

1

Program evaluation in the health field

Myriam Khat

Institut National d'Etudes Démographiques (INED)
Paris, France

The great leap forward in health conditions dates from the 1950s, with the discovery of antibiotics and the start of wide-ranging programs transferring modern medical technology to the Third World. Some of these interventions have proved to be more effective than others, and at the present time, given the shortage of resources and the need to ensure that action is targeted in the proper direction, it is crucial that health programs be evaluated before, during, and after their implementation.

Nowadays, the art of evaluation is a discipline in its own right which consists of applying a scientific procedure in order to judge the relevance and the effects of an intervention. The purpose of this introduction is to establish a framework for evaluation research, by defining its purposes in the health field more specifically, by clarifying the vocabulary which is attached to it, and by discussing the concepts which underlie the different types of evaluation.

The example of lung cancer screening

Screening programs are a popular type of health program, which are based on the expectation that early detection and early treatment of diseases is associated with better prognosis. This expectation may or may not be met, depending on the disease and on the screening test, and at the population level screening programs can be extremely costly. For these reasons, the benefits which are attached to this type of health programs have to be demonstrated before they are implemented on a large scale.

The example of lung cancer screening is particularly instructive, as this cancer is the commonest cancer in the world today, and a public health problem of growing importance in a number of developing countries. The major preventable risk factor for

lung cancer is tobacco smoking, and health education programs do lead to a reduction of incidence, but only in the long-term. In the short-term, reduction of mortality by early detection is a lung cancer control strategy potentially appealing to health planners.

Cases can be screened by chest X-ray, and there are serious reasons to believe that screening of lung cancer by chest X-ray, and early treatment of screened cases could result in better prognosis and survival. Several trials to investigate the effectiveness of lung cancer screening have been reported. One of them began in Czechoslovakia in 1976 (Kubik, Parkin, Khlát et al., 1990). For the purpose of this trial, 6,364 high-risk cigarette smoking males aged 40-64 were randomized into an intervention group which received 6-monthly screening by chest X-ray, and a control group of the same size which received no asymptomatic investigation. The two groups were followed up for a period of 6 years, and lung cancer cases were followed up until early 1989, a period of at least 5 years from diagnosis, or until death.

One of the findings in this study was that lung cancer cases detected by screening at an early stage had a significantly better survival rate than cases diagnosed after having reported symptoms. But longer survival is in itself related to earlier diagnosis, and the key outcome variable to consider is in fact mortality. A comparison of the two groups in terms of their lung cancer mortality indicated no significant difference, after a follow-up of more than 10 years. The authors therefore concluded that:

- the results of therapy for this disease are not improved by early diagnosis, and programs of screening by regular X-ray examination confer no benefit, and;
- in terms of prevention, efforts should be devoted to primary prevention, by preventing young people from taking up tobacco smoking, and persuading established smokers to stop.

This example highlights the role of health program evaluation as a key element to consider in health planning decision-making, and also the relevance of mortality as an outcome measure in impact assessment. A few basic questions concerning evaluation are addressed below. What is evaluation? Why evaluate? What are the different types of evaluation? What are the strategies for impact assessment?

Evaluation activities

Evaluation can be looked upon as the final step in the process of program implementation. The first step is planning, which involves the definition of the health problem to be tackled and the layout of a program to deal with it; the second step is program operation; and the third step is evaluation, the aim of which is to measure whether the program has fulfilled its objectives.

Generally speaking, evaluation is defined as "the gathering, analysis and interpretation of information to judge the worth of existing or projected programs or

interventions to improve the life of humankind" (Rossi and Freeman, 1993). Alternatively, evaluation can be viewed as "the systematic gathering of information in order to make choices among alternative courses of action" (Borus, Buntz and Tash, 1984). Evaluations of social and health programs have now become commonplace.

There are two main types of evaluations, each of which addresses different questions, but which have to be considered in turn when assessing a program's effects (Borus, Buntz and Tash, 1984):

- **process evaluation**, sometimes referred to as "program monitoring", asks the question: "how did or does the program operate"? In other words, it can be defined as the systematic examination of the program operation, in order to find out: whether the program is reaching the appropriate target population, and; whether its delivery of services is consistent with the plan for the program. Process evaluation is a prerequisite to impact evaluation, given that no mortality impact can reasonably be expected if the program implementation is unsatisfactory. Process evaluation may be done by the program managers or by members of the program's administrative staff.
- **impact evaluation** asks the question: "what difference has the program made"? It puts emphasis on the changes brought about by the existence of the program, to estimate whether or not the intended effects have been produced. This activity requires comparing what has occurred, given the existence of the health program, with what would have occurred had the program not existed. In principle, impact evaluation seeks to assess the entire spectrum of changes related to the program, and not only those which were originally defined as program objectives, as it is quite likely that health programs will have outcomes other than those expected. Impact evaluation in the health field requires a strong background in epidemiology and demography, and is likely to conflict with the program managers' own interests, which is why evaluations are often contracted out to private researchers or universities.

Issues in impact evaluation

The two key concepts in impact evaluation are effectiveness and efficacy, in response to the two basic questions: can it work, does it work?

- **Efficacy** refers to the program effects in optimal conditions, in which all the target population is reached by the program, and in which the observance is perfect (can it work?).
- **Effectiveness** refers to the program effects in real-life conditions, in which part of the target population is not reached, and in which compliance with the program is more or less strict (does it work?). A high level of efficacy is necessary but not sufficient: a health program which has high efficacy but low efficiency is of little use.

One important point is that all impact evaluations are comparative: as explained by Rossi and Freeman (1993), determining impact requires comparing, with as much rigor as possible, targets who have experienced an intervention (“experimental”) with equivalent targets who have experienced something else (“control”), be it alternative treatments, or simply going untreated (placebo).

It is crucial that control targets be equivalent to experimental targets, in order to be able to establish the causal link between intervention and outcome: if they are not, then the difference between the first and second sets of targets may be due to other factors. The notion of internal validity of an evaluation study precisely reflects the degree to which external factors have been accounted for.

The relevance of mortality as an outcome measure of evaluation studies in the health field is questionable (Fournier, 1994). Morbidity data have to be considered in the first place, but in many instances a very convincing indication of the effectiveness of a health program is the occurrence of a mortality decline. It has also been suggested that programs should be assessed on the basis of their reduction of mortality differentials within a population, which in itself is a direct indicator of the broader goal of equity (Graham, 1989).

The feasibility of measuring the mortality impact depends mainly on two aspects of the problem (Ewbank, 1995; Ewbank and Gribble, 1993):

- the **magnitude of the expected mortality reduction**: it is possible to measure the effects of interventions which can produce substantial reductions in mortality, whereas if the expected reduction is too small the required sample size may be unattainable.
- the **pace of the expected mortality reduction**: if the program reduces mortality very slowly, it becomes very difficult to isolate the program effect from other long term factors affecting mortality.

The demonstration of an actual drop in death rates can be very difficult, due to lack of adequate data in many settings where evaluation is needed, and demographers play a crucial role in producing measurement instruments applicable for use within program budgets, and at the same time sensitive to short-term changes in survival. In this perspective, demography can contribute significantly to increased knowledge in the field of evaluation and be of considerable help to health planners in making sound policy decisions.

From clinical trials to large scale programs

The logical sequence in health program implementation may be briefly described as follows (Ewbank, 1995). First, the efficacy of the intervention is supported by a clinical trial which measures the impact on mortality and other indicators such as incidence of disease or serum antibody level, or on the basis of simple biological or

epidemiological models. Second, field trials allow the demonstration that the uptake of the recommended practices (in terms of delivery of health services and coverage of target population) is feasible, and measure the epidemiological effects on populations. There may be large differences between clinical trials and field trials. For instance, in clinical trials it is the clinic staff who generally prepare and administer the medicines to the patients, whereas in field trials these tasks are performed by the patients themselves without supervision. In addition, programs to prevent infectious diseases can reduce the number of active cases, and thus change the epidemiology of these diseases by changing the risk of infection for the entire population, including those going untreated.

It is worth noting that the effect of large-scale interventions which have already been evaluated in field trials still needs to be measured, as the benefits to populations can be very different from the benefits to individuals. Indeed, large scale programs rarely achieve the coverage rates found in well supervised field trials, due mainly to improper procedures, incomplete observance, or selective factors. For example, if those who receive services belong to the upper socioeconomic groups, and are therefore at lower risk of infection or complication, then the effects of large scale programs might be smaller than the effects of field trials. In particular, when interventions are implemented in government programs, the quality of services can be considerably affected, as can the coverage rates.

Relevance of different types of health programs

Disease-targeted interventions based on technologies are increasingly being questioned (Graham, 1989), and their relevance has been discussed in a recent IUSSP-sponsored workshop held in Brazil on the evaluation of the impact of health interventions (Rashad, Gray and Boerma, 1995). Experts argued that the risks of dying are unevenly distributed in large populations, and that individuals who die are a particular group in the population characterized by a higher than average exposure to health hazards. In addition, it has been pointed out that interventions centered around diseases are not able to modify the conditions which produced the ill-health in the first place.

Overall, the dominant opinion was that interventions should derive from comprehensive approaches with a full appreciation of the various dimensions (social, behavioral, economic,...) of the health problems, rather than mechanically targeted at specific diseases.

Seminar outline

The Committee for International Cooperation in National Research in Demography (CICRED) convened a meeting in Paris on February 26-28, 1996, to stimulate involvement of demographers in evaluation research, and help increase their role in the evaluation of health programs. About twenty centers were invited to participate in this meeting, and the representatives who attended belonged to one of three groups:

- experts in health program evaluation, put in charge of discussing the methodological issues attached to impact evaluation in the health field (data sources, mortality indicators (direct vs. indirect), study designs and methods for data collection, alternative indicators, etc);
- demographers and epidemiologists experienced in evaluation research, invited to present a project design and substantive findings;
- demographers involved in an ongoing project, invited to give a brief presentation of their protocol and objectives for discussion and advice.

This volume assembles the contributions of the first two groups of participants, divided into three sections: the first comprises the papers focused on methodology, the second includes various types of health program evaluation projects, grouped according to the data source (survey data, demographic surveillance systems, death registration data), and the last touches on a topic of more general interest, namely “who benefits from health programs?”, with detailed presentations of two health programs, one of which generates additional health differentials along social lines, and the other which leads to a reduction of the differential mortality between the sexes at early ages. The various presentations and the main points of the discussions are summarized below.

In the first section of the seminar, the specific role of evaluation was positioned in the wider context of health planning, and the role of demographers in health program evaluation was clearly delineated:

- provision of information on health indicators, to draw a baseline picture of the population and pinpoint needs and priorities;
- provision of information on the determinants of health and survival of individuals, to elaborate appropriate interventions;
- provision of appropriate measurement tools to monitor changes related to interventions.

The distinction between process evaluation and impact evaluation was stressed, and one important point which has been raised is that, although death is the ultimate outcome to be considered in evaluations, efforts should be made to also collect data on morbidity, physiological parameters, behavioral variables, and even quality of life to complete the picture.

The issue of methods of data collection in evaluation studies was tackled, and the point was made that moving away from the randomized trial gold standard was feasible, provided sufficient care was exercised in order not to over-interpret the studies. A very useful framework was proposed for the design and evaluation of child survival programs, according to the intended uses of the study. Underlying the discussion on this topic was the crucial problem of causal inference, and the idea that the stronger the inference needed, the heavier and the more costly the design, keeping in mind that an evaluation study should not cost more than the program it is intended to assess.

The methods of mortality measurement (direct vs. indirect) were examined, and the poor performance of indirect estimations in reflecting short-term changes was stressed. More generally, it was argued that indirect estimates do not seem to be appropriate in situations where mortality is changing rapidly. The findings presented illustrated very well the reversibility of the child mortality decline in a developing country, in response to political upheaval, or to worsening economic conditions. All progress may ultimately be called into question and reversals are likely to occur, which fully justifies the importance of long-term monitoring of mortality trends.

Another idea which was supported was that qualitative research, and more precisely in-depth open interviews are potentially very fruitful in the context of evaluation studies. Knowing the exact circumstances of a death, be it a maternal or a child death, can certainly help identify behavioral or sociological factors amenable to interventions, or help understand why a program has not functioned properly.

The question of the relevance of mortality as an outcome measure in evaluation studies was raised and safe motherhood programs were presented as one example where morbidity indicators and in particular "near-miss death" morbidity (a severe clinical condition preceding death) could be more appropriate to reflect changes.

As an alternative design, the case-control method is potentially promising in the field of evaluation. A number of selection biases are inherent to this retrospective type of study, but can be overcome provided sufficient information is collected. The main advantage is that this study design is quick and relatively cheap to implement, and, when properly analyzed, can be very informative. The main challenge now is to move these studies out of the hospital setting into the community, and especially, to use the death registers as a potential source of cases.

The Demographic and Health Surveys and the CDC-assisted reproductive health surveys are extremely useful in providing baseline data and a potential for international comparison. In addition, the availability of surveys at different points in time in the same countries can serve to assess behavioral changes and changes in health status, possibly in response to health programs. One point which was considered concerns the under-utilization of the DHS surveys in general, and the communication gap between the demographic and public health communities.

With regard to vaccination against measles, the study presented revealed a huge reduction in juvenile mortality (roughly 50 %) after this intervention. The data collection method may be questionable, however more controlled studies have in the past shown a reduction of about 20%, which in any case is greater than would be expected on the basis of deaths directly related to measles. This raises a number of questions concerning the repercussions of vaccination against measles on the incidence of other diseases, and about the inter-relations between measles and malnutrition.

The study in Madagascar is particularly instructive, and clearly illustrates the wealth of information which can be extracted at very low cost from the local health office registers in cases where death registration is almost complete. The processing of this information provides an excellent monitoring instrument in the capital city of Madagascar. Coupled with regular cross-sectional surveys at a national level, it can even generate a complete picture of mortality trends and differentials in the entire country, and reflect the effects on mortality of the national health programs.

The demographic surveillance system set up in Matlab, and, on a smaller scale, the population laboratories functioning in rural Senegal provide wonderful opportunities for evaluation studies, using either the Intervention/Control contrast, or the Before/After contrast. Regardless of the fact that ideal designs of this type are very costly, it has to be pointed out that when programs are expanded beyond limited populations, the quality of services tends to decline, which means that the outcomes of the evaluation should be treated with caution.

Although the demographic impact of health programs is generally expressed in terms of change (mortality change or fertility change) in the population as a whole, a concern with equity also leads to the examination of differentials within the populations. If certain socio-economic sections of the population benefit more than others from health programs, these may result in either a narrowing or a widening of differentials: while the former is certainly desirable, the latter indicates that the program is not reaching those who need it most and that definite action is needed to improve coverage.

Two papers illustrated the potential impact of health interventions on health differentials within the population according to social lines or to gender, with opposite effects depending on the context: a widening of socio-economic-based differences with a very modern health technology in a Western country (France), and a reduction in differentials between the sexes with a multi-purpose health program in a developing country (Bangladesh).

Proposals for evaluation of specific health programs were presented, and these concerned a wide variety of interventions, ranging from the introduction of running water to training of midwives and the setting up of social services for the elderly (see List of other contributed papers). The discussions which followed helped to identify the objectives, select the appropriate indicators, and set out the analysis.

As already mentioned, the objectives of the CICRED in convening this meeting were to contribute to a transfer of technical know-how and expertise between centers, and to provide a discussion forum in order to stimulate and initiate the development of collaborative projects. The usefulness of the meeting can only be evaluated over the next five years; in the meantime it is hoped that this volume will further these objectives and give them a wider audience.

The CICRED wishes to thank the United Nations Population Fund and the French Ministry of Cooperation, whose generous financial assistance made it possible to hold the seminar in excellent conditions, and to publish the proceedings both in English and in French. Dr. Philippe Collomb, Executive Director of CICRED, and Drs. André Quesnel and Jacques Véron, officers in charge of priority demographic orientations in CICRED, provided valuable advice during the planning stage of the seminar. The practical organization of the seminar was in the competent hands of Mrs Silvia Huix-Adamets and Nelly Puyraud from the CICRED, and the technical editing of the proceedings was carried out by Dr. Godfrey Rogers.

References

- Borus ME., Buntz CG. and Tash WR. (1984) *Evaluating the Impact of Health Programs: A Primer*. The MIT Press, Cambridge, Massachusetts.
- Ewbank DC., Gribble J. (eds.) (1993) *Effects of Health Programs on Child Mortality in Sub-Saharan Africa*. National Academy Press, Washington, D.C., pp. 5-10.
- Ewbank D. (1995) Evaluating Large Scale Health Interventions: Methodological Issues. pp. 277-295 in Rashad H, Gray R and Boerma T (eds.) *Evaluation of the Impact of Health interventions*. Derouaux Ordina Editions, Liège.
- Fournier C. (1994) *La place de la démographie dans l'évaluation des programmes de santé*. Bamako: Centre d'Etudes et de Recherche sur la Population pour le Développement, Working Paper.
- Graham W. (1989) Measuring the impact of health interventions on mortality in developing countries: Why bother? *Journal of Biosocial Science*, Suppl. 10: 69-78.
- Kubik A., Parkin DM., Khlát M. et al., (1990) Lack of Benefit from Semi-annual Screening for cancer of the lung: follow-up report of a randomized controlled trial on a population of high-risk males in Czechoslovakia. *Int. J. Cancer*, 45: 26-33.
- Rashad H., Gray R. and Boerma T. (1995) From clinical trials to large scale applications of health interventions: facts and issues of concern. pp. 1-10 in Rashad H., Gray R. and Boerma T. (eds.) *Evaluation of the impact of health interventions*. Derouaux Ordina Editions, Liège.
- Rossi PH. and Freeman HE. (1993) *Evaluation. A systematic approach*. Sage Publications, Newbury Park, London, New Delhi.