IV. USES OF MORTALITY DATA FOR EVALUATING THE SUCCESS OF SPECIFIC HEALTH AND DEVELOPMENT PROGRAMMES*

Douglas C. Ewbank**

The goal of health and development programmes is improvement of the health and well-being of a population, and one measure of their success is changes in mortality. Evaluation of programmes that affect mortality (by design or otherwise) should therefore take into account changes in mortality. An evaluation must include two components: measurement of the change; and a test for a causal relationship between the change and the programme. These components will draw on techniques from both demography and epidemiology. Demographers tend to study the overall level and pattern of mortality, while epidemiologists generally study case-fatality rates and mortality from specific diseases. Between these two disciplines lies the study of refined measures of mortality for selected groups that may have been affected by a specific programme.

Evaluation studies must be designed specifically for the purpose of testing for a causal relationship between a programme and any observed change in mortality. The test of causation should consider four factors. First, did the observed mortality change, whether absolute or in relation to existing trends, occur after the programme had begun and when its impact was expected? Secondly, was the mortality change observed for various causes of death consistent with the expected impact of the programme? Thirdly, was the mortality change observed only among, or more pronounced for, population subgroups served by the programme? Fourthly, were there other changes (e.g., economic) closely related to the mortality change and therefore likely alternative contributing causes? An evaluation must also be designed in such a way that programme impact can be compared with stated programme goals and with the impact of other programmes. This comparison may have to be based on different indices of mortality change than those used in the test of causation. Section A of this chapter discusses the role of national data for evaluation studies. Sections B and C discuss the specification of expected programme impact and measurement of actual impact relevant to establishing a link between a programme and subsequent events. Section D considers the special problems associated with assessing the health impact of general development programmes.

Section E reviews the advantages and disadvantages for different purposes of different study methodologies and some examples of evaluations are discussed in section F. Issues of sampling are discussed in the annex.

The discussion is limited to issues relevant to health and development planning in developing countries. Although comparisons of the costs and benefits of health interventions in developed countries are of increasing importance in a period of escalating costs and diminishing marginal benefits, the evaluation strategies in developed and developing countries are sufficiently different to require separate treatment.

A. ROLE OF NATIONAL MORTALITY DATA FOR EVALUATION

Few programme evaluations in developing countries have been based on national data even though their results are frequently used for national planning. There are several reasons for this situation. First, national data are rarely accurate enough or available at intervals that are short enough to catch the impact of a specific programme. Apart from a few countries in Asia and Latin America, mortality estimates in developing countries are generally based on data from periodic censuses or sample surveys. Secondly, it is difficult to design research on national data because of the unavailability of control areas. Although national programmes are frequently introduced gradually on a regional basis, the differences between areas and the problem of the frequency of data availability reduce the opportunities for controlling for exogenous changes. Thirdly, there are such a large number of changes that might affect mortality rates at the national level that it is often impossible to single out the effects of a given programme. Lastly, many projects have only a localized impact, either because of limited geographical scope of the programme or because of variations between areas in the severity of the health problems to which the programme is directed.

Although national mortality data are rarely useful for the evaluation of the impact of individual programmes, they play an important role in the planning process. First, national data provide a measure of the country’s overall progress in health. Secondly, the level of mortality and patterns of mortality by age, sex

---

*The original version of this chapter appeared as document IESAI/PIAC.17/3.
**Department of Population Sciences, Harvard School of Public Health, Cambridge, Massachusetts, United States of America.
and cause revealed by national data can provide important clues to the health problems that must be attacked by health programmes. For example, high mortality at ages 1-4 indicates the need to improve the nutritional status of children. Once this need has been established, evaluations of programmes in small areas can be used to determine which approaches (e.g., nutrition education, nutrition surveillance, oral rehydration to reduce the impact of diarrhoea or agricultural programmes) are most cost-effective. Thirdly, planners need national data in order to determine the implications of small studies for national policy formation. This information should include a comparison of the levels and patterns of mortality in the study area with similar measures for the entire country.

The following discussion focuses on the evaluation of specific programmes in small areas. In some studies, national data have been used as baseline data for studies in "typical" areas. In a few cases (e.g., the evaluation of the effect of eliminating malaria in Sri Lanka), it has been possible to base evaluations on national data.

B. Specification of Expected Impact

The first step in the evaluation of the mortality impact of a programme is to determine its likely nature. This determination includes specification of: (a) the subpopulations that should be most affected (e.g., age and sex groups, farmers or coal-miners); (b) the causes of death that should change most; (c) the expected size of the impact; (d) the causal mechanisms of the changes (e.g., a change in the case-fatality rate or change in incidence); and (e) the expected timing of the impact (e.g., periodic or constant, immediate or delayed, continued or transitory). It is then possible to design an evaluation strategy that focuses on those measures which are most sensitive to the hypothesized impact. Each of these components is discussed below.

Specification of Target Subpopulations

For some programmes it is possible to specify which age/sex groups will be most affected by the programme. A simple example of this is a programme of tetanus immunization of pregnant women. (Immunity to tetanus is transmitted to the foetus through the placenta.) Virtually the only mortality impact of such a programme will be on the neonatal mortality rate at ages 5-15 days because the tetanus mortality rate amongst women of reproductive ages is quite low. Similarly, it is frequently possible to specify the population subgroups for which the impact will be largest, as with water-supply programmes, which will affect only those using the new source, and occupational health programmes, which will affect only those in specific occupation groups.

Specification of Causes of Death Most Affected

Even the simplest programmes can affect several causes of death. For example, a measles vaccination programme can reduce mortality from pneumonia and malnutrition because a case of measles can seriously affect the overall health of a malnourished child. Therefore, the specification of the causes of death most likely to be affected by a programme must take into account the synergistic relationships among diseases. (Problems with the use of cause-of-death information in the evaluation of programme impact are discussed in section C.)

Specification of Likely Magnitudes of Effects

The best way to estimate the likely magnitude of the change in mortality is to examine the mortality rates for those causes of death which are most likely to be affected. In some programmes, only a few causes will be affected, in which case the expected mortality decline would be some reasonable fraction of deaths to those few causes. For example, tetanus immunization programmes are not likely to affect any other cause of death because tetanus has a high case-fatality rate. A tetanus immunization programme can eliminate virtually all tetanus deaths; therefore, the expected programme impact would be equal to the tetanus mortality rate. However, a midwife training programme might be expected to reduce mortality by only half as much. A rough estimate of the expected magnitude is generally needed for the calculation of required sample size (discussed in the annex).

Specification of Mechanisms for Change

Demonstration of a causal relationship between a programme and a change in mortality rates is strengthened by specification of the mechanisms through which the programme affects mortality. For example, if the introduction of a spraying campaign to combat malaria is accompanied by a reduction in mortality, it is unlikely that the campaign was responsible for the reduction unless there was also a reduction in the incidence (and subsequently the measured prevalence) of malaria.

Specification of Timing of Impact

Timing is crucial in the establishment of causal links. The effect should not precede the cause, although it need not follow immediately. Many types of programmes cannot be expected to have an immediate impact. For example, general nutritional education will not achieve its full impact on mortality immediately since it does not save those who are already severely malnourished. The mortality impact may, therefore, only become apparent after several months. Nutrition programmes directed towards the most severely malnourished (e.g., rehabilitation centres) will have a more immediate impact. In the case of occupational health programmes designed to reduce contact with carcinogenic materials, there may not be any noticeable effect for decades.

In addition, there are differences in the length of time for which noticeable mortality effects might be
observed. For example, a single round of measles immunization would only be expected to reduce measles cases for a few years, until the population of susceptible children was renewed by subsequent births. The impact on mortality might be observable in the immunized cohorts for a few more years because the prevention of measles in childhood may improve nutritional status and subsequent survival probabilities.

The impact of some programmes may vary by season. For example, immunization programmes against influenza and programmes to stabilize food supplies will have a large impact only during certain seasons. The mortality impact of these programmes can be studied, if data permit, using changes in the seasonal pattern of mortality rates.

### C. PROBLEMS IN MEASURING IMPACT

Once the potential impact of a programme has been specified, several central issues still must be resolved before the evaluation strategy can be devised. One is the choice of the appropriate indices of mortality change for measuring programme impact and for comparing the effects of different programmes. Many alternatives are available, including the crude death rate, the expectation of life at birth \((e_o^a)\) and the person-years of life saved. The second is definition of the study population. Although the selection of the study population must be related to the target population of the programme, this may not provide sufficient guidance for the researcher if flows of migrants into and out of the area and diffusion of information spread the benefits to other areas. The third is the use of cause-specific mortality rates involving the design of methods for collecting usable cause-of-death data at reasonable cost and the determination of the way in which it can best be used once collected.

### Indices for documentation of programme impact

The indices chosen to document the impact of the programme should focus on the expected programme impact in order to strengthen the case for causality. A carefully chosen index can also greatly reduce the sample size needed to demonstrate programme impact. The question of required sample size is discussed in the annex but an example is useful. In 1974, the infant mortality rate in the Matlab study area in Bangladesh was 138 per 1,000 live births and the neonatal mortality rate was 78, including about 20 deaths per 1,000 live births due to tetanus. Evaluation of the impact of a programme to eliminate neonatal tetanus would require a sample size of 1,946 live births in both the test and the control areas to be 80 per cent sure of finding a result that was significant at the 5 per cent level if the true decline in the neonatal rate was 20 per 1,000 (see equation (5) in the annex). Using the infant mortality rate, 3,444 live births in both the test and the control areas would be needed to establish such a 20-point drop. By focusing on the neonatal rate rather than the infant rate, one can thus reduce the required sample size for the study by a factor of almost two.

Using a more narrowly defined age group, from 5 to 15 days, during which 90 per cent of all neonatal tetanus deaths occur, a sample size of only about 700 live births would be required. Limiting the test of causation to reported deaths from neonatal tetanus reduces the required sample size even more. Researchers faced with a small sample size and rare events frequently opt for the rate for the larger age group in order to increase the number of observed events. However, this procedure reduces the chance of documenting an actual impact and increases the risk of falsely concluding that there was an impact.

### Indices for comparison of programmes

In the context of national policy-making, it is not sufficient to show that a particular programme has an effect on mortality; it is also necessary to make comparison among programmes so as to select those best suited for implementation. For a comparison of programmes that affect different age groups, it is often not possible to use the same indices that are optimal for the tests of causation. For example, a comparison of a programme to prevent neonatal tetanus with a tuberculosis treatment programme can be made using change in the crude death rate or the expectation of life at birth, but not with changes in the death rate at 5-15 days.

The choice of a summary measure for comparing programmes is often based on tradition or on the type of data available. However, this choice can affect the ranking of programmes. For example, Chen shows that a family planning programme can produce a large change in the crude death rate by reducing the number of high-risk infants and children but it is unlikely to have a substantial impact on life expectancy at birth. A similar comparison of the number of lives saved and the person-years of life saved shows that the latter figures gives more weight than the former to the saving of infant lives, a weighting which may not be consistent with social costs.

The weightings implicit in the different indices can be examined using life tables and stable population relationships. A reduction of the mortality rate at age \(x\), \(\mu_x\), by an amount \(k\) will change life expectancy at birth by:

\[
de_o = e^a \left[ \exp (k \mu_x) - 1 \right] l_x \]

(1)

where \(e^a\) is the expectation of life at age \(a\); and \(l_x\) is the number of life-table survivors at exact age \(a\).

Similarly, the effect of this change on the person-years of life saved per capita is

\[
d_{PYL}^{POP} = e^a \left[ \exp (k \mu_x) - 1 \right]
\]
where \( p_x \) is the proportion of the population between ages \( x - \frac{1}{2} \) and \( x + \frac{1}{2} \). In a stable population, \( p_x \) is determined by the life table and population growth rate. Therefore,

\[
d\left(\frac{P_{\text{YL}}}{\text{Pop}}\right) = e_x [\exp (k \mu_x) - 1] b \int_0^1 \exp (-rx) \, dx
\]

where \( b \) is the crude birth rate; and \( r \) is the stable growth rate. In all stable populations with positive growth rates, the use of equation (2) will give more weight to changes at the youngest ages than the use of equation (1) because of the \( \exp (-rx) \) term.

In developing countries, the choice of the appropriate index for comparing programmes is a matter of serious concern because the selection of health programmes strictly on the basis of cost-benefit considerations leads to a heavy emphasis on programmes for infants and young children and the relative neglect of adults. For most indices, the weight given to reductions of mortality at the youngest ages is greater in developing countries than in developed. This factor is most evident for such indices as the crude death rate and the person-years of life saved, which depend upon the age distribution. The high population growth rates common in developing countries lead to large relative weights being applied to improvements at the youngest ages. Even the change in the expectation of life at birth give larger relative weights to the youngest ages in developing countries than in developed countries because the ratios \( l_x/l_{x+1} \) decline more rapidly with age in high-mortality life tables than in low-mortality tables. One justification for giving relatively large weights to mortality changes at the youngest ages in developing countries is that at high levels of mortality, mortality declines at the youngest ages are often associated with improved nutritional status and the general health of children, with implications for general development and health in later adult years.

One flaw in the use of life expectancy at birth for measuring programme impact is the implicit assumption that those children whose lives are saved by the programme this year will also benefit from the programme throughout the remainder of their lives. Measures of programme impact are usually compared with the costs of the programme, and gains in life expectancy are not ideal for this purpose. One can utilize instead the person-years of life saved, using as the weights the values of \( e_x \), in the absence of the programme. This use assumes that if the programme ends, mortality rates will return to the old level. In this way, the measured impact of the current year's programme does not include the results of next year's expenditures.

For many programmes, the evaluation of the impact on mortality is only one aspect of the evaluation. In some cases, the impact is to be added to other measures of impact through a calculation of the monetary value of the benefits. In this case, special attention can be given to the saving of years of working (i.e., "productive") years of life. One measure of mortality impact that is compatible with this concept is working years of life saved, which is very similar to the person-years of life saved. The index estimates the number of expected years of life saved between ages 15 or 20 and age 65, a range taken as an approximation of the working years. It still gives a substantial weight to the saving of infant and child lives since survival to age 5 virtually ensures survival to age 20. However, the weight given to deaths prevented at the youngest ages is substantially less than the weight given in the calculation of person-years of life saved.

The concept of person-years of life saved can be modified to incorporate the economic concept of current value or social discounting, that is person-years of expected life saved have less value to the society today if the enjoyment of those years will occur in the future. For example, the saving of the lives of 1,000 infants will save some expected years of life at age 1 next year and at age 51 many years from now. The expected savings 50 years hence have less value to the society today because they will not be enjoyed for a long time. Therefore, less weight should be given to the expected years of life saved many years from now. This calculation would provide an index that is much more compatible with the economic concepts of the benefits of programmes.

The person-years of expected life saved is defined as

\[
P_{\text{YL}} = \sum e_x \exp (k \mu_x - 1)
\]

where \( k_0 \) is now the proportional reduction of mortality at age \( x \). The current value of person-years of life saved, \( PV_{\text{PYPYL}} \), is therefore defined as

\[
PV_{\text{PYPYL}} = \sum \epsilon_x \exp (k_0 \mu_x - 1)
\]

where \( \epsilon_x \) is the current value of the expected years of life for persons aged \( x \). Using a standard life-table notation,

\[
e_x = T_x/l_x = \sum_{y=x}^w n_{x+y}/l_x
\]

where \( w \) is the oldest age in the life table. Following this, one can define \( \epsilon_x \) as

\[
\epsilon_x = \sum_{y=x}^w n_{x+y} \exp [-r(y + \frac{n}{2} - x)]/l_x
\]

where \( r \) is the social rate of discount.

The current value of expected years of life saved has several advantages. First, it is more consistent with
the concepts that economists use for project evaluation. Secondly, it lessens the weight given to the saving of lives at the youngest ages, thus perhaps better reflecting the priorities of many societies. This concept can easily be extended to the calculations of the current value of person-years of working life saved.

A full cost-benefit analysis of health programmes would include the relative economic contributions of various age groups. With this approach, the person-years of life saved at each age are weighted by the difference between economic production and consumption at that age in addition to the discounting of future benefits to current values. For example, the saving of an infant life involves increased costs to society until the age at which a child’s economic production exceeds his consumption.

The Ghana Health Assessment team developed an index of programme impact that combines the reductions in morbidity with those of mortality. The index is the number of healthy days of life that could be saved by eliminating a given disease. The index combines the person-years of expected life saved and the person-years of illness from a given cause. For example, they have calculated that in Ghana, malaria is responsible for the largest number of healthy days of life lost; of these days, only 54 per cent are days of life lost. The second largest cause of healthy days of life lost is measles, for which 97 per cent of the loss is days of life lost. A comparison of malaria and measles in terms of lost person-years of expected life would reverse the order of importance of these diseases.

Since the impact of a health programme on a particular index may be affected by the age distribution of deaths or the population age distribution, it can be difficult to use the results of a study in one population for planning in a second population. For this reason, published studies should include the changes in the age-specific and, whenever possible, the age/cause-specific death rates so that the potential impact can be calculated for other populations. This information would also be helpful in the comparison of two programmes first evaluated using different indices.

**Definition of the study population**

The population to be studied for an evaluation of programme impact can be defined either in terms of the programme recipients or in terms of the population of a geographical region covered by the programme. In many instances, these two populations are quite similar, unless the population is subject to high rates of migration. One example is an immunization campaign in an urban slum with very high rates of in-migration and out-migration. Measuring the change in the mortality rate for the population living in the slum will certainly understate the impact of the programme. Some of the children whose lives were saved will have out-migrated and been replaced by children who were never immunized. Even if the analysis could remove the deaths and years of risk of children who entered the area after the immunization campaign or if the programme continued to immunize new migrants, the evaluation would still miss the impact on those who out-migrated.

The definition of the population to be studied also must depend upon the means for carrying out the study. In this regard, in-migrants and out-migrants are quite different. Although it might be possible to exclude the experience of in-migrants or to analyse them separately, it is very difficult to follow the progress of those who out-migrate. Because of the problem of following out-migrants, many evaluations miss some of the programme impact. If the understatement of programme impact is likely to be severe, it may be necessary to follow some or all of the out-migrants.

The inclusion of in-migrants in an evaluation area presents another serious problem: specification of the baseline rates for in-migrants. If the in-migrants come from different areas or belong to different socio-economic groups than the stable residents, their pre-programme mortality rates might be substantially different from those measured in a baseline survey. For example, the evaluation of an agricultural programme might be biased if baseline data are not available for the migrant labourers entering the area in order to benefit from the new opportunities. One way to deal with this problem is to use retrospective techniques for examining differences in pre-programme mortality rates for migrants and non-migrants. Although such techniques might not provide very accurate results, they may be better than any attempts to measure prospectively the rates of potential migrants.

**Advantages and disadvantages of cause-specific mortality rates**

In programmes that are designed to reduce the incidence or severity of a small number of diseases, the demonstration of a causal relationship between the programme input and mortality trends is strengthened if the target causes of death decline more than other causes. This approach can be very misleading, however, if there are significant synergistic relationships between the target diseases and other causes of death. This is certainly the case with childhood mortality in developing countries.

One approach to handling interactions between causes of death is to classify the deaths according to primary, associated and underlying causes. For example, a child death might be attributed to measles with malnutrition as an underlying cause. Although in theory this helps to solve the problem, it frequently proves to be difficult to apply. First, when causes must be determined from retrospective interviews with surviving relatives, it is difficult to establish even a primary cause of death. Secondly, the introduction of a chain of underlying causes and primary causes complicates the demonstration of causation. For example, a decline in the number of deaths due to influenza with malnutrition as an underlying cause could be produced by either a nutrition programme or the appearance of a less virulent strain of influenza virus. In the absence of
One advantage of examining cause-specific mortality rates is that it may be possible to demonstrate programme success even if the all-cause mortality rates show little or no increase. For example, since maternal mortality is only a part of the mortality among women aged 15-45 years, a reduction of the maternal mortality rate might not show up clearly in the overall mortality trend for this age group. Examination of causes of death could demonstrate that the programme actually had reduced maternal mortality.

The problem associated with cause-of-death information is that it is very difficult to collect. In developing countries, where few deaths occur in hospitals or under the care of a physician, cause-of-death information from registration data is very weak. The experience of the Cholera Research Laboratory with a prospective vital event survey using physicians or specially trained field-workers to carry out interviews of the family immediately after deaths has shown that useful cause-of-death information can be collected, at least for broad categories, but the cost is high and the detail limited.4

D. EVALUATION OF MORTALITY IMPACT OF DEVELOPMENT PROGRAMME

It is harder to evaluate the impact on mortality of development projects than that of health programmes, since they influence mortality only indirectly. It is difficult both to specify the target population, because the effects of a development project may diffuse well beyond the project area, and to predict the nature and timing of any impact on mortality rates. Well-known examples are irrigation programmes or dam projects, which may raise the incomes and nutritional status of the local population, while altering the ecology both inside and outside the project area in ways that lead to changes in established patterns of disease incidence and prevalence.

Moreover, the major mortality impact of some development projects, such as flood control, road-building, introduction of drought-resistant crops or provision of adequate food-storage facilities, may be the reduction of crisis mortality, that is, short, sharp increases in mortality associated with epidemics, war or natural disasters. The evaluation of such an impact raises special problems because such crises occur at irregular intervals, making it impossible to ensure that a potential crisis would occur during an evaluation study, and because, in general, there is no basis for estimating how severe a crisis would have been in the absence of the project. Most retrospective measurement procedures are of little use in the assessment of the severity of past crises, since they provide only averages of mortality experience over time; and even maternity histories, which appear to provide detailed child mortality sequences, tend to smooth out irregularities as a result of reporting errors. However, the reduction of crisis mortality will also reduce average mortality, so measures of average mortality before and after the institution of the programme will, if they cover sufficiently long periods, indicate the general effect of the programme, if not its specific impact. Similarly, a sequence of annual mortality rates for the periods before and after the institution of the programme may demonstrate its effect if the irregularities in the sequence prior to the programme are much more marked than those of the sequence after the programme.

E. SELECTION OF STUDY METHODOLOGY

There are numerous approaches to the collection of data for the study of mortality, each with its own advantages and disadvantages in terms of cost, precision, simplicity and ease of linkage with other records. These methods have been used in various combinations and in many versions. The following simple classification of the approaches is used here:

(a) Prospective:
(i) National registration system, possibly on a sample of localities;
(ii) Vital events registration and periodic censuses in a special study area for research purposes;
(b) Retrospective:
(i) Fertility or pregnancy histories;
(ii) Reports of deaths in the preceding year or series of years;
(iii) Brass type of approaches using data on survival of children, parents or spouse.

Each of these approaches can be used for the evaluation of programmes, but no one approach can be regarded as ideal for all circumstances. The following sections discuss the relative advantages and disadvantages of each approach with regard to the definition of the population; the detail of information concerning age, sex and cause of death; the age range covered; the completeness of recording or accuracy of the information; the cost and the required sample size; and efficiency in assessment of crisis mortality.

Prospective approaches

National (sample) vital events registration

Vital events registration involves the registration of demographic events at the time of or shortly following their occurrence. The usefulness of data on registered events depends upon prompt and complete tabulation, preferably with details on the age, sex and place of usual residence of the deceased, and on the cause of death. The calculation of mortality rates requires information on the population at risk, which must be based on a different data-collection system, usually a national census. Because of the difficulties in developing effective, complete systems for collecting and tabulating vital events data, several countries have
attempted to achieve complete registration in a national sample of geographical areas.

The first problem with the use of vital registration data is that few developing countries have systems that approach complete reporting of deaths. The usefulness of registered deaths is severely limited if fewer than approximately 60 per cent of the events are registered. A number of techniques have been developed for estimating the completeness of coverage of registered deaths by age by comparing the age distributions of deaths with the overall population age distribution. However, these techniques are generally ineffective for the adjustment of reported child deaths and for such populations as those of small areas that experience extensive migration. The problem of completeness is especially difficult since coverage often varies with age, sex, cause and area.

A second problem with vital statistics data is the accuracy of reporting. For example, the reporting of cause of death may be poor except for those deaths registered by trained medical personnel, failure to specify the usual area of residence may make it difficult to match reported deaths with the population at risk; and misreporting of characteristics can be a serious problem, particularly if the pattern of misreporting in the death records differs from that in the census data.

Two problems specifically hamper the use of vital registration data for studying programme impact. First, it is often difficult to get tabulations of registered deaths and base populations for areas covered by special research projects. Secondly, registration data provide little information on socio-economic variables and participation in a specific health programme, so that it is often not possible to study rates for those population subgroups most likely to be affected by a given programme.

Despite these problems, vital registration data can be useful for programme evaluation. First, registered deaths can often provide minimum estimates of rates for some age groups or causes and can indicate the presence of a disease as an important cause of death even without appropriate denominators. Secondly, registration data may provide the best basis for estimating the frequency of episodes of crisis mortality. Lastly, it may be possible in some cases to use local death registration files for verifying or correcting retrospectively reported dates of birth and death.

**Prospective vital events and population surveys**

In some cases, it is necessary to design a prospective study of mortality in the project area to record deaths through periodic surveys of households. When the surveys are frequent enough, this becomes a vital events registration system with active registrars. If the periodic surveys are carried out by a special survey team, this can be an expensive approach to programme evaluation. The intensive surveillance of the population of the Matlab study area cost about one dollar per person per year. Although this cost is far less expensive per interview than that of the World Fertility Survey, it does involve substantial resources. In some community health programmes, community workers collect information on vital events and household residents to ensure complete coverage of the population with health services. In this case, the research costs are only those of extra supervision, coding and tabulation. This approach was used quite successfully in an evaluation of a community health programme in Haiti.

The advantage of periodic surveys is that the data collected can be tailored to match the requirements of the evaluation. Information can be collected and tabulated on relevant socio-economic characteristics, programme participation, health practices and mortality by selected age groups. Data collection can also be linked with nutritional surveillance or with sample epidemiological or socio-economic surveys. Although reporting of cause of death is still a problem, the reporting of causes and symptoms can be adapted to maximize the usefulness of the data for programme evaluation. Another advantage of this approach is that it ensures comparability of numerators and denominators in terms of age, place of residence and other characteristics. The study population can also be defined so as to include or exclude various subgroups, such as temporary migrants.

Successful collection of vital events information depends upon the quality of supervision, the abilities of the interviewers and the co-operation of the population. If complete coverage is not achieved, it is likely that those persons missed in the events survey were also missed by the health programme, leading to a bias in the estimated effects on the health of the entire population.

A frequent objection to this approach is that the sample sizes required to evaluate health programmes are so large that the cost of population surveillance methods becomes prohibitive. The annex to this chapter discusses methods for estimating the sample size requirements based on the size of the expected impact. However, it is important to take note that the cost of a three-year or five-year prospective study of a population of 10,000 is generally small, compared with the amount that may be spent on the basis of the programme evaluation.

**Retrospective approaches**

**Maternity histories**

One problem with prospective reporting systems is that they often begin at the same time that programme services begin and therefore provide little or no baseline data. One solution to this problem is to collect maternity histories with dates of births and deaths (or age at death) for all children born to women in the population. The apparent simplicity of this approach is misleading. Maternity histories are subject to several
sources of error, including date and age misstatement and omission of children. These reporting errors can lead to serious distortions of levels and trends in both fertility and mortality rates. For example, one common pattern of misreporting leads to a false rise in fertility during the 5-14 years preceding the survey. To some extent, these errors can be reduced through careful interviewing, extensive supervision of the interviewers, matching of reported events with birth registration documents, etc. However, the data from maternity histories should always be approached with considerable scepticism.

In addition to the problems of reporting errors, maternity histories only provide useful information for the first 10-20 years of life and provide little useful data on cause of death. Although maternity histories ensure comparability of numerators and denominators, they do not always provide data for the appropriate population. The tabulated rates include the deaths of some children who died before their mothers migrated into the study area and exclude deaths of children whose mothers have since left the study area. Therefore, maternity histories do not provide exact historical data for an area that has experienced extensive migration. This problem can be lessened by comparing the maternity history for each woman with her migration history, but information on deaths of children of out-migrants is virtually impossible to collect.

Reports of recent deaths

Another approach to the collection of baseline data is to interview each household about deaths of family members during the previous one or two years. Data collected in this way frequently show substantial underreporting of deaths. The completeness of reporting can be studied using the same techniques as those used to analyse vital registration data, although the results are often unsatisfactory due to differences in coverage by age and the methods are not readily applicable to data for small areas affected by migration. Reports of recent deaths are also affected by recent migration and by the frequent dissolution of households following the death of the head of household.

In summary, retrospective reports of recent deaths cannot be regarded as a suitable basis for the evaluation of health projects and are unreliable for collection of baseline data. Ensuring adequate coverage is almost impossible, and the nature of the errors limits the usefulness of the standard adjustment procedures.

Information on survival of close relatives

Survey reports of the survival of close relatives provide indicators of the level of mortality and some indications of the trend in mortality if the age pattern of mortality is known and the trend has been long-term and steady. The most useful of these approaches is the Brass child-survival method, which involves asking women to report their total number of live births and the number of their children who have died. Unlike maternity histories, this approach does not rely upon dates of birth and death for each child, only upon the current age of the mother; the reference period is the mother's lifetime. Therefore, the Brass child-survival method is not affected by reference period errors or dating problems and is less sensitive to age-misreporting. However, it is sensitive to differential under-reporting of deceased children, a serious problem in many surveys. The cost of not asking about date of birth and age at death for each child is that the method provides little information on trends and age patterns of mortality. Similar methods involve questions about the survival status of parents and first spouse which provide information on adult mortality. These methods also provide little information about trends and age patterns of mortality.

In summary, reports of the survival of close relatives can provide useful baseline data, especially in situations in which mortality has remained relatively constant or changed steadily for the past 10-20 years.

F. EXAMPLES OF EVALUATION STUDIES

There are surprisingly few published evaluations of the mortality impact of programmes in developing countries, largely because few such studies have been carried out, although also because some completed studies have never been published in readily accessible forms. This section reviews some important published evaluations, indicating ways in which the issues described in earlier sections have been approached, although without describing the individual studies in detail.

The studies can be classified into three groups. The first group is studies of health interventions directed towards a particular cause of death, generally based on data from special study areas where data-collection systems have been devised to test the impact of a specific programme. The second group covers special studies for evaluating the impact of integrated or general health programmes; these studies are few in number, but they are important because of the current trend towards integration of health services. The third group includes some studies based on national data; few such studies have been made because few countries have data of the quality required for this kind of evaluation, and because few health and development programmes can be expected to have impacts that are large enough and sudden enough to be observable in national mortality rates. A fourth kind of study is unrepresented in the examples given below: studies of the mortality impact of economic development programmes. To the present author's knowledge, there are no good studies of this type, although a few are currently under way. Also unrepresented are attempts to measure the impact on mortality of general development using indicators such as per capita income; such attempts fall outside the scope of this paper.
Health interventions directed towards particular causes of death

The double-blind study of tetanus toxoid and cholera vaccine conducted by the Cholera Research Laboratory (now the International Centre for Diarrhoeal Disease Research) in Matlab Thana, Bangladesh, represents an example of the use of special study areas; the incidence of tetanus is highly age-specific and is best examined with this type of survey. The difference between the neonatal mortality rates for children born to women receiving the different vaccines (cholera and tetanus) is taken as an estimate of the amount of tetanus mortality. This estimate of neonatal mortality from tetanus is slightly less than the estimate from the reports on cause of death. Some of this difference is caused by false attribution of cause; some deaths attributed to diseases described using the same local names as tetanus occurred among children born to women who had received the tetanus vaccine. However, some of the differences may also be due to self-selection of the women who entered the vaccine trials.

An example of the usefulness of mortality data for two study areas at different periods comes from the study of mortality among children aged 1-4 years in the Hanover district of Jamaica, during the introduction of a nutrition programme. The programme was introduced into two study areas at different times; in both cases, the mortality rate at 1-4 years declined substantially immediately following the introduction of the programme but at different time periods. By staggering the beginning of the programme in the two areas, the second area serves as a control area for the first and vice versa.

Another study that made good use of highly age-specific mortality rates best provided by small-scale follow-up surveys evaluates a spraying campaign against malaria in Kenya. In their analysis, the authors examine mortality rates for infants by month of age. They did not expect any impact on mortality during the first months of life since new-borns are protected by temporary (passive) immunity to malaria provided by the mother. During the later part of the infant year, they expected that eliminating malaria might affect the mortality rates, because the risk from a first infection is more serious than an infection in someone with some acquired immunity. Their data on infant deaths in the study area and the control areas show that death rates were the same during the first three months of life (29 per 1,000 live births). However, at 3-11 months of age, the death rate in the control area (132) was much higher than the rate in the study population (66). The case for the impact of the malaria eradication programme is strengthened by the fact that mortality rates in the two areas only differed for the age group of infants where the programme impact was expected, although it should be noted that the accuracy of the data is open to some doubt.

The study of three villages in Guatemala demonstrates the advantages of mixing retrospective data with data from death registration to study the impact of health programmes. By combining the results of prospective recording of infant and child deaths with the retrospective studies carried out by the Institute of Nutrition of Central America and Panama (INCAP), the authors compared the mortality declines with the declines in each of the three villages during the years preceding the introduction of several health and nutrition programmes. The retrospective data and data from official reports aid the interpretation of the results because mortality had been declining at different rates in the three villages before the programme. However, the extra data complicate the analysis and the conclusion. For example, the authors note: "Although the gain was greatest in the treatment village, 36 per cent, this was no more than expected from the decline during the baseline period". Repeated statements of this type reflect both the difficulty of ascribing causation and the problem of combining prospective and retrospective data. Although careful evaluation of programme impact often must rely on retrospective baseline data, it is important to recognize that extensive efforts to document differences in past trends between test and control areas may lead to less satisfying, though possibly less misleading, conclusions. For example, one report summarizes the results of the Guatemala study by stating that "The numerous complications encountered prevented the investigators from reaching what they considered to be unambiguous conclusions about the project's impact on mortality trends". In this case, the complications also included relatively small numbers of prospectively recorded deaths due to the restricted sample size.

An example of what might be termed "evaluation by indirect attribution of cause of death" is the study of the impact of diarrhoea treatment centres in the Cholera Research Laboratory study areas. It was estimated that the case-fatality rate for patients in the centres was less than 1 per cent. This rate was compared with an estimate in 1980 that about 47 per cent would have died without treatment. This estimate was based on the assumption that all patients who had lost at least 10 per cent of their body weight in fluid loss would have died without treatment. The estimate was then adjusted for readmissions and for some competing risks. Although this approach can provide only a rough estimate of the number of lives saved by the programme, it is useful for estimating the magnitude of the expected effect of the programme on the overall mortality rates of the area; this estimate can then be compared with the actual trends in mortality.

Another study of cause-specific data from the Cholera Lab areas shows that there was a close relationship between distance to a treatment centre and mortality levels. Communities that were within four miles of a centre had annual mortality from diarrhoea of 100-150 per 100,000 population, while villages that were six miles from a centre had rates of 270 per 100,000. Areas that did not have any medical facilities had rates of 250-300. There was also a significant rela-
tionship between distance to a centre and utilization of a centre. This close association with distance to the centres increases the likelihood that the centres were responsible for a drop in mortality. It is interesting to note that the relationship between distance and the crude death rates was not significant; the effect of the programme was only apparent when the cause-specific rates were examined.

Lastly, a note of caution about the use of small-area studies is in order. If the results of such studies are to be used for national planning, it is necessary to try to determine how the programme results would differ if applied nationally as a result of differences in health status, age structure and other relevant differences between the study population and the population as a whole. Since the data on national mortality and health are generally quite limited, careful use must be made of whatever information is available.

**Evaluation of integrated health projects**

The tendency in health planning during the past decade has been towards health programmes that integrate maternal and child health, nutrition, malaria control, other health programmes and, frequently, family planning. Complete evaluations of this type of programme require two different approaches: (a) what might be termed an epidemiological approach, evaluating the impact of each programme component; and (b) a demographic approach, evaluating the overall impact.

Results from studies of the impact of programmes directed towards specific diseases have to be used with great care in estimating the overall impact of an integrated health programme. It is not possible to estimate the expected impact of an integrated programme by adding together the expected impacts taken from specific studies of each element of the integrated programme. Because of the complex interactions among diseases, the overall impact of an integrated programme may be less than the sum of their expected individual impacts. The overall impact of an integrated programme thus needs to be measured directly, and it is often very difficult to evaluate the contribution of individual elements of the programme.

The usefulness of the demographic approach in the evaluation of integrated health programmes is demonstrated by an evaluation of an integrated health programme in rural Haiti. That study demonstrates that a programme combining immunization, nutritional surveillance, targeted supplemental feeding, nutritional education, nutritional demonstration, oral rehydration, screening for tuberculosis, deworming and support for traditional birth attendants reduced mortality significantly over a five-year period. Although no attempt was made to measure the individual impact of these separate programme components, examination of trends in cause-specific mortality rates strengthened the contention that the mortality trends were a result of the programme efforts. The demographic approach is most useful for studying the impact of integrated programmes or for cases where several causes of death may be affected by the programme. The epidemiological approach is most important for testing the value of one kind of intervention (e.g., a vaccine) or for comparing two approaches to the same problem (such as malaria prophylaxis and insecticide spraying).

**Evaluation of studies based on national data**

It has already been noted that national data have not been widely used in the evaluation of health programmes in developing countries, generally because the available data are of insufficient quality, or provide insufficient detail, to support a thorough study. However, two examples of the use of national data are given below, one for a developing country with rather good registration data and one for a disadvantaged subgroup of the population of a developed country.

The first example is the evaluation of the malaria education campaign in Sri Lanka in the late 1940s. The number of deaths registered declined by 37,000 between 1946 and 1947, and the number of deaths from malaria declined by some 8,000. Malaria is a debilitating disease which might be expected to have important secondary effects on death rates from other diseases, so a full evaluation would require some attempt to assess such secondary effects. In a more detailed analysis of the Sri Lanka case, Gray uses additional information from sources other than death registration and attributed 23 per cent of the national post-war mortality decline to the control of malaria. Regression analysis was used to demonstrate that: (a) there was no correlation between the availability of health services and district mortality rates before the malaria campaign; (b) nutritional differences were not responsible for the mortality differences between malarious and non-malarious zones; and (c) the mortality declines were concentrated in the age groups that were most heavily affected by malaria. He thus was able to eliminate health services and nutritional differences as major factors behind mortality differences between malarious and non-malarious areas and to conclude that the apparent effect of eliminating malaria on mortality was real.

The second example is the study by Pool of the Maori in New Zealand. The crude death rate of the Maori declined by 45 per cent between 1945 and 1956, largely as a result of declines in tuberculosis deaths. In 1945, tuberculosis was the leading cause of death between the ages of 5 and 45, whereas by 1956, death rates from tuberculosis had declined to very low levels at all ages under 45. He attributes the decline to the introduction of mobile mass X-ray units, which were testing 10 per cent of the population annually by 1958.

**Summary and conclusions**

This chapter has considered the difficulties that beset the use of mortality data in the evaluation of health and development programmes. One difficulty is that mortality data alone may not be sufficient for a
thorough evaluation; such is clearly the case with development programmes, for which a much wider range of criteria would be required for evaluation. It is also the case even for specific health programmes, for which changes in morbidity not necessarily reflected in mortality figures would be an important element in a complete evaluation. A further problem is in the choice of summary measures to represent programme impact; different indices might give quite different pictures of the impact of a programme, both in absolute terms and in relation to the performance of other programmes. This problem is particularly acute when comparing the value of competing development projects; for such purposes, some attempt must be made to calculate the current social utility of a health programme.

Perhaps the most crucial problem affecting the evaluation of health programmes is that of establishing a causal link between the programme and subsequent mortality changes. Although no complete answer is available, solid support for a causal link in the case of highly specific programmes can be built up by examining changes in mortality by age and sex, by cause and by timing. For such purposes, detailed follow-up surveys covering the period of the programme are the only realistic study design. For more general health programmes, expected impacts at particular ages or on particular causes of death will be less clear-cut; and the use of cheaper survey procedures, although providing much less detail but preferably including the coverage of a control area, may be sufficient. National statistics on deaths rarely provide the basis of more than very impressionistic assessments, in part because of problems of data accuracy and in part because of a severe lack of detail.

Despite the large number of health projects implemented in recent years, careful evaluations are disappointingly thin. This lack undoubtedly arises in part from the fact that expenditure on evaluation is not directly productive and is therefore not accorded high priority. It is no coincidence that several of the most thorough evaluations grew out of small-scale surveys to test particular clinical procedures, where the evaluation was the object of the exercise rather than a sideline. Such surveys are obviously important, providing invaluable information about the potential efficiency of different clinical procedures, but for general monitoring of health projects, something less demanding and less expensive is required. Thus, every health project should include some attempt at assessment, but not every project should try to emulate the work of the Cholera Research Laboratory. The design of the evaluation should take into account the objectives of the project and its likely effects. It is hoped that the discussion in earlier sections may prove useful in the selection of an appropriate evaluation programme.

NOTES

4. Stan D'Souza, "Small-area intensive studies for understanding morbidity and mortality processes: two models from Bangladesh—the Matlab Project and the Companian Health Project, chap. XIV of the present volume.
5. For an excellent review, see Samuel H. Preston, "Use of direct and indirect techniques for estimating the completeness of death registration systems", chap. VIII in the present volume.
7. For a thorough account, see Manual X. Indirect Techniques for Demographic Estimation (United Nations publication, Sales No. E.83.XIII.2).
13. Ibid., p. 446.
Because of the expense and organization required to conduct prospective studies of programme impact, most such studies cover relatively small populations. Even for populations of only 10,000 or so, the sampling error associated with total deaths in a year will not be large. However, evaluations of programme impact require more detailed measures, specific by age, sex or causes, than simple changes in the actual number of deaths. For example, in a population of 10,000 with a crude birth rate of 40 and an infant mortality rate of 150, one would expect to find about 60 infant deaths in a year. However, because of the small number of births, the actual number of observed deaths would fluctuate from year to year in the range of from 45 to 75 infant deaths during most years even if the risk of dying stayed constant, so the number of observed deaths could drop from 70 to 50 without a change in the underlying risk. It is therefore necessary to calculate a confidence interval for the risk of dying.

For some mortality indices it is easy to apply standard statistical formulae to calculate sampling variances. For example, rates can be assumed to be distributed binomially and the appropriate tests used to test significance of observed changes in these rates. Keyfitz\(^a\) presents the following equation for the variance of the life expectancy at age \(x\), \(e_x^o\):

\[
\text{Var}(e_x^o) = \sum_{y=x}^{n} \left(\frac{l_x}{l_y} \right)^2 (n - a_x + e_y^o + n) \text{Var}(e_y^o)
\]

where \(l_x\) is the proportion surviving to age \(x\); \(n\) is the width of the age interval; \(a_x\) is the average age of death of those dying between \(y\) and \(y + n\) (usually estimated as \(n/2\)); and \(q_y\) is the proportion dying between exact ages \(x\) and \(x + n\) in the life table. The variance of \(e_x^o\) can be approximated as:

\[
\text{Var}(e_x^o) = n^2 e_{x}^o \left(M_x - e_{x}^o \right) / n \text{P}_x
\]

where \(M_x\) is the age-specific mortality rate for the age group from \(x\) to \(x + n\); and \(\text{P}_x\) is the population on which \(e_{x}^o\) is based. If the observed number of deaths in an age group is small (say, less than 5), then \(\text{Var}(e_x^o)\) can be estimated using the Poisson distribution for \(e_{x}^o\) rather than the binomial:

\[
\text{Var}(e_x^o) = n^2 e_{x}^o \left(M_x - e_{x}^o \right) / n \text{P}_x
\]

Table IV.A.1 presents the variances of the \(e_x^o\) values for the life table for males in Matlab in 1974, collected as part of the Cholera Lab studies. The population size has been scaled to a total of 100,000. As can be seen from equations (1) and (2) (or (1) and (3)), the variance of \(e_x^o\) is proportional to the reciprocal of the population size given the same age distribution and the same estimated life table, so a sample size of 10,000 in the Matlab example would give rise to a standard error of \(e_x^o\) of about 1.9 years, larger by a factor of \(\sqrt{10}\) than that for the 100,000 sample.

The formulae for the sampling variances can be used to determine the sample size needed to test the impact of a programme. In order to calculate the required sample size, it is necessary to specify: (a) the level of significance to be used in the statistical test (usually 95 per cent); (b) the expected change in the index of impact; and (c) the desired probability of finding a significant result assuming that the expected change occurs (the power of the test). If the estimate of the measure of impact can be assumed to be normally distributed around the true value, the required sample size is:

\[N = \left(Z_{\alpha} + Z_{\beta}\right)^2 2\sigma^2 / \delta^2\]

where \(Z_{\alpha}\) is the value of the standard normal deviate corresponding to the selected significance level; \(Z_{\beta}\) is the value of the standard normal deviate corresponding to the desired power; \(\sigma^2\) is the population variance of the measure; and \(\delta\) is the postulated change in the measure. The resulting \(N\) is the sample size needed in both the test and the control; therefore, the total sample size is \(2N\).

This formula can be used to choose the sample size for an evaluation of a programme that is expected to increase the value of \(e_x^o\). The evaluation should be based on a one-tailed test because one would only be interested in the programme if it increased life expectancy. An appropriate value of \(Z_{\alpha}\) is 1.645, which corresponds to a 95 per cent significance level with a one-tailed test. \(Z_{\beta}\) can be taken as 0.841, which corresponds to an 80 per cent chance (\(\beta\)) of finding a significant result if the life expectancy really does change by \(\delta\) years. The choice of \(\beta\) is somewhat arbitrary since, unlike \(\alpha\), there are no firmly established conventions. The value of \(\alpha^2\) is equal to \(Var(e_x^o)\) times the total population size on which the life table is based. It is useful to note that an \(\alpha\) of 0.90 and a \(\beta\) of 0.90 give an \(N\) only 6 per cent larger than an \(\alpha\) of 0.95 and a \(\beta\) of 0.80.

Table IV.A.2 presents the sample sizes needed to test the impact of programmes using the Matlab life table given in table IV.A.1 and a variety of different assumed impacts. The resulting sample sizes are in terms of person-years of observation. For example, the 445,000 corresponding to a \(\beta\) of 0.8 and a change in the life expectancy of one year can be one year of observation for a population of 445,000 in the test area or two years of observation for a test population of 222,500.

If the programme evaluation is to be based on changes in a proportion, e.g., the infant mortality rate or an age-specific mortality rate, a different formula is needed to determine the required sample size. In this case the required sample size can be calculated using an

| Table IV.A.1 | Standard Errors of Estimated Life Expectancy Values from a Population of 100,000 with Life-Table and Age Distribution of Males, Matlab, Bangladesh, 1974 |
|---|---|---|
| Age | Life expectancy at age | Standard error of \(e_x^o\) |
| 0 | 53.4 | 0.60 |
| 1 | 59.5 | 0.62 |
| 5 | 59.8 | 0.55 |
| 10 | 56.7 | 0.53 |
| 20 | 46.7 | 0.53 |
| 30 | 37.5 | 0.51 |
| 40 | 28.9 | 0.48 |
| 50 | 21.0 | 0.45 |
| 60 | 14.2 | 0.39 |

| Table IV.A.2 | Sample Sizes for Test Area Required for Various Levels of Change in Life Expectancy at Birth with a Significance Level, \(\alpha\), of 95 Per Cent and a Power, \(\beta\), of 80 Per Cent |
|---|---|---|
| Change in life expectancy at birth, \(e_0^b\) (years) | Required sample size |
| 0.5 | 1781000 |
| 1.0 | 445000 |
| 2.0 | 111000 |
| 3.0 | 49000 |
| 4.0 | 28000 |
| 5.0 | 18000 |

\(^a\)Total required sample for test and control is twice this value. Calculations based on Var \(e_0^b\) = 0.60, taken from Table IV.A.1, rounded to the nearest thousand.
equation derived by Miettinen. This equation allows for different sample sizes in the test and control areas. The ratio of \( N_0 \) and \( N_1 \), the sample sizes in the control and test areas, is set at \( R \). With a postulated change in the rate from \( P_0 \) to \( P_1 \), and given values of \( R \), \( Z_a \) and \( Z_b \), the required sample size is:

\[
N_0 = \frac{Z_a^2 K + Z_b^2(A + B)^2 + 2 Z_a Z_b (A + B) \sqrt{K}}{(P_1 - P_0)^2(A + B)}
\]

where \( K = (A + B) (RA - B) - R(P_1 - P_0)^2 \),
\( A = P_0 (1 - P_0) + P_1 (1 - P_1) \); and
\( B = (R - 1) P_0 (1 - P_0) \).

Table IV.A.3 gives some examples of the sample sizes required for various levels of impact. The baseline values are similar to the actual rates for the Matlab study area in 1974. This table demonstrates the importance of the expected impact of the programme. To be 80 per cent sure of detecting a significant programme impact if the real impact is a 20-point reduction in a rate of 140 per 1,000, one needs a sample size of about 3,500 in the test area. However, with a real change of 30 points the sample can be limited to 1,500. Table IV.A.3 also makes clear the importance of the level of the rate. Detecting a change of 5 points per 1,000 is easier when the change is from 25 to 20 than when it is from 80 to 75.

In some instances, it can be cost-effective to use a larger sample size in the control than in the test area even though the minimum total sample size is achieved when the two are equal. For example, if existing vital statistics records are to be used for the baseline (i.e., the control), the cost of data collection in the test area might be much higher, even if only in terms of the length of time before sufficient person-years of risk are observed. With an estimated change from a rate of 140 to 120, a sample size of 3,500 is required in the test area if equal sample sizes are used. If the control area has a sample size twice the test area, then the sample in the test area can be 2,600.

Although the total sample size increases from 7,000 with equal samples to 7,800 with unequal, the costs of the programme or the differences in the costs of data collection might justify this increase. On the other hand, if there is little time for the collection of baseline data, it might be necessary to set the test sample at twice the control.

More complicated study designs, such as a comparison of two study areas with each other and with a control, require more elaborate formulae for the calculation of required sample sizes. Cohen provides a useful set of tables and numerous examples for use in studies based on correlations, chi-square comparisons, F tests and regression analysis.

<table>
<thead>
<tr>
<th>Base rate per 1,000</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Unit for sample size)</td>
<td>IMR (^b) = 140 (live births)</td>
<td>58 650</td>
<td>10 550</td>
<td>3 500</td>
<td>1 500</td>
</tr>
<tr>
<td></td>
<td>Neonatal = 80 (live births)</td>
<td>35 350</td>
<td>6 550</td>
<td>2 000</td>
<td>800</td>
</tr>
<tr>
<td></td>
<td>6-23 months (^c) = 40 (survivors to 6 months)</td>
<td>17 850</td>
<td>4 150</td>
<td>850</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>1-4 years = 25 (person-years)</td>
<td>10 850</td>
<td>2 350</td>
<td>400</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^a\)Sample sizes rounded to nearest 50.
\(^b\)Infant mortality rate.
\(^c\)Cohort mortality between ages 6 months and 2 years, per 1,000 survivors to 6 months.