-based intervention cannot be based on the same clear evidence, however, and it might be more appropriate to put a lower subjective probability on the intervention having an effect, for example, p = 0.65. The potential effect is estimated to 600 YLS, where the expected utility will be 390 (600 × 0.65). This completely hypothetical example shows that sometimes a weaker basis for decisions must be accepted if the potential effects are much greater. The concept of expected utility can be developed by using more complicated decision trees where also quality of life or potential negative effects of an intervention are included in the analysis.

Another interesting observation from reviewing the literature on population-based CHD interventions is that the two projects in Scandinavia, North Karelia and Norsjö, achieved more significant risk factor reductions than the studies from the United States. A comparison of the intervention design and the process indicates that the Scandinavian projects had higher local involvement, a deeper commitment from medical care and primary health care in particular, and a higher penetration rate per inhabitant. The US projects were much larger in terms of both costs and intervened populations, but much of the efforts seem to have focused on the evaluation and less on intensive interventions. The total estimated costs in the Norsjö project were two to four times greater per inhabitant than, for example, those of the Minnesota project.[26] The participation rate in the screening was 95% in the Norsjö project and 60% in the Minnesota project. These differences in intervention design and process between projects from different settings have so far aroused little attention.

Conflicts of interest: none.

1 Drummond MF, Heyse J, Cook J, *et al*. Selection of end points in economic evaluations of coronary-heart-disease interventions. *Med Decis Making* 1993;**13**:184–90.
2 Nutbeam D. Evaluating health promotion - progress, problems and solutions. *Health Promotion* 1998;**13**:27–44.
3 Susser M. The tribulations of trials-intervention in communities. *Am J Public Health* 1995;**85**:156–8.
4 Fishbein M. Great expectations, or do we ask too much from community-level interventions? *Am J Public Health* 1996;**86**:1075–6.
5 Last JM. *A dictionary of epidemiology*. 2nd ed. New York: Oxford University Press, 1988.
6 Oliver M. Coronary risk factors: should we not forget about mass control? *Lancet* 1983;ii:37–8.
7 McCormick J, Schrabanek P. Coronary heart disease is not preventable by population interventions. *Lancet* 1988;ii: 839–41.
8 Salonen JT. Did the North-Karelia Project reduce coronary mortality? *Lancet* 1987;i:269.
9 Hopkins PN, Williams RR. A survey of 246 suggested coronary risk factors. *Atherosclerosis* 1981;**40**:1–52.
10 Brännström I, Weinehall L, Persson L-Å, *et al*. Changing social patterns of risk factors for cardiovascular disease in a Swedish community intervention programme. *Int J Epidemiol* 1993;**22**:1026–37.
11 Lindholm L, Rosén M, Weinehall L, *et al*. Cost effectiveness and equity of a community based cardiovascular disease prevention programme in Norsjö, Sweden. *J Epidemiol Community Health* 1996;**50**:190–5.
12 Weinehall L, Westman G, Hellsten G, *et al*. Shifting the distribution of risk: results of a community intervention in a Swedish programme for the prevention of cardiovascular disease. *J Epidemiol Community Health* 1999;**53**:243–50.
13 Multiple Risk Factor Intervention Trial Research Group. Multiple risk factor intervention trial. Risk factor changes and mortality results. *JAMA* 1982;**248**:1465–77.
14 Luostarinen T, Hakulinen T, Pukkala E. Cancer risk following a community-based programme to prevent cardiovascular diseases. *Int J Epidemiol* 1995;**24**:1094–9.
15 Puska P, Salonen JT, Nissinen A, *et al*. Change in risk factors for coronary heart disease during 10 years of a community intervention programme (North Karelia project). *BMJ* 1983;**287**:1840–4.
16 Tuomilehto J, Geboers J, Salonen JT, *et al*. Decline in cardiovascular mortality in North Karelia and other parts of Finland. BMJ 1986;**293**:1068–71.
17 Nutbeam D, Smith C, Murphy S, *et al*. Maintaining evaluation designs in long term community based health promotion programmes: Heartbeat Wales case study. *J Epidemiol Community Health* 1993;**47**:127–33.
18 Lindholm L, Janlert U, Karlsson K-E, *et al*. "Gröna nyckel-hålet" projekt för folkhälsan (The "Green Keyhole" project for public health. The effects of society's impact on health is difficult to measure.) [In Swedish]. *Läkartidningen* 1997; **94**:1905–7.
19 Rosén M. On randomized controlled trials and lifestyle interventions. *Int J Epidemiol* 1989;**18**:993–4.
20 Aaro LE, Hauknes A, Berglund E-L. Smoking among Norwegian schoolchildren 1975–1980. II. The influence of the social environment. *Scand J Psychol* 1981;**22**:297–309.
21 Kannel WB, Wolf PA, Garrison RJ, eds. *The Framingham study. An epidemiological investigation of cardiovascular disease. Section 37: The probability of developing certain cardiovascular disease in eight years at specified values of some characteristics*. NIH Publication No 87–2284, 1987.
22 Law MR, Wald NJ, Thompson SG. By how much and how quickly does reduction in serum cholesterol concentration lower risk of ischaemic heart disease? *BMJ* 1994;**308**:367–72.
23 The Multiple Risk Factor Intervention Trial Research Group. Mortality rates after 10.5 years for participants in the multiple risk factor intervention trial. Findings related to a priori hypotheses of the trial. *JAMA* 1990;**263**:1795–801.
24 Farquhar JW, Fortmann SP, Flora JP, *et al*. Effects of communitywide education on cardiovascular disease risk factors. The Stanford Five-City Project. *JAMA* 1990;**264**: 359–65.
25 Winkleby MA, Taylor CB, Jatulis D, *et al*. The long-term effects of a cardiovascular disease prevention trial: the Stanford Five-City Project. *Am J Public Health* 1996;**86**: 1773–9.
26 Lindholm L. *Health economic evaluation of community-based cardiovascular disease prevention. Some theoretical aspects and empirical results*. [Thesis]. Umeå: Department of Epidemiology and Public Health, Umeå University, 1996.
27 Luepker RV, Råstam L, Hannan PJ, *et al*. Community education for cardiovascular disease prevention. Morbidity and mortality results from the Minnesota Heart Health Program. *Am J Epidemiol* 1996;**144**:351–62.
28 Multiple Risk Factor Intervention Trial Research Group. Coronary heart disease death, nonfatal acute myocardial infarction and other clinical outcomes in the Multiple risk factor intervention trial. *Am J Cardiol* 1986;**58**:1–13.
29 Rose G. Strategy of prevention: lessons from cardiovascular disease. *BMJ* 1981;**282**:1847–51.