World Health Organization
Regional Office for Europe
Copenhagen

# Methods for cohort studies of chronic airflow limitation

WHO Regional Publications, European Series No. 12

World Health Organization
Regional Office for Europe
Copenhagen

# Methods for cohort studies of chronic airflow limitation

C. du V. Florey

*Department of Community Medicine*
*St Thomas's Hospital Medical School*
*London, United Kingdom*

and

S.R. Leeder

*Faculty of Medicine*
*University of Newcastle*
*New South Wales, Australia*

WHO Regional Publications, European Series No. 12

This publication is dedicated to the memory of David Maddison, late Dean of the Faculty of Medicine at the University of Newcastle, NSW, who provided inspiration to the many members of the Faculty, and especially to the authors during the preparation of the book.

ISBN 92 890 1103 3



TYPESET IN INDIA
PRINTED IN ENGLAND

# ACKNOWLEDGEMENTS

We should like to take this opportunity of thanking the many people who gave advice. In particular, we thank Associate Professor A. J. Woolcock for her comments and for providing Annex 6 describing the test for bronchial reactivity. We also thank Miss S. Chinn, Dr A. Dobson, Professor W. W. Holland, Dr M. Karvonen, Professor P. Macklem, Mr A. Swan and Dr R. van der Lende for their searching and most constructive criticism. Mr Swan also kindly provided Fig. 2. We are very grateful to Mrs Terese Alder for the many patient hours spent typing and retyping the manuscript.

C. du V. Florey is also grateful to the Faculty of Medicine, University of Newcastle, New South Wales, Australia, for providing support during his sabbatical leave, when this book was written.

# CONTENTS

# Introduction

With the advent of antibiotics in the 1940s the importance of acute respiratory infections as a cause of morbidity and high mortality declined rapidly. However in some countries, particularly the United Kingdom, the burden of chronic disease of the respiratory system has persisted, with substantial death rates ascribed to "chronic bronchitis". Because of the importance of chronic respiratory conditions as a cause of lost productivity, lower quality of life, and a drain on health service resources, a large number of studies have been carried out to investigate their etiology in a search for methods of prevention and a more precise definition of their natural history.

The search for a working definition of chronic lung disease started with the term "chronic non-specific lung disease" (*1*). This was then superseded by "chronic obstructive lung disease" and "chronic airways obstruction". Recently it has been suggested (*2*) that the term "chronic airflow limitation" best describes the common problem of patients with a number of underlying abnormalities, including emphysema, bronchiolitis, chronic bronchitis and chronic asthma. Emphysema is enlargement of the distal air spaces accompanied by the destruction of alveolar walls, and bronchiolitis is inflammation, obstruction or obliteration of the bronchioles. Both conditions can be diagnosed with certainty only on biopsy or at autopsy. Chronic bronchitis, better described as mucus hypersecretion, is characterized by chronic sputum production, and the term should be used specifically for this clinical situation; by itself it does not necessarily cause airflow limitation. Patients with long-standing, severe asthma may have chronic airflow limitation due to residual obstruction, the pathology of which is poorly understood but includes smooth muscle hypertrophy and bronchiolitis with mucus plugging.

Chronic airflow limitation may be partially reversible in response to bronchodilator treatment in some patients, but the term implies that, in spite of maximal therapy, there is residual limitation to airflow on forced expiration. There is no agreement as to which test best demonstrates the inability to blow air rapidly from the lungs. However, it is generally recognized that the volume/flow/time relations of a

forced expiration will reveal established abnormalities causing airflow limitation and possibly also subtle changes when the underlying abnormalities are at an early stage.

Etiological factors found to be related to chronic airflow limitation include, most importantly, cigarette smoking and high levels of particulate air pollution from industrial, domestic and occupational sources. Chronic airflow limitation also occurs more frequently in males than in females at all ages. It is a major health problem, especially in older people, when the clinical manifestations become serious, demanding increasing medical care and causing absenteeism from work.

Despite the large number of studies of chronic airflow limitation, many questions remain unanswered. We do not know what the origins of the obstruction are. Does the process start in childhood? Indirect evidence suggests that it does, but detailed long-term studies have not yet been carried out. Does it spring from susceptibility to viral infection in the first two years of life, or is there another, endogenous, factor that is switched on by viral infection? Can the potential patient be identified in time to prevent disease?

From another point of view, the effects of environmental pollution may be monitored or examined through changes in, or contrasts between, respiratory illness rates of people exposed to different concentrations of the pollutants concerned. This approach may be used to study either acute episodes of pollution suspected of having long-term effects, or chronic lower levels of pollution. Opportunities for such studies arise wherever there is industrial development. Epidemiological studies are essential whenever there is a catastrophe in which chemicals are released that may have long-term consequences for the health of the exposed population.

But why is an epidemiological approach needed? Studies on animals have contributed substantially to the understanding of the effects of a variety of factors on the respiratory system. However, the results are not directly applicable to man because of species differences and because, in most studies, the concentrations required to elicit a response within a reasonably short space of time have been far in excess of those experienced by human populations in the general environment. Laboratory experiments on human volunteers, in which control over the levels of factors of specific interest allows precise estimates of exposure to be made, have also contributed, but have been limited because of ethical and practical considerations; in particular, they are rarely conducted under natural conditions.

Since definitive answers to many of the problems posed by chronic airflow limitation have not been forthcoming from animal or human laboratory experiments, epidemiological studies have played a major part in defining the roles of putative causes.

This publication describes the epidemiological methods for studying the evolution of chronic airflow limitation, and is a contribution to the WHO programme for the control of chronic diseases. The investigation of the origins of respiratory illness has been strongly supported by

WHO in the past through its sponsorship of a multinational study of the relation between air pollution and respiratory symptoms and disease in primary school children, but there is now a need to develop a general methodology for research on chronic airflow limitation. This need has also been recognized by the WHO European regional programme on environmental health aspects of the control of chemicals, in which a multifaceted approach to environmental problems has been used. As one part of that programme, epidemiological investigations are to be carried out to study the toxicity of groups of chemicals in specific situations or from specific sources.

More experience has probably been gained worldwide in the epidemiological investigation of chronic airflow limitation through studies of the effects of air pollution, including tobacco smoke, than of any other risk factor. A variety of study designs and measurements of the respiratory system have been used, so that it is now possible to describe a general plan and the options within it for studies of factors predictive of chronic airflow limitation. Furthermore, considerable experience has already been gained of international cooperative studies in the field, so that their strengths and weaknesses are well known (even if the solutions to some of the problems are not yet clear). It is our purpose to identify the successful methods from past work and to describe how a cohort study of the evolution of chronic lung disease might be conducted.

The aim of this book is to assist research workers in the design and execution of their own studies of chronic airflow limitation. Chapter 1 is concerned with the general design of cohort studies, and with their advantages and disadvantages in comparison with other epidemiological methods. The foundations are laid here, but further introductory reading will be essential for readers unfamiliar with the epidemiological approach. In Chapter 2 we discuss the types of population suitable for study and the difficulties of interpreting the data obtained from them. Chapter 3 covers in detail the methods of obtaining a history of respiratory illness and of measuring lung function. It emphasises the need for simplicity and for the rejection of complex measurements, which are unreliable in the field. Chapter 4 describes the statistical analysis of the data. This is not comprehensive—the mark of the good research worker is the ability to marshal the data in unusual patterns to give insight into a problem—but it sets out the essential basic approaches and draws attention to the analytical power provided by the relatively new computer programs for generalized linear models. Chapter 5 deals with documentation, and stresses the importance of the protocol.

All the methods described have been tested in the field and are known to work in principle. They will need modification and adaptation to local situations. Their ultimate success in helping to unravel the remaining questions posed by chronic airflow limitation will depend on the skill of the investigator.

*CHAPTER 1*

# Principles of epidemiological studies

Most epidemiological studies make use of naturally occurring situations rather than designed experiments, and several methods have been developed to take advantage of such situations. The simplest approach to the investigation of problems in the community is *descriptive*: the characteristics of people and of their environment are counted or measured and their frequency or some statistical quantity is used to portray the situation. This is a sound approach in the early stages of an epidemic of an infectious disease or when the health of a community is being mapped in order to assist in planning the delivery of services. Plotting the variations in mortality by month or year of birth is another example. The data may also be used to suggest hypotheses.

## ANALYTICAL STUDIES

A somewhat more complex approach is that of analytical studies, in which hypotheses may be generated from the results of a variety of analyses of the interrelations between variables, or hypotheses may be tested in suitable naturally occurring circumstances or in the contrived circumstances of an experiment. Four general designs are used in epidemiological studies: the cross-sectional, the case-control, the randomized controlled and the cohort. The last is our concern here, but a brief review of the first two will show how they relate in practical terms to the cohort study. (The randomized controlled trial is a highly specialized design and is not considered here.)

### Cross-sectional studies

The cross-sectional study is one in which selected attributes of a population (or a sample of the population) are measured at one point in time. The data may be used purely descriptively, to show how much of some measured characteristic there is in the population, or to help to

develop hypotheses. The frequency of qualitative characteristics (such as symptoms or diseases) is usually given as a prevalence rate (the number of people with the characteristic divided by the number of people examined or, alternatively, in the sample). Quantitative characteristics, such as vital capacity, can be expressed in the form of histograms or in terms of their means and standard deviations. Cross-sectional studies may also be analytical, since the interrelations between the variables can be examined. Hypotheses of cause and effect and of association can be developed from an assessment of the relations between variables. However, the design allows only the *testing* of rather simple hypotheses, e.g. does the prevalence rate or the mean of some characteristic differ from some selected value or from the values obtained in other populations? Even these "tests" lead only to further hypotheses to explain any difference found. This limitation of the design – namely, that the studies do not provide the data for determining the direction of influence between correlated variables – stems from the fact that the sample is composed of survivors of some original but unspecified cohort, the remainder having emigrated or died, and from the lack of data from the past to determine the order of events in a cause-and-effect model. For data obtained by recall, this problem may be partially overcome, but the reliability of the information must be suspect.

### Case-control studies

The case-control design is used for testing hypotheses of cause and effect. Both the cause and the effect must be specified in advance so that at least two groups, one with and the other without the effect (or disease), may be questioned to determine whether or to what extent the putative cause was encountered in the past. This is the design frequently used in clinical epidemiological studies, because both cases and controls can be drawn from hospital patients, and details of past events may be obtained either during a consultation for other purposes or from written records. The speed of execution and the need only for small numbers of observations, as compared with the cross-sectional or cohort studies (an obvious advantage for the study of rare diseases), make this design particularly attractive. However, there are many drawbacks because of biases that can affect the results.

These biases have been described by Sackett (*3*) and are mentioned briefly here because some of them are applicable to cohort studies (most can, however, be avoided in such studies provided that their existence is appreciated). The major biases can be divided into those that affect the sampling and those that affect the measurement, either of the event or condition, or of the predisposing factor.

#### *Prevalence–incidence bias*

This bias (*4*) is encountered in all epidemiological studies. It relates to the effect on the analysis of people missing from the sample because

they failed to survive long enough to take part, their episodes of illness were too short to be recorded during the survey, evidence for their exposure to the causal factors was lost, or the condition, though present, was clinically silent. Some of these people never appear in the sample because of the consequences of the severity of their condition, thereby probably weakening the association with the causal factor, whereas others with the condition may be placed in the control group because of failure to diagnose. Failure to determine past exposure will affect the correct classification of cases and controls and may bias the result in an unmeasurable way, either in favour of or against an association.

*Admission rate bias*

This bias (*5*) relates to the effect of different entry rates into the sample of people in the four basic categories (case, control, exposed, non-exposed). This is simply illustrated by a theoretical example. It might be of interest to see whether chronic bronchitis (CB) and carcinoma of the lung (CaL) tend to be mutually exclusive conditions. In some defined population of 4000 people, let us suppose that 15% have CB and 10% have CaL and that the two diseases are independent. Of the 600 people with CB, 60 (10%) have CaL and of the 3400 people without CB, 340 (also 10%) have CaL. If the mortality rates (which may be thought of as admission rates to post-mortem examination) are 40% for CB, 80% for CaL and 20% for "not CB", the pathologist will find that 53/269 people with CB have CaL (19.7%) and 286/898 "not CB" people will have CaL (31.9%). The chi-square value (14.2) indicates that the difference in the percentages is extremely unlikely to have occurred by chance ($P < 0.001$), which points to a protective effect of CB for CaL, whereas the truth is that the two diseases are independent of each other. Since there is no way of preventing this bias in case-control studies, nor can it be measured, the value of the results of a single study or of several using the same technique is limited. The calculations are illustrated in Tables 1–3.

Table 1. Distribution of chronic bronchitics (CB), lung cancer patients (CaL) and those without these conditions ($\overline{CB}$, $\overline{CaL}$) in a general population of 4000 people

| | CB | $\overline{CB}$ | Total |
|---|---|---|---|
| CaL | 60 | 340 | 400 |
| $\overline{CaL}$ | 540 | 3060 | 3600 |
| Total | 600 | 3400 | 4000 |
| With CaL | 10.0% | 10.0% | |

No association between CB and CaL.

Table 2. Calculation of deaths from CB plus CaL (given independence of effects). Mortality rates: CB = 40%; CaL = 80%; $\overline{CB}$ = 20%

| CaL | CB | | |
|---|---|---|---|
| | Alive | Dead | Total |
| Alive | 7 | 5 | 53 |
| Dead | 29 | 19 | |
| Total | 36 | 24 | 60 |

Of the 60 people with CB and CaL, 40% die of CB. Of the 36 remaining, 80% die of CaL. A similar calculation for the 340 people with CaL and $\overline{CB}$ yields 286 deaths.

Table 3. Numbers coming to post-mortem. Mortality rates: CB = 40%; CaL = 80%; $\overline{CB}$ = 20%

| | CB | $\overline{CB}$ | Total |
|---|---|---|---|
| CaL | 53 | 286 | 339 |
| $\overline{CaL}$ | 216 | 612 | 828 |
| Total | 269 | 898 | 1167 |
| With CaL | 19.7% | 31.9% | |

Highly significant negative association between CB and CaL.

*Unmasking bias*

This bias arises when the factor under consideration is innocent but causes a symptom of the disease in question that leads to the unmasking of the disease through subsequent clinical investigation. In other words, some of the cases may be detected because their contact with the factor caused them to have symptoms of their disease, not because the factor caused the disease. Since the cases and controls are selected by different processes (only one of which includes provocation by the factor) they cease to be suitable for comparison (*6*). Thus exposure to an occupational factor causing wheezing, but not asthma, may lead to investigations that "unmask" a group of latent asthmatics, who are then

classed together with other asthmatics discovered in the normal way. Exposure to the occupational factor may then be found more frequently among the cases than in the non-asthmatic control population. This bias tends to increase the perceived strength of association between the case condition and the "cause". It may be prevented by matching cases and controls on the method of detection, though this can lead to overmatching so that a real effect is missed.

*Non-response bias*

The effects of this bias on the results cannot readily be predicted, but characteristics of the non-responders can sometimes be obtained from other sources and compared with those of the responders to give clues as to its strength. Assessment of the bias may also be possible by selecting a random sample of the non-responders (to reduce the number of subjects involved) and seeking their cooperation with particular diligence.

*Membership bias*

This, the last of the major sampling biases, is caused by non-random allocation of cases and controls into groups exposed and not exposed to factors believed to be causative. The name is derived from the fact that the members of each group have themselves selected their exposure to factors such as cigarette smoking, physical exercise or dietary indiscretions. This bias cannot be prevented but evidence can be obtained to show whether there are important differences between the groups (such as age) that might lead to inappropriate conclusions.

*Diagnostic suspicion bias*

Measurement biases in case-control studies are mostly peculiar to that particular design, but the diagnostic suspicion bias may also occur in cohort studies. This bias arises when, in the course of investigation, either in an interview or in the assessment of physical or laboratory measurements, the investigator becomes suspicious of a respondent's diagnosis. This may lead him to order further investigations or be influenced in his interpretation in a non-standard way. Provided the existence of this bias is appreciated, the study can be designed to eliminate it by using "blind" techniques for assessment.

*Calculation of relative risk*

On the assumption that the biases have been avoided or accounted for in some way, the data from case-control studies may be used to calculate relative risk. This is the ratio of two absolute risks, the first being that of contracting the condition when exposed to the suspected causal factor and the second that of contracting the condition in the absence of the

factor. Only the relative risk can be obtained from case-control studies and this only when the condition in the general population is rare (say $< 1\%$) and the cases and controls are representative of their respective populations (the derivation of relative risk or relative odds is treated in many elementary textbooks of statistics, such as that of Hills (*7*). Case-control studies do not provide the data necessary for the derivation of absolute risk, a statistic of greater interest since it indicates how frequently the condition occurs after exposure, without reference to any other group.

The disadvantages of case-control studies clearly make them unsuitable for drawing definitive conclusions about cause and effect. They are, however, valuable as a first step in testing hypotheses because of their practical simplicity and rapid execution. For very rare diseases they may be the only feasible, though not the ideal, design.

It might be well to remember the comment made on case-control studies by William Farr, first Superintendent of the Statistical Department in the General Register Office in England and Wales: "the replies will be general, vague and I fear of little value". It is essential to use cohort studies to test the results obtained from case-control studies whenever the problem is of sufficient importance to warrant the time and expense.

### Cohort studies

A cohort is used here to mean a group of people defined at some point in time by certain characteristics, such as age, sex, race, or geographical location. For example, a cohort might be defined as all infants born in a country between specified dates (a birth cohort), or a group of 50-year-old executives living in Paris. If the cohort is randomly selected from the population with the defining characteristics, any observations made on it are referable to that population.

The aim of the cohort study is to determine whether characteristics observed at the start or appearing during the course of the study are related to subsequent events, such as myocardial infarction or death from respiratory illness. The study may start with a cohort of people who show no evidence of the disease or diseases of interest. This healthy cohort is obtained from a sample of unselected people by using the results of an examination to exclude those with disease. It is wise, however, to follow those who are excluded, because the results of their examinations may provide insight into the natural history of advanced disease that the healthy group cannot provide for many years. The data for the diseased group and the healthy cohort can be analysed separately without prejudice to either. The first examination is similar to a cross-sectional study, but arrangements have to be made so that the respondents can be followed up at some future time(s). At each examination every attempt must be made to obtain a $100\%$ response. At the first examination not less than $90\%$ of those in the sample should be seen because some of them will fail to come to follow-up examinations

as a consequence of death, departure or disinclination. So important is the maintenance of a high response rate for the success of a cohort study that special and sometimes costly methods of keeping in contact with the cohort members are required. These include, for example, an annual letter with reply-paid card to check changes of address, and the building up of a feeling among respondents of belonging to something special and worth while. The extent to which contact need be made between examinations depends very much on the type of population under study, its stability, its geographical boundaries, and the importance it attaches to the research.

Three variations on the basic cohort design have been used to reduce the length of time required and to overcome the problem of increasing non-response with time.

If poor follow-up response is expected, the use of routinely collected data may be helpful. It may be possible to obtain data on mortality, hospital admissions or sickness absence for members of the cohort. The cohort might be defined *retrospectively* by using employment records in an industry, and information from regular medical check-ups and on retirement and mortality. Provided that the data are of good quality, it is then only necessary to examine the cohort once—a final follow-up so to speak. Although this can be a much faster and cheaper approach than the truly prospective one, it suffers from lack of standardization of past measurement and of data on some variables of interest, thus limiting its usefulness. This retrospective cohort design may be altered so that, in place of a single final survey, the defined cohort is examined several times over the following years. Thus the length of time required for the study is reduced, as before, but a more precise definition of the changes during the latter part of the period is obtained.

The second variation consists of using several cohorts of different ages and following them for a defined period, so that the age of each cohort reaches that of the initial age of the next older cohort. For example, two groups of children aged five and ten years at the start of a study might be followed for five years. In this way a picture of the development of disease can be built up relatively rapidly. However, the assumption that changes observed in a younger cohort were also experienced earlier by an older cohort must be assessed in some way, such as by taking into account changes with time in other factors that may have affected the disease process. Although this design has the advantage of speed in execution, it may lead to erroneous conclusions about the evolution of disease because of differences in lifetime experiences of the cohorts. For example, the 10-year-olds in the hypothetical study mentioned above might have suffered from a particularly severe epidemic of influenza before the birth of the five-year-olds.

The third variation may be used when the non-response over the period of study is expected to be high. If the duration of the study is to be, say, 15 years but non-response is expected to be substantial after five years, three cohorts might then be used, the first to be followed for five

years, the second for the next five years and the third for the last five years. To avoid the biases introduced by different life experiences, each successive cohort should have the same age and sex structure as the preceding cohort would have had in its sixth year if there had been 100% response throughout. Moreover, the second and subsequent cohorts should consist of people who have remained in the area (or industry, etc.) for the entire period of the study. This design suffers from the lack of continuity of measurement on the same person, but may be useful in monitoring the effects of changing environmental pollutant levels over long periods, for which continuity may be less critical than in studies of natural history.

The *incidence rate* is the statistic unique to cohort studies, and is defined as:

$$\frac{\text{Number of new cases occurring in a given period}}{\text{Number of people at risk for that period}}$$

This may be multiplied by ten raised to a suitable power to give a value greater than one (more appealing to the eye than a probability value) and may be expressed in terms of fixed periods, e.g. per week, month or year. For example, there may be 50 new cases of lower respiratory infection in 3000 schoolchildren in a three-month period. The incidence rate is then 50 cases per 3000 children per three months. This can be presented in a more standard form as follows:

$$\frac{50}{3000} \times 1000 = 16.67 \text{ per } 1000 \text{ per quarter year}$$

If the incidence rate did not vary with time it could be expressed per month by dividing by 3, or per year by multiplying by 4. This rate can also be calculated for different subgroups of the cohort, defined by their initial or subsequent values of selected variables, so as to determine the size of the risk associated with those values. Relative risks for pairs of subgroups can be calculated from the incidence rates.

Incidence rates may give a fairly precise estimate of incidence for easily recognized and reliably diagnosed diseases, but are less precise for most chronic non-infectious diseases. These degenerative diseases may run a long silent course before they become clinically overt, so that it is difficult to pinpoint the time when a healthy person becomes a new case. Because of its indeterminate nature, a new case must be defined for the purposes of the study, and in the same way as in other studies, if comparisons are to be made. Diagnostic criteria must be specified. These define the presence of the disease in terms that can be measured in an epidemiological study – some diagnostic tests are too complex to apply to large numbers of people – but they should also be clinically relevant, so that the results of the study can be used by clinicians as well as by community physicians.

The advantage of the cohort study as compared with cross-sectional and case-control studies, apart from the estimation of incidence, is that the natural history of a disease can be studied. Causes may be