

the squamous metaplasia was seen in large portions of the polyp as well as in the adjacent endometrium. Morules usually disappear with conservative treatment and cannot be seen in the subsequent hysterectomy specimens.^{1,3} The presence of morules in our patient might be due to incomplete removal of the stalk of the polyp during the curettage, or her hormonal status (early pregnancy) may have led to recurrence of the previous metaplastic process.

We thank Dr J. P. van der Westhuyzen, Chief Medical Superintendent of Tygerberg Hospital, for permission to publish, and Mrs J. Oosthuizen for typing the manuscript.

REFERENCES

1. Bomze EJ, Friedman NB. Squamous metaplasia and adenoacanthosis of the endometrium. *Obstet Gynecol* 1967; **30**: 619-625.
2. Crum CP, Richart RM, Fenoglio CM. Adenoacanthosis of the endometrium. *Am J Surg Pathol* 1981; **5**: 15-20.
3. Dutra FR. Intraglandular morules of the endometrium. *Am J Clin Pathol* 1959; **31**: 60-65.
4. Kurman RJ, Norris HJ. Endometrial neoplasia: hyperplasia and carcinoma. In: Blaustein A, ed. *Pathology of the Female Genital Tract*. 2nd ed. New York: Springer-Verlag, 1982: 313-315.
5. Lane ME, Dacalos E, Sobrero AJ, Ober WB. Squamous metaplasia of the endometrium in women with an intrauterine contraceptive device. *Am J Obstet Gynecol* 1974; **119**: 693-697.
6. Blaustein A. Morular metaplasia misdiagnosed as adenoacanthoma in young women with polycystic ovarian disease. *Am J Surg Pathol* 1982; **6**: 223-228.
7. Marcus SL. Adenoacanthoma of the endometrium. *Am J Obstet Gynecol* 1961; **81**: 259-267.
8. Wilson JR, Du Bois RO. Report of a fatal case of keratomalacia in an infant with postmortem examination. *Am J Dis Child* 1923; **26**: 431-446.
9. Sarlin JG, Mori K. Morules in epithelial tumors of the colon and rectum. *Am J Surg Pathol* 1984; **8**: 281-285.
10. Lauchlan SC. Metaplasias and neoplasias of Müllerian epithelium. *Histopathology* 1984; **8**: 543-557.
11. Demopoulos RI. Normal endometrium. In: Blaustein A, ed. *Pathology of the Female Genital Tract*. 2nd ed. New York: Springer-Verlag, 1982: 266-267.
12. Fluhmann CF. Comparative studies of squamous metaplasia of the cervix uteri and endometrium. *Am J Obstet Gynecol* 1954; **68**: 1447-1463.
13. Liu FTY, Lin HS, Burich RL, Wagner JE. Effects of some oral contraceptive steroids on the development of endometrial squamous metaplasia and cysts in rats. *Am J Obstet Gynecol* 1972; **114**: 685-690.
14. McEuen CS. Metaplasia of uterine epithelium produced in rats by prolonged administration of estrin. *Am J Cancer* 1936; **27**: 91-94.
15. Schardein JL, Kaump DH, Woosley ET, Jellema MM. Longterm toxicologic and tumorigenesis studies on an oral contraceptive agent in albino rats. *Toxicol Appl Pharmacol* 1970; **16**: 10-23.

Epidemiological research methods

Part II. Descriptive studies

J. L. BOTHA, D. YACH

Descriptive studies are used to quantify the extent of a health problem in a population.¹ Their use, and the types of research questions that can be answered by them, are illustrated by two local studies.

In the first, 'A comparison of the mortality rates of various population groups in the RSA', health problems of the national population were of interest:² 'Before health priorities can be determined and the health resources deployed to the best advantage, it is necessary to know what the major disease problems are. This should be based on a knowledge of the pattern of mortality and morbidity in the population.'

In the second study, 'Hypertension management and patient compliance at a Soweto polyclinic', problems in health care delivery to a specific patient population were of interest:³ 'Before starting to look for undiagnosed hypertensives in the community we decided to determine whether the service was dealing satisfactorily with its current hypertensive patients.... We therefore designed a study to answer the following specific questions....'

1. Do the Senaoane polyclinic staff measure the blood pressures of adult patients attending for the first time?

2. Do primary health care nurses manage these "first-visit"

patients in accordance with the blood pressure management protocol they have been trained to use?

3. Do patients who have started antihypertensive drug treatment return regularly for blood pressure measurement and treatment?

4. If they return regularly, are their blood pressures lowered?

In a descriptive study, therefore, the magnitude and distribution of a health problem in a specified population is studied in terms of TIME (when did it occur?), PLACE (where did it occur?) and PERSON (which groups are affected?). The design starts with an idea that occurs to the researcher about a particular problem. This is followed by selecting a group of individuals to be studied (sampling), considering which attributes to measure (measurement), describing the findings, and finally drawing conclusions on the basis of the findings. Commonly, new ideas or hypotheses are generated in this final stage, usually regarding possible explanations for the health problems described (cause-effect relationships). Such relationships may be attempts to explain the aetiology of diseases or the effect of preventive, curative or rehabilitative measures.

Important issues affecting the reliability of the sampling and measurement processes are discussed, some descriptive statistical measures demonstrated and how conclusions are affected by these, are indicated.

Sampling

It is usually not practical or financially feasible to carry out measurements on the entire population. Therefore epidemiologists usually make their measurements on a sample or subset

Institute for Biostatistics of the South African Medical Research Council and Departments of Community Health, University of Stellenbosch, Parowvallei, CP and University of Cape Town

J. L. BOTHA, M.B. CH.B., M.S.C. (CLIN. EPID.)

D. YACH, M.B. CH.B., B.S.C. HONS (EPID.), M.P.H.

of individuals from the given population and summarize their findings using descriptive statistical measures. From the sample, therefore, an estimate of the magnitude and distribution of the health problem (the effect of interest) in the population is obtained, and used to describe (or make inferences about) that effect in the population. In a general sense, any instance of disease can be regarded as the effect of a given cause or set of causes.

Careless sampling carries two penalties, reflected in the estimated measure of effect: the measure may be biased, and/or the measure may lack precision.

Sampling bias (systematic sampling error) occurs if the selected sample is unrepresentative of the population of interest, which means that sampled individuals differ systematically from those not sampled. For valid inferences to be made in a descriptive study about a population effect from an estimate of that effect in a sample, the sample must be representative of that population.

Random sampling error reflects the variation in estimated measures of effect that would occur if many different random samples were taken from the target population and a summary measure estimated from each: some estimates would be higher and others lower than the true effect in the population. Confidence limits (CL) are used as a measure of precision for estimates obtained.⁴ Generally the imprecision resulting from random sampling error is smaller the larger the sample size.

Selecting a sample for a descriptive study involves using a sampling technique that avoids systematic sampling error and minimizes random sampling error, therefore yielding an estimate of effect that is both valid (unbiased) and precise.

We mention here four commonly used sampling techniques, illustrated by sampling with each, from the same hypothetical population of interest: children in 8 rural villages, each containing 8 homesteads, with 1 child per homestead (Fig. 1). The target population (N) is therefore 64 children. Villages A, B and C have taps supplying water, villages D, E, F, G and H

have no taps. We have the resources to investigate a total sample (n) of 16 children (sampling fraction 16/64 or 25%) from which an estimate (d/n) of the incidence rate of diarrhoea (D/N) in the week preceding our survey may be obtained.* Obviously in practice both the population and sample are usually larger and the sampling fraction usually less than 10%.

Simple random sampling

A simple random sample of n individuals is a sample drawn in such a way that all sets of n individuals in the target population of N individuals have the same chance of being in the sample. A sampling frame, consisting of a list of all individuals in the population numbered from 1 to 64 (in our example), is drawn up. Drawing 16 numbers from a hat containing the 64 numbers, each on a separate bit of paper, would yield a simple random sample. The random process of drawing numbers from a hat can be replaced with a computer program, or by using a table of random numbers, but a numbered list of all individuals in the population is always necessary.

In our example, it is possible to draw many ($4,88 \times 10^{14}$) different samples of 16 children from the 64 in the population. Using tables of random numbers,⁵ we obtained sample 1 in Table I, yielding an estimated incidence rate in the preceding week of 8/16 or 500/1000, which is reasonably close to the population value of 530/1000. (Rates are commonly expressed per 1000 or 100 000 population.) However, we might have drawn sample 2 (estimated incidence rate 0/16) or sample 3 (estimated incidence rate 16/16), both obviously yielding biased or invalid estimates. The likelihood of drawing such unrepresentative samples with simple random sampling is small, but real.

In Table I the precision of these point estimates is also shown as 95% CL (interval estimates). They indicate the range of incidence rates that contain the population incidence rate with 95% certainty. In these three samples of size 16 discussed above, the estimates are all imprecise even though their estimated values vary widely. Increasing the sample size evidently improves the precision. Our example is a hypothetical one with a very small population. Much greater precision is possible in real studies with larger populations from which larger samples can be drawn.

Stratified random sampling

Although randomness in drawing a sample is necessary to increase the likelihood of obtaining a representative sample and therefore an unbiased summary measure, it results in random sampling error which reduces the precision of the summary measure. If prior information about attributes related to the effect being measured is available for the target population, it can be used to sample in such a way that the variability can be reduced (and precision improved) without introducing bias. Such attributes typically include age, sex or socio-economic status. This is done in stratified random sampling, where the target population is divided into strata and a random sample of individuals drawn from each stratum. A set of strata is a collection of subsets of individuals in the target population such that each individual belongs to one and only one such subset. In stratified sampling, a sampling frame is constructed for each stratum, so the size of each stratum in the population must be known.

In our example, 3 villages have taps ($N_1 = 24$ children) and the other 5 do not ($N_2 = 40$ children). The diarrhoea incidence

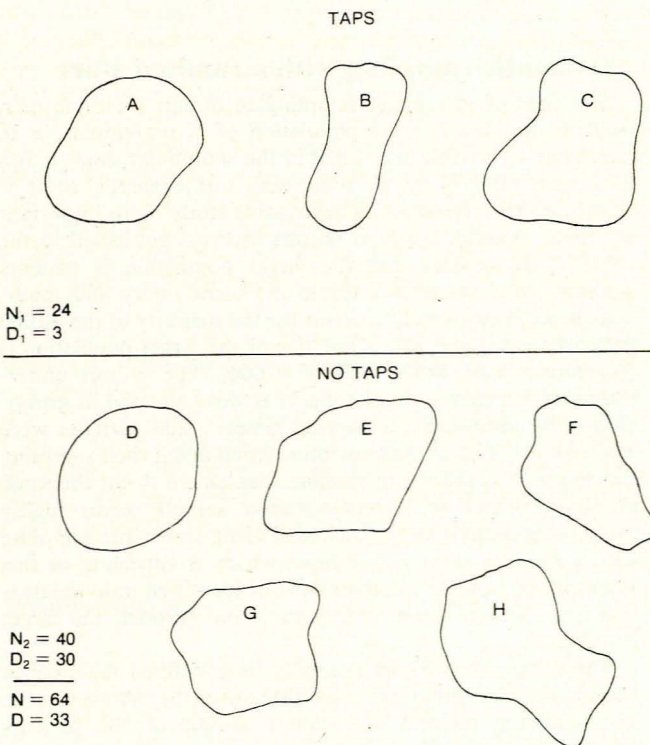


Fig. 1. Hypothetical target population used to illustrate sampling techniques (N_1 , N_2 , N are the number of children at risk in villages with taps, without taps and all villages respectively; D_1 , D_2 , D are the numbers of children with diarrhoea).

* 'D' refers to the total number of children with diarrhoea in the target population of size N; 'd' refers to the number of children with diarrhoea in a sample size 'n'.

TABLE I. VARIATION IN POINT AND INTERNAL ESTIMATES OF DIARRHOEA INCIDENCE RATES WITH VARYING SAMPLE SIZES (INCIDENCE RATE/1 000 CHILDREN)

Sample (size)	Point estimate	Validity*	Exact 95% CL ⁶	Precision
1 (16)	500	Good	247 - 754	Bad
2 (16)	1 000	Bad	794 - 1 000	Bad
3 (16)	0	Bad	0 - 206	Bad
4 (32)	500	Good	319 - 681	Better
5 (100)†	500	Good	398 - 602	Better yet
6 (1 000)†	500	Good	469 - 532	Good

*By comparison with population incidence rate (here known to be 530/1 000, but usually unknown).
†If such large samples were possible.

rate is likely to be higher in those villages without taps. If a 25% random sample of children from each stratum of villages were taken, a weighted average estimate of incidence rate for the total sample can be obtained. In this technique sampling is proportionate to the population distribution: selecting 6 children (n_1) from the villages with taps and another 10 (n_2) from those without taps, an estimated incidence rate in the previous week of 8/16 or 500/1 000 was obtained. This point estimate was obviously as unbiased as that from simple random sampling, and was more precise (Table II). The sampling variability of the stratified random sample is less than that of a simple random sample, because the variability within strata is less than among all patients. If prior information about other attributes related to the incidence rate of diarrhoea were known, more strata could have been used.

TABLE II. VARIATION IN THE PRECISION OF POINT ESTIMATES OF DIARRHOEA INCIDENCE RATES WITH VARIOUS SAMPLING TECHNIQUES (INCIDENCE RATE/1 000 CHILDREN)

Sampling technique	Sample size	Point estimate	Approximate 95% CL ⁴
Simple random	16	500	247 - 754
Stratified random	16	500	290 - 710
Cluster	16	500	10 - 990

Apart from the improvement in precision of the overall summary measure, it is possible with stratified sampling, to obtain stratum-specific estimates of the effect of interest, to describe its variability according to strata. In our sample 1, our estimated incidence rate in villages with taps was 1/6 or 167/1 000 (95% CL 4-641/1 000) and in villages without taps 7/10 or 700/1 000 (95% CL 348-933/1 000). This variation is as expected, although the estimates are imprecise. The precision of stratum specific estimates can be improved either by selecting equal sample sizes per stratum (balanced) or by selecting sampling sizes proportional to the stratum specific variances (optimal), if these are known.⁷

Cluster sampling

Cluster sampling refers to sampling clusters (usually self-defined subsets) of individuals in the target population randomly and then sampling individuals within the selected clusters. When clusters represent geographically compact sets of individuals, the interviewers may spend more time interviewing than travelling. For this reason, it is often used in rural research,^{8,9} where travelling can be particularly time-

consuming. In our example, we may regard each village as a cluster, and may wish to select 2 of the 4 clusters and 2 children in each cluster. Doing this, we selected villages A and E, and 8 children in each, yielding an overall incidence rate of 8/16 or 500/1 000. Here, an unbiased point estimate was also obtained, probably because 1 village with and 1 without taps was selected, by chance. It is possible with cluster sampling, though, to select both villages with or without taps, in which case a biased estimate is likely.

In cluster sampling, sampling variability arises from two sources (the sampling of clusters and the sampling of individuals within clusters) and it is generally larger than in a simple random sample. The precision of the summary estimate is therefore usually worse than in a random sample (Table II). To achieve the same precision with cluster sampling as would be possible with simple random sampling, previous studies suggest that a sample size of approximately twice the size (design effect = 2) of the simple random sample would be needed.⁹

Systematic sampling with a random start

The idea of systematic sampling to obtain a sample of n individuals from a target population of N individuals, is to select every (N/n) th individual in the sampling frame. It is a technique unlikely to be used with our example, so it is illustrated by referring to a descriptive study of an important problem, 'Alcohol levels in trauma victims' published in the *SAMJ*.¹⁰ It appears that the target population is patients admitted to hospital as a result of violent injury and motor vehicle accidents which 'account for the majority of the 40 000 patients treated annually'. The size of the target population is therefore at least 20 400 (51% of 40 000). The authors understandably selected a sample: 'patients were assessed in groups of 5 - 10 admissions at varying times... 200 patients were assessed...'. This is the only information about their sampling technique: it appears non-random (haphazard is not the same as random) and an unrepresentative sample seems highly likely. This could have been avoided using systematic sampling with a random start, a technique which is appealing in this situation, because no numbered list of identified individuals is necessary, and it spreads the sample out through the target population.

The authors could, for example, have defined their target population as trauma victims admitted over the following year. The sampling interval to obtain a sample of 200 patients would be $20\,400/200 = 102$. The starting point is selected randomly by generating a random number between 1 and 102. If this number is, say, 21, the first patient to be selected is the 21st admission after the start of the defined year, the next is then the 123rd ($21 + 102$) patient, and so on until 200 patients

have been sampled. If the investigators decide that a year is too long for the intake of patients and specify a 2-month period ($N = 20\,400/6 = 3\,400$ patients) instead, the sampling interval for a sample of 200 patients will be $3\,400/200 = 17$. This sample will not be representative of the target population if these 2 months are atypical, e.g. December and January.

Using systematic sampling with a random starting point a representative sample will be highly likely, if the sampling frame is constructed as the result of a random process (here order of admission to Tygerberg Hospital Trauma Unit). If the order of individuals (and therefore the variation of the attribute being measured) in the sampling frame is related to the sampling interval, and depending on the exact relationship, a biased summary measure may be obtained and its precision may be more or less than with a random sample.

Combined techniques

It is sometimes possible and indeed advisable to combine different sampling techniques, e.g. stratified random and cluster sampling. When this seems necessary, or indeed, when any technique other than simple random sampling is used, advice on sampling and estimation procedures should be obtained, from biostatisticians, epidemiologists or by referring to standard texts.^{7,11,12}

Non-response after sampling

Again we can refer to the article on alcohol in trauma victims:¹⁰ '200 patients were assessed with complete data being available on 142.' Here the authors tell us that their 'response' rate was 142/200 or 71% and this implies that the summary estimate of alcohol consumption reported may be biased, if those on whom complete data were not available ('non-respondents') differed systematically from the rest ('respondents') in attributes related to alcohol assumption. If the non-respondents were more commonly those who were unconscious on admission, and therefore more (or less) likely to be heavy drinkers, the summary measure of alcohol consumption will be accordingly biased (too low or too high). In another article published in the *SAMJ*¹³ ('Smoking habits and attitudes of Durban metropolitan anaesthetists'), the respondents (78% of those surveyed), may be mostly those anaesthetists who do not smoke, particularly because they are well aware of the dangers of smoking. If we assume the worst case (all non-respondents smoke) the authors' estimate that 58% (59/102 respondents) had smoked at some time, could increase to a maximum of 67% (87/130).

The so-called non-response bias¹⁴ is one of the most commonly occurring biases in epidemiological studies, and it can invalidate results, especially if the response rate is less than 80%. The same applies to volunteer bias,¹⁴ which may occur if those who refused to give consent for a blood sample to be taken, differed systematically from those who 'volunteered'. No information about the consent rate is given in the publication, although consent was obtained.

Measurement

Lord Kelvin's statement, quoted by Beiser,¹⁵ on the importance of measurement in science applies equally well to epidemiology: 'I often say that when you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of Science, whatever the matter may be.'

In both local studies referred to earlier,^{2,3} health problems were expressed as counts. Events or people were counted: deaths, obviously, in the study of mortality patterns,² and, *inter alia*, clinic visits in the study of hypertension management and patient compliance.³ Such simple counting procedures are commonly used, and as a result people can usually be classified into discrete categories, either simple nominal classes such as for population groups² or gender,³ or classes indicating some order, e.g. socio-economic classes I - V, or the age groups used in both studies.^{2,3} Apart from counting events, attributes can also be measured by a process of mensuration: e.g. the blood pressure measurements in the hypertension study. Here the measurement is on a continuous scale, and not in discrete categories. Such continuous measurements can, however, be categorized into discrete categories, as was done to summarize the distribution of blood pressure measurements in the hypertension study.³

Careless measurement on sampled individuals carries two penalties similar to those resulting from careless sampling: measures may be invalid, because the measurements were made incorrectly, and/or measures may lack precision, because measurements were made unreliably.

Systematic measurement error or measurement bias occurs when, e.g. an observer or instrument measures an attribute in the same individual repeatedly higher, or repeatedly lower, than the real value. Random error occurs when measurements are spread around the 'true' value being estimated and results in imprecise (or unreliable) summary measures of effect. Random measurement error refers to the concordance reached in repeated measurements (in this pure sense, irrespective of how close to the true value they are). Inter-observer reliability refers to the concordance between two (or more) observers if they use the same instrument to measure an attribute in the same individual.

In Fig. 2 the way systematic error affects the validity of, and random error affects the precision of, estimated effects is shown. The centre of the dart board represents the true population effect, and each dart in the board represents an attempt to estimate that effect in a sample.

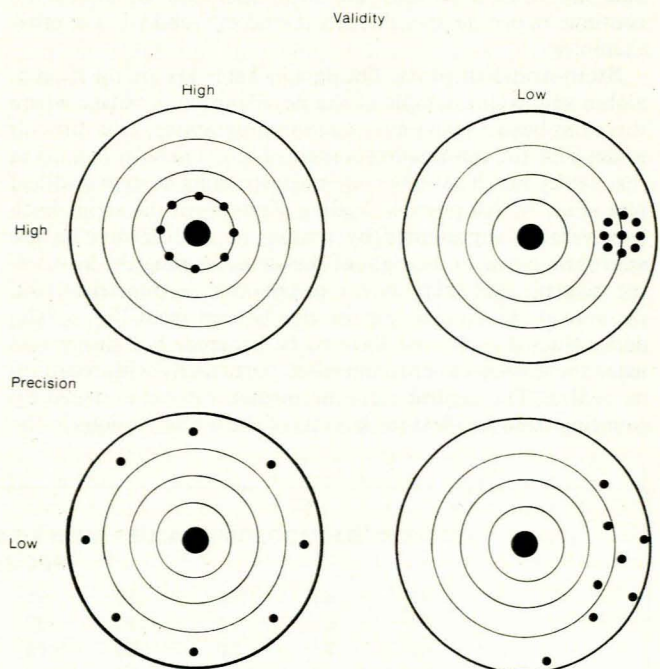


Fig. 2. The effect of systematic and/or random error on the precision and validity of measures of effect estimated from samples (adapted from Ahlbohm and Norrell,¹⁶ with permission from the publishers).

In any descriptive study, it is important to identify potential sources of both systematic and random measurement error. Their impact on the estimation of the effect of interest should then be predicted. This can usually only be done by measuring the impact and its direction (obscuring or magnifying the effect) in a pilot study. If the impact is trivial, the source of error can be ignored. If not, the error should be minimized by training or excluding the observers responsible.

Questionnaires should be regarded as survey instruments which are also subject to systematic and random measurement errors. Interviewers need to be trained (calibrated) and the instrument needs to be checked for reliability and validity.

Descriptive statistical measures

In a descriptive study the effect of interest is usually the magnitude and distribution of a health problem in a specified population. This effect is then estimated from measurements made on a sample of individuals. Descriptive statistical measures of effect are used to summarize the findings so that conclusions can be drawn about the effect in the population. Epidemiologists, therefore, first explore the distribution of the raw measurements made on the sampled individuals, and secondly, they use those measurements to estimate measures of effect to describe the population of interest.

Exploring raw measurements

Epidemiologists are usually interested in three key properties of the distribution of their continuous measurements: the centre of the distribution, its spread and its shape. Three types of graphical display, and summary measures relevant to them, will be discussed here: stem-and-leaf plots, box plots and histograms. These graphical displays all have the advantage of simplicity (as few lines or symbols as possible), and should be self-explanatory (clear title, labelling, footnotes). In stem-and-leaf plots all the data points are displayed, while not all data points are displayed in box plots or histograms. Because raw data are needed to illustrate these displays, we could not continue to use the two published studies^{2,3} and will use other examples.

Stem-and-leaf plots. The data in Table III are the haemoglobin values in a sample of the population of a village where there has been a malaria eradication programme. It is difficult to describe the information contained in the data by looking at that table, but it is easier after constructing a stem-and-leaf plot (Fig. 3). All possible leading digits form the stem. Each data value is represented by writing its trailing digit in the appropriate row to the right of the stem, forming the leaves of the display. The shape of the distribution (symmetrical) and the spread (35 - 123%) can be clearly seen from Fig. 3. The data values do not first have to be ordered, but this would make the display easier to interpret, particularly with regard to its centre. The central value or median can be obtained by counting from smallest (or largest) of the n observations in the

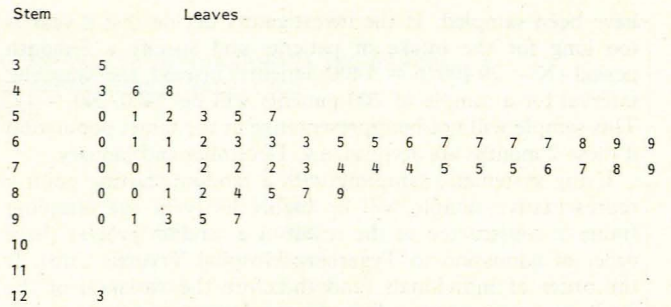


Fig. 3. Ordered stem-and-leaf diagram of the percentage haemoglobin values of a sample of a village population where there has been a malaria eradication programme.

data set until the $((n + 1)/2)$ th value is obtained. In Fig. 3, with $N = 60$ and counting down from 35%, the median is the $(61/2)$ th value, or that value that would lie midway between the 30th (71%) and 31st (70%), i.e. 70.5%. Other summary information can also be obtained from Fig. 3. The most commonly occurring values can clearly be seen, and the value 123% appears removed from the rest. The latter is technically called an 'outlier' and it should be investigated as a possible coding or transcribing error, and corrected if needed. The median divides the data into halves and the values that split each half into quarters are called quartiles, lower (25%) and upper (75%) quartiles. The distance between these, the interquartile range, is another measure of the spread of the distribution. The median and the quartiles are examples of quantiles, where a quantile refers to a fraction of the data, e.g. the lower quartile refers to 0.25 of the data. Percentiles are also quantiles, the fractions being transformed to percentages, e.g. the median is the 50th percentile.

Box plots. In the box plots of the same data set (Fig. 4), the upper and lower quartiles are displayed by the top and bottom of a rectangle (which represents the central 50% of the data), and the median is portrayed by a horizontal line within the rectangle. Dotted lines extend from the ends of the box to defined points, usually the extremes. The box plot gives a quick impression of the key properties of the distribution: the median shows the centre, and the length of the box or the extremes the spread of the distribution. If the median divides the box into two equal lengths, the distribution is likely to be symmetrical, otherwise asymmetrical. Box plots are particularly useful for displaying the distributions in several groups for comparison.

Histograms. A histogram is a commonly used display, and to construct it, the range of data has to be grouped into classes. The shape of the histogram and the detection of outliers depend on the number and the width of intervals chosen. The number of observations in each class is plotted as the height of the bar (Fig. 5).

Other. A distribution should always first be displayed graphically and inspected, before summary measures are calcu-

TABLE III. HAEMOGLOBIN VALUES (%) IN A SAMPLE POPULATION IN MALARIA ERADICATION PROGRAMMES

43	63	63	75	95	75	80	48	62	71
51	61	74	123	93	82	74	65	63	53
80	77	60	69	73	76	91	55	65	69
50	68	72	89	75	57	66	79	85	70
87	67	72	52	35	67	99	81	97	74
67	46	84	59	61	90	67	78	70	72

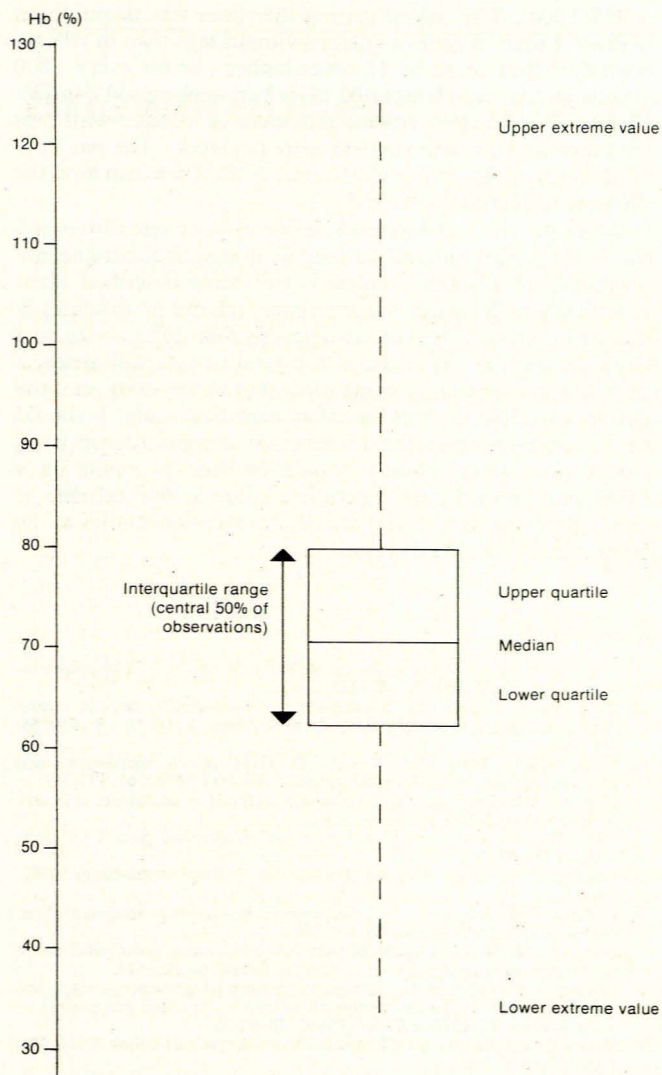


Fig. 4. Box plot of haemoglobin levels (%).

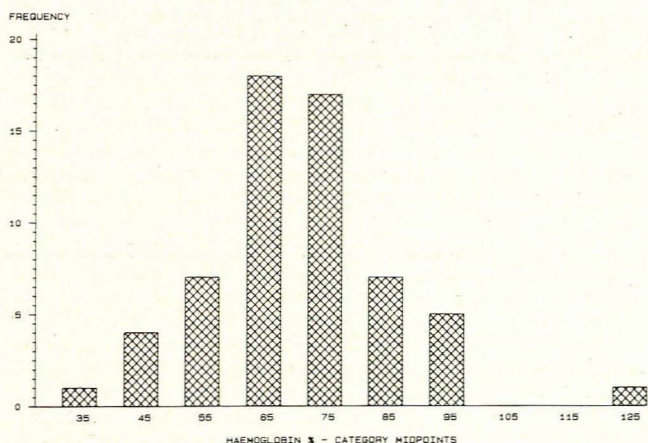


Fig. 5. Histogram of haemoglobin levels (%).

lated, because the shape of a distribution may determine which summary measures are to be used. The median and the interquartile range can be used irrespective of the shape of the distribution. Because they are measures based on the rank order of observations, they are less affected by outliers or the shape of the distribution than other commonly used summary

measures such as the arithmetic mean (measure of centrality) or the standard deviation (measure of spread) which are based on the intrinsic value of the observations. These latter two measures should not be used if the distribution is asymmetrical, and should, strictly speaking, only be used if the distribution is Gaussian.

Estimated measures of effect

In exploring the raw data with graphical techniques, information about some measures of effect may be obtained directly. For example, if the attribute of interest was the median haemoglobin level in the population, a point estimate of 70,5% could be obtained from the stem-and-leaf plot in Fig. 3. Alternatively, the arithmetical mean haemoglobin value could have been estimated (71,1%), and the variance (a measure of spread)⁷ of the sample distribution used to estimate the random error of that point estimate, or conversely, its precision (95% CL 67,2 - 75,0%).

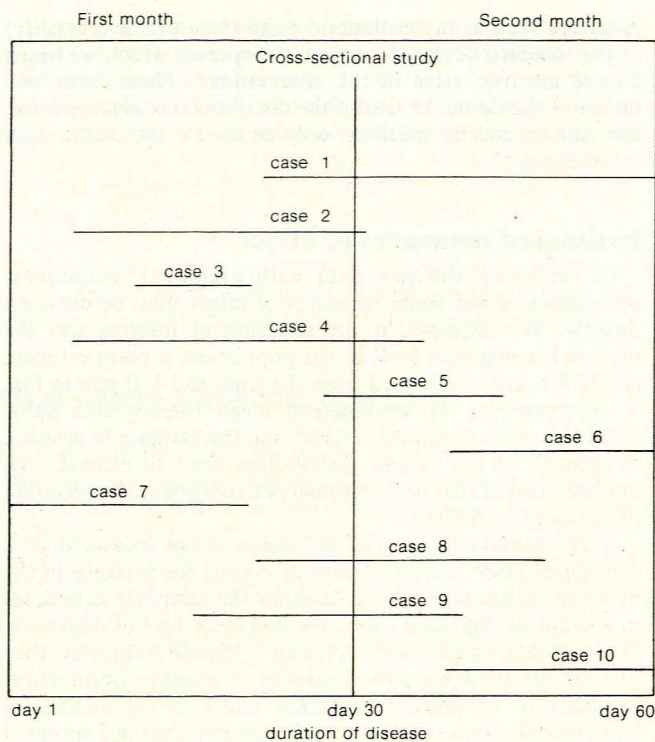
Very commonly the effect of interest is not measured on a continuous scale, but as a count of events, for instance in the hypothetical example used to illustrate the sampling techniques as a count of diarrhoea cases, the incidence rate of diarrhoea. When epidemiologists are interested in disease frequency, they distinguish between two measures of disease occurrence: incidence or prevalence. Incidence refers to the number of longitudinal events (e.g. new cases) reported during a specified period in a defined population. The source of these events can be follow-up studies or routine surveillance systems (notifications, registration of births and deaths). Incidence is thus a dynamic measure, in contrast to prevalence, which is a static one. The prevalence of a disease refers to the number of people found to have a particular condition at a specified time. Prevalence usually refers to point prevalence (i.e. the number of cases found at a particular time), but in practice, prevalence studies are carried out over a short period of time.

Epidemiologists are usually interested in comparing populations (or the same populations at different times) with regard to the incidence or prevalence of a condition. In descriptive studies comparisons with other studies are usually implied. To make valid comparisons, incidence or prevalence rates are used, in which the numerator is incidence or prevalence as defined in the previous paragraph. The denominator of the incidence rate is derived from studies of the average population at risk or the census and refers to the number of people at risk of an event during a specified period. To calculate the diarrhoea incidence rate in our example, not only the number of cases occurring in a certain time and at a particular place was required, but also the number at risk. To calculate the measles incidence rate, all those immunized against measles and those who have previously had measles should be excluded from the population at risk. In contrast, the denominator of the prevalence rate is derived from a population count estimated at the time of the survey, and refers to all people examined for presence or absence of a condition or capable of having it.

In reality, only incidence rates derived from deaths represent a true incidence. Usually rates are derived from a combination of sources (a screening programme and hospital records for example) and are more correctly termed occurrence rates. The difference between incidence and prevalence is graphically illustrated in Fig. 6. It is clear that a disease of short duration is more likely to be under-represented in the prevalence rate than one of long duration. The relationship between prevalence, incidence and duration can be stated algebraically by:

$$\text{prevalence} \approx \text{incidence} \times \text{duration}$$

Comparison of rates can be made as a rate ratio, indicating how many times more frequently the disease occurs in one group than in another, or as a rate difference, indicating the excess of occurrence in the group at a higher risk. The rate



Denominator = for the whole time period
 Prevalence on day 30 = $6/100 = 0.06\%$
 Incidence for day 1 to day 30 = $8/100/\text{month} = 0.08\%/\text{month}$
 Incidence for day 31 to day 60 = $2/100/\text{month} = 0.02\%/\text{month}$

Fig. 6. Graphic demonstration of incidence v. prevalence. (The length of the line under the case indicates the duration of the disease.)

ratio, or rate difference can be seen as the effect of interest, also in a descriptive study. For example, using a stratified random sampling technique, we obtained stratum-specific diarrhoea incidence rates in the preceding week of 167/1 000 and 700/1 000 in villages with and without taps, respectively. The estimated rate ratio for villages without taps relative to those with taps is therefore 4.2 (95% CL 1.53 - 11.56) and the estimated rate difference 533/1 000 in that week (95% CL 121

- 945/1 000). The risk of getting diarrhoea was therefore on average 4 times higher in villages without taps than in villages with taps (but could be 11 times higher), or for every 1 000 people at risk, on average 500 more per week would contract diarrhoea in villages without taps than in villages with taps (but it could be as many as 945 more per week). The rate ratio is dimensionless, whereas the rate difference conveys the absolute magnitude of the risk.

When the effect of interest is a rate ratio or rate difference, the study design can still be seen as descriptive, because the magnitude of a health problem is still being described albeit specifically with regard to a difference (relative or absolute) in disease occurrence between two groups that differ owing to a defined exposure. By using a rate ratio or rate difference, a comparison is obviously being made within the study, and the design is tending towards that of an analytical study. It should be remembered that the three basic designs (descriptive, analytical or intervention¹), should be seen as points on a continuum, with a pure descriptive study at one extreme of that continuum. In the next article, intervention studies as the other extreme, will be discussed.

REFERENCES

1. Yach D, Botha JL. Epidemiological Research Methods: I. Why epidemiology? *S Afr Med J* 1986; **70**: 267-270.
2. Wyndham CH, Irwig LM. A comparison of the mortality rates of various population groups in the Republic of South Africa. *S Afr Med J* 1979; **55**: 796-802.
3. Saunders LD, Irwig LM, Wilson TD. Hypertension management and patient compliance at a Soweto polyclinic. *S Afr Med J* 1982; **61**: 147-151.
4. Gore SM. Statistics in question: assessing methods — confidence intervals. *Br Med J* 1981; **283**: 660-662.
5. Fleiss JL. *Statistical Methods For Rates and Proportions*. New York: John Wiley, 1973: 35-38.
6. Lentner C, ed. *Geigy Scientific Tables*. Vol. 2. Basle: Ciba-Geigy, 1982: 89-102.
7. Anderson TW, Sclove SL. *An Introduction to the Statistical Analysis of Data*. Boston: Houghton Mifflin, 1978: 638-659.
8. Irwig LM, Ingle RF. Childhood mortality rates, infant feeding and use of health services in rural Transkei. *S Afr Med J* 1984; **66**: 608-613.
9. Lemeshow S, Robinson D. Surveys to measure programme coverage and impact: a review of the methodology used by the expanded programme on immunisation. *World Health Stat Q* 1985; **38**: 65-76.
10. Muller R, Van Rensburg LCJ. Alcohol levels in trauma victims. *S Afr Med J* 1986; **70**: 592-593.
11. Levy PS, Lemeshow S. *Sampling for Health Professionals*. Belmont, Calif.: Lifetime Learning Publications, 1981.
12. Kish L. *Survey Sampling*. New York: John Wiley, 1965.
13. Callander C, Locke DA. Smoking habits and attitudes of Durban metropolitan anaesthetists. *S Afr Med J* 1986; **70**: 589-591.
14. Sackett DL. Bias in analytic studies. *J Chronic Dis* 1979; **32**: 51-53.
15. Beiser A. *The World of Physics*. New York: McGraw-Hill, 1960.
16. Ahlbohm A, Norell S. *Introduction to Modern Epidemiology*. Chestnut Hill, Md: Epidemiology Resources, 1984.