4 Collection of Statistics: Forms of Record and Inquiry

In all scientific work we are involved in asking questions. In medicine, for example, we may seek to learn the effects of a specific treatment used upon patients with a specific disease. But whether we require the population concerned to make reply themselves to a questionnaire, whether we seek the information by oral inquiry through ourselves or trained social workers, or whether we make a clinical examination or adopt some laboratory means of procedure, we are, in the last analysis, always asking questions. Accordingly, we need a form of record upon which to ask those questions and to record the answers. One of the first and, indeed, most decisive steps in any inquiry, therefore, is the construction of that form – *what* should be included and *how* it should be included. Each question must be given the closest thought to see whether it is clear and definite; what the possible answers are; whether the answers can be adequately, if not wholly accurately, obtained; how they can be analysed and put into a statistical table at the end of the inquiry or experiment. If the questions are incomplete, ill-conceived, or inadequately answered no statistical analysis, however erudite, can compensate for those defects or produce the answers that the worker had hoped to get. The time to remember that, is not at the end of an investigation but at its beginning.

We must also remember two other things – on the one hand that the drafting of clear and unambiguous questions is an extremely difficult task and, on the other hand, that many persons find the completion of any form an extremely difficult task. As an early report on the Census of England and Wales correctly emphasised, 'those who are conversant with forms and schedules scarcely realise the difficulty which persons, not so conversant, find in filling them up correctly.'

Questions and Answers

In formulating questions, or headings, for inclusion on a form there are a number of basic principles to bear in mind.

(i) To begin with one should consider closely whether there is any

ambiguity in the question and, consequently, in the answers received. A very simple example can be found in that innocent question which appears on so many forms – age. Age last birthday? Or age nearest birthday? Generally, one might expect to be given the age last birthday. But at any moment of time, about one person in 12 is within one month of their *next* birthday and might therefore consider that that was the more appropriate age to give. Perhaps it does not matter. But that certainly does not mean that preliminary consideration should not be given to the question; *a decision should be taken that it does not matter.* In certain circumstances it certainly would matter. For instance, in a group of 100 children aged 5 years *last birthday*, the average age would be $5\frac{1}{2}$ years, or close to it (the individual ages would run from 5 to less than 6). In a group of 100 children aged 5 years to the *nearest birthday* the average age would be 5 years, or close to it (the individual ages would run from $4\frac{1}{2}$ to less than $5\frac{1}{2}$). The bodily measurements of the two groups would obviously differ appreciably since the average ages differ by 6 months. With a form on which 'age' is asked for we shall be unaware of what in fact is given. We should, therefore, specify 'age last birthday' (or, better still, ask for 'date of birth' and then make the required calculation ourselves).

Ambiguity in the question also arises if use is made of what may be called 'overlapping groups.' Thus, in a British inquiry into prematurity, the recipient of a form was asked to state the number of babies born within three birth weight groups (measured in pounds), (*a*) $2\frac{1}{2}$ to $3\frac{1}{2}$ lb, (*b*) $3\frac{1}{2}$ to $4\frac{1}{2}$ lb, and (*c*) $4\frac{1}{2}$ to $5\frac{1}{2}$ lb. Suppose a baby is born weighing precisely $3\frac{1}{2}$ pounds; does it belong to group (*a*) or to group (*b*)? Probably most persons would answer (*b*). Suppose a baby is born weighing precisely $4\frac{1}{2}$ pounds, does it belong to group (*b*) or to group (*c*)? Probably most persons would answer (*c*). But suppose a baby is born weighing precisely $5\frac{1}{2}$ pounds. Does it belong to the table at all? On the previous decisions one must say no. But since the definition of prematurity is $5\frac{1}{2}$ pounds or *less* it certainly was intended to belong. The form of inquiry is therefore poor since it is ambiguous. It will be interpreted and answered differently by different recipients.

(ii) As far as possible *every question should be self-explanatory* and not require the respondent to consult a separate sheet of instructions. The importance of this principle will, of course, vary with the circumstances. If a few highly trained persons are making the observations, conducting the clinical examinations, whatever it may be then clearly they can be relied upon to turn to, and follow, detailed instructions. But if the form is to be completed by large numbers of less trained and less

interested persons, then experience shows that they cannot be relied upon to read and remember extensive footnotes or instructions. For example, in a trial of a vaccine against an infectious disease of childhood details might be sought for each affected child of any known exposure to another case. Such definitions of the possible varieties of exposure which are of interest must be given or some of the answers will undoubtedly be vague and uncertain. If possible the definitions should be incorporated in the question on the form itself. Thus for each case occurring the following alternatives of exposure might be specified *within the question*:—

(*a*) Within the child's own home.
(*b*) At a day nursery.
(*c*) At school.
(*d*) Elsewhere: specify place.
(*e*) No known exposure.

The respondent is thus shown on the form itself the categories of information that are sought and can answer clearly and without undue trouble.

(iii) Almost invariably *every question should require some answer.* Without that precaution it is often impossible to know for certain whether a person did not possess some characteristic or whether no information was in fact sought or obtained. For instance, the question at issue may be 'did the patient during pregnancy suffer an attack of rubella?' An answer 'yes' is a clear positive but no answer at all, or merely a dash (—), is by no means a clear negative. It *may* mean that, but it may, on the other hand, mean that the question was never asked or that no certain information was forthcoming. One does not know. The question should be given in such a form as invariably to require a clear answer, such as (i) Yes; (ii) No; (iii) Not Known. Every question, therefore, needs some final category to make certain that some answer must be given – whether that final category be 'not known,' 'no information,' 'unspecified,' or 'other.' Sometimes it can be useful to present a series of alternatives and ask the subject to select the one which he or she regards as nearest to the truth. For example:—

Do you have a headache? { Never / Very rarely / Occasionally / Frequently / Very frequently

(iv) *The degree of accuracy to which measurements are required* should be specified. Should, for example, blood pressures be recorded to the nearest millimetre, to the nearest 5, or to the nearest 10? If no specification is given the recorders will inevitably vary amongst themselves in the accuracy to which they work. It may also be necessary sometimes to specify the nature of the measurement required – whether, for instance, body temperatures are to be taken orally or per rectum.

(v) It would seem needless to say that much thought should be given as to *whether it is likely that a particular question can be answered adequately by anyone at all*; yet much experience of forms suggests that the advice is not so needless. For example, very few persons can give the certified cause of death of their parents or, *a fortiori*, of other relations. It is very unlikely that persons can remember accurately the minor illnesses (cold in the head, etc.) they have suffered in a previous 12 months. On the other hand it certainly does not follow that questions which it is obvious cannot be answered with complete precision are thereby rendered valueless in all circumstances. The amount of inaccuracy may be unimportant to the problem at issue or it may still allow broad conclusions to be reached. Each situation must be carefully weighed on its merits. For instance, one can be sure that few persons in their 50's or 60's can give an exact statement of their past smoking habits, the habits of a lifetime. It does not follow that they cannot give an answer which is sufficiently accurate to allow them to be classified into a few broad categories – heavy, moderate, light, or non-smoker – and for the frequency of some other characteristic, e.g. cancer of the lung, to be usefully examined within those broad categories. It should be remembered, too, that if the errors are unbiased, then the degree of association found between the two characteristics is likely to be *less* than the degree that exists in reality. (If in reality the A's differ from the B's to a certain degree, that degree will be made less if through unbiased errors of memory, etc., we have included some A's with the B's and some B's with the A's. The contrast has been rendered less 'pure' and clear-cut.)

Under this heading there will also be occasions on which close attention must be paid to the problem of *'observer error or variation.'* Will doctors, or other workers, surveying the same patients classify them in the same way? For example, it is well known that experienced school medical officers will differ among themselves in the assessment of a child's nutritional condition. After a short interval of time, they will also differ from themselves. It is equally well known that experienced readers of X-ray films will differ over the interpretation of, or even over the presence or absence of, lesions in the chest. And according to the

examiner involved, it has been shown that clinical histories will differ in the frequency with which somewhat ill-defined conditions, such as chronic bronchitis, are discovered. It is not really the observers who are at fault; it is the method. It is not precise enough to allow uniform and clear-cut decisions. It follows that close thought should be given to the problem of observer error or variation, whether it is likely to be an important feature and, if so, whether there are any means of reducing it.

(vi) Any form of record which must be completed by many persons should, to the utmost extent, be made *simple in wording and logical in the order of its questions*. The amount of work required of the respondent may often be reduced by putting the question in such a form that the answer demands only a cross or a tick or the ringing of one specified category (as in number (ii) above). Such answers are also very easily tabulated (and the categories on the form can be already numbered if punched cards are subsequently to be used).

(vii) Finally, much attention should be given to *the number of questions*. Obviously this must vary widely with circumstances, but at the same time there are probably no circumstances in which the constructor of a form should not ask himself of *every* question: 'Is this question essential? Can I obtain useful answers to it? Can I analyse them usefully at the end?' Such a self-discipline is likely to reduce the size of any form. The temptation to collect information in case it might be useful should be resisted.

There are also circumstances, as described in Chapter 2 (see p. 19), in which it may be wise to distribute a large number of questions over different samples and thus avoid too heavy a burden being placed upon any one respondent. On the other hand there are some circumstances in which it may be profitable deliberately to include questions for purposes of checking the nature of the response or to encourage an unbiased response. For instance, in approaching a sample of children aged 5 years to study, say, their previous attacks of infectious diseases, it might be useful to include a question on vaccination against measles. If the frequency of vaccination is known for the whole population from national statistics then, from its frequency in the sample, one may judge whether the sample is likely to be well chosen and representative.

In certain circumstances in approaching by questionnaire a population, marked or not marked by some characteristic, it is not unlikely that the marked persons will tend to reply and the unmarked not to reply. For example, patients who have undergone a blood transfusion are followed up by post some months later and asked, with appropriate questions, whether they have had symptoms of jaundice. Those who have had

symptoms, the positives, may well be inclined to reply more readily than those who have not, the negatives, so that a false measure of the incidence of serum hepatitis is reached. It may therefore be profitable to put the question in such a form as to give everyone some reason to reply, e.g. by listing the symptoms of various common diseases (rheumatism, etc.) and asking the respondent to put an X appropriately, or by adding some very general question applicable to everyone such as 'Since leaving hospital have you been in good or poor health?' Except to encourage an unbiased response to the real question at issue, the nature of the extra query is, of course, unimportant.

Finally, it should be remembered that a form containing many questions may lead to less care on the part of the recipient; a shorter form, in other words, may promote greater accuracy of reply.

The Pilot Inquiry

In many inquiries – and particularly very large-scale or expensive inquiries – it will be extremely profitable to conduct a small pilot investigation. A small sample of the population can be drawn and approached – and that procedure itself will reveal any difficulties in the sampling method and in reaching the respondents. The responses themselves will give a measure of the non-response rate to be expected and throw light on questions which prove to have been ill-worded or ambiguous or which cannot be answered adequately at all. Revision can then be made before the inquiry itself is set in train. Thus it has been reported that a preliminary survey in the United States showed that a proposed new income tax form was 'incomprehensible to a substantial part of the public. . . . As a result of this survey a new form was devised which everyone could understand and the Treasury gained millions of dollars from the increased revenue' (P.E.P. Broadsheet on the Social Use of Sample Surveys, 1946). Apart, perhaps, from some legitimate doubts as to whether *any* income tax form could be devised which would be understood by everyone, the moral of the small pilot inquiry is clear.

The same principle may well apply in studying past records, e.g. clinical case notes. A small random sample will rapidly show whether the information required is available or whether there are too many gaps and omissions. It may also be made to show the scale of the proposed work in time and manpower.

Type of Inquiry

Leaving aside the experimental approach (the strongest weapon in the

scientist's armoury) there are, of course, in detail very many ways of investigation by observation. There are, however, two broad categories of approach, each with its own merits and defects, which are worthy of consideration – namely, the *case/control* or *retrospective* and the *cohort* or *prospective.*

In the *case/control, retrospective,* inquiry the starting-point is the affected person, e.g. the patient with cancer of the lung, and the investigation lies in the uncovering of features in his *history* which may have led to that condition, e.g. cigarette smoking, industrial hazards, air pollution, etc. Does one (or more) of those features appear more frequently in the histories of affected persons than in the histories of an unaffected normal population? This, indeed, is the classical method of epidemiology which seeks to show that of persons infected with typhoid fever, or cholera, most had consumed a particular supply of water while of those who were not attacked relatively few had done so. It is very rarely a question of all versus none since invariably some consumers are not attacked and frequently some non-consumers are attacked, e.g. through secondary infection. Thus, it is a comparison of *relative* frequencies and for this purpose not only the histories of the cases are required but the histories of some 'controls.' The choice of appropriate controls demands careful thought to ensure that the comparisons are valid. Usually the control group should be as similar as possible to the affected group except for the presence of the disease in question. Sometimes, indeed, it may be possible and valuable to pair each affected individual with a control individual of the same age, sex, etc., and thus deliberately equalise the groups in some features that will influence the comparison. Often, however, it will be highly profitable to seek more than one control group. If a whole series of control groups, e.g. of patients with different diseases, give much the same answer and only the one affected group differs, the evidence is clearly much stronger than if the affected group differs from merely one other group.

We must also bear in mind that the evidence is often based upon past records and is dependent upon the completeness and accuracy of those records.

The *cohort, prospective,* method, on the other hand, starts with an unaffected sample of the population, e.g. without cancer of the lung, characterises each member of the sample by one or more features, e.g. smoking habits, occupation, place of residence, and then records the *future* occurrence of an event (the development of cancer of the lung) in relation to those features. Does the disease appear more frequently in some groups than in others?

The Pros and Cons

The advantages and disadvantages of the two approaches are these. With the cohort method the sample under study can usually be clearly defined and it is easier to consider whether it is likely to be representative of some universe or is biased. It may be very difficult to identify similarly the nature of a case/control sample or the nature of its very possible bias, e.g. are patients in hospital with a coronary infarct likely to be representative of all such patients in regard to some particular constitutional or environmental factor such as dietary habits? Are the controls, selected for comparison with them, representative of the general population (or some specific population) without a coronary infarct? What selective influences may bring the affected and the unaffected into observation? For instance are physicians more likely to elicit and record particular features in certain specified patients than in other patients being used as controls? For instance would they probe more closely into the smoking habits of a patient with cancer of the lung than of a patient with appendicitis? Those are difficult questions to answer. On the other hand, following up the same example, the difficulties of interpretation are much less if we take every nth person in a defined normal population, determine their dietary habits, and then record subsequent events. The nature of the sample is clearer. We may, however, note that in this cohort or prospective inquiry there may be refusals to co-operate, persons lost sight of and other forms of non-response, all of which will make the sample less certain in its nature.

Neither method, of course, can provide 'proof' of cause and effect. We are always seeking the *most reasonable* interpretation of an association. And usually we shall have to consider the most probable *order* of events in an association – do the dietary habits tend to lead to an infarct or does the person liable to an infarct tend to have certain dietary habits?

One great advantage of the cohort method is that, knowing the number of persons at risk in each group, the incidence rates of the events subsequently observed can be easily calculated and compared. Such a calculation is more difficult with the case/control method. For example, we could observe pregnant women suffering an attack of rubella in the first, second, and third and later months of pregnancy and subsequently observe the incidence of congenital defects in their babies and calculate its incidence in each group. On the other hand, starting with the defective babies, we might find retrospectively that all the mothers had had rubella whereas relatively few mothers of normal control babies had been attacked. To measure from such data the actual risks of a defective child might be impossible or, at least, very difficult.

The case/control method, however, is likely to give an answer more speedily than the cohort with its prolonged follow-up and, in some circumstances, it is likely to be the only possible approach. For example, with a relatively rare condition like multiple sclerosis it might be quite impossible to categorise a sufficiently large population to give incidence rates from future occurrences within a reasonable span of time. It would not be at all difficult to accumulate a large group of cases and retrospectively to explore their past.

The cohort method does not, however, always involve a subsequent waiting period. It can be applied so long as a population can be defined *at any specific time* and then its subsequent events noted, e.g. in records already accumulated. For instance, for the live births that took place in a given hospital over some earlier years it might be possible to determine from the available clinical records whether or not the mother had an X-ray of the abdomen. One might then determine by inquiry or other already available records, the health of the child 5 years later. In other words, the cohort method has thus been applied to existing records.

In some circumstances one might well choose to make a pilot case/control inquiry before embarking upon a more arduous cohort investigation.

In conclusion, though the prospective approach must usually be the 'method of choice,' there can certainly be no *one* right way in which to make every investigation.

Summary

One of the most decisive and difficult tasks in any inquiry is the construction of an appropriate form of record. Care must be taken to ensure that the questions are clear and unambiguous and, as far as possible, self-explanatory. Each question should require some answer and the standard of accuracy necessary for the purpose in hand should be considered. To ensure a high rate of return a form may need to be kept short. On the other hand, to ensure an unbiased return there may be occasions when extra questions are useful. Pilot inquiries can be invaluable in revealing the difficulties and defects of a proposed large-scale investigation.

Many inquiries follow one of two forms of approach – the case/control or retrospective (looking backwards) and the cohort or prospective (looking forwards). The latter has much in its favour but with rare events may be impossible. There can be no one right or wrong way in all circumstances.

5 Presentation of Statistics

Once a number of observations or measurements has been made, or collected, the first object must be to express them in some simple form which will permit, directly or by means of further calculations, conclusions to be drawn. The publication, for instance, of a long series of the responses of patients to a specific treatment is not particularly helpful (beyond providing material for interested persons to work upon), for it is impossible to detect, from the unsorted mass of raw material, relationships between the various factors at issue. The worker must first consider the questions which he believes the material is capable of answering and then determine the form of presentation which brings out the true answers most clearly. For instance, let us suppose the worker has amassed a series of after-histories of patients treated for gastric ulcer and wishes to assess the value of the treatments given, using as a measure the amount of incapacitating illness suffered in subsequent years. There will be various factors, the influence of which it will be of interest to observe. Is the age or sex of the patient material to the upshot? Division of the data must be made into these categories and tables constructed to show how much subsequent illness was in fact suffered by each of these groups. Is the after-history affected by the type of treatment? A further tabulation is necessary to explore this point. And so on. The initial step must be to divide the observations into a relatively small number of groups, those in each group being considered alike in that characteristic for the purpose in hand. To take another example, relevant to the remarkable and fascinating history of scarlet fever with its fluctuating virulence, Table 1 shows some past fatality-rates from scarlet fever in hospital; for this purpose children within each year of age up to 10 and in each five-year group from 10 to 20 are considered alike with respect to age. It is, of course, possible that by this grouping we are concealing real differences. The fatality-rate at 0–6 months may differ from the fatality-rate at 6–12 months, at 12–18 months it may differ from the rate at 18–24 months. To answer that question, further subdivision – if the number of cases justifies it – would be necessary. In its present form (accepting the figures