3 Collection of Statistics: Selection

In the previous chapter attention was devoted mainly to the situations in which a sample of observations could be deliberately drawn for study from some known population. The basic problems are then to define the population and to find an appropriate means of drawing a random sample from it. Often in medicine we are not in that situation at all. We have to accept whatever sample of observations (persons, records, etc.) may present itself in the natural course of daily events. *What is the nature of the sample* then becomes the crucial question. What are its characteristics? Are we entitled to argue from the particular to the general? For that is what we are invariably hoping to do. In seeking to advance knowledge we are, for example, not very interested in the fact that a particular and relatively small number of patients with a specific disease rapidly recovered when treated in some defined way. We are intensely interested in knowing whether that form of treatment is the method of choice for the generality of patients with that disease. We must then consider very carefully whether the sample is representative of all such patients and not in any way biased or 'selected.'

It is important to be clear on the meaning the statistician attaches to the word 'selected.' By a selected sample he denotes a sample which is not representative of the parent population of which it is a part. The selection may have been deliberate, in which case the form of selection is known and the lack of comparability between the sample and the population is usually perfectly clear. For instance, if the treatment of some form of cancer by means of surgery were confined to those patients without metastases, then it is obvious that these patients are *not* a representative sample of all patients with this form of cancer, but are selected on the criterion of no detectable metastases. To compare their subsequent mortality experience with that of all patients is, therefore, a most doubtful procedure, for we are clearly not comparing like with like in all respects except with regard to surgical treatment. Even without that treatment the death-rate of patients without metastases is likely to differ materially from the death-rate of the general run of patients.

More often, however, the 'selection' is not deliberate but is quite unforeseen or is unrealised. To say, therefore, that Mr. So-and-so's figures relate to a selected sample of patients is not an aspersion on Mr. So-and-so's scientific honesty; the statement implies only that owing to the method of collection of the figures, or to the limited field in which Mr. So-and-so was able to operate, it is quite impossible for his sample to be representative. It may be that with care that selection might have been avoided; often it is unavoidable. Its possible presence cannot be too carefully remembered or taken into account in interpreting all statistics. As, however, it is frequently overlooked by the authors of studies in medicine, this chapter is devoted to a series of examples.

Examples of Selection

Sex Ratio at Birth

As a simple illustration the author once took from *The Times* newspaper the frequency with which male and female births were recorded in the birth column. The number of male births was 3304 and female births 3034, so that the sex ratio was 1089 males for each 1000 females. According to the Registrar-General's figures for England and Wales at that time, the sex ratio of births in the country as a whole rarely exceeded 1050. It is clear that from the point of view of sex ratio the births recorded in *The Times* were unlikely to be representative of the births in the country as a whole. It is possible that first births are more frequently recorded in those columns than births of a later order, and that such births have a different sex ratio; or that proud parents are more likely to record their heirs than their daughters; or that the sex ratio differs between social classes. Whatever the explanation, with such a sample of births, if that were all that were available, one could not generalise about the population of the whole country with any security.

Hospital Statistics

Turning to a more medical problem, hospital statistics can very rarely be regarded as unselected. The patients are frequently drawn from particular areas and from particular social classes. Still more important, in many diseases only those patients who are seriously ill are likely to be taken to hospital. It is obvious that we cannot determine with any approach to accuracy the fatality-rate of any disease, say measles at ages 0–5 years, if our statistics are based mainly upon the seriously ill – patients, for example, in whom a secondary pneumonia has developed – and ignore the mass of children whose symptoms are so slight that they

can safely be treated in their own homes. Of *all* children with measles those in hospital would then form only a small and stringently selected group; our deductions from such a group are correspondingly limited, especially with regard to such factors as the incidence of complications and the rate of fatality or recovery. It is not too much to say that there is hardly any disease in which a hospital population must not initially be regarded with suspicion if it is desired to argue from the sample to the universe of all patients. No such argument should be attempted without a preliminary and rigorous examination of the possible ways in which selection may have occurred.

The same difficulty arises with secular comparisons – e.g. when we wish to see whether the fatality from some disease has changed from one year to another. In each year the fatality-rate is measured from the experience of the patients admitted to hospital, and in each year those patients are a sample of all patients with the disease in question. It must be considered whether that sample has changed in type. In both years the sample may be a selected sample but the selection may not be identical. The kind of patient admitted may have changed. For example, in a group of American hospitals it was once reported that the fatality-rate from appendicitis declined from 6 per cent in one year to 3·5 per cent four years later. Was that a 'real' decline, due to more efficient treatment maybe, or was there a concurrent change in the types of patients admitted? Examination of the basic figures showed that in these hospitals some 2500 patients were operated upon in the earlier year, while in the later year the number had risen to 3500, an increase of 40 per cent. It is impossible to believe that an increase of 40 per cent in four years was a real increase in the incidence of appendicitis. It is more likely that the desire for admission to these particular hospitals or some criterion of admission had changed, that some patients who were admitted in the later year would not have entered them in the earlier year. It is possible, therefore, that the type of entry has changed as well as the volume – perhaps that milder cases were admitted and operated upon in the later series which were not present in the earlier series. In the absence of positive evidence on that point the change in fatality cannot be accepted at its face value or as satisfactory evidence of the effect of a change in some other factor – e.g. the benefit of earlier admission to hospital in the later year. The following two questions must always be considered. Has there been a change in the population from which the samples are drawn at two dates – i.e. a change relevant to the question at issue? At each date was there an equal probability that a particular type of patient would be included in the sample?

Day of Treatment

In measurements of the value of some form of treatment, statistics of the following type are frequently given:—

Day of Disease upon which Treatment was first given	Fatility-rate per cent of Treated Patients
1	1·3
2	3·6
3	7·5
4	9·3
5	12·8
6 or later	16·4

It is possible that the level of this fatality-rate at the different stages is seriously influenced by selection. Let us suppose, as is often the case, that the treatment is given to patients brought to hospital and that *all* patients do not necessarily go to hospital. Then on the first day of disease a variety of patients will be taken to hospital, in some of whom, in the absence of the special treatment, the disease is destined to run a mild course, in others a severe course. The presence of a proportion of mild cases will ensure a relatively low fatality-rate, even if the special treatment has no specific effect. But as time passes this proportion of mild cases in the hospital sample is likely to decline. By the time, say, the fourth day of disease is reached, a number of patients treated at home who were not seriously ill will have recovered or be on the way to recovery. Their removal to hospital is unnecessary. On the other hand, patients who have made a turn for the worse or whose condition has become serious are likely to be taken to hospital for immediate treatment. Thus on the later days of disease the sample removed to hospital for treatment may well contain an increasing proportion of persons seriously ill, it obviously being unnecessary to transfer those who are making an uninterrupted recovery. In other words the patients removed to hospital on the fourth day of disease may not be a random sample of all patients who have reached that day of the disease but may consist rather, perhaps mainly, of patients still seriously ill. Such a group will certainly have a relatively high fatality-rate.

Another example of this statistical difficulty may be taken from some fatality-rates recorded for appendicitis. It was reported that in a group of cases 2 per cent died of those admitted to hospital within 24 hours of the

onset of symptoms compared with 10 per cent of those whose admission was delayed till after 72 hours. But it is likely that the group of patients admitted early is composed of a proportion of the seriously ill and a proportion that would do well whether admitted to hospital or not. On the other hand, those who are admitted after a delay of three days from onset are likely to be patients whose condition is serious, for clearly those whose condition has become quiescent are unlikely to be taken to hospital at that point of time. If such a sequence of events occurs, it is clear that the group of patients admitted early is not *in pari materia* with the group of patients admitted late. Selection may not be the whole explanation of the difference between the fatality-rates; indeed it is not likely to be, for there are excellent reasons for the early treatment of appendicitis. But it is a possible factor with statistics such as these, which makes it very difficult to measure accurately the *magnitude* of the advantage.

Post-mortem Statistics

In an attempt to obtain more accurate data emphasis is often placed upon post-mortem statistics. One has, however, to remember that this increased accuracy is gained at the risk of using material that may be highly selective. It is rare for every death to be subjected to autopsy and those that are chosen are by no means chosen randomly. They are more likely to be chosen *because* the cause of death is obscure or because the case presents features of special interest. No measure of these or other selective factors is likely to be made easily.

In this field of autopsy statistics it is, too, important to note that the interpretation of the observed frequency of occurrence of *two* disorders in the same person is very difficult indeed. Thus it may appear that the number of persons found to have the two conditions is appreciably more than would be expected on a purely chance basis, i.e. based upon the known frequency of the occurrence of each condition separately. But the occurrence of two disorders in the same person not only may make it more likely that he will seek hospital care, it may also make it more likely that he will die and thus come to autopsy. Such a possible selective mechanism must always be thought upon with much care in the specific study.

Follow-up Studies

Because of their incompleteness, follow-up studies of patients are often subject to selective influences. In basing conclusions upon the

patients who *are* successfully followed-up we are presuming that the results we record for such a group would be unchanged if we succeeded in tracing and adding in the lost-to-sight patients; in other words, we presume that the characteristic 'being followed-up' is not correlated with the characteristic that we are measuring, e.g. survival. Is that likely to be true? On the one hand it might be easier to learn that a patient is dead – through the Registrar-General or other official records – than that he is alive and gone to Australia. On the other hand the living may answer inquiries more readily than the relatives of the dead. Once more all we can do is to consider the likelihood and nature of selective factors in any specific situation, or, of course, better still, avoid them at all costs by making the follow-up comprehensive.

Infant Feeding

That on *a priori* reasoning there is a very strong case for the breast feeding of infants is obvious, but to secure a measure of the degree of its advantages has always been, owing to selective factors, extremely difficult.

Under actual conditions of life we cannot obtain two groups of infants, exclusively breast-fed and exclusively bottle-fed from birth, who will remain in their respective groups whether their progress is good or bad. If a baby on the breast is not thriving, its diet is likely to be changed to partial or complete artificial feeding. Thus, a purely breast-fed group would contain mainly babies who are doing well; those who are doing badly would often be transferred at different ages to the artificially fed group to the detriment of the latter in statistical comparisons. Even if one were to preclude any such selective additions to the artificially fed group, one would still have by the withdrawals from the breast-fed group a differentially selected group of breast-fed babies – unless the babies selected for artificial feeding were deliberately retained in the breast-fed group in spite of the change of feeding; but such a group could not truly be designated 'breast-fed' and a comparison would be valueless.

Such selective factors must be closely considered in interpreting results in this field.

Self-selection

An interesting real-life example of what may be termed *self-selection* is worth notice for it is characteristic of many a modern problem in occupational medicine. In its early work the Industrial Pulmonary Disease

Committee of the Medical Research Council was concerned with the problem as to whether the working health and capacity of coal-miners were impaired by the inhalation of anthracite dust. To begin with, a study was made of the size and age constitution of the working population at the South Wales anthracite collieries. Underground workers had to be excluded from the investigation, since such workers, as well as being exposed to various concentrations of anthracite dust in the atmosphere, may in addition have been exposed to silica or other dust. If the health of these workers was found to be impaired it would be impossible to implicate the anthracite dust as the responsible agent. The impairment might equally well be due to exposure to stone dust containing silica, which, it is well known, can produce serious damage to health. In addition, it was considered necessary to exclude surface workers who had at any time worked underground, since the effects of exposure to silica dust will not necessarily be immediately apparent and also because impaired health may have been the reason for transference from underground to surface work. This was, in fact, known to be the reason in numerous cases, so that such workers would be a highly select group. Attention was therefore turned to workers who were exposed to anthracite dust on the surface and had *always* worked on the surface. Such workers, it was found, are employed on a relatively light task. Not only, as stated, was there a tendency to draft to it operatives who had previously worked underground and had for one reason or another become partially incapacitated, but in addition it was clear that a large number of boys were initially employed upon this work but rapidly moved away to other work. In the main these boys were drafted underground where the physical labour was heavier but the rate of pay superior.

The inevitable inference is that the healthy and strong individuals will transfer to underground work while those who remain on the surface are likely to be of under-average physique and health. In other words, there has been a form of self-selection. If the examination of such surface workers showed that they included a high proportion with impaired health or that they suffered an unduly high rate of sickness in comparison with some standard, this result could not with security be ascribed to the effects of dust inhalation. It might be considerably influenced by the fact that these surface workers were, through the operation of selection, initially less healthy than a random sample of all surface workers. Such features of physical (and mental) selection must necessarily bear heavily upon the characteristics of persons entering, and remaining in, all specific occupations.

This investigation in South Wales also revealed an example of selection through volunteers being accepted for examination. For the reasons outlined above no inquiry was at that time made at the collieries and the field of study was transferred to dock-workers exposed to anthracite dust. Of 250 such workers it was arranged to examine, clinically and radiologically, a sample of 40 operatives – namely, 15 workers employed for only 3–4 years and 25 older workers with 15–40 years' service. These two groups were selected at random from the complete list of operatives, to ensure, as far as possible with such small numbers, a representative sample. At the examination eleven of these men were absent and to make good the deficiency in numbers volunteers were secured in place of the absentees.

The results of the examinations suggested that there was a readiness to volunteer on the part of individuals who, on account of some known or suspected disability, desired to be medically examined. Such substitutions may, therefore, result in the sample ceasing to be random and representative of the population from which it was drawn. The stipulated *quantity* had been maintained but the *quality* had been lost.

It should also be noted that if no substitutes had been accepted we still would not know that the 29 operatives who did attend were a cross-section of the originally chosen 40. Why did the 29 choose to come, why did the 11 stay away? If it was for any reasons connected directly or indirectly with their state of health, then the sample attending *must* give a biased picture. The neurotically inclined may more readily seek medical examination or they may more readily avoid it. The physically fit man may regard it as a waste of time; or, having nothing to fear he may be quick to submit. We do not know (or very rarely know) the motives at work and the greater, therefore, the departure from the original sample, the greater must be our doubts in interpreting the results we record.

Volunteers for Treatment

For very similar reasons a sample which is composed of volunteers, i.e. self-selected individuals, is not likely to be representative of the population at large. If, for example, the prevention of the common cold by a vaccine is offered to some specified group of persons, the volunteers are likely to belong mainly to that section of the group which suffers most frequently or severely from the complaint. They hope for some advantage from the treatment. Those who have been free from colds for a long time are unlikely to come forward. Those vaccinated are, therefore, a select group, not comparable with the remainder of the population from which

they were drawn (or, more strictly, drew themselves). In such a position the question must always arise: is the act of volunteering correlated with any factor which may influence the final results of the experiment? In the present example, if the vaccinated volunteers are mainly the common cold 'susceptibles' and the non-volunteers are mainly the 'resistants,' then clearly the contrasts between their attack rates that we might make to measure the value of the vaccine are quite meaningless.

Questionnaires

Inquiries carried out by means of questionnaires are *par excellence* those in which selection must be suspected. In all such inquiries replies to the questions put – even to the simplest question – are received from only a proportion of the individuals to whom the form is sent. There can never be the slightest certainty that the individuals who choose to reply are a representative sample of all the individuals approached; indeed very often it is extremely unlikely that they are representative. For a simple example, one may take a careful inquiry made by the Editor of *The Lancet* into the Present-day Openings of Medical Practice. To measure the success which recent graduates had achieved in their profession a questionnaire of three relatively simple questions was addressed to the 1490 men and women who in a specified year registered their names with the General Medical Council, viz. (1) What branch of medicine have you taken up? (2) What led you to this choice? and (3) What was your approximate income from professional work last year? To overcome objections to providing such personal information no clue to the identity of the correspondent was required. Yet of the individuals approached only 44 per cent replied. Are these persons a representative sample of the 1490? It is possible, as is clearly pointed out in the report, that there might be a tendency for those who have been successful in their profession to be more eager to register their success than for those who have failed to register their failure. Alternatively, the latter might under the veil of secrecy be glad of the opportunity of stating frankly the drawbacks of the profession. Those who have turned to other professions might tend not to reply, under the impression that the inquiry cannot concern them. Successful and busy individuals might be unwilling to give time to the inquiry. It is impossible to determine whether any such factors are operative in the determination to answer or not to answer. The difficulty is inherent in all inquiries carried out by this method and must never be ignored.

It almost invariably *is* ignored by the daily press, which will report that 70 per cent of some group think, for example, this or that about

atomic warfare, the National Health Service, or the President of the United States, and overlooks the fact that the 70 per cent is based upon the quite small proportion of the group who chose to answer – and almost invariably it never reports the percentage who chose to answer. The figures are utterly misleading.

It is possible sometimes, however, to see whether the final sample is or is not biased in certain known respects. For example, suppose the population to be approached consists of all the persons on the medical register at a given time. For each of these persons we may know such characteristics as sex, age at qualification, degrees or other qualifications obtained, type of medical work upon which the person is engaged – general practice, public health, etc. Only 50 per cent of the total population, let us suppose, answer the questionnaire addressed to them all. In the statistical analysis of the available answers we can at least see, and it is of course essential to do so, whether the sample is representative of the total population in relation to the *known* characteristics of the latter. If 50 per cent of the men and 50 per cent of the women answered, then the sample obtained is not biased in its sex ratio; but if 60 per cent of the men and only 25 per cent of the women answered, then a bias has been introduced, for the ratio of males and females in the sample is different from the real ratio in the population. We must make some allowance for that fact in analysing the results and cannot merely use the sample as it stands. Similarly, we may see whether older and younger persons answered to an equal degree and whether those engaged upon different types of work were equally forthcoming. By such means we can then determine whether or not certain classes of persons have tended to answer more or less readily than others, and thus know whether or not our sample is biased in these known respects and, if necessary, make allowances for it. While such a check is highly important, indeed essential, it cannot be entirely conclusive. Even if the sample *is* representative in the known respects, we cannot be sure that those who chose to answer were in other respects representative of the total. For instance, 50 per cent of men and 50 per cent of women may answer, but in *each* group those who answer may mainly consist of those who feel more deeply upon the questions addressed to them, or be those, *mirabile dictu,* who like filling in forms. In other words, the sample is correct in its sex proportions, but for neither sex do we know that the sample is such that it will accurately express the views of the total men and women originally approached.

In reporting these, or similar, inquiries a statement should always be made of the number of missing questionnaires or items. Clearly if 90 per cent of the required data were obtained, more reliance can be placed upon

the results than if the proportion were only 45 per cent, and the reader should be in a position to judge.

The type of simple correction one can sometimes make for a known bias can be demonstrated arithmetically from the following hypothetical figures:—

	Male	Female	Total
Number of persons to whom questionnaires were sent	10 000	2000	12 000
Number of persons who answered	6000	500	6500
Mean Income reported by those who answered	£5000	£4000	£4923

Using the sample as it stands, we see that the mean income reported by the 6000 men who answered was £5000, and of the 500 women who answered was £4000. The mean income of all persons in the sample is, therefore,

$$(6000 \times £5000 + 500 \times £4000) \div 6500,$$

which equals £4923. But this figure is clearly too high, since men, on the average, earned more than women, and in the sample we have a ratio of 12 men to 1 woman, whereas in the total group the real ratio is only 5 men to 1 woman (due to the fact that 60 per cent of the men answered and only 25 per cent of the women). If we are prepared to believe that for both sexes those who answered were a representative cross-section of the total approached, then the correct estimate of the average income of a person must be obtained by 'weighting' the observed mean incomes of the sexes by the correct numbers of persons of each sex. Thus we have $(10\,000 \times £5000 + 2000 \times £4000) \div 12\,000 = £4833$. In other words, we are accepting the sample figures as giving a true picture for each sex, but must combine them by using the known true proportion of men to women in the population sampled, in place of the untrue proportion given by the sample.

House Sampling

Finally an interesting and still highly topical example of selection in taking a sample of houses is suggested in a report on the historic influenza pandemic of 1918–19 (Ministry of Health's Reports on Public Health and Medical Subjects, No. 4). To obtain facts as to the incidence and fatality from influenza in that great epidemic a house-to-house inquiry was undertaken in five areas of a large city, information being obtained *so far as possible* at every fifth house. However, houses which were found closed at the time of visit were ignored. But houses in which there are young children are less often found closed and this would tend to affect the age-distribution of the population recorded in the sample. Compared with the population from which it was drawn the sample would be likely to contain an undue proportion of young children and a deficit in the number of adults. Any substitution of another house for the original randomly chosen one cannot correct for this bias in the sample, and such substitutions are to be avoided in sampling inquiries.

It will be noted that a selection of this type would be difficult to foresee. It is here that the statistician has some advantage, for his experience of such inquiries makes him familiar with the methods that are likely to ensure a random sample and those that are likely to lead to one that is unrepresentative of the population from which it is taken. Workers who are unfamiliar with sampling inquiries but wish to embark upon one may, therefore, find his advice of assistance.

Summary

If we wish to generalise from some sample group of observations – which is invariably the case in real life – we must possess a sample which is representative of the population to which it belongs. In taking samples, deliberately, or in accepting samples that arise in the daily course of events, we must realise that 'selection' may occur through the operation of various factors. A selected sample is one which is not representative of the total population, in which one member of that population sampled has more chance of appearing than another, whether that bias be due to deliberate choice or unconscious selection of the members incorporated in the sample. Self-selection of the members of a group is a common form of bias, e.g. in the physical or mental status of those who follow a certain occupation. Another exceedingly common form of bias lies in the absence of some of the required records, e.g. by persons (uninterested, busy, or lazy, whatever the reason may be) who do not reply to a questionnaire. In generalising from a sample, or in making comparisons between one sample and another, the possible presence of selection must always be very closely considered.