

*This report contains the collective views of an international group of experts and does not necessarily represent the decisions or the stated policy of the World Health Organization.*

WORLD HEALTH ORGANIZATION  
TECHNICAL REPORT SERIES

No. 510

**STATISTICAL PRINCIPLES IN  
PUBLIC HEALTH FIELD STUDIES**

**Fifteenth Report of the WHO Expert Committee  
on Health Statistics**

WORLD HEALTH ORGANIZATION

GENEVA

1972

© World Health Organization 1972

Publications of the World Health Organization enjoy copyright protection in accordance with the provisions of Protocol 2 of the Universal Copyright Convention. For rights of reproduction or translation of WHO publications in part or *in toto*, application should be made to the Office of Publications and Translation, World Health Organization, Geneva, Switzerland. The World Health Organization welcomes such applications.

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Director-General of the World Health Organization concerning the legal status of any country or territory or of its authorities, or concerning the delimitation of its frontiers.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

PRINTED IN SWITZERLAND

## CONTENTS

	Page
1. Introduction . . . . .	5
2. General principles of research design for field studies . . . . .	6
2.1 Statement of objectives . . . . .	6
2.2 Measurement of disease frequency . . . . .	6
2.3 Identification of modifying factors . . . . .	7
2.4 Operational definitions . . . . .	8
2.5 Choice of research method . . . . .	8
2.6 Selection of the study population . . . . .	9
2.7 Estimating the scale of a study . . . . .	9
2.8 Development of methods of measurement . . . . .	9
2.9 Pilot studies . . . . .	10
2.10 Monitoring research . . . . .	10
2.11 Administration . . . . .	10
3. Developments in sampling and measurement . . . . .	11
3.1 Recent advances in sampling techniques . . . . .	11
3.2 Problems in complex sampling . . . . .	11
3.3 Sampling in developing countries . . . . .	12
3.4 Measurement errors in sample surveys . . . . .	12
4. The experimental approach in field studies . . . . .	14
4.1 Field experiments involving randomization . . . . .	15
4.2 Selection of a research design . . . . .	16
4.3 Quasi-random or systematic methods of allocation . . . . .	18
5. Non-experimental studies on etiology or intervention . . . . .	18
5.1 Types of study . . . . .	19
5.2 Comparability . . . . .	20
5.3 Representativeness . . . . .	22
5.4 Efficiency . . . . .	22
5.5 Causal inference . . . . .	23
6. The application of field studies in public health services . . . . .	23
6.1 Community diagnosis . . . . .	24
6.2 Preliminary research on public health action . . . . .	25
6.3 Evaluation of public health services . . . . .	26
7. Data handling and analysis . . . . .	27
7.1 Data systems . . . . .	27
7.2 Editing . . . . .	27
7.3 Preliminary analysis . . . . .	28
7.4 Data reduction and analysis . . . . .	28
8. Recommendations . . . . .	30

## WHO EXPERT COMMITTEE ON HEALTH STATISTICS

Geneva, 21-27 March 1972

### *Members : \**

Dr W. G. Cochran, Professor of Statistics, Department of Statistics, Harvard University, Cambridge, Mass., USA (*Chairman*)

Dr K. El Din Shawky, Professor of Preventive and Social Medicine, Faculty of Medicine, Ein Shams University, Cairo, Egypt

Dr M. Masuyama, Professor of Statistical Mathematics, Science University of Tokyo, Japan

Dr D. D. Reid, Professor of Epidemiology, Director, Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine, London, England (*Rapporteur*)

Dr Z. Roth, Head, Statistical Section, Institute of Hygiene and Epidemiology, Prague, Czechoslovakia

Professor D. Schwartz, Director, Statistical Research Unit of the National Institute of Health and Medical Research, Villejuif, France (*Vice-Chairman*)

### *Secretariat :*

Dr W. P. D. Logan, Director, Division of Health Statistics, WHO, Geneva, Switzerland

Dr R. B. McHugh, Professor and Director of Biometry, School of Public Health, Health Science Center, University of Minnesota, Minneapolis, Minn., USA (*Temporary Adviser*)

Dr O. S. Miettinen, Associate Professor of Epidemiology and Biostatistics, Department of Epidemiology, School of Public Health, Harvard University, Boston, Mass., USA (*Temporary Adviser*)

Mr K. Uemura, Chief Statistician, Health Statistical Methodology, Division of Health Statistics, WHO, Geneva, Switzerland (*Secretary*)

Dr A. Weber, Statistician, Regional Office for Europe, WHO, Copenhagen, Denmark

---

\* Unable to attend: Dr N. Marchevsky, Director, Centre for Biostatistics and Demography, Faculty of Medicine, University of Buenos Aires, Argentina.

# STATISTICAL PRINCIPLES IN PUBLIC HEALTH FIELD STUDIES

## Fifteenth Report of the WHO Expert Committee on Health Statistics

The WHO Expert Committee on Health Statistics met in Geneva from 21 to 27 March 1972 to discuss the role of statistical principles in the design and analysis of public health field studies. Dr P. Dorolle, Deputy Director-General, opened the meeting on behalf of the Director-General.

### 1. INTRODUCTION

In this report, "public health field studies" means research projects conducted in the community, usually outside the hospital, and directed towards objectives of public health importance. These studies may be broadly classified as follows:

(1) *Descriptive*. Here the aim is to give an account of the frequency of a disease and its related factors in the population or to outline its natural history or evolution. Such field studies may also be used in medical administration to enumerate health resources and to measure the demands made upon them.

(2) *Explanatory*. In these studies, an understanding of a particular public health problem—e.g., the causes of chronic disease or the dynamics of acute infections—is sought.

(3) *Operational*—i.e., concerned with the need for public health initiative, deciding between alternative courses of action, and appraising the functioning of health services.

(4) *Technical*. The immediate aim of these studies is to develop or test diagnostic methods or other techniques of measurement—e.g., of social or cultural conditions relevant to health—that can be used effectively in field enquiries.

The development in recent years of field studies as a reliable method of research has shown that scientific rigour need not be confined to the laboratory or even the hospital ward. Moreover, the principles and

methods of research design and data analysis involved should be widely applicable in public health practice in both developed and developing countries.

## 2. GENERAL PRINCIPLES OF RESEARCH DESIGN FOR FIELD STUDIES

Proposals for research in public health often have both originality of approach and obvious relevance to practical problems; but they may lack clarity of concept and scientific rigour in their design. The general principles that underlie good research design are broadly statistical in nature and statistical techniques are essential to analyse the data collected. At the planning stage of a public health study, however, the main needs are for a clear definition of objectives, an outline of the strategy to be adopted, and a justification for the choice of a study population and of the clinical and other methods to be used.

### 2.1 Statement of objectives

The objectives of a piece of research in public health can best be expressed in the form of the precise question or questions to which an answer is sought and the operational terms in which the relevant variables are to be defined. In practice it is helpful, in making the crucial comparison that seeks to resolve the major point at issue in the investigation, to outline the final tabulation and methods of analysis of results. Although there are certain to be subsidiary tables, required for answering other, less important questions, the design of the study must ensure that the primary question is answered as clearly and decisively as possible.

### 2.2 Measures of disease frequency

The special focus of interest in respect of the personal, social, or other results of the disease must be reflected in the choice of the measure of its frequency used in the study. Here, the operational terms are clinical and statistical rather than demographic. If concern is with loss of life alone, death is the endpoint of main interest. The possibility, particularly in international studies, of bias in reporting the underlying cause of death makes it essential to include the "all causes" or total death rate as well as the rate for the cause in which one is particularly interested. In either case, it is essential to define the cause in standard terms, such as those of the International Statistical Classification of Diseases, Injuries, and Causes of Death. Less serious forms of diseases can be defined in either clinical or functional terms—e.g., by using standard questionnaires on symptoms or

physical tests. The impact of sickness may also be measured in social terms—e.g., the frequency of absence from work or of prolonged disability attributed to a specific disease.

The choice of the clinical endpoint and the etiological or administrative nature of the study will determine the statistical expression of the frequency of disease. Thus, the effect of disease on mortality may be measured by the death rate, and its effect on morbidity either by prevalence, as determined by questionnaire and/or functional test, or incidence, in terms of the number of new cases of disease. All these rates will be expressed as ratios to the defined population exposed to risk. Studies of etiology and health intervention often require the incidence of disease to be measured in terms of disability or death; the need for medical care might be ascertained more easily by measuring the prevalence of disease for which treatment has not yet been given. Whatever the choice, its justification and operational definition are basic.

### **2.3 Identification of modifying factors**

Consideration of the ultimate outline tabulation will usually prompt further questions relevant to the planning of research. These concern factors that may affect the presentation of the results in that table and the way in which they are measured and taken into account in the analysis. The identification of such factors is usually based on prior clinical or other experience or on the preliminary analysis of routinely collected vital statistics. For example, clinical experience might point to the dominant role of smoking in the genesis of chronic obstructive lung disease, whereas an urban excess in the rate of recorded deaths from the disease points to the possible influence of air pollution.

These variables, together with others, such as age and sex, must therefore be reviewed to assess their relative importance and the need for including them in the study, and to see how they might be measured. Because of the limits imposed by the costs of investigation and the willingness of the population to be investigated, it is often inevitable to establish priorities for inclusion. Thus, in the above-mentioned example, age, sex, and smoking habits are crucial and can usually be readily ascertained and recorded. Other factors, such as the social circumstances of the individuals studied, are relevant but less easily and precisely assessed, so that their inclusion is less important. Yet others—e.g., a family history of respiratory disease—may be of potential interest from an etiological point of view but irrelevant to the major etiological comparison envisaged in the study design. If it were to burden respondents unduly and jeopardize the prospect of high response, their inclusion would not be justified by the incidental interest that they might provide.

#### 2.4 Operational definitions

Once selected for inclusion, all factors must be defined in the research plan in operational terms. Thus, smoking habits might be assessed in terms of the number of cigarettes of a filter or non-filter type smoked daily; local levels of air pollution, in terms of the concentration of smoke ( $\mu\text{g}/\text{m}^3$ ) at street level; and social and economic class according to some classification of occupations or levels of education.

#### 2.5 Choice of research method

The choice of method in field studies will depend on which of the main objectives described in the introduction is being pursued. In the field of public health administration, for example, descriptive studies usually aim at precise estimates of disease frequency in samples that are representative of the whole population served by a particular medical service. On the other hand, in explanatory or etiological studies, emphasis is placed on the detection of major differences in the disease experience of populations that differ in their degree of exposure to the disease in question. But in both types of enquiry field trials of an experimental rather than observational nature may also be used to test etiological hypotheses or compare the relative value of alternative methods of providing medical care.

In descriptive studies, the aim is usually to obtain reasonably precise estimates of disease prevalence in a whole population rather than to look for informative contrasts. Representativeness, and the sampling methods used to ensure it, are here of vital importance. The procedures to be employed must therefore be clearly stated in the research plan. These sampling methods have already been discussed in the Committee's tenth report (WHO Expert Committee on Health Statistics, 1966), but recent developments in this field are reviewed in section 3.

In research on causation, only experimental studies can give answers that can be interpreted with confidence. Subjects drawn from the same basic population are allocated at random either to an "experimental" or to a "control" group. Only the former is exposed to the prophylactic, therapeutic, or administrative procedure being tested, but both groups are followed up with equal assiduity. Equality in this and in the application of the criteria of measurement—e.g., of disability due to disease, or of social functioning—is essential.

When random allocation of individuals is impracticable, non-experimental evaluation of the causal connexion can be effective.

The choice of approach in etiological enquiries depends largely on the current state of knowledge. Initially, vital statistics and other sources can provide suggestive but not conclusive evidence. The first stage of a field study may be the relatively rapid and cheap reconnaissance afforded by

a case-control study, perhaps in hospital patients. Should these preliminary findings support the initial suggestion, a more ambitious long-term cohort study may be planned. The important practical point is that the cost and effort involved in public health studies impose a cautious progressive approach in the process of investigation.

## **2.6 Selection of the study population**

Whatever the method chosen, the study population must be appropriate to it and to the ultimate objective of the study. In the descriptive type of study, representativeness of the target population is essential, as stressed in the tenth report of the Committee (WHO Expert Committee on Health Statistics, 1966). In etiological studies, where the emphasis is on contrast rather than on representativeness, restricted population groups may be both useful and convenient. For example, children in school are an accessible population for studies of infectious diseases. Groups sharing the same occupation and social status but differing widely in their exposure to, say, air pollution in different parts of the country are useful in the study of chronic diseases. On the other hand, they may be less representative of the total population in some respects. They are thus inappropriate for studies of medical care problems since it is not the occupied section of the population but those persons who are unable to follow an occupation who represent the greatest burden on the medical services of the community. In the final selection of a population for study, other considerations—e.g., the need for cluster sampling to minimize travel time for clinical teams—enter into the decision.

## **2.7 Estimating the scale of a study**

There is no simple way to determine the optimum number of subjects that must be included in a study in order to ensure a proper balance between costs and returns in terms of useful information. As before, one must concentrate on the ultimate tabulation. The study population must be large enough to yield sufficiently precise estimates of disease frequency and offer an acceptable chance of detecting epidemiological differences of practical importance between population groups. The minimum number of individuals that will enable such a degree of precision to be achieved can be calculated with reasonable accuracy if the expected incidence rates can be derived, e.g., from routine morbidity statistics.

## **2.8 Development of methods of measurement**

In the Committee's eleventh report (WHO Expert Committee on Health Statistics, 1967) attention was drawn to the importance of the appropriateness

of measures of disease and of consistency in standards among observers in epidemiological studies involving comparisons. The methods discussed in section 3 may be of value in developing such measures for operational use.

### **2.9 Pilot studies**

Where a small representative sample of the target population is subjected to all the procedures envisaged in the full-scale investigation, a pilot study is essential for developing a research plan. Such a preliminary study tests in the field the acceptability and practicability of the methods of sampling, of enlisting cooperation, and of clinical assessment that it is proposed to use in the study population. It also gives an opportunity for training field staff in these procedures. Moreover, the data collected, may be helpful in taking the ultimate decision on the optimum size of the full-scale study.

### **2.10 Monitoring research**

Consistency in standards of observation is crucial to any comparative study. In large-scale or continued surveys, these standards must be monitored so that deviations can be detected and remedied. For example, in a study involving the measurement of blood pressure or the serum cholesterol level, unusual fluctuations in standards may be revealed by plotting on a control chart the mean levels found each week. In order to find out a trend, the use of a cumulative sum (or cusum-) control chart is recommended.

Monitoring may also be needed for ethical reasons. In a field trial—e.g., of specially equipped ambulances in the care of patients who have acute myocardial infarction while at home—major differences in death rates between contrasted groups may appear. In such circumstances, it would be desirable to study the results and apply appropriate sequential tests to detect any serious added hazard in a particular group. Committees responsible for the ethical approval of research proposals would certainly require specific precautions to be described in the research plan.

### **2.11 Administration**

The submission of detailed research plans is likely to involve important administrative matters, such as the structure or membership of the guiding committee; the number and qualifications of field staff and the availability of specialist advice; the definition of clinical responsibility, particularly for the work of technical personnel; and access to computer facilities. A chronology of the stages of preparation, of the pilot study, and of the execution of the full-scale investigation will show whether there is a reasonable prospect that the study will be completed in a reasonable time. These

considerations should be taken into account in making detailed cost estimates under separate headings, such as staff salaries and equipment.

### 3. DEVELOPMENTS IN SAMPLING AND MEASUREMENT

Sampling, a basic technique in public health studies, was the subject of the Committee's tenth report (WHO Expert Committee on Health Statistics, 1966). That report pointed out the usefulness of sampling in evaluating the health status of a population group, investigating factors affecting health, studying the administration of health services and the effectiveness of health measures, controlling standards of hygiene, and processing statistical information. The report went on to discuss the planning of sample surveys in public health; the technical aspects of sample design, including the handling of non-sampling errors; and the analysis, evaluation, and reporting of the results. It concluded with a recommendation that WHO should undertake the production of a manual on sampling methods in public health investigations. This manual is expected to be completed soon.

#### 3.1 Recent advances in sampling techniques

Since the tenth report of the Expert Committee on Health Statistics was published (1966), most of the improvements in sampling have concerned techniques that are already widely used but are known to have deficiencies—for example: (a) in the estimation of ratios, the use of adjusted ratios and of new methods—less subject to bias—for estimating their variances; (b) the introduction of more convenient methods of sample design and estimation in multi-stage sampling with unequal probabilities of selection; and (c) further work on the stratification introduced after selection of the sample. These methods apply mainly to cross-sectional surveys. New developments in sampling at repeated intervals in longitudinal surveys have been limited. Some work has been done on the determination of the optimum length of time during which respondents remain in a sample; on replacement policy; on methods of estimating the current level, and on time-changes in the level and longer-term average levels. Techniques for revising a sample when selection probabilities or sampling fractions within strata need to be up-dated have also been discussed.

#### 3.2 Problems in complex sampling

Great progress has been made in complex sampling, in which the study of relations between variables is one objective and the analysis may involve non linear indices, such as multiple or partial regression and correlation

coefficients, and ratios of rates. In the past, there has been a lack of mathematical techniques for calculating standard errors and confidence intervals for such indices, since theory based on a simple random sample does not apply. Two general methods of remedying this deficiency are in process of development: the ingenious use of Taylor series expansions (Tepping, 1968), and comparisons of the measure in question when calculated from sub-samples drawn from the sample. Examples of the latter are: (a) balanced repeated replications, primarily for samples with two units per stratum, in which a carefully chosen set of half-samples with one unit per stratum is used (McCarthy, 1969), and (b) the method sometimes called the "jack-knife" (by analogy with the multipurpose knife with many blades). With this method, the calculation is made by successively dropping independent groups of the same size from the sample. Large sample theory is available for a broad class of non-linear statistics; it also indicates the types of statistics—e.g., medians and order statistics—for which the method is inappropriate. Studies on the performance of these methods in sample surveys have given encouraging results so far (Frankel, 1971). These methods are already in use in the health field and may become important as non-linear statistics, such as probability distributions and life-table functions, come to be estimated more frequently from health surveys.

### 3.3 Sampling in developing countries

It is often said that, in countries where almost the entire population is literate and ample technical resources exist, the main problems in the planning and conduct of a sample survey have been solved. This is not so in many developing countries, where the handicaps include the absence of a complete and convenient frame or list of the units for sample selection; difficulties of travel and communications; and the problem of devising questionnaires and interviewing techniques that respondents will accept and can understand. Such problems are unlikely to be solved by research in technologically advanced countries: their solution demands accurate knowledge of the culture and of planning and training facilities in the country in question. An example of useful research is a review (Brass, 1971) of the relative usefulness in developing countries of 5 methods that have been proposed for measuring population change in order to appraise the need for a fertility control programme. More studies of this kind, related to health information, with pooled experience from different developing countries, would be helpful to public health statisticians.

### 3.4 Measurement errors in sample surveys

The problem of measurement errors has come to attract increasing attention from both the theoretical and the practical points of view.

#### 3.4.1. *Theoretical development*

Mathematical research has dealt with the effects of measurement errors of different sizes and types on statistical estimates—for instance, the effects of errors of misclassification (false positive and false negative results) on the estimation of rates, on differences between rates, and on relative risks, and the effects of quantitative measurement errors in the variables on multiple and partial regression coefficients (Cochran, 1968). The consequences of measurement errors can be serious, as they affect the interpretation of the statistics, particularly in screening surveys. Further work has been done on a theoretical model that subdivides the mean square errors of estimates into components due to sampling variance, response variance, and the square of the systematic bias. In turn, the response variance is subdivided into the contributions from different sources—respondents, interviewers, coders, and supervisors. The objective of this model is to draw attention to the most important contributors to the mean square error as a guide to the allocation of resources in reducing measurement errors (Hansen & Waksberg, 1970).

#### 3.4.2 *Practical applications*

The next step is the difficult one of trying to estimate from field studies how large specific measurement errors actually are for different important estimates made from surveys. An obvious method is to find an alternative accurate measuring process and use it as a standard for appraising the routine measurements in a small study; but such accurate measurements are often lacking. Something can also be learned by repeated application of the routine process according to an experimental plan, followed by a "components of variance" analysis. However, a thorough investigation of possible approaches awaits future research.

These two steps—mathematical analysis and estimation of the sizes of measurement errors in practice—may tell how important specific measurement errors are in a survey. The next step—techniques to cope with these errors—is also proving slow and expensive. In an interview survey of families, for instance, measurement errors may be affected by the questionnaire, the choice of respondent, the respondent's answer to the question, his motivation to respond carefully, his memory, and the role and effect of the interviewer. The following examples illustrate the kinds of research that have been conducted.

In family expenditure studies where outside check data were available, it has been possible by randomized experiments to study the effects of the following: choice of family respondent; periods during which the respondent was asked to recall the events; reporting periods; repeated interviewing; and different degrees of probing. Telescoping—the shifting of an expenditure from one reporting period to another—was found to be a serious source

of errors, but ways of reducing this were devised. In regard to health expenditure, a short questionnaire was found to be accurate enough, but limited probing for certain specific items was advisable. Some experience indicates that probing by a detailed questionnaire in surveys is not necessarily more effective than limited questions, so that each case must be examined on its merits.

Some health surveys have included studies on the effects of the respondents' motivation and attitude to the survey, and of the behaviour of the interviewer. These indicated, to the surprise of the investigators, that the attitudes of respondents and interviewers to the survey did not appreciably affect the quality of the data supplied. However, the amount of "interactive behaviour" was important—namely the extent to which respondent and interviewer showed interest in one another, for instance by asking questions about each other's personal circumstances, family, and friends; by jokes; and by conversation after the interview. Further research will be required to discover whether methods effective in one country are effective also in other countries.

When a question concerns sensitive topics such as sexual behaviour and the respondent is asked if he has taken a specific action A or possesses a characteristic C, he may refuse to answer or deliberately answer incorrectly. An ingenious technique currently being explored is called the "randomized response" method. By a randomizing device, either the statement "I have taken action A" or the statement "I have not taken action A" is put to the respondent, who is requested to say whether it is true or false. The interviewer does not know in any individual case which question has been asked, but can control the proportion P of cases in which the first question is asked. From a knowledge of P and of the proportion of affirmative answers (if truthful) an unbiased estimate of the proportion of persons who have taken this action can be obtained (Greenberg et al., 1971).

In summary, rapid progress is being made in the theory and practice of sampling. Of particular importance in the design and conduct of sampling surveys is the increasing emphasis placed on the identification of major sources of error in clinical and other methods of measurement in the field. Some of the problems associated with response in sensitive areas of enquiry are being vigorously tackled. More attention should be devoted to assessing and developing techniques of sampling and measurement in developing countries.

#### **4. THE EXPERIMENTAL APPROACH IN FIELD STUDIES**

In this section, experimental studies are defined as studies in which the allocation of subjects to particular exposures is under the control of the investigator. Such studies are of two types: those involving a process of random allocation and those in which allocation is predetermined but

may be considered to be quasi-random. In the latter case, departure from strict randomization is made for practical reasons but in such a manner that it does not seriously affect the theoretical basis of the conclusions.

#### 4.1 Field experiments involving randomization

The principles of trials in which allocation is strictly at random are well known in the context of clinical trials of different modes of therapy. With a single-factor intervention, such as treatment with only one drug, the subjects are arranged into clinically homogeneous groups or "blocks" of subjects thought likely to respond to treatment in a similar manner as far as possible. Within such blocks, patients are allocated at random either to treatment by the drug (when this is ethically and practically appropriate) or to the administration of a placebo. The numbers required to detect significant differences will depend on the variability of individual responses and the degree of precision desired.

When combinations of more than one factor are being tried at the same time, more complex factorial designs can be used. If, for example, 2 factors  $F_1$  and  $F_2$  are involved, 4 different treatment combinations are formed: neither  $F_1$  nor  $F_2$ ;  $F_1$  only;  $F_2$  only; and both  $F_1$  and  $F_2$ . For example,  $F_1$  might represent an antihypertensive drug (spironolactone) and  $F_2$  a diuretic used in the treatment of malignant/hypertension. The data may be presented as follows:

Diuretic ( $F_2$ )	An antihypertensive drug ( $F_1$ )	
	No	Yes
No		
Yes		

The advantages of this method are that (1) the average effect of each factor is estimated with the same precision as if the experiment had been devoted to this factor alone; (2) it is possible to examine whether the effect of the antihypertensive drug differs according to the diuretic, and *vice versa*. The same principle may be applied in examining whether a single factor has the same effect on different types of patient (e.g., an anticoagulant for the secondary prevention of heart disease) as follows:

Age of patients	Anticoagulant	
	No	Yes
Under 50 years		
50 years and over		

A handicap to the widespread use of factorial designs is that, with  $m$  factors at 2 levels replicated  $r$  times, the number of subjects,  $2^m r$ , is often much too large. Ingenious devices enable the number of subjects to be reduced while still providing estimates of the effects of individual factors. These plans can be developed from arrangements known as orthogonal arrays associated with linear graphs (Masuyama, to be published).

The basic principles of experimental design are as applicable to public health field trials as they are to clinical trials, although there may be a greater number of practical problems in their organization and execution and the number of treatment combinations must often be severely restricted.

It may be necessary to apply a treatment to groups of subjects instead of to individual subjects. For instance, when testing the spraying of houses as an antimalarial device, 10 villages selected at random out of 20 may be sprayed, the remainder being left unsprayed. Similarly, in investigating the effects of milk on the growth and health of children, free milk may be offered to children in some schools but not in others. Pairing (or blocking if there are more than 2 treatments) may be used in such trials if it is thought that homogeneous pairs (or blocks) of villages or schools can be constructed; each member of the pair (or block) will then be allocated to a different treatment at random. Although in such cases the basic measurements may still be made on individuals (e.g., village inhabitants or school-children), it should be noted that the basic unit of analysis for calculating tests of significance, standard errors, and confidence limits is the group of subjects (village or school) to which a treatment was applied. Thus, in the examples cited, the data for analysis would be the malaria case rates in the 10 sprayed and 10 unsprayed villages, and the average growth rates of children in schools with and without free milk. To treat villagers or school-children as independent observations in the analysis may lead to a serious underestimation of standard errors.

So far, most experiments in public health research have been of a relatively simple type. However, the increasing use of factorial designs in appropriate circumstances is to be recommended.

#### 4.2 Selection of a research design

As mentioned in section 2, the selection of a research design depends in part on its objectives. If the aim is to verify or refute some hypothesis about a relatively simple biological relationship, such as a defined clinical response to a particular drug, the experimental plan follows the classical lines. Where the object is to test the effect on the public health of a complex programme in order to establish its feasibility and utility in practice, the trial becomes "pragmatic" rather than "explanatory" in nature.

In an "explanatory" trial the aim is to acquire general scientific knowledge at the biological level. The treatments to be compared are

defined clearly and strictly—for example, the use of a chemically defined drug and a placebo. The criteria utilized for the comparison are few, limited if possible to one alone, reflecting a definite biological response. On the other hand, a certain freedom may be possible in the selection of subjects if it is thought that the biological property to be studied can be extrapolated from one population to another. As far as possible, however, recourse will be had to a homogeneous population. The method of analysis usually consists in testing the null hypothesis and in estimating the difference between the groups compared.

The purpose of a “pragmatic” trial, on the other hand, is to study the effects of a single treatment or to decide between various curative or preventive treatments used in current practice, possibly in a more complex situation—e.g., where treatment in hospital is being compared with treatment at home.

The judgement criteria are usually multiple; thus they may include the biological action, side effects, and cost in the broad sense of the word, so that an exhaustive summing-up can be made. The choice of subjects is of major importance, because direct extrapolation to a target population is aimed at and because it may not be possible to extrapolate the conclusion drawn from the trial (A is better than B) if the participants are not a representative sample of subjects to whom the treatment is to be applied in the future. Finally, since a decision is involved, the method of analysis attempts to utilize procedures taken from cost-benefit analysis and decision theory, in addition to hypothesis testing and estimation.

Questions appropriate to these two approaches might be:

- (1) Is a particular drug efficacious in lowering blood cholesterol levels?
- (2) Is a programme in which this drug is used in conjunction with antihypertensive drugs and propaganda against smoking capable of reducing the incidence of ischaemic heart disease?

Although this dichotomy is seldom clear-cut, the distinction is useful in that it brings out some of the problems associated with the large-scale trials of preventive methods that include more than a single active element (Schwartz et al., 1970). The following considerations are relevant:

(a) It is seldom possible to establish which of the reductions in risk factors involved (smoking, cholesterol, or blood pressure) is responsible for the change in disease incidence. On the other hand, a negative result can be held to imply that no single factor has affected the incidence of the disease, so that it is unlikely to be worth investigating each separately.

(b) In a study in which many factors are involved in an unsystematic manner, differences in response to the lowering of one factor such as cholesterol in relation to simultaneous changes in, say, blood pressure

cannot be estimated with confidence. However, if only simultaneous reductions in all risk variables produce changes in disease incidence, a single-factor trial could not detect this potentially important fact.

(c) Single-factor trials may indicate the influence of that factor acting in isolation, but such a situation may not arise very frequently in practice.

(d) Although "multi-factor" trials are more realistic, there is no guarantee that similar results will be obtained in the circumstances of another population, in which the balance of risk factors may be very different.

At this stage of the development of research methods in these complex conditions, no precise advice can be given on the choice of an optimum study design. More needs to be done to assess the theoretical problems involved in the conduct of field studies of the value of the joint action of two or more public health measures when used concurrently and the possible role of factorial experiments.

#### **4.3 Quasi-random or systematic methods of allocation**

What may be termed "quasi-random experiments" may be of special value in public health research. These involve some systematic method of allocation that is administratively convenient and yet apparently does not lead to any consistent difference between the two series being compared. If, in a study of the use of a special ambulance for the emergency care of cardiac patients, this ambulance is available only on alternate days, allocation can be made on the basis of the day on which the call for the ambulance is made. In such circumstances, although the comparability of the two groups would first be checked as far as possible, the conventional tests of significance could usually be applied with reasonable confidence.

### **5. NON-EXPERIMENTAL STUDIES ON ETIOLOGY OR INTERVENTION**

In public health research, experimentation in the study of etiology or in evaluating intervention is often either impossible or very costly. In such situations, a non-experimental approximation to a randomized trial may be attempted. The principles of this technique are discussed here.

The characteristic of these non-experimental studies is that it is not possible—for ethical, administrative, or other reasons—to expose the subjects randomly to the potential etiological factor or intervention measure. An example is the study of the effects of smoking on human beings. Such research must then be based simply on observations in a natural setting. This

can entail various types of study, some of them rather far removed from the experimental ideal; it also necessitates the pursuit of validity by special techniques of subject selection and data analysis. Although this type of field research can be conducted more cheaply than experimental studies, the validity of causal inference remains largely a matter of extra-statistical judgement.

### 5.1 Types of study

The closest non-experimental counterpart of a randomized trial involves the selection of subjects in different categories of exposure and their follow-up to determine outcome. In non-experimental research it is also possible to use the reverse approach, selecting subjects in different categories of outcome and enquiring about their previous experiences with respect to exposure or intervention. In either type of study, there are two or more series to be compared.

The above-mentioned approaches to the selection of subjects concern "cohort" and "case-control" studies, respectively. In both cohort and case-control studies it may be possible to elicit histories of exposure (e.g., to a potential toxic hazard in industry) over the whole period before the clinical onset of the disease—i.e., the observation is *longitudinal*. When information about both exposure and outcome refers to the same point in time, as when current exposure to that hazard is taken to indicate the relative degree of exposure in the past, the observations are *cross-sectional* rather than longitudinal. In *retrospective* studies, both exposure and outcome have occurred before the start of the study; in *prospective* studies the exposure is known at or before the start, but the outcome is established by follow-up. When the follow-up period is long and appropriate data are available from existing records, a retrospective longitudinal cohort study offers a great advantage over the prospective approach in terms of cost and speed of execution. For example, the long-term effects of stress in terms of coronary heart disease may be evaluated in retrospect by identifying, from military records, men who have been exposed or not exposed to captivity, and by assessing subsequent manifestations of heart disease by reference to routinely maintained personal medical records. When the disease in question is rare (e.g., leukaemia in adolescents) but the exposure under study (e.g., prenatal X-ray exposure) can be determined either from existing records or by history-taking, then a case-control study may be desirable on the grounds of practicality.

Often the data are collected on an *ad hoc* basis, but increasing use is being made of disease registers and surveillance systems (e.g., for drug evaluation) as sources of data for the non-experimental evaluation of etiology and intervention.

## 5.2 Comparability

The validity of the estimation of the effect at issue depends upon two elements: the degree of comparability of the series—say, cases and “controls”—and the representativeness of the study subjects and setting in relation to the types of subject and setting to which the results are to be extrapolated.

Both the observations and the relevant characteristics of subjects should be comparable. In experiments this is achieved mainly by randomization (possibly with blocking) and “blind” assessment, respectively. In non-experimental research, comparability of subjects is sought by deliberate control of certain confounding factors and by trying to secure comparability of selection. Comparability of observations cannot usually be attained by blind assessment but has to be based simply on the standardization of observation procedures.

### 5.2.1 Comparability of subjects

(a) *Control of confounding factors.* Confounding factors are associated with both the exposure and the outcome criterion. More specifically, a confounding factor is one that, although associated with the “exposure” under investigation, is itself, independently of any such association, a “risk-factor” for the disease. For example, in the study of the role of alcohol in the etiology of oesophageal cancer, smoking is a confounding factor if it meets the following criteria: it is associated with the consumption of alcohol; and it must be associated, within alcohol consumption categories, with the risk of oesophageal cancer. In these conditions, the effects of alcohol consumption can be determined only if the influence of smoking is taken into account. Unless appropriately controlled, the “effects” of such factors remain inseparable from those of the exposure under study.

In the control of confounding factors one or more of the following techniques may be used:

*at the stage of subject selection:*

- (i) restriction of the type and source of study subjects;
- (ii) matching in the selection of the control series;

*at the analysis stage:*

- (iii) analysis by stratification, with pooling of comparative information over all strata, or with the computation of standardized measures of risk;
- (iv) multivariate analysis.

These techniques are elaborated below.

The use of a restricted, relatively homogeneous series of subjects is a well-established principle in experimentation. In non-experimental studies

this principle is also commonly applied, mainly by limiting the age range and by excluding minor categories, such as minority ethnic groups.

Matching is the procedure of selecting the control series in such a way that for certain factors it will have the same distribution as the population. Thus, in a case-control study, the control series may be chosen so as to have the same age-sex distribution as the cases. Matching has the advantage of clarity but is often an expensive procedure. In case-control studies it also involves the disadvantages that there are limits to the evaluation of matching factors as possible risk factors for the outcome and that efficiency (information per subject) tends to be reduced by matching. On the other hand, matching may afford the only effective way to control some important but ill-defined and complex factors. A common example of this in public health field studies is the use of neighbourhood controls.

When a confounding factor is measured on a simple scale—e.g., age or sex—and its distribution in the compared series is expected to be reasonably similar, its control may be deferred to the analysis stage of the study. The classical approach to the control of confounding factors in the analysis involves stratification of the subjects by the factor(s) to be controlled. Elementary estimates are obtained within the strata, such as age-sex categories, and compared between the series. These are often summarized over the strata into indices such as standardized rates. This method of control has the advantage of being inexpensive and yet informative. However, when several factors are to be controlled simultaneously, this cross-classification approach becomes complicated and inefficient. In these situations multivariate analysis (section 7.4.2) is often a useful substitute for the more classical method.

(b) *Subject selection.* The need for comparability in selection as well as in the control of confounding factors may create problems in non-experimental research, particularly in case-control studies. For example, owing to selective recognition, venous thrombosis may be reported more often among women who use oral contraceptives than among those who do not, since women using this mode of contraception may be under closer medical observation for the detection of thrombosis and other ailments, or may be drawn from a more intelligent and better informed group of women who readily note and report such complications.

Non-response is common source of selection bias because it tends to relate to the disease pattern in cohort studies and to exposure or intervention status in case-control studies. Every effort must be made to minimize non-response.

### 5.2.2 *Comparability of observations*

Comparability of observations is just as necessary in non-experimental studies as it is in experiments, but can be much more difficult to attain. In

cohort studies the subjects usually are not "blind" with respect to their exposure or intervention status, and this can bias the ascertainment of outcome when dealing with criteria of outcome whose assessment involves subjective elements. This problem is generally accentuated in case-control studies, where awareness of the outcome may substantially influence the memory of the subject with regard to his history of exposure or intervention. In both types of study, bias resulting from investigative procedures, particularly those used in interviewing, should be eliminated as far as possible by applying carefully standardized methods.

### 5.3 Representativeness

As was indicated in section 4 in connexion with experimental studies, it is pertinent in planning a non-experimental study to distinguish between general scientific objectives and the pursuit of information more directly and specifically relevant to the planning of health programmes in a given administrative region. When the emphasis is on the former, it is usually desirable to study individuals of a restricted and carefully specified type with relatively little concern for their *source*. But a study directed mainly towards local decision-making in health programmes would cover individuals of different types and a well defined, reasonably representative target population. Strict representativeness, based on appropriate sampling procedures, is necessary also in field studies of the effectiveness of public health programmes (section 6).

### 5.4 Efficiency

As already noted, comparability and representativeness ensure that the results of a study are valid. A study must also be efficient—i.e., yield the maximum of information in relation to its cost. Efficiency is affected mainly by (a) the basic approach; (b) the types of individual to be studied; (c) their source; (d) the means of checking the validity of the results; (e) the relative sizes of the groups to be compared; and (f) the methods of observation.

(a) The different approaches in non-experimental field research vary greatly in efficiency, but it is usually a simple matter to choose between them, since feasibility is the basic consideration.

(b) As to the individuals to be studied, efficiency is best served by selecting these from groups particularly exposed to the disease in question, where the effect is expected to be large and subjects are both readily available and easily studied. When there is interest in the relationship of the effect to a factor such as age, an efficient design tends to be one in which the extremes of the range of interest are included; often it is also desirable to add another group somewhere in the middle of that range.

(c) In the specification of the source of the defined types of subjects the predominant consideration should be practicality, except when the study is aimed at serving decision-making in a particular community (*vide supra*).

(d) With regard to the means of controlling the confounding factors, efficiency is usually best served by avoiding matching and simply recording the information on potential confounding factors. Their confounding effect can then be assessed in the analysis, and control exercised by stratification or, if need be, by multivariate analysis.

(e) The determination of the relative sizes of the compared groups involves an important, often easily exploitable, principle of optimizing efficiency: the sizes of the groups are made proportional to  $\sigma/\sqrt{c}$ , where  $\sigma$  is the within-group standard deviation and  $c$  is the cost per subject.

(f) The decision what observations to make is guided not only by what is essential for the ultimate and crucial tabulation and the accuracy of measurement, but also by the cost of such measurement. In public health field studies it is often efficient to collect simplified information even if this requires the use of larger series to compensate for the relatively low degree of accuracy in measurement.

### 5.5 Causal inference

Causal inference in non-experimental research is essentially extra-statistical and judgemental in nature, requiring knowledge of the biomedical issues involved. Usually the question of whether or not major confounding factors were left uncontrolled is especially relevant, but selection and observation biases also deserve careful review.

Despite difficulties in their interpretation as well as in their design and conduct, non-experimental field studies have a definite place in the evaluation of etiology or intervention. The validity of the information that they yield is intermediate between the validity of mere informal experience and that of highly formal experimental trials.

## 6. THE APPLICATION OF FIELD STUDIES IN PUBLIC HEALTH SERVICES

The methods of field investigation outlined above can be readily applied to solve some of the problems involved in the development of health services—a process consisting of three phases :

(1) *Community diagnosis*. In this phase, the frequency of disease and the health resources in the community are ascertained, with emphasis on the extent to which the needs for prevention or cure are met.

(2) *Research.* Here, new methods of disease control or health care are rigorously assessed to determine their efficacy in producing the specific prophylactic or curative action required.

(3) *Evaluation of services.* The last phase consists in special studies of the effectiveness with which health programmes based on preliminary research are being applied in practice in the community and of the results achieved. It also entails striking a balance between the benefits accruing to the community as a result of disease control or better care and the costs in terms of resources or incidental hazards to health.

There have been attempts in recent years to construct mathematical models to solve problems relating to the organization of health services and to the choice of optimum programmes for specific disease control activities. The satisfactory construction of such models, which are based on operational research techniques, seems to be heavily dependent on public health field studies for providing the necessary basic data. This is an area requiring further investigation.

Field studies are particularly important during the later phases of public health service development and evaluation, but they also contribute much to the earlier stages of planning and research.

### 6.1 Community diagnosis

Sample surveys may be of special value in estimating the needs for medical care and the health resources in the community currently available to meet them. Useful as vital statistics and other routine returns may be, surveys carried out on representative samples have, as emphasized in the Committee's tenth report (WHO Expert Committee on Health Statistics, 1966) a potential that has not been sufficiently exploited.

Sample surveys are of value not only for determining the frequency of disease but also for eliciting the opinions of the populations to whom new public health measures are to be applied. In all such surveys, the development of measurement techniques, as discussed in section 3, is fundamental. However, whereas the use of standardized procedures of clinical assessment is now widely accepted as essential in surveys of disease frequency, standardization of measurement of opinions that can decide public acceptance of new public health programmes has been less satisfactory. In areas where psychological sensitivities may exist—e.g., family planning—the newer techniques, described in section 3, of obtaining frank responses while maintaining confidentiality merit development. Another measurement that is urgently required is that of social conditions or cultural background, particularly in countries where no occupational classification based on census returns is readily available. Not only are such measurements required for etiological studies of the relation of disease to environment but public health planning requires knowledge of the social and cultural

groups that require medical care most but that are the least enterprising in seeking it.

Although universal literacy is an aid to the conduct of sample surveys, it is not essential to their use. Similarly, current developments in simple and robust, but adequate, procedures of clinical interrogation and assessment will lead to an economical use of professional staff. Thus the use of sample surveys in countries where there is little awareness of the extent of disabling and debilitating conditions, such as chronic cervicitis and cervical erosion in women, can lead to a productive reorientation in health service planning.

## 6.2 Preliminary research on public health action

In the preliminary assessment of methods of disease prevention, the past history of immunization has underlined the need for the rigorous use of carefully controlled field trials (Pollock, 1966). The same caution must be applied to the preliminary testing of newer public health measures, such as changing or supplementing the diet. The principles enunciated in section 4 can also be applied here. In a simple case—e.g., where the question arises whether it is advisable to treat persons suffering from mild hypertension found by screening—the model of the clinical trial in which individuals thus discerned are allocated on a random basis can be followed.

In the context of other public health measures, however, randomization of individuals may not be practicable—e.g., when methods of vector control have to be applied to whole areas if they are to have any reasonable chance of affecting the spread of the disease. In field studies of the use of molluscicides in the attempted control of schistosomiasis, the method of allocating units such as villages in a random fashion to treated and untreated areas should be used, bearing in mind the cautions given in section 4 about the analysis of results.

Where randomization, on either an individual or a unit basis, is impracticable, the quasi-random experimental methods cited in section 4 can be employed. In trials of diet in relation to thromboembolic complications of surgical operations among hospital patients, for example, wards or hospitals might be supplied with different diets during successive periods planned so that the series are comparable. In before-and-after studies, as when the effects of fluoridation of water supplies are being assessed, concurrent surveys in matched control towns are necessary.

The controlled trial approach to assessment could be much more widely used than at present in solving problems of decision in medical care and public health. A recent example of its use is the comparison of case-fatality rates in patients suffering from myocardial infarction and randomly allocated to care at home or in specially equipped hospitals. Where at all practicable, the fully randomized experiment is the ideal. However, where this is not possible, adequate substitutes in trial design can often be found.

### 6.3 Evaluation of public health services

#### 6.3.1 *Monitoring*

Sound evaluation of the functioning of public health programmes is a pressing need. In this connexion, field studies can usefully supplement the statistical use of routine records discussed in the Committee's fourteenth report (WHO Expert Committee on Health Statistics, 1971). Sampling surveys, for example, are almost essential for monitoring a health programme. Before the programme is put into effect, they can be used to establish a baseline for the population to which the programme is to be applied. For instance, the level of disease prevalence may be measured in an intervention programme against heart disease and, at the same time, the population's attitude to crucial factors such as cigarette smoking and regular exercise can be assessed. Equally important is the use of serial subsampling both of the original sample and of those not concerned in the baseline survey. In this way, trends in attitude, opinion, and response (in terms of personal action) to health education through the mass media may be watched. Such monitoring is essential if up-to-date information is to be gleaned about the social groups reached by health propaganda and the degree to which they follow the advice given.

Similarly, when physical assessment by clinical examination is required, sample surveys economize the use of staff and yet give sufficiently clear indications of trends in physical conditions to serve as a basis for action. Thus, in the evaluation of a child care service in a country where malnutrition is rife, sampling of the children attending the clinics could reveal the success or failure of the service in terms of changes in body weight or some other objective index long before grosser disease became apparent. In the ultimate stage of before-and-after experiments, sample surveys would again be employed as an effective way of measuring the prevalence of the more flagrant pathological results of malnutrition.

#### 6.3.2 *The experimental approach*

Essential as observation by sample surveys may be, no opportunity should be lost for introducing an element of experimentation into the process of evaluating health services. Where resources are limited and the demand great—e.g., where special units are used for the care of children recovering from rheumatic fever—allocation by strict randomization is no more arbitrary than haphazard policy of admission. Furthermore, it offers advantage that the comparison of those cared for outside and inside the special units may be based on series selected in a completely unbiased fashion, thus enabling the merits or drawbacks of the service to be determined with relative ease and confidence.

## 7. DATA HANDLING AND ANALYSIS

The principles of data handling and analysis in public health are not unique to field studies but do have some characteristic features. The main problem in handling data is usually their complexity and volume. Field studies are also often of long duration and can involve several collaborating centres. The design of the data handling system is thus of crucial importance. It may be necessary, for example, to provide for constant monitoring of the data and for feedback of information to the participating investigators. The analysis of data from field studies requires a special approach, not only because of the characteristic parameters of disease frequency, but also because such studies are often non-experimental and generally require the control of confounding factors.

### 7.1 Data systems

The primary recording of data in public health field studies usually involves a carefully designed data form, questionnaire, or similar record. The principles of designing such forms have been discussed in many contexts, for example by Rose & Blackburn (1968). One characteristic feature of such forms is the use of sequentially numbered main items, which may include conditional subquestions. Another important feature is that the structure of the response to each question is predetermined by the use of mutually exclusive categories covering all possibilities. For example, swelling of ankles may be recorded, by an appropriate check mark, as either "none", "questionable", "mild", "moderate", "severe" or "unknown".

Often the complexity of the data calls for processing either by mechanical aids—e.g., the card sorter—or by electronic computer. The data may be transferred in various ways, most commonly by manual coding and keying on to punched cards, paper tapes, or magnetic tapes, which are used in mechanical machines or as input for the computer. The planning of a computer file structure is often a difficult problem, because this structure has an important bearing on the ease of up-dating, editing, and analysis. The specifications depend on the tasks envisaged as well as on the available computing facilities. Advice from competent consultants on this part of the planning for data handling is often called for.

### 7.2 Editing

The first task in data handling is editing—i.e., searching for and correcting omissions and apparent errors. With a simple set of data, this editing may be based on listings of the data. More generally it involves the pro-

duction of frequency distributions of the individual items of information and a check for their consistency.

### 7.3 Preliminary analysis

Once the data have been corrected as far as possible, the next task is to gain broad familiarity with the details. This is essential for intelligent planning of the ultimate analyses, which are expressly oriented to the objectives of the study.

The main feature of the preliminary analysis is again the computation of a frequency distribution or percentiles for each item of information. Pairwise relationships of variates are also often of interest. Their display involves cross-classifications, correlation matrices, and scattergrams.

In non-experimental studies on etiology or intervention there is a need to assess the possible confounding effect of various factors by determining whether they are associated with the exposure or intervention, and also with the outcome—even within categories of exposure or intervention (cf. section 5). When considering one factor at a time, computations of frequency distributions or percentiles may be used. Sometimes, however, a large number of factors have to be considered as potentially confounding, and the evaluation of a joint confounding effect may call for multivariate analysis (cf. section 7.4.2). Step-by-step reduction in the number of variables involved in the multivariate model leads to the ultimate few factors that are most relevant and have to be controlled in evaluating the effect of the exposure or intervention at issue.

### 7.4 Data reduction and analysis

#### 7.4.1 *General principles*

The problem of data reduction in experimental studies is generally quite straightforward. For the different groups to be compared, measures such as rates, means, and medians characterizing the various outcome criteria can be readily computed.

In non-experimental cohort studies, the basic reduction of data is usually complicated by the need to control various confounding factors. A first display of such data is usually given separately for each of the strata defined by the confounding factors. Further reduction is usually desirable, and this may be done in terms of the computation of standardized rates. When only two groups are compared, either the “direct” or the “indirect” method of standardization may be used. When three or more groups are involved, the usual “indirect” approach of computing ratios of observed and “expected” numbers of events does not lead to perfectly comparable values. In this situation, the direct method of standardization is preferable.

Such standardized rates may then be compared between the groups, perhaps by converting them into risk ratios.

In case-control studies, the problems of data reduction are rather different (MacMahon & Pugh, 1970). The first display of reduced data typically involves two-by-two tables showing frequencies of exposure or intervention for cases and "controls", separately for each stratum defined by the confounding factor. The "odds ratio" is usually computed for each stratum and is based on the exposure rates because it serves as an estimate of the respective risk ratio ("relative risk"). A risk ratio estimate may be derived in a variety of ways. For example, if the risk ratio is thought to be uniform over the strata, the maximum likelihood estimation is the procedure of choice (Gart, 1970). Otherwise, a standard risk ratio may be computed according to a recently described technique (Miettinen, 1972). This mode of data reduction is readily extended to the case of more than two levels of exposure.

When using the risk ratio as a measure of "effect", the change in its value in successive age groups can be important, for the increasing "baseline" frequency of disease with age makes the relative effect of the factor under study less pronounced. Thus, it is generally desirable to explore the magnitude of the risk ratio in relation to age and perhaps other factors. This also provides for making age-specific comparisons between results in different populations.

#### 7.4.2 *Multivariate analysis*

The classical modes of data analysis in public health field studies are now being complemented by multivariate procedures. The inherently multivariate character of research problems has always called for such analyses, which have become feasible with the advent of the computer era. Moreover, experience during the last few years indicates that multivariate procedures generally provide reasonable results in practice, even though the procedures themselves have been developed under rather restrictive assumptions about the variables involved.

In the analysis of etiological and intervention studies, multivariate procedures permit simultaneous control of a number of confounding factors. This is achieved by entering the various confounding factors and the exposure of special interest into the same multivariate function to explain either a qualitative or a quantitative outcome. The procedures usually applied are those of the discriminant function and general linear model, respectively. When the model is reduced by successively deleting noncontributory factors, a few relevant factors may remain, and their ultimate control may be carried out by the use of classical methods involving stratification.

Analysis of this type also leads naturally to a description of the outcome in relation not only to any exposure of prime interest but also to a variety

of other factors. When dealing with the risk of disease, the risk functions—namely, the mathematical expressions of the risk—so far considered have included the linear model (with a binary dependent variate) and the multiple logistic function. When using a linear model, the estimation of the parameters has so far involved both weighted and unweighted least squares fitting. With a nonlinear model, the estimation has been based on discriminant function analysis (Truett et al., 1967) and, alternatively, on an iterative maximum likelihood procedure (Walker & Duncan, 1967). The parameters of the logistic function could also be estimated by the use of the log-linear model proposed by Birch (1963). Experience so far has not revealed any major differences in the performance of these different approaches.

An estimated risk function may be evaluated in terms of its performance in describing the morbidity or mortality within the study material and in predicting morbidity or mortality in other populations. The fit within the study series may be assessed by stratifying the study subjects according to the value of the estimated risk function and comparing, within each stratum, the typical risk estimate derived from the function with the actual morbidity or mortality rate observed among the subjects in it. The adequacy of prediction beyond the series of model fittings may sometimes be evaluated within the total study series at hand by fitting the model to a subset of the subjects and evaluating the fitted function in the remaining subjects. For example, the model may be fitted to subjects from one geographical area and its predictive performance may be evaluated in those from another. Alternatively, the function from one study may be applied to the data from another. In both these approaches, the methods of evaluation may be based on the comparison of observed and predicted rates of morbidity or mortality within strata defined by the predicted values for the individuals involved. Whereas the fit within the study series has been good in the limited published experience with both linear and logistic functions, the prediction for a new series can be unsatisfactory. There are purely statistical reasons for expecting this. For example, it has been shown that discriminant functions tend to show an exaggerated risk separation within the study series, and that this bias depends on the number of variables in relation to the number of subjects in the study (Cornfield, 1967). Moreover, there may be actual differences between the populations in estimation and application, and therefore the ultimate criterion is performance in practical application to new subjects.

## 8. RECOMMENDATIONS

The application of statistical principles to public health field studies is a rather recent development which, with modern emphasis on health care, is rapidly gaining ground. Descriptions of these principles are scattered

throughout the literature and are known only to experienced workers in the field. At the same time, these principles are still inadequately developed. Therefore, it is recommended that the following lines of action should be considered :

1. Early publication of the WHO manual on sampling already under preparation.

2. Preparation of manuals or other compendia on principles and special statistical problems in : (i) field studies on disease etiology and the effects of control measures ; (ii) long-term, large-scale, collaborative experiments on intervention measures ; and (iii) subjects presenting specific problems—e.g., cancer treatment, coronary heart disease, and mental health.

3. Encouragement of research into the statistical principles and methods of public health field studies, particularly in the area of evaluating the efficiency of newly implemented and current public health programmes.

4. The convening of a meeting of experts to discuss the statistical aspects and practical utility of mathematical models in the planning and administration of health services.

#### REFERENCES

- Birch, M. W. (1963) Maximum likelihood in threeway contingency tables. *J. roy. statist. Soc., B*, **25**, 220–233
- Brass, W. (1971) A critique of methods for estimating population growth in countries with limited data, *Bull. int. statist. Inst.* **44**, 397–415
- Cochran, W. G. (1968) Errors of measurement in statistics, *Technometrics*, **10**, 637–666
- Cornfield, J. (1967) Discriminant functions, *Rev. int. statist. Inst.*, **35**, 142–153
- Frankel, M. R. (1971) *Inference from survey samples*, Ann Arbor, Mich., Institute for Social Research, University of Michigan
- Gart, J. J. (1970) Point and interval estimation of the common odds ratio in the combination of  $2 \times 2$  tables with fixed marginals, *Biometrika*, **57**, 471–475
- Greenberg, B. G., Kuebler, R. R., Abernathy, J. R. & Horvitz, D. G. (1971) Application of the randomized response technique in obtaining quantitative data, *J. Amer. statist. Ass.*, **66**, 243–250
- Hansen, M. H. & Waksberg, J. (1970) Research on non-sampling errors and censuses and surveys, *Rev. int. statist. Inst.* **38**, 317–330
- MacMahon, B. & Pugh, T. F. (1970) *Epidemiology : principles and methods*, Boston, Mass., Little, Brown
- McCarthy, P. J. (1969) Pseudo replication. Further evaluation and application of balanced half-sample technique, *US Public Health Service Publication* No. 1000, Series 2, No. 31
- Miettinen, O. S. (1972) Components of the crude risk ratio, *Amer. J. Epidem.*, **96**, 168–172

- Pollock, T. M. (1966) *Trials of prophylactic agents for the control of communicable diseases*, Geneva (World Health Organization : Monograph Series, No. 52)
- Rose, G. A. & Blackburn, H. (1968) *Cardiovascular survey methods*, Geneva, World Health Organization (Monograph Series, No. 56)
- Schwartz, D., Flamant, R. & Lellouch, J. (1970) *L'essai thérapeutique chez l'homme*, Paris, Flammarion
- Tepping, B. J. (1968) Variance estimation in complex surveys, *Proc. Soc. statist. Sect. Amer. statist. Ass.*, 11-18
- Truett, J., Cornfield, J. & Kannel, W. (1967) A multivariate analysis of the risk of coronary heart disease in Framingham, *J. chron. Dis.*, **20**, 511-524
- Walker, S. H. & Duncan, D. B. (1967) Estimation of the probability of an event as a function of several independent variables, *Biometrika*, **54**, 167-179
- WHO Expert Committee on Health Statistics (1966) Sampling methods in morbidity surveys and public health investigations, *Wld Hlth Org. techn. Rep. Ser.*, No. 336
- WHO Expert Committee on Health Statistics (1967) Epidemiological methods in the study of chronic diseases, *Wld Hlth Org. techn. Rep. Ser.*, No. 365
- WHO Expert Committee on Health Statistics (1971) Statistical indicators for the planning and evaluation of public health programmes, *Wld Hlth Org. techn. Rep. Ser.*, No. 472
-