*Bulletin of the World Health Organization*, 64 (2): 311-320 (1986)


# Evaluating primary prevention programmes against cancer

JAMES A. HANLEY[1]

*Based on current knowledge, roughly one third of all cancers worldwide are preventable, and primary prevention is increasingly seen as an important cancer control strategy. Interventions to reduce the exposure to known causes can be effected through legislation or education, or by means of vaccination or chemoprevention. Since primary prevention actions can be costly and will compete for resources needed for other disease control activities, and since there is no guarantee that they will be successful, they should not be introduced haphazardly but on the basis of scientific evaluations. This paper presents the main principles to be followed in designing such evaluations; the illustrations often, of necessity, come from other diseases (particularly cardiovascular disease), where there is considerably more experience. Because the interventions involve changes in lifestyle and behaviour, and because a long time is necessary to observe the ultimate endpoints, controlled intervention studies against cancer present many scientific and logistical difficulties. Some interventions, such as vaccination and chemoprevention (to test suspected protective agents) may be evaluated by traditional clinical trial methodology, using intermediate as well as final (cancer incidence and/or mortality) endpoints. Active, target-directed and preferably controlled health service research studies will definitely be needed to assess community or population interventions based on legislation or education.*

## INTRODUCTION

The number of new cancers worldwide in 1975 was estimated to be approximately 6 million, more than half of them in the developing countries (*1*). Cancer is the third leading cause of death in persons aged over 5 years and its incidence is increasing, both because of an aging world population and because of higher age-specific risks for some tumours, notably lung cancer: an epidemic of cancer is predicted in the majority of the developing countries by the year 2000 (*2*). Two of the three approaches to reduce morbidity and mortality from cancer are by treatment and aftercare of diagnosed cancers (so-called "tertiary" prevention), and by early detection coupled with effective therapy ("secondary" prevention). However, treatment is costly and often unavailable or given too late; promotion of early detection is possible for some cancers but for others there are many technical and financial barriers (*3*) to its widespread use.

The third approach, which focuses on eliminating the conditions that cause a cancer to develop ("primary" prevention), is increasingly being advocated as an important control strategy (*4–7*); not only is the cancer prevented, but the same measures also reduce the risk of other diseases. Unfortunately, prevention is often guided by the "heart rather than the head" and its effectiveness tends to be poorly evaluated (*8*). Because the concept of primary prevention of cancer and its evaluation are relatively new, there has not been much discussion of the scientific challenges involved. The present paper considers the methodological problems in designing primary prevention studies in cancer, and in interpreting the resulting data.

### *Steps in primary prevention research*

Preventing any cancer from developing requires two distinct phases: the first identifies the causes and the second alters the exposure to them. The first phase can be effected by carcinogenicity testing on animals, or by observational studies and experimental interventions on humans to test if altering the exposure to an agent actually changes the natural history of a cancer, or of precursor lesions such as colonic polyps, oral leukoplakia, and gastrointestinal

[1] Consultant, Cancer Unit, World Health Organization, Geneva, Switzerland; on sabbatical leave from the Department of Epidemiology and Biostatistics, McGill University, 1020 Pine Avenue West, Montreal H3A 1A2, Quebec, Canada. Requests for reprints should be sent to Cancer Unit, World Health Organization, 1211 Geneva 27, Switzerland.

Table 1. Agents/conditions which cause (+ and + +) major cancers or protect (−) against them[a]

| | Site of cancer | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Stomach | Lung | Breast | Colon/ rectum | Cervix uteri | Mouth/ pharynx | Oeso- phagus | Liver | Prostate | Bladder |
| Worldwide incidence (million)[b] | 0.7 | 0.6 | 0.5 | 0.5 | 0.5 | 0.3 | 0.3 | 0.3 | 0.2 | 0.2 |
| Tobacco/alcohol: | | | | | | | | | | |
| Tobacco smoking | | + + | | | + | + + | + + | | | + + |
| Chewing betel/tobacco/lime | | | | | | + + | | | | |
| Passive smoking | | + | | | | | | | | |
| Alcoholic drinks | | | | | | + + | + + | + + | | |
| Diet/nutrition: | | | | | | | | | | |
| Dietary fat | | | + | + | | | | | + | |
| Nitrate/nitrite | + | | | | | | + | | | |
| Obesity | | | + | | | | | | | |
| Dietary fibre | | | | − | | | | | | |
| Vitamin A | | − | − | − | | | − | | | − |
| Vitamin C | − | | | | − | | − | | | |
| Green vegetables | − | | | | | | | | | |
| Viruses/parasites/toxins: | | | | | | | | | | |
| Viruses | | | | | + | | | + + | | |
| *Chlonorchis* | | | | | | | | + + | | |
| *Schistosoma* | | | | | | | | | | + + |
| Aflatoxin | | | | | | | | + + | | |
| Sexual/reproductive habits: | | | | | | | | | | |
| Sexual behaviour | | | | | + + | | | | | |
| Late 1st pregnancy | | | + + | | | | | | | |

[a] + denotes a suspected carcinogenic connection between the agent/condition and the cancer; + + denotes an established carcinogenic connection (*10*); − denotes a suspected protective connection.

[b] Estimates for 1975 from reference *1*.

or cervical dysplasia. The costs and expected yields from different types of research to identify preventable causes of cancer (e.g., carcinogenicity testing versus chemoprevention trials) cannot properly be compared without formal analysis of the data (*9*).

The second phase tests ways to implement existing knowledge about cancer-related causative and protective factors. Since one does not need to know the exact causative mechanism or active constituent in order to make an intervention, the knowledge already available to us has considerable potential for prevention; thus, if some of the established causes of cancer were removed, a considerable number of cases would be prevented. For example, tobacco alone is responsible for an estimated 30% of all cancers in the USA (*10*) or 20% worldwide (*1*); other agents and conditions known to cause, or protect against, cancer (see Table 1) are believed to be responsible for at least another 10% of cancers in the USA, and for more than this in the developing countries where cancers produced by viruses are prominent. If one included in this list those agents and conditions (such as dietary factors) where there is still some doubt about the evidence, the majority of cancers would be avoidable. Many argue that even with these latter agents and conditions, we should intervene opportunistically.

This second phase (or implementation step) is necessary because knowing the etiology does not guarantee that we can prevent the disease, particularly in the case of factors connected with lifestyle such as tobacco, alcohol, exposure to ultraviolet

light, and sexual behaviour, where complex psychological, physiological, as well as cultural and commercial forces are formidable barriers to change. However, some important life-style changes have been achieved, e.g., in cigarette smoking by British and U.S. physicians and by male persons in the USA and some Nordic countries, in traditional tobacco use among certain groups in rural India (*11*), and in reducing the risk of herpes and AIDS infection (*12*). These examples suggest that sociologists, behavioural scientists, and marketing specialists should also be involved in designing active disease prevention programmes aimed at changing life-style behaviour so that individuals may reduce their exposure to cancer-causing agents for cancer prevention and other health reasons. The jump from etiology to prevention is no less a challenge in exposures of a more physiological nature, as evidenced by the many research questions that need to be answered regarding vaccination against the hepatitis B virus (HBV) carrier state (*6*). Two second-phase studies of such vaccination of infants have been planned for China and the Gambia, (see below) and will, if they document a change in the carrier state, continue for up to 35 years to determine whether it is indeed followed by the expected reduction in cancer incidence.

Some topics related mainly to prophylaxis trials, such as ethical issues and the choice of study endpoints, have been discussed elsewhere (*13*); this paper will focus, under two main headings, on the types and levels of intervention, and on data acquisition and measurement of the effects of intervening (costs will not be considered explicitly). The illustrations will often, of necessity, come from other diseases (particularly cardiovascular disease), where there is considerably more experience.

## INTERVENTIONS: WHAT KIND AND ON WHOM?

Studies to evaluate the following modes of intervention will now be considered:

— Legislation and regulation: mandatory labelling of products with health warnings, discouraging and restricting the promotion of carcinogenic products (e.g., by taxation), regulating the contents of these products, regulating workplace exposures, and restricting smoking activities;

— Education: informing the public about cancer risks and helping them to change their life-style, e.g., concerning smoking, alcohol, tobacco chewing, and diet;

— Chemoprevention: use of 13-cis benzoic acid (to prevent recurrent skin cancer), retinyl acetate (cervical dysplasia), and anti-schistosomiasis drugs (bladder cancer), as well as dietary supplements such as beta-carotene, vitamin E, and selenium;

— Vaccination: against infections due to hepatitis B virus or, in the future, human papilloma virus.

An intervention can be targeted at a whole population or part of it, or at selected high-risk individuals, depending on the composition of the population and the available channels, personnel and resources. Legislative intervention is generally promulgated at the national level, although local authorities may have certain powers, e.g., to restrict cigarette smoking in public buildings. When the intervention is directed at an entire population, it is more difficult to know if the results following the intervention are due specifically to it; varying the intervention in different subdivisions of the population can help measure the true effect (see below).

Interventions can be applied through education on a nationwide scale (e.g., health messages given to the entire population by the health ministry or a voluntary organization), or locally (within a community, workplace, school or family), or individually (counselling by physicians or local health care workers). Personalized education, even if feasible, is usually too costly and difficult to evaluate owing to possible "contamination" of individuals in non-intervention groups through chance contacts with recipients of the intervention or with educators. In a community setting, this type of contamination can be turned to good use: as the social influences that determine a person's habits lie in his home and work environment, education should optimally be inserted into as many aspects of his life as possible. Interventions in communities, rather than among high-risk individuals, pose other constraints and challenges such as complexity of implementation; need for cooperation of agencies; follow-up of individuals who have moved; impossibility of blinding, or of concentrating on a single risk factor; low penetration; lack of interest; and inability to study a sufficient number of communities (*14*).

In principle, chemoprevention agents can be allocated alternately to one individual and not the next. Even though such allocation generally yields the most "statistically efficient" evaluation of an intervention, there may be logistical difficulties such as individual randomization. Vaccinations can also be given to individuals by alternate selection, but groups may participate more readily and the administration of vaccines to entire groups is simpler. This is the approach proposed for studies of HBV vaccination in Qidong (China) and the Gambia, where entire birth cohorts in different communes, and in areas served by different vaccination teams, respectively, will be vaccinated and compared with others that are not. To increase acceptability, it may help if each individual or group was made the subject of an intervention

against one cancer, while serving as a "control" in an intervention against another, as proposed for evaluation of cancer screening (M. Zelen, 1982, unpublished WHO report). Other design variations are also possible (*15–17*).

With some interventions, there may be a choice of strategies. In the "medical" model, the intervention is aimed at those who are thought to be at higher risk (what upper fraction of the risk distribution should be given the intervention is discussed in reference *18*), while in the "public health" model the intervention is aimed at all individuals in a community or workplace. Several considerations favour a mass approach: (i) the higher cost of identifying and intervening on individuals, (ii) the inability of indices to predict which individuals would develop cancer, (iii) the naturalness of intervening in a social setting, (iv) the greater likelihood that the changes achieved in the general population will endure (this is particularly relevant in choosing between smoking-cessation and smoking-prevention strategies), and (v) the fact that although risks are higher at higher levels of the risk factor, only a small fraction of the population may have the highest risk levels. Since the majority of disease cases then arises from those with lower, but still elevated, risk levels, the aim should be to shift the entire risk distribution downwards, rather than just those in the tail of the distribution curve (*19–22*).

## EVALUATION

In this section we shall discuss different comparison groups and data acquisition and interpretation. We shall begin with experimental designs involving individuals or small groups, and how they can be adapted to long-term interventions on large numbers of persons, and then consider community or larger interventions. Although the latter will often be only "quasi"-experimental in design, they can still, by choosing suitable comparison data, provide a critical evaluation (*23*). Evaluation of only a single intervention and a single "non-intervention" will be considered, although it may be useful to study graded interventions or a second "active" control group (to counteract placebo and measurement artefacts).

### *Interventions allocated to individuals*

*Design.* The most likely use of this method of intervention is in the study of vaccines, chemoprevention, and personal counselling, i.e., where the intervention can easily be applied to any one individual and where there is no possibility of it being shared by those who are not to receive it. Informal methods of individual allocation, such as selection of alternate names from a list, or use of the digits of a serial number or birthday, can be abused; interventions should therefore be assigned by a mechanism outside the control of the subject and the investigator, both of them not knowing which intervention will be used until after the subject has been deemed eligible and been registered in the trial (*24*). Since sample sizes are likely to be large (see below), unstratified randomization can be used with virtually no loss in statistical power.

*Data analysis: behavioural change studies.* If possible, individual behaviours should be recorded on a quantitative rather than a "yes–no" scale, should be verified by objective measures (e.g., urinary or serum levels of constituents linked with smoking or other behaviour, or questioning of others to verify the reported behaviour), and should be obtained before as well as after the intervention. Post-intervention behaviour, or, if measured, the change from before to after the intervention (the "pre–post" change), can be compared using standard statistical methods.

The number of subjects required depends on the nature of the changes that are sought, quantitative changes (e.g., in the number of cigarettes smoked) being generally easier to demonstrate statistically than qualitative ones (e.g., changing from smoking to nonsmoking); for the latter, the proportion of subjects likely to change their behaviour determines the sample size needed (see below). If the intervention is simple and inexpensive, one can only expect, and must be satisfied with, small changes. For example, even if the success of physicians' advice against smoking were small, the intervention can be directed at a large number of individuals and prevent many cases of induced disease (*25*). Such intervention studies must be large enough to accurately measure small effects in order to guard against concluding that because a "statistically significant" result is not observed, a reduction of public health importance is not achievable.

*Data analysis: cancer precursor and cancer studies.* The state of precursor lesions should, if possible, be recorded on a graded rather than a simple "present–absent" scale, and (if the procedure does not pose excessive risk) both before and after the intervention in order to allow comparison of progression rates (*11*). Studies aimed at preventing the appearance of a cancer precursor require the examination of individuals at regular intervals, with methods for identification geared to local conditions. Already diagnosed (or fatal) cancers can be ascertained from population or hospital-based registries, if they exist; otherwise, a special investigation will be needed. In more developed countries, follow-up will be easier if the participants can be cross-linked with

national data-banks or are chosen from professional bodies that maintain directories of their members.

In the calculation of incidence rates, the choice of time-scale for the "events" (e.g., cancers diagnosed) and the person-years at risk must be guided by the object of the intervention. If the latter is thought to modify an early phase of cancer development, those lesions appearing soon after the intervention should not be allowed to dilute the comparison, while actions that might influence the later stages of cancer development should be tested in the early post-intervention period (it is the belief in the latter mechanism that makes the 5-year trial of beta-carotene in 20 000 physicians aged 50–75 feasible (*17*)).

Sample sizes needed for cancer incidence studies are derived from the Poisson distribution, emphasizing that statistical power is more related to the number of "events" than the size, *per se*, of the denominator. In order for a one-sided statistical test with a $P=0.05$ level of significance to have a reasonable (80%) chance of detecting a 25% reduction in the incidence rate, using equal-sized experimental and comparison groups, one requires sufficient person-years to generate approximately 175 "events" in the comparison group. Detecting a 50% reduction in incidence would require less than 40 "events", i.e., a study less than a quarter the size, but considering factors such as incomplete compliance, diminishing effects, migration, and the fact that attributable risks are less than 100%, planning for more than a 25% risk reduction (thus decreasing the required sample size) is unrealistic. A possible exception is vaccination against hepatitis B virus infection where the two proposed studies in China and the Gambia, involving the follow-up for 30–35 years of 100 000 and 60 000 newborns, respectively, expect protection rates of 60–80%. Other statistical aspects, including the effect of volunteers, pre-trial measurements of risk, lag time and compliance, have been discussed in other studies (*26*, *27*).

Consideration must be given, when determining sample size, to the possibility that an intervention may not produce the same reductions in cancer incidence across all risk strata and to whether the reductions in incidence will be proportional (multiplicative) or absolute (additive). This latter issue is also important when intervening on multiple factors, e.g., smoking and alcohol consumption (*10*).

If the number of "events" (diagnosed cancers) in incidence studies is small, it is possible to economize on data collection and processing by employing a case–control analysis (*28*), only using the data for those who developed the cancer and for a sample of those who did not (*29*, *30*). It also allows more thorough rechecking of group membership, using vaccination scars, serology, etc. This method is stronger if used in a closed cohort formed by a randomized trial, and offers an inexpensive way to use already completed intervention trials, e.g., of diet changes (to prevent coronary heart disease (CHD)) or BCG vaccination (tuberculosis), to assess their impact on cancer endpoints. CHD prevention trials have been of short duration by comparison with the longer latency and lower "event" rates in cancer, but the marginal cost to continue to trace cancer events and the quality of the additional data that could be obtained should be considered. Ongoing non-experimental cardiovascular disease studies (*31*) could also be extended to include cancer risks.

*Interventions allocated to collections of individuals*

In many instances, it is neither feasible nor desirable to allocate one individual to one intervention and an immediate "neighbour" to another, and it may be more appropriate to allocate related individuals (e.g., a family, a class or the entire school, the clients served by a primary health care worker, or the inhabitants of an area) as a "unit" to the same intervention; if resources are limited, or if the individuals in a population (e.g., adult males) are not easily reached by other methods, a mass media campaign may be necessary.

The statistical analyses and inferences from studies that allocate interventions to entire clusters, or in which individuals undergo communal treatment, have often been inappropriate and overoptimistic (*32*). The correct interpretation of such trials has been fully treated in the literature on: testing teaching innovations using entire classrooms of students (*23*); health care research studies of interventions involving providers of care (*32*); the use of villages as experimental units in tuberculosis prophylaxis trials (*33*); and the use of single randomization to decide the treatment which a physician would offer to all his eligible patients (*34*). Applications of these principles include: CHD prevention in factory workers (*35*), hepatitis prophylaxis in army units (*36*), educational intervention at air force bases (*37*), and a re-analysis of the data from the Stanford three-community project (*38*), contrasting the "proper" analysis, which treats the community as the unit, with that which treated individuals as units.

*Use of many small units.* Examples include families, classes, schools, or patients in physicians' practices. The large number of units allows matching and random allocation to equalize the average baseline risks of intervention and non-intervention units, and makes it more difficult for after-the-allocation perturbations to selectively occur in some of the intervention or non-intervention units.

In educational interventions, variations in the abilities of instructors or counsellors may be greater than variations in the recipients, producing more than random concordance of results within each unit. To analyse the data from such studies (*39*) one should (*a*) ignore the within-unit variation in response and consider each unit as just one observation, (*b*) compute the average response in each unit, and (*c*) judge the differences between the results in the intervention and non-intervention units against those seen among the units receiving the same intervention. Cluster randomization requires larger sample sizes: Donner's investigation (*40*) of intra-unit concordance in spouse pairs and physician practices, and for both binary and continuous responses, illustrates the "inflation factors" involved.

*Use of fewer large units.* If several (say, $2k$) large units, such as the inhabitants of provinces, counties, health districts, towns, or villages, are available, the intervention can be carried out in half (say, $k$) of these, with the remaining $k$ serving for comparison. For example, a US National Cancer Institute study of smoking cessation methods called for 8 matched pairs of communities, with communities within each pair randomized to either an intervention or control condition. When $k$ is only 2 or 3 (as was typical in community trials in CHD prevention (*22*)), it is difficult to ensure comparability; deliberate balancing of units is critical, while randomization is of more limited use. Even if matched from the start, subsequent unexpected developments in a unit, such as legislative or other administrative changes, or publicity (e.g., a temporary court ban on a TV message about cigarettes (*41*)) can have a major impact.

Data on behavioural changes in each unit may be obtained from production or consumption statistics or collected in survey samples. Assessing changes in cancer incidence is more complex, particularly if the units are ill-defined; as a minimum, the age, sex and residential history of each case (numerator), along with the age and sex distribution of the population in the unit (denominator), are required.

As before, the appropriate analysis is by unit, giving a total of $2k$ observations; indeed, studies with a very small number of units more properly belong under quasi-experimental studies (see below). A justification for this seemingly stringent approach is provided by an example: in a study attempting to reduce non-attendance at exercise classes (M. Belisle & E. Roskies, personal communication, 1984), prophylactic intervention was carried out in four ($k=4$) classes of 25 students each ($n=25$), with four other classes serving as "controls". The four experimental classes ranked 1st, 2nd, 3rd, and 8th in average attendance. Upon probing why one class was ranked 8th, it was revealed that it was the one that met at 7 in the morning. Even if one could be assured that all the class directors provided the same degree of motivation, or even if each participant were individually randomized, or if there had been not $n=25$, but $n=250$, in each class, a small $k$ cannot reduce the risk of perturbations and the difficulty of interpretation.[a]

*Quasi-experimental studies*

In experimental studies the intervention is allocated to units that are selected by a mechanism under the control of the experimenters, and the effects are observed. Although many interventions have to be introduced as part of a regular administrative routine, which makes this kind of experimental allocation impossible, there may still be an opportunity at least to measure the intervention process and outcome, and to do this at other times or on other populations for comparison. This limitation on the choice of when and on whom to intervene, but with the freedom to choose when and whom to measure, distinguishes the quasi-experimental from the truly experimental study. Quasi-experimental studies often take advantage of whatever is available, e.g., in personnel and materials, as shown by the examples.

*Serial data from a single intervention.* The most convincing experimental data concerning the effect of an intervention are obtained, where it is practical, by repeatedly applying and withdrawing the intervention from the same individuals. This design might possibly be used to compare smoking-cessation techniques or to study the prevention of certain recurrent tumours (e.g., of the urinary bladder, skin and mucosa), but it is suitable only for short-term treatments with short-term effects and reversible outcomes.

The usual tactic employed when one can only study a single group or unit is to compare the cancer-producing behaviours and/or actual cancers ("events") both before and after the intervention. The danger of concluding *post hoc, ergo propter hoc* has repeatedly been pointed out. The procedure, when there is a single group, can be improved by computing not just one but a series of data readings (say, one per year) during the pre- and post-intervention periods (*42*) (see Fig. 3 on page 38 of reference *23* for hypothetical examples of "interrupted" time series, and references *43* and *44* for

[a] A large number of individuals in an intervention unit allows one to judge if this unit had a different response from the non-intervention unit; a large number of intervention and non-intervention units should allow one to judge whether a different response will be true of units in general (at least on average) and whether the response might be due to the intervention.

concrete examples with national anti-smoking legislation). Data measured in this way over a period of time can help avoid another artefact of this "pre–post" design, namely regression to the mean, which occurs if a particular group is chosen for study because it presented the most extreme "event" rate or behaviour pattern. Unless these "extreme" data are an enduring feature of the group, an observed change post intervention might well mean that it was selected on the basis of a randomly extreme pre-intervention experience.

*Serial data from several (staggered) interventions.* While serial data may indicate a change following an intervention, they cannot rule out the possibility that this was due to some other concurrent factor, such as another programme, or a change in the method of record-keeping. This uncertainty is lessened if the intervention is carried out in several places (*11*), and reduced still further by introducing the intervention at a different time in each place. This is illustrated by considering the effect of introducing seat-belt legislation to reduce motor vehicle fatalities. If one evaluated the effect of a 1973 seat-belt law in a single state, one would have to contend with the widespread introduction, around the same time, of reduced speed limits. Data from several states, each of which introduced seat-belt legislation at a different time, would help to separate the two effects. Several studies have used changes in the time and space pattern of BCG vaccination to assess its effect on cancer risk (*45*).

The staggered introduction of interventions, if feasible, is most useful in assessing rapid outcomes, such as behavioural changes following legislation or education. Also, compared with the "parallel" design (see below), this approach spreads out the work of implementation and data collection, an important consideration if a large number of highly trained staff are needed. However, for studying long-term cancer incidence, it may be difficult to interpret the changes, even if one has several staggered series: many other uncontrolled changes may take place concurrently over the follow-up period (which will be longer than the period over which the interventions were introduced).

*Two-group (parallel) designs.* The above-mentioned single-group, "pre–post" design, which may be denoted schematically as O—X—O (using "X" for the intervention and "O" for the observation or measurement before and after it (*23*)), is vulnerable to the effects of uncontrolled factors. An alternative is to use a concurrent group as a parallel non-intervention or "control" unit. The data can be of the form O—X—O versus O——O (if one can obtain pre-intervention data (*22*)) or the simpler, but much weaker, X—O versus —O (if one cannot).

If one relies on existing data sources, both the intervention (X) group and the comparison population may well have had the pre-intervention measurements made in the same way. However, the very fact of intervening may change data recording after the intervention in the X group, in which case the post-intervention measurements must be standardized across both groups. It is not critical that the two groups should have the same baseline values, but the data for comparison should be measured in the same way. Also, as with individuals, one has to avoid a spillover effect of programmes from one group to another; otherwise, if changes occur in the comparison as well as the intervention group, as in the Multiple Risk Factor Intervention Trial (MRFIT) study (*27*), critics of the intervention are quick to label the results as "negative".

Parallel groups, as in the following three examples, were used in several quasi-experimental studies of the effect of BCG on cancer incidence: children in Jerusalem were compared with other Israeli children who were subject to different vaccination policies; some 40–50% of vaccinated infants in a Chicago hospital were compared with the remaining ones who were not vaccinated; and Quebec nurses who had been vaccinated as schoolchildren were compared with others who were not (*45*). In all of these studies, particularly as they lacked pre-intervention data, there was obvious concern that the experiments were not perfectly "natural", i.e., the vaccinated individuals were different in an important way from those not vaccinated. It might have been possible in the Israeli and Quebec studies to assemble corresponding pre-intervention counterparts of the vaccinated and unvaccinated groups, since the allocation was by geographical area and by school, respectively. This would have been impossible in the Chicago study. Efforts were made in these studies to compare groups on other, "dummy" outcomes (i.e., outcomes not expected to be related to the intervention), such as trauma, in order to provide reassurance that they were similar in all other (measured) respects.

The issue of whether an entire group, or each individual in it, is to be considered as the appropriate unit for statistical analysis is even more relevant in quasi-experimental designs. The correct choice depends on the degree to which subjects are affected individually or communally by both the disease and the intervention. The 1954 U.S. poliomyelitis vaccine trial (*46*) is a case in point. In several areas, instead of being individually randomly allocated to either placebo or vaccine, all children in grades 2 were offered the polio vaccine while those in grades 1 and 3 were simply observed. On the assumption that

poliomyelitis is not contagious within a classroom or does not appear to otherwise cluster in time or space, and that none of the batches of vaccine was defective, then the individual child is probably the appropriate unit of analysis. To the extent that these assumptions are violated, the classes, or possibly the groups of vaccinated and unvaccinated children within a community or area, become the units of analysis. Fortunately, the results were sufficiently convincing whether viewed on an aggregate basis, or compared separately within each area; the findings were also corroborated by those in the areas where individual random allocation was used.

A second pointer from this same portion of the poliomyelitis trial concerns the pitfalls of self-selection (*47*). Those grade 2 children whose parents did not permit them to be vaccinated had different attack rates from the children in grades 1 and 3, all of whom, by design, were denied vaccination. It is difficult to construct a group of children in grades 1 and 3 to serve as a comparison group for those grade 2 children who received the vaccination.

*Several (parallel) intervention and non-intervention groups.* In the same way that several serial comparisons offset the weaknesses of a single group, so too will several parallel (intervention and control) groups strengthen a comparison made within a single pair. This applies especially if the groups have not been formed by randomization and are strongly clustered in their initial characterisitics or in their treatment or assessment.

A second advantage in having several groups is the opportunity to study systematic variations in the intervention. In the situation where one has experimental control over the allocation, one might allocate two or more versions or "doses" of the intervention, and possibly even study them in each of several cohorts in a counterbalanced design; in a less controlled situation, one may be able to take advantage of the fact that different groups received different amounts of the intervention and thus capitalize on what, it is hoped, is a natural experiment.

## CONCLUSION

This article has listed the possible modes of controlling cancer by primary prevention and discussed ways to evaluate their effects. While it may be tempting to simply introduce an intervention without any provision for evaluating its impact because of a mistaken belief that "it is bound to work" or because of the many difficulties involved in the evaluation itself, there is considerable evidence that the attempts to prevent acute and chronic diseases have not always worked, or worked as well as was hoped. Thus, it is as important to evaluate an intervention as it is to actually intervene. It is only through proper attention to prospective evaluation that one can determine how well a primary intervention worked, or how well it is likely to work in other settings, and how cost-effective it is when compared with other control strategies.

## ACKNOWLEDGEMENTS

I am grateful to J-F Boivin, W. F. Forbes, K. Stanley, J. Stjernswärd and S. Wacholder for comments and suggestions. The work was supported by the Cancer unit, World Health Organization, Geneva, Switzerland, and the Natural Sciences and Engineering Research Council of Canada.

## RÉSUMÉ

### EVALUATION DES PROGRAMMES DE PRÉVENTION PRIMAIRE ANTICANCÉREUSE

Chez les individus qui ont survécu aux cinq premières années de la vie, le cancer est l'une des trois principales causes de décès, tant dans les pays développés que dans les pays en développement. A l'heure actuelle, pour bon nombre des cancers les plus courants le dépistage et le traitement précoces efficaces ne sont pas réalisables, ni sur le plan technique ni financièrement, pour une grande partie de la population mondiale. La prévention primaire (c'est-à-dire le fait d'aller à la source du problème et d'empêcher le cancer de se développer) est considérée de plus en plus comme une stratégie importante dans la lutte anticancéreuse. Elle comporte deux étapes: identifier les agents nocifs et les agents protecteurs et modifier les conditions d'exposition à ces agents.

Une partie importante de la première phase a déjà été menée à son terme en ce sens que, sur la base des connaissances actuelles, on estime qu'un tiers de tous les cancers dans le monde sont évitables; le but devrait consister désormais à élaborer et à mettre à l'épreuve les moyens d'appliquer cette connaissance par le truchement de la législation, de l'éducation, de la vaccination et de la chimioprophylaxie.

Du fait que les actions de prévention primaire doivent soutenir la concurrence d'autres activités de lutte contre la maladie, il convient d'en évaluer le coût et les avantages sur une base scientifique. Le présent article expose les grands principes à observer dans la conception de telles évaluations; malheureusement, comme on n'a prêté que peu d'attention à l'exécution et à l'évaluation des actions de prévention primaire du cancer, les exemples portent nécessairement le plus souvent sur d'autres maladies (en particulier les maladies cardio-vasculaires) au sujet desquelles on a beaucoup plus d'expérience. Etant donné que ces interventions contrôlées impliquent des changements de comportement et qu'il faut un long délai pour observer une réduction du cancer, ces actions soulèvement nombre de difficultés sur le plan scientifique et logistique. Peut-être certaines interventions telles que la vaccination et la chimioprophylaxie (pour tester des agents présumés protecteurs) pourraient-elles faire appel à une méthodologie traditionnelle d'essais cliniques. Elles devront peut-être considérer comme points intermédiaires, non plus une réduction de l'incidence du cancer, mais plutôt des modifications du niveau des nutriments (par exemple les vitamines) ou des niveaux d'infection, ou bien de l'état des lésions servant de précurseurs telles que les polypes du côlon, la leucoplasie buccale ou la dysplasie du col utérin.

Des recherches sur les services de santé actives et orientées vers un but déterminé, et de préférence effectuées avec des groupes témoins, au sein de communautés ou d'une population entière, seront très nettement nécessaires pour évaluer les mesures de lutte dans d'autres secteurs tels que la législation et l'éducation. Ces évaluations lancent bien des défis nouveaux, notamment l'inexistence ou la "contamination" des groupes de comparaison traditionnels, l'absence de comparabilité ou l'impossibilité d'éliminer certaines variables, et le fait qu'il faille s'en remettre aux statistiques disponibles ou aux échantillons provenant d'enquêtes. Quoi qu'il en soit, en suivant les principes de conception quasi-expérimentale qui ont été couramment appliqués à l'évaluation des interventions sociales, on peut encore mettre sur pied des mécanismes permettant de surveiller les effets des efforts de prévention primaire.

Même si l'on est parfois tenté de simplement appliquer une intervention sans envisager l'évaluation de son incidence parce qu'on croit à tort qu'elle donnera sûrement des résultats satisfaisants, ou en raison des nombreuses difficultés que soulève l'évaluation elle-même, nombreux sont les indices qui prouvent que beaucoup d'interventions préventives n'ont pas donné les bons résultats escomptés. Par conséquent, il importe tout autant d'évaluer une intervention que de l'exécuter. C'est uniquement en veillant comme il convient aux principes de l'évaluation prospective qu'on pourra déterminer l'efficacité d'une intervention primaire, la comparer avec d'autres stratégies de lutte et savoir dans quelle mesure elle donnera vraisemblablement de bons résultats chez d'autres sujets, à d'autres époques et en d'autres lieux.

## REFERENCES


1. PARKIN, D. M. ET AL. Estimates of the worldwide frequency of twelve major cancers. *Bulletin of the World Health Organization*, **62**: 163–182 (1984).
2. Cancer increases in developed countries. *Weekly epidemiological record*, **60** (17): 125–129 (1985).
3. *Screening for cancer* (UICC Technical Report Series, Volume 78). Geneva, International Union against Cancer, 1984.
4. A WHO MEETING. Reappraisal of the present situation in prevention and control of lung cancer. *Bulletin of the World Health Organization*, **60**: 809–819 (1982).
5. A WHO MEETING. Control of oral cancer in developing countries. *Bulletin of the World Health Organization*, **62**: 817–830 (1984).
6. A WHO MEETING. Prevention of hepatocellular carcinoma by immunization. *Bulletin of the World Health Organization*, **61**: 731–744 (1983).
7. STJERNSWÄRD, J. ET AL. Cancer control: strategies and priorities. *World health forum*, **6**: 160–164 (1985).
8. BANTA, H. D. & LUCE, B. R. Assessing the cost-effectiveness of prevention. *Journal of community health*, **9**: 145–165 (1983).
9. WEINSTEIN, M. C. Cost-effective priorities for cancer prevention. *Science*, **221**: 17–23 (1983).
10. DOLL, R. & PETO, R. *The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today*. Oxford, Oxford University Press, 1981.
11. MEHTA, F. S. ET AL. An intervention study of oral cancer and pre-cancer in rural Indian populations: a preliminary report. *Bulletin of the World Health Organization*, **60**: 441–446 (1982).
12. Self-reported behavioural changes among homosexual and bisexual men—San Francisco. *Journal of the American Medical Association*, **254**: 2537–2538 (1985).
13. MAGNUS, K. & MILLER, A. B. Methodology of controlled prophylactic trials in cancer. In: Armitage, P. et al., ed. *Methods and impact of controlled therapeutic trials in cancer, Part I* (UICC Technical Report Series, Volume 36). Geneva, International Union against Cancer, 1978, pp. 97–128.

14. FARQUHAR, J. W. ET AL. Community studies of cardiovascular disease prevention. In: Kaplan, N.M. & Stamler, J., ed. *Prevention of coronary heart disease*. Philadelphia, Saunders, 1983, pp. 170–181.
15. Report of MRC Committee on Influenza and other Respiratory Virus Vaccines, 1960–1963. *British medical journal*, **2**: 267–271 (1964).
16. PAL, S. C. ET AL. A controlled field trial of an aluminium phosphate-adsorbed cholera vaccine in Calcutta. *Bulletin of the World Health Organization*, **58**: 741–745 (1980).
17. STAMPFER, M. J. ET AL. The $2 \times 2$ factorial design: its application to a randomized trial of aspirin and carotene in U.S. physicians. *Statistics in medicine*, **4**: 111–116 (1985).
18. SONDIK, E. ET AL. High-risk subjects and the cost of large field trials. *Journal of chronic diseases*, **27**: 177–187 (1974).
19. ROSE, G. Relative merits of intervening on whole populations versus high-risk individuals only. In: Lauer, R. M. & Shekelle, R. B., ed. *Childhood prevention of atherosclerosis and hypertension*. New York, Raven Press, 1980, pp. 351–356.
20. ROSE, G. Strategy of prevention: lessons from cardiovascular disease. *British medical journal*, **282**: 1847–1851 (1981).
21. ROSE, G. Sick individuals and sick populations. *International journal of epidemiology*, **14**: 32–38 (1985).
22. PUSKA, P. ET AL. The community-based strategy to prevent coronary heart disease: conclusions from the ten years of the North Karelia project. *Annual review of public health*, **6**: 147–193 (1985).
23. CAMPBELL, D. T. & STANLEY, J. C. *Experimental and quasi-experimental designs for research*. Chicago, Rand McNally & Co., 1963.
24. CHALMERS, T. C. The control of bias in clinical trials. In: Shapiro, S. H. & Louis, T. A., ed. *Clinical trials, issues and approaches*. New York, Marcel Dekker, 1983, pp. 115–127.
25. RUSSELL, M. A. H. Effect of G.P. advice against smoking. *British medical journal*, **2**: 231–235 (1979).
26. WHO EUROPEAN COLLABORATIVE GROUP. An international controlled trial in the multifactorial prevention of coronary heart disease. *International journal of epidemiology*, **3**: 219–224 (1974).
27. THE MRFIT GROUP. Statistical design considerations in the NHLI Multiple Risk Factor Intervention Trial. *Journal of chronic diseases*, **30**: 261–275 (1977).
28. MANTEL, N. Synthetic retrospective studies and related topics. *Biometrics*, **29**: 479–486 (1973).
29. HOROWITZ, R. I. & FEINSTEIN, A. R. Improved observational method for studying therapeutic efficacy. *Journal of the American Medical Association*, **246**: 2455–2459 (1981).
30. SHAPIRO, C. ET AL. A case-control study of BCG and childhood tuberculosis in Cali, Colombia. *International journal of epidemiology*, **14**: 441–446 (1985).
31. TUNSTALL-PEDOE, H. Monitoring trends in cardiovascular disease and risk factors: the WHO "Monica" project. *WHO Chronicle*, 39: 3–5 (1985).
32. WHITING-O'KEEFE, Q. E. ET AL. Choosing the correct unit of analysis in medical care experiments. *Medical care*, **22**: 1101–1114 (1984).
33. HORWITZ, O. & MAGNUS, K. Epidemiologic evaluation of chemoprophylaxis against tuberculosis. *American journal of epidemiology*, **99**: 333–342 (1974).
34. SIMON, R. Composite randomization designs for clinical trials. *Biometrics*, **37**: 723–731 (1981).
35. ROSE, G. ET AL. U.K. heart disease prevention project: incidence and mortality results. *Lancet*, **1**: 1062–1066 (1983).
36. KARK, J. D. Pre-exposure prophylaxis of viral hepatitis with ISG in an endemic area. *Scandinavian journal of infectious disease*, **15**: 3–6 (1983).
37. COTE, L. J. Evaluation of the U.S. Air Force Heart Program demonstration. *Preventive medicine*, **10**: 287–291 (1981).
38. WILLIAMS, P. T. ET AL. A comparison of statistical methods for evaluating risk-factor changes in community-based studies; an example from the Stanford three-community study. *Journal of chronic diseases*, **34**: 565–571 (1981).
39. CORNFIELD, J. Randomization by group: a formal analysis. *American journal of epidemiology*, **108**: 100–102 (1978).
40. DONNER, A. An empirical study of cluster randomization. *International journal of epidemiology*, **11**: 283–286 (1982).
41. EGGER, G. ET AL. Results of a large-scale media antismoking campaign in Australia: north coast "Quit for Life" programme. *British medical journal*, **287**: 1125–1128 (1983).
42. CAMPBELL, D. T. Measuring the effects of social innovations by means of time series. In: Tanur, J. M., ed. *Statistics, a guide to the unknown*. San Francisco, Holden-Day, 1972, pp. 120–129.
43. DOXIADIS, S. A. ET AL. Impact of a nationwide antismoking campaign. *Lancet*, **2**: 712–713 (1985).
44. WARNER, K. E. Cigarette consumption in the 1970s: the impact of the antismoking campaign on consumption. *Science*, **211**: 729–731 (1981).
45. HOOVER, R. N. BCG vaccination and cancer prevention: a critical review of the human experience. *Cancer research*, **36**: 652–654 (1976).
46. FRANCIS, T. ET AL. An evaluation of the 1954 Poliomyelitis Vaccine Trials: summary report. *American journal of public health*, **45** (supplement, part II): i–xiv and 1–63 (1955).
47. MEIER, P. The biggest public health experiment ever: the 1954 field trial of the Salk poliomyelitis vaccine. In: Tanur, J. M., ed. *Statistics, a guide to the unknown*. San Francisco, Holden-Day, 1972, pp. 2–13.