JIA 79 (1953) 0129-0169

THE INSTITUTE OF ACTUARIES

THE USE OF SAMPLING IN PENSION FUND VALUATIONS

By F. GORDON SMITH, M.A., F.I.A.

of the Government Actuary's Department

[Submitted to the Institute, 26 January 1953]

INTRODUCTION

THE application of sampling technique to statistical data, which a few decades ago was a rarity, has now become common practice. Our past President, Mr F. A. A. Menzler, in his presidential address said:

Sampling, in all its manifestations, has become almost a social and industrial routine.

In this paper, by examining in some detail the application of sampling to a particular branch of actuarial work, namely pension fund valuations, an attempt is made to investigate the possibilities of its use in the actuarial field. By the choice of pension fund valuations, which are a particularly complicated example of the actuarial technique, the investigation will, it is hoped, give a good insight into the sort of problems involved.

THE PROBLEM DISCUSSED

2. If a sample is contemplated in an actual pension fund valuation, the results for the whole will not be known and the error in using a sample will remain an inherent uncertainty. Experiments were made with pension fund data which, by using first the whole data and then sample data, give some guidance on the extent of the errors introduced. A sample may be used in the actuarial valuation of a fund, provided that the knowledge derived from the result materially assists in the management of the fund concerned. If this is accomplished, then progress has been made. If it were possible to reduce the degree of uncertainty introduced by sampling sufficiently to accomplish all that is really required of a valuation, then a case would be made out for the sample investigation being an adequate substitute for a full investigation, providing that there was a real saving of labour.

One of the first steps in a valuation is to compare the actual retirements, deaths and withdrawals over the inter-valuation period with those expected according to the last valuation basis. At the last valuation the basis may have been obtained directly from the experience of the fund—this is likely only with a large fund—or may have been derived from a consideration of the experience of the fund in relation to rates obtained from other sources. The valuation basis is not usually changed on account of minor deviations of the actual experience of the fund from that expected; the effect of such minor deviations would be re-examined at subsequent valuations, when it might be found that they had either been swallowed up by other changes or had grown sufficiently to require adjustment of the valuation basis. On the other hand, changed conditions and circumstances may have caused such large deviations that a new valuation basis is necessary. When a suitable basis has been determined the valuation of the tabulated data as at the valuation date proceeds. No reference is made to errors in the valuation result caused by random

deviations in the experience underlying the valuation basis. It is assumed that these would not be large enough to alter the conclusions which might reasonably be drawn from the results of the valuation, and that they would be comparable in size with the minor deviations in experience already mentioned for which no change in valuation basis would be made.

3. The data an actuary requires in order to proceed with the valuation can be divided conveniently into two parts:

(i) the statistics relating to the inter-valuation period;
(ii) the statistics of members and pensioners existing on the valuation date.

The compilation of the necessary data for either part is usually based on actuarial advice, but the actual work is the concern of those administering the pension fund and involves non-actuarial clerical labour. With a large pension fund the volume of work may be heavy and much of it of a mechanical kind. Sampling will effect a considerable saving in this labour, although closer supervision will be necessary.

There is not likely to be any saving in the computing work of the valuation if the data are supplied in sample form. Once data are tabulated the actual size of the numbers does not materially affect the amount of time involved in the work and, indeed, as the sample data are subject to wider marginal errors than the full data would be, the actuarial side of the valuation work may be made more difficult because the crude rates of decrement and of salary progression are more uneven.

THE EXPERIMENTAL APPROACH

4. A valuation group of 80,073 male lives in service on the valuation date was available as a universe, together with corresponding movements during the inter-valuation period. Random sampling was effected by means of the reference number of each life. These numbers had been allocated in sequence according to the order in which particulars of each new entrant had reached a central office—except that the first two digits represented the year of birth. A 10% random sample was drawn by selecting all lives whose last digit was a 4. This sample totalled 8053 lives, differing from 8007 (one-tenth of 80,073) because the movements of the '4's' since entry into service had not corresponded exactly with those of the universe. However, the sample is a random one and there should be no bias.

The sample of 8053 lives was itself used as a universe for the drawing of 10% samples. A pilot sample was first drawn by selecting all lives whose tens digit was a 3 (all last digits of this universe were 4's). This sample was subsequently replaced and the 8053 lives were sorted into 100 groups by means of the tens and hundreds digits. Ten further samples were then selected by means of sets of ten random numbers between 0 and 100 taken from the tables prepared by Prof. M. G. Kendall and Mr B. Babington Smith. In order to assess with any confidence the limits of the errors likely to be encountered experiments are necessary with a number of samples. If all the samples had been selected by means of the tens digits (as the first one was) ten of such samples would have amounted to the whole universe and samples subject to

this constraint would not have given such a good estimate of the limits of likely error as the samples actually chosen.

A further universe of 2482 lives was obtained by selecting a particular subgroup from the universe of 80,073 lives; the characteristics of this subgroup were known to differ significantly from the remainder of the universe. From this subgroup eleven 10% samples were selected by a similar method to that used for sampling from the group of 8053 lives.

The work was concentrated on the sample from the group of 8053 lives as being the most suitable size of group for experiment; valuation groups as large as 80,073 lives are rarely met in practice and the preliminary investigation with the third group showed such large sampling errors that the amount of experiment with this group was limited.

A further completely independent universe relating to an overseas pension scheme for native lives was also available and was used solely for calculating standard deviations of salary and duration at various ages (Appendices I and IV).

In testing the effect of sampling on valuation results a simple pension scheme was assumed where a pension of $\frac{1}{60}$th of pensionable salary for each year of service was payable on age or ill-health retirement, and the rate of contribution was 10% of salary. Pensionable salary was taken as the average salary earned over the last 5 years of service (the repercussions of sampling on a scheme where pension depended on salary earned over the whole of service would be different). Subsidiary benefits were excluded but, as the value of such benefits is small compared with the value of the main benefits, the conclusions to be drawn from the results should not be materially altered by their exclusion. It was assumed that the valuation could be performed satisfactorily by using a mean duration at each age. The valuation decremental rates were those published in the Appendix to the Government Actuary's Report dated 22 December 1950, made under the Teachers (Superannuation) Acts—H.C. 128/1950–51—which are reproduced for convenience in Appendix II. The valuation rate of interest assumed was $3\frac{1}{2}$%.

5. Actuaries have used various methods which shorten the valuation calculations as distinct from the preparation and tabulation of the data. Data may be grouped and valuation factors constructed applicable to the central ages of groups. This question was examined by H. Freeman (*J.I.A.* LXI, 9) and by W. E. H. Hickox (*J.I.A.* LXI, 86). Saving by approximations in valuation technique is, however, quite different from saving by sampling technique, and in considering the application of sampling it has been assumed that normal valuation procedure is followed. Nor was it thought justifiable to cut down the work involved in computing the valuation results shown later by using approximate methods, because the errors introduced by such methods might mask the true position in regard to sampling errors.

6. The following summary sets out the main course of the inquiry as conducted by means of experiments:

DATA RELATING TO THE INTER-VALUATION PERIOD

(*a*) *Decrements:* illustrations from a model valuation group and from actual pension fund data—Table 1 and §§ 8–10.

(*b*) *Exposed to risk:* Table 2, theoretical approach and Table 3, experimental results (§ 11).

DATA RELATING TO MEMBERS ON THE VALUATION DATE

(*a*) *Salary:*
- (i) examples of standard deviations (Appendix I);
- (ii) errors in valuation results due to sampling for salary—Table 4, and analysis of these errors—Table 5 (§ 16);
- (iii) salary ratio scales derived from samples of a group of 2482 lives—Table 6 (§ 16);
- (iv) error in rate of contribution if the salary scale is obtained from a sample (§ 16).

(*b*) *Age distribution:* analysis of percentage errors in valuation results due to sampling for age distribution—Table 7 (§ 17).

(*c*) *Duration of service:*
- (i) examples of standard deviations (Appendix IV);
- (ii) errors in valuation results due to sampling for duration—Table 8 (§ 18).

(*d*) *Sampling for combinations of characteristics:*
- (i) valuation results for various combinations of sample characteristics (Appendix III);
- (ii) estimates of the standard error in the valuation results (expressed as a percentage) due to sampling for various combinations of characteristics for a group of 8053 lives—Table 9 (§ 20).

(*e*) *Summary:* Summary of standard deviations in net liability (expressed as percentages) due to sampling separately for salary, duration etc., and estimates of the standard error due to sampling for combinations of characteristics—Table 10 (§ 21).

SAMPLING FOR THE DATA RELATING TO THE INTER-VALUATION PERIOD

Decremental rates experienced by those in service

7. For practical reasons it is usually necessary to separate pension fund data, dividing them, for example, into data for hourly paid staff, clerical staff, etc., as wide differences are found in the decremental rates. The different classes are then subdivided according to sex and year of birth and possibly length of service. The resulting groups will still contain employees earning different salaries and employed on different jobs, but who will be treated together for actuarial purposes. It is apparent, therefore, that, even if the original data were numerous, by the time they have been divided into their valuation groups the numbers will be quite moderate—and if the data are confined to a sample the numbers in these groups may be small. When the experience is being examined and the number of exits by cause, e.g. death, retirement or withdrawal, corresponding to each valuation group is needed, the number in some groups may be too small to be significant and this will be particularly so if only sample data are available.

8. In order to give an idea of the number of exits which might be experienced in an actual case a 'model office' valuation group of those in service was constructed. The model group is assumed to have an all-male membership and to be supported by an annual intake of 1000 new entrants who enter on

their twentieth birthday. The group is assumed to have reached a stationary condition. While in service the members are assumed to be subject to mortality and ill-health retirement rates which approximate to the combined experience of a number of local government superannuation funds as disclosed in their 1939 valuations. The withdrawal rate is arbitrary and the expected withdrawals correspond to only a proportion of those actually experienced by these funds. Age retirements are assumed to occur at age 65. The number of members in service totals 35,685, a number greatly in excess of the number which would be found in a valuation group of most pension funds. The expected numbers of exits each year are:

Deaths in service	208
Ill-health retirements	180
Age retirements (at age 65)	478
Withdrawals from service	134

The expected death and retirement experience of the model group over 5 years is analysed in five-year age-groups in Table 1. It is assumed that the square root of the expected number of exits in each group is a measure of the standard error. As the withdrawal rate is arbitrary no useful purpose would be served by applying the tests to this decrement.

Table 1. Expected deaths and retirements in five years for a model valuation group of 35,685 lives

Age-group (1)	Deaths		Retirements	
	Expected deaths (2)	Standard error (3)	Expected retirements (4)	Standard error (5)
20–24	25	5 (20)	—	—
25–29	25	5 (20)	—	—
30–34	50	7 (14)	25	5 (20)
35–39	60	8 (13)	25	5 (20)
40–44	85	9 (11)	25	5 (20)
45–49	125	11 (9)	50	7 (14)
50–54	165	13 (8)	70	8 (12)
55–59	225	15 (7)	195	14 (7)
60–64	280	17 (6)	510	23 (4)
65	—	—	2390	49 (2)

Note. The figures in brackets express the standard error as a percentage of the expected deaths or retirements.

If a model valuation group of one-tenth of the size of the first group is considered it will have a membership of 3569 and the percentage random errors likely to arise would be about $\sqrt{(10)}$ or, say, three times as great as those appearing in Table 1. Similarly, if a group of one-hundredth the size of the first group is considered it will have a membership of 357 and the percentage random errors would be about ten times as great as those shown in the table. The errors shown in the table are of the order of 10% or more for deaths or retirements. Even with full use of graduation technique the preparation of suitable tables of rates of decrement for the valuation directly from the experience of a fund will, therefore, be difficult or impossible unless the membership is large.

9. In practice, however, the decremental rates applicable to those in service cannot be considered in isolation. For example, the effect on the valuation of changes in the rate of mortality in service or of withdrawal from service, is likely to be much less than the effect of changes in the salary scale or in the rate of interest. If the benefit on death is not very different from, or on average is slightly less than the actuarial reserve, the decrement of death in service may even be ignored. Withdrawals are usually a source of profit, and in many funds it may be convenient not to allow directly for a rate of withdrawal at all, but to deal with withdrawal profits as they emerge and are disclosed at each valuation, or alternatively to make only a moderate or nominal allowance for withdrawals, any withdrawals in excess of this allowance contributing to profits. Such profits may be useful as a means of offsetting in some measure losses which may arise from increases in salary levels or increased longevity among pensioners. Another reason for adopting such a course is that withdrawals tend to fluctuate widely from year to year (being sensitive to many influences), and it may be unwise to capitalize future withdrawal profits by direct allowance in the valuation basis when they may never materialize. The dangers are illustrated in the local government superannuation schemes where employees will enjoy in the future much greater facilities for retaining their pension rights on transfer than they did in the past. When considering suitable decremental rates for use in valuing a particular fund an actuary is able to refer to other tables of mortality, whether population, life assurance, annuitant or pension fund mortality, as well as to various decremental rates used in pension fund valuations; the appropriate valuation rate of ill-health retirement or age retirement needs to take account, not only of past experience but of the future policy of the employer in regard to these matters. It is against this general background, and having in his mind a fairly clear idea of the effect of varying the valuation rates, that the actuary decides on suitable rates to use. He is, therefore, in practice, able to arrive at suitable valuation rates which may reflect the particular experience of the fund, even when in theory the paucity of numbers makes the problem difficult.

10. The use of sampling methods for obtaining the numbers of decrements, however, would so increase the standard errors that, unless the valuation groups contained considerably more lives than the model group of 35,000, data limited to a sample of the 'in service' decrements during the inter-valuation period would not be acceptable. This was confirmed in an actual investigation relating to a 10% sample of some recent experience of male lives. The valuation group totalled 20,000 and the exits during five years of a 10% random sample were:

Deaths	26 spread over all ages;
Retirements	150 spread over all ages, 46 of which occurred at age 65;
Withdrawals	256 spread over all ages and combining the experience of all ages at entry.

These exits would only supply limited information towards determining rates for valuation purposes. The rate of decrement of age retirement, limited to a short range of ages, has such an important bearing on the valuation result that in any case it would be advisable to have full particulars of age retirements.

Apart from the relative increase in random errors due to sampling there is also the possibility of other errors being introduced. With full data, checks

on the total numbers of movements can often be applied. When records are kept for only a sample of the movements it is not easy to detect errors due to the operators neglecting to select all sample cases, or duplicating some data. This was confirmed from actual experience in obtaining sample data.

11. Although the exits, which make up the numerators of the 'in service' decremental rates, are too few to lend themselves to sampling methods, the possibility remains of sampling for the 'exposed to risk' which provide the denominators for the rates. E. S. Andersen (*Transactions of the Thirteenth International Congress of Actuaries*, 1, 593) has examined theoretically the effect on a single decremental rate of using a sample for estimating the exposed to risk. Table 2 is derived from one which appears in his paper.

Table 2. Table showing the ratio of the standard error of the rate of decrement derived from data where the exposed to risk is estimated from a sample to the standard error of the rate derived from full data

q	1 % sample	5 % sample	10 % sample	20 % sample
·001	1·05	1·01	1·00	1·00
·005	1·22	1·05	1·02	1·01
·020	1·73	1·18	1·10	1·05
·100	3·32	1·73	1·41	1·22

In building up a service table for a pension fund we are concerned with multiple decrements, and Table 2 is not directly applicable. The effect of estimating the exposed to risk from a sample was calculated by sampling from the experience data relating to the group of 80,073 lives and also for the group of 8053 lives, the sample exposed to risk being multiplied by ten to give an estimate of the full exposure. The results are shown in Table 3.

Table 3. The effect of estimating the exposed to risk over a period of three years from 10% sample data

Age last birthday at beginning of year of decrement	Central exposed to risk estimated from sample data expressed as a percentage of that calculated from full data	
	Valuation group of 80,073 male lives	Valuation group of 8053 male lives
	%	%
25–29	99	112
30–34	101	105
35–39	98	95
40–44	101	100
45–49	95	117
50–54	100	103
55–59	99	103
60–64	99	92

The errors would correspond to some extent with those obtained by entering Table 2 with a decremental rate equal to the sum of the multiple decremental rates.

In practice, the age distribution of those in service on the valuation date is usually available, and from this and the full particulars of exits and entrants it would be possible to build up the full exposed to risk and to obviate the necessity for sampling. Accordingly, the use of sampling for the exposed to risk would seem to be limited to those few cases where the full age distribution is not known and sampling would also have to be used for the valuation itself.

Mortality experience of pensioners

12. The question of sampling has not been studied in relation to age or ill-health pensioners' experience for various reasons, among which the following are the most cogent:

(i) Changes in pensioners' mortality have considerable effect on the financial position of a pension fund and alter the values both of existing and of future pensions.

(ii) There are rarely sufficient numbers of deaths for sampling to be considered; this objection applies with even greater force when the investigation is on a select basis.

(iii) The rates of mortality at the older ages are high and the relative errors introduced by sampling for the exposed to risk are likely, on theoretical grounds, to be large. Also, at the advanced ages even the full data will be scanty, so that sampling would not be feasible.

(iv) The records relating to pensioners are simpler than those for active staff and are usually readily available.

It is doubtful whether sampling methods would save any time. Probably, with the numbers involved, it would be easier to supply full rather than sample data, and with full pensioners' data the actuary can often make useful checks with statistics which appear in the annual reports of the fund.

Similar considerations apply to the question of sampling for particulars of pensions in payment on the valuation date.

Salary experience

13. Salary levels have been so fluid in recent years that the valuation salary scale is likely to be fixed on the basis of the latest information available about salaries, salary scales, and promotion and future policy—although past history would need to be considered. The investigation has, therefore, been limited to sampling for salary at the valuation date and this is dealt with in §§ 15 and 16.

SAMPLING FOR THE DATA RELATING TO MEMBERS ON THE VALUATION DATE

14. Calculations of the errors in the valuation results were made with sampling technique applied separately to the characteristics of salary, duration and age distribution and also to various combinations of them. The sampling errors were assessed by multiplying by ten the values obtained from 10% sample data and comparing with the full valuation results.

Salary

15. Let it be assumed that particulars of salary are to be obtained from a sample of the 'in force' at the valuation date. The magnitude of the standard error of the mean salary of a sample at a particular age will depend on the

standard error of salary at that age. Salary is a characteristic varying with the particular staff or valuation group concerned and it is not to be expected that any theoretical formula for the standard error of salary would be evolved. The standard deviation of salary at different ages was calculated from data derived from actual pension schemes and the results are tabulated in Appendix I. The standard deviation of salary is of the order of 20% of mean salary, except for Group C which relates to an overseas scheme for native lives. The data available were limited to public employees, but for other pension schemes it should not be difficult from general considerations to gauge whether the variation in salary would be greater or less than that indicated in these examples, e.g. for works staff and hourly paid employees a small spread of salaries in relation to mean salary would be expected.

The tables given in Appendix I show estimates of the standard errors of mean salary expressed as percentages for 10% samples taken from valuation groups of various sizes. The standard error of mean salary varies from an average of about 1% for sampling from a group of 80,073 lives to 4% for a group of 8053 lives and to 8% for a group of 2482 lives.

Table 4. Errors in valuation results due to sampling for salary: 10% samples

	Group of 80,073 lives		Group of 8053 lives						
	Full data	Sample salary data	Full data	Sample I	Sample II	Sample III	Sample IV	Sample V	S.D.
Pensions:									
Value in £000's	164,043	163,718	16,284	15,943	16,928	16,024	16,261	16,406	350
Percentage error	—	−·2	—	−2·1	+4·0	−1·6	−·1	+·7	2·2
Contributions:									
Value in £000's	63,424	63,269	6,382	6,261	6,390	6,384	6,434	6,486	75
Percentage error	—	−·2	—	−1·9	+·1	+·0	+·8	+1·6	1·2
Net liability:									
Value in £000's	100,619	100,449	9,902	9,682	10,538	9,640	9,827	9,920	325
Percentage error	—	−·2	—	−2·2	+6·4	−2·6	−·8	+·2	3·3

The percentage error in the mean increases at the older ages owing to the smaller numbers of cases at those ages. This points to one of the dangers of using a sample salary scale for estimating the value of future pensions. The salaries at the last few ages, in a final average salary scheme, determine the major part of the pension liability for all lives and an error at these ages would not be balanced by similar percentage errors in the contrary direction at younger ages. The larger percentage errors at the older ages are therefore unfortunate, but could be avoided if full data were available at these ages.

16. The valuation results where salary particulars alone are determined from 10% sample data, full data being used otherwise, are set out in Table 4 for the groups of 80,073 and 8053 lives. Owing to the amount of work involved the calculations were confined to five of the samples from the group of 8053 lives. The crude salary scales obtained from the mean salaries at each age were graduated by summation formulae; in each case the same formula was used for the scale derived from the full data as from the sample. Summation formulae were used in order to minimize the introduction of bias into the results due to the method of graduation itself. The sampling errors for

salary are subdivided in Table 5 into those due to distortion of the salary ratio scale and those due to errors in the value for mean actual salary at each age.

Table 5. Analysis of errors due to sampling for salary

	Group of 80,073 lives	Group of 8053 lives					
		Sample I	Sample II	Sample III	Sample IV	Sample V	S.D.
Pensions:	%	%	%	%	%	%	%
Percentage error due to sampling for:							
(i) Salary ratio scale	+·1	− ·7	+3·7	−1·1	−·3	− ·4	1·8
(ii) Mean actual salaries	−·3	−1·4	+ ·3	− ·5	+·2	+1·1	0·8
(iii) (i) and (ii)	−·2	−2·1	+4·0	−1·6	−·1	+ ·7	2·2
Contributions:							
Percentage error due to sampling for:							
(i) Salary ratio scale	+·1	−1·2	+ ·9	+ ·5	+·7	+ ·9	·9
(ii) Mean actual salaries	−·3	− ·7	− ·8	− ·5	+·1	+ ·7	·6
(iii) (i) and (ii)	−·2	−1·9	+ ·1	+ ·0	+·8	+1·6	1·2
Net liability:							
Percentage error due to sampling for:							
(i) Salary ratio scale	+·1	− ·5	+5·5	−2·1	−·8	−1·2	2·7
(ii) Mean actual salaries	−·3	−1·7	+ ·9	− ·5	+·0	+1·4	1·1
(iii) (i) and (ii)	−·2	−2·2	+6·4	−2·6	−·8	+ ·2	3·3

In the sample valuations for the group of 80,073 lives the view was taken that although details of the full population were supposedly unknown, a smoother progression of mean salaries and durations from age to age would be expected from the full population than from the sample. The sample mean durations were accordingly smoothed and graduated salary scale values were taken as mean actual salaries. In the result it was found that the smoothing decreased the value of pensions by ·06% and increased the value of contributions by ·20% and as it happened the final result would have been better had unsmoothed values been taken. Accordingly, in the sample valuations relating to the groups of 8053 and 2482 lives, mean durations were taken from the sample unsmoothed, and mean actual salaries were used ungraduated to compute actual salaries at each age. Where a salary ratio scale was needed in calculating commutation functions for five samples from the group of 8053 lives, the crude values were graduated as described at the beginning of this paragraph.

For the third group of 2482 lives the progression of mean salary with age obtained from the samples was very irregular, even when the data were grouped in five-year age-groups. The progression also showed marked divergencies between one sample and another. The figures for grouped data, related to a radix of 100 at age 22, are shown in Table 6.

The standard deviations of mean salary shown in Table 6, having regard to the number of lives in the age-groups of the samples, were found to be consistent with the standard deviations of salary shown in Appendix I for Group B (allowance having been made for the fact that the deviations of the salary ratio depend on the deviations of mean salary in the lowest age-group as well as the age-group in question).

The marked deviations in the ratio of mean salary in the higher age-groups to that in the lowest groups are partly due to the data being more scanty in these groups than in the middle of the age range and hence the characteristics more susceptible to chance fluctuations. Unfortunately the salaries in the higher age-groups are the important ones in determining the value of pension benefits. If it be assumed that a deviation of up to twice the standard deviation may arise in the sample selected, the salary ratio in the highest age-groups may be over 20% in error. Errors of such magnitude would vitiate the valuation results.

Graduation would smooth out to some extent the errors due to chance fluctuations, and in the last two columns of Table 6 the graduated salary ratio scales obtained from the full data and sample no. XI are shown. It was estimated that the error in the net liability due to errors in the graduated scale obtained from this sample would be about 15%. Such an error is greater than can be accepted.

Table 6. Salary ratio scales derived from 10% samples from a group of 2482 lives

Age last birthday, central age in group	Sample mean salary in group, ungraduated												Mean salary for groups, graduated	
	Full data	Sample data												
		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	S.D.	Full data	Sample XI
22	100	100	100	100	100	100	100	100	100	100	100	100	—	100	100
27	117	122	112	124	126	108	131	110	125	124	106	114	8 (7)	120	117
32	136	141	135	142	148	130	150	129	149	140	121	133	9 (7)	141	140
37	163	157	163	167	167	148	171	164	165	160	144	154	8 (5)	169	160
42	178	176	176	185	205	152	197	166	192	203	155	181	17 (10)	186	195
47	189	168	168	183	196	203	190	174	222	183	187	207	16 (9)	195	213
52	190	183	210	191	231	154	203	205	212	177	168	198	21 (11)	196	187
57	197	191	187	200	225	192	240	187	209	220	180	191	19 (10)	204	205
62	191	162	189	192	223	178	200	185	204	195	167	240	22 (12)	201	233

Note. The figures in brackets express the standard deviation as a percentage of the mean salary for the full data.

The salary ratio scale besides being used for estimates of the future salaries of existing staff is needed to estimate the progression of salary for a new entrant. The error in the rate of contribution for a new entrant due to sampling for salary, expressed as a percentage of the correct rate of contribution, amounted to ·1% for the one sample from the group of 80,073 lives, 2·0% for the group of 8053 lives (a standard deviation calculated from five samples) and about 10% for sample XI selected from the group of 2482 lives.

Age distribution

17. In order to assess the sampling error due solely to age distribution full particulars of salaries and durations have been assumed in calculating the results shown in Table 7. In practice, sampling for age distribution would necessitate sampling for actual salaries and durations as well, even if the valuation salary ratio scale were fixed, and the error which would be introduced must be measured by adding the separate errors due to sampling for

Table 7. Analysis of percentage errors in valuation results due to sampling for age distribution: 10% samples

	Group of 80,073 lives	Group of 8053 lives											
		Sample data											
	Sample I	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	S.D.
Percentage error in value of:	%	%	%	%	%	%	%	%	%	%	%	%	%
Pensions due to:													
(i) total number	+ ·6	− ·8	+ ·6	−3·9	−1·6	−1·4	+3·7	+5·3	+3·8	+2·8	−3·1	+2·2	3·0
(ii) spread	− ·8	+1·5	+ ·3	+1·1	+2·0	− ·2	− ·8	+ ·0	+1·8	+ ·9	+1·1	− ·2	1·1
(iii) age distribution, i.e. (i) and (ii)	− ·2	+ ·7	+ ·9	−2·8	+ ·4	−1·6	+2·9	+5·3	+5·6	+3·7	−2·0	+2·0	3·0
Contributions due to:													
(i) total number	+ ·6	− ·8	+ ·6	−3·9	−1·6	−1·4	+3·7	+5·3	+3·8	+2·8	−3·1	+2·2	3·0
(ii) spread	+ ·4	− ·9	− ·2	− ·8	−1·8	+ ·0	+ ·8	+ ·1	− ·9	−1·4	− ·7	+ ·2	·9
(iii) age distribution, i.e. (i) and (ii)	+1·0	−1·7	+ ·4	−4·7	−3·4	−1·4	+4·5	+5·4	+2·9	+1·4	−3·8	+2·4	3·3
Net liability due to:													
(i) total number	+ ·6	− ·8	+ ·6	−3·9	−1·6	−1·4	+3·7	+5·3	+3·8	+2·8	−3·1	+2·2	3·0
(ii) spread	−1·6	+3·0	+ ·6	+2·3	+4·5	− ·3	−1·8	− ·1	+3·5	+2·5	+2·2	− ·4	2·4
(iii) age distribution, i.e. (i) and (ii)	−1·0	+2·2	+1·2	−1·6	+2·9	−1·7	+1·9	+5·2	+7·3	+5·3	− ·9	+1·8	3·5

age distribution, actual salaries and duration. For this reason, as the errors due to salary alone were so considerable, no detailed valuations were made for the samples from the group of 2482 lives to determine the errors due to age distribution, although the percentage error in the total number of each sample was calculated.

The error due to incorrect age distribution has been subdivided into that due to the total number in the sample not being exactly 10% of the total number in the full data and that due to the numbers in the sample not being distributed between the various age-groups in the same proportions as in the full data. These subdivisions of the error are shown in the table as the errors due to 'total number' and 'spread' respectively. The number in the sample from the group of 80,073 lives was 8053—an error of ·6%. The numbers in the samples from the group of 8053 lives ranged from 774 to 848, i.e. from about 4% below to 5% above an exact tenth, with a standard deviation of (3·0 ± ·6)% of the mean.

If the chance of any particular life being selected for the sample of the 8053 lives is one-tenth, the distribution of numbers in the samples would be represented theoretically by the expansion of $(\cdot 1 + \cdot 9)^{8053}$, the standard deviation being 26·9 or 3·34% of the mean. The standard deviation calculated from the eleven samples is therefore consistent with that of the theoretical distribution. For the group of 2482 lives the numbers in the samples ranged from 225 to 284, i.e. from 9% below to 15% above the mean with a standard deviation of (7·7 ± 1·7)% of the mean. The standard deviation of the corresponding binomial distribution is 15 or 6·0% of the mean which again is consistent with the calculated value. For the group of 80,073 lives the one sample showed a deviation of about half the standard deviation of the corresponding binomial distribution.

Duration of service

18. Duration of service is the main characteristic determining the liability in respect of past service. The liability for future service may, however, also be affected by past duration to an extent which varies as between one scheme and another. For instance, in some schemes 40 years' service is a qualification for age retirement which may take place at any age between 60 and 65, in others the maximum pension is earned after 40 years' service. Further complication arises if there is correlation at each attained age between salary and duration of service. Accordingly, in the valuation of some funds, grouping together all durations corresponding to each year of birth and working on an average duration would not give a sufficiently close approximation to the valuation liability. Division of the valuation data according to duration of service would be necessary, e.g. into those who entered pensionable service at ages 20, 21, 22, 23, 24 and at ages over 25. For such a case sampling for duration of service would be limited to the last group, in which the numbers may be too small for the duration determined from a sample to give a good indication of the true average duration at each attained age. In other funds, however, subdivision by duration is not necessary.

The standard deviation of duration at various ages was calculated for some actual examples, and the results are shown in Appendix IV. In schemes where recruitment is normally over a narrow range of ages, duration of service at the higher ages would not be expected to vary widely for a given attained age.

However, economic and financial disturbances are ever present to affect the flow of recruitment, which is heavy in some years and light in others, and entrants are spread over a wide range of ages at entry. A standard deviation of duration which is high in relation to the mean must be expected whatever the normal recruitment policy of the employer.

The tables in Appendix IV give estimates of the standard errors of mean duration, expressed as percentages, for 10% samples taken from various valuation groups. The standard error varies from an average of about 4% for sampling from a group of 80,073 lives to 12% for a group of 8053 lives and to 25% for a group of 2482 lives.

The estimated standard errors of mean duration exceed those of salary but, nevertheless, in considering the effect on valuation results it was thought that sampling for mean duration was more likely to produce satisfactory results than sampling for salary. The errors in the mean salary at the highest ages disturb the salary ratio scale at its most important point and affect the value of pensions at all ages, whereas in the case of mean duration, errors in one direction at some ages tend to offset errors in the other direction at other ages. Also high relative errors in duration occur at the young ages where the value of past service is small for two reasons, first the small duration of service and secondly the powerful discounting and survivorship factor. For these reasons, and because the effect of duration on the value of future service pensions is slight and the value of contributions unaltered, it would be reasonable to expect the errors in the valuation results due to sampling for duration to be smaller than those due to sampling for other characteristics. This was confirmed by the experiments made.

In the pension fund data which were available for investigation no correlation between duration and salary at each attained age could be traced. The effects on the valuation results of determining the average duration at each age from a 10% sample, full particulars of salaries and age distribution being assumed, are shown in Table 8.

The error due to sampling for duration in the case of the one sample from the group of 80,073 lives was ·5% in the value of pensions and ·9% in the net liability.

Accumulated contributions

19. If the past contributions for each life are accumulated at interest up to the valuation date this may involve a considerable amount of computing work. It may be pointed out by the administrators of the scheme that this benefit will become payable in only a few cases and even then it has to be computed only once, that is at the time of exit.

In most schemes the value of the returns of past contributions on withdrawal, the returns usually being confined to employees' contributions, is small compared with the value of the main benefits, say of the order of 1%, and a margin of error could be accepted in the computation of this item without sensibly altering the valuation results. Accordingly, the total accumulated contributions at each age may well be estimated from a sample of the 'in force'. Assuming that a 10% random sample is taken, the sample at each individual age may be considerably greater or less than 10% and some method of weighting needs to be used in rating up the sample figures. The weighting may be by numbers, salaries, length of service or a combination of these items.

Table 8. Errors in valuation results due to sampling for duration: 10% samples

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	S.D.
Group of 8053 lives:												
Pensions:												
Value in £000's	16,181	16,372	16,168	16,324	16,102	16,311	16,440	16,357	16,257	16,292	16,150	103
Percentage error	−·6	+·5	−·7	+·2	−1·1	+·2	+1·0	+·4	−·2	+·0	−·8	·6
Contributions:												
Value in £000's	6,382	6,382	6,382	6,382	6,382	6,382	6,382	6,382	6,382	6,382	6,382	—
Net liability:												
Value in £000's	9,799	9,990	9,786	9,942	9,720	9,929	10,058	9,975	9,875	9,910	9,768	103
Percentage error	−1·0	+·9	−1·2	+·4	−1·8	+·3	+1·6	+·7	−·3	+·1	−1·4	1·0
Group of 2482 lives:												
Pensions:												
Value in £000's	5,005	5,050	5,066	5,168	4,929	4,873	5,037	4,965	5,003	4,979	5,064	75
Percentage error	−·2	+·8	+1·1	+3·1	−1·7	−2·8	+·5	−1·0	−·2	−·7	+1·0	1·5
Contributions:												
Value in £000's	2,168	2,168	2,168	2,168	2,168	2,168	2,168	2,168	2,168	2,168	2,168	—
Net liability:												
Value in £000's	2,837	2,882	2,898	3,000	2,761	2,705	2,869	2,797	2,835	2,811	2,896	75
Percentage error	−·3	+1·3	+1·9	+5·4	−3·0	−4·9	+·8	−1·7	−·4	−1·2	+1·8	2·6

When the numbers are small at some ages it may be necessary to take a larger sample or to accumulate contributions for all lives at those ages.

In calculating the total accumulated contributions for each age, whether for a sample or for the whole data, it may be convenient, especially when punched cards are being used, not to calculate for each life, but to tabulate the total contributions paid in each past year for all lives of the same attained age and to apply accumulation factors to these totals. Similarly, it may be found in the valuation of other subsidiary benefits (e.g. the return of contributions on death in service, or the payment of the balance of contributions on the early death of a pensioner) that sampling may give sufficiently accurate estimates for the valuation.

Sampling for several characteristics

20. If it be assumed that the interaction of the errors due to the separate characteristics upon each other would produce adjustments of a smaller order of magnitude than the errors themselves, the total error in the value of pensions or contributions for a particular combination of sample characteristics could be obtained by adding the errors attributable to the separate sample characteristics (with due regard to sign).

Valuation results for various combinations of sample characteristics are set out in Appendix III. The resulting errors correspond closely with those obtained by adding the separate errors due to sampling for salary, duration or age distribution. The assumption of an additive property for the errors is, therefore, a reasonable one for practical purposes. These valuations have been calculated for only one sample from the group of 80,073 lives and one from the group of 8053 lives so that an estimate of the standard errors for a combination of characteristics cannot be calculated directly. However, in view of the additive property of the respective errors and assuming no appreciable correlation at each age between salary, duration and numbers of lives, an estimate of the standard error for a combination of sample characteristics would be $(\Sigma S_i^2)^{\frac{1}{2}}$ where S_i is the estimated standard error due to sampling for the characteristic i alone. The S_i's have been calculated for the group of 8053 lives and for this group the estimated standard error for various combinations is set out in Table 9. The estimated standard error is shown together with an upper limit which allows for errors in the estimates of the separate standard errors themselves due to the limited number of samples taken. The upper limit exceeds the calculated figure by twice the estimated standard error of the standard error obtained by the formula $S/(2n)^{\frac{1}{2}}$, where n is the number of samples.

If data relating to only 10% of the cases are available, sampling for all characteristics is necessary (case A). The errors are reduced slightly in case B where the total number of cases is known and the sample numbers at each age are adjusted by a constant factor so that they amount in total to 10% of the full data. For case C the circumstances are the same as in case B, except that the valuation salary scale is assumed to be given, so that sampling for salary only affects the actual salaries at each age. In case D the valuation salary scale is assumed to be given and full particulars of numbers of lives at each age to be known. In case E a full age distribution is assumed but the valuation salary scale, actual salaries and duration are all estimated from a sample.

Table 9. Estimates of the standard error in valuation results due to sampling for various combinations of characteristics for a group of 8053 lives: 10% samples

Characteristics for which sampling is employed	Standard error in valuation results, expressed as a percentage of the value from the full data		
	Pensions %	Contributions %	Net liability %
All characteristics: case A	3·8 (5·7)	3·5 (5·0)	4·9 (7·5)
All characteristics but adjusting numbers: case B	2·5 (4·0)	1·5 (2·3)	4·2 (6·5)
Mean actual salaries, duration and age distribution (but adjusting numbers): case C	1·5 (2·3)	1·1 (1·6)	2·8 (4·1)
Mean actual salaries, duration: case D	1·0 (1·6)	0·6 (1·0)	1·5 (2·3)
Salary ratio scale, mean actual salaries, duration: case E	2·3 (3·7)	1·2 (1·9)	3·4 (5·5)

Note. The percentages in brackets are the estimated upper limits of the standard error.

Summary of standard deviations in net liability due to sampling separately for each characteristic and for various combinations

21. Standard deviations in net liability, together with upper limits representing the standard deviations increased by twice the estimated standard error of the standard error, are summarized in Table 10.

Table 10. Standard deviations in net liability due to sampling separately for salary, duration, etc., and for combinations of them: 10% samples

Characteristics for which sampling is employed	Standard deviation in net liability, expressed as a percentage of the value from the full data %
Group of 8053 lives:	
(*a*) Salary:	
(i) salary ratio scale	2·7 (4·4)
(ii) mean actual salaries	1·1 (1·8)
(iii) (i) and (ii)	3·3 (5·3)
(*b*) Duration	1·0 (1·4)
(*c*) Age distribution:	
(i) total number	3·0 (4·3)
(ii) spread	2·4 (3·4)
(iii) (i) and (ii)	3·5 (5·0)
(*d*) Salary, duration and age distribution	4·9 (7·5)
(*e*) Salary, duration and age distribution (spread only)	4·2 (6·5)
(*f*) Salary (mean actual salaries only) and duration	1·5 (2·3)
Group of 2482 lives:	
(*g*) Duration	2·6 (3·7)

Note. The percentages in brackets are the estimated upper limits of the standard error in net liability.

The square of the standard deviation in net liability due to using sample data to determine both the salary ratio scale and mean actual salaries exceeds the sum of the squares of the deviations due to sampling for salary ratio scale and mean actual salaries separately. On the face of it this would indicate positive correlation between the two separate errors (not unexpected from general considerations) but owing to the small number of samples the standard deviation is not a sufficiently close estimate of the standard error for the comparison to have any significance. Nor is it possible to make a valid comparison between the standard deviation due to sampling for age distribution as a whole and the standard deviations of its constituent parts. As mentioned in § 20 in estimating the standard errors due to sampling for combinations of salary, duration and age distribution it has been assumed that there is no appreciable correlation between them.

Table 10 shows, for the group of 8053 lives, that if data relating to only 10% of the cases are available and, therefore, sampling applies to all characteristics, the standard deviation in net liability is 4·9% (upper limit of standard error 7·5%) of the value determined from full data. As it is not unreasonable to expect errors amounting to twice the standard error to arise, and these would be greater than could be accepted in an actuarial valuation, sampling for all characteristics is not feasible. If the total numbers in service on the valuation date are known, and the numbers in the sample are adjusted by an over-all factor so that they amount in total to exactly 10% of the whole data, the upper limit of standard error is reduced to 6·5%, but the errors might still be dangerously large. In practice the case of sampling separately for age distribution cannot arise, as sampling for salary and duration would necessarily follow. If sampling is confined to the obtaining of salary particulars the upper limit of standard error is 5·3% and possible errors are still too large.

However, if the circumstances are such that the valuation salary ratio scale is the same as that previously used or can be determined from other considerations than a study of the salaries in payment on the valuation date, and the full numbers of lives at each age are known, sampling being used solely to determine actual salaries and duration, the table shows, for the group of 8053 lives, a standard deviation in net liability of only 1·5% (upper limit 2·3%). If the correct durations are known these figures are reduced to 1·1% (upper limit 1·8%). If sampling is confined to duration the errors are smaller still, the standard deviation being 1·0% with an upper limit of 1·4%.

The samples taken from the group of 2482 lives show that a 10% sample is not satisfactory as a means of obtaining a valuation salary ratio scale (see Table 6). Sampling for age distribution implies sampling for salary as well and would therefore also be unsatisfactory. If sampling is confined to duration the standard deviation of net liability is 2·6% with an upper limit of 3·7% (Table 10).

CONCLUSION

22. The view is sometimes put forward that conditions affecting pension funds are always changing, that the amount of contribution by the employee is determined by his capacity to pay—the employer meeting the balance of cost—and that consequently no great degree of accuracy in the investigation of the past experience or of the current financial position is called for. Whatever the type of scheme, funded or unfunded, the importance of knowing what it costs and will cost in the future is obvious; the layman rarely appreciates

the magnitude of pension scheme commitments especially if they are unfunded. Having been supplied, *inter alia*, with particulars of the members and pensioners existing on the valuation date, together with data relating to past experience, the actuary advises on the financial position of the fund. If the particulars or data relate to a sample only his task becomes difficult or impossible and his estimates may be unreliable. Unless he can be confident that this will not be so, he will strongly resist the suggestion that full information should not be supplied, especially as most of the particulars required for the periodic actuarial valuations have to be kept in any event, as part of the normal administrative records of a pension scheme.

The numerical results relate to 10% random samples, a standard type of pension scheme, and valuation groups taken from data for long-established schemes for public or semi-public employees, where variations of salary and length of service at a given age are not likely to be excessive. If actual problems were being considered the size of the sample would need to be chosen appropriately, the larger the group, the smaller the percentage sample. For certain valuation groups it may be helpful to sample for a larger proportion of the data for the lowest and highest age groups, where the numbers are small, than for the main body of the data. In special cases stratified sampling may be indicated as likely to give better results. In some circumstances practical considerations may limit the data that can be obtained to full particulars of certain sections of the membership selected as a reasonable cross-section of the whole; the errors introduced by such sampling may be distinctly greater than those from random sampling and it may be difficult to put any limit on the extent to which they are increased. Care must be taken to avoid bias in selecting a sample, the method of sampling depending on the particular circumstances of the fund. The use of an adjusting factor so that the number of lives in the sample amount in total to the correct proportion of the full data has the advantage of correcting errors due to the omission of lives that should be included in the sample and that would otherwise cause an underestimate of the liabilities—a disturbing possibility.

The additional errors due to sampling are likely to be of the same order whether the valuation is a first one or not. The first valuation of a pension fund has, however, particular importance as it provides the first opportunity for a real check on the assumptions made in choosing the actuarial basis for the scheme. Also at the inception there is often an initial liability due to the granting of back-service rights or to the entry to the scheme of staff at the older ages who pay inadequate rates of contribution. In the case of large funds relating to public employees, although some estimate may have been made at the inception, the initial liability often falls to be assessed at the first valuation. This is therefore doubly important and it may not be prudent to use sampling methods, which inevitably introduce errors into the valuation, even if small. At subsequent valuations, however, when the scheme is well established, errors are less likely to alter the position sufficiently to affect the conclusions that may be drawn from the valuation results.

Although an actuarial valuation is not an instrument which measures liabilities with a high degree of precision, and in pension-fund valuations compared with life-office valuations there are additional uncertainties introduced by the salary and ill-health retirement factors, in weighing up the effects of sampling it is important to remember that we are in the final result concerned with a balance item, the difference between the funds in hand and

the net liability. Is the fund paying its way or not? The result will accordingly be very sensitive to errors in the net liability.

For the great majority of private pension funds the size is too small to justify sampling of any kind and the question need only be considered for the relatively few large funds. The particulars of actual funds met with in practice, however, vary so widely that any conclusions need to be interpreted cautiously if the corresponding effects on other funds are being estimated—and circumstances may prevent any parallel being drawn. The problem of the valuation of each fund needs to be treated on its own merits, and particularly so regarding any suggestion of sampling. Certainly the experiments have not proved the case for the general use of sampling methods in actuarial valuations of large pension funds. On the contrary they indicate that to base the valuation results on data relating to a sample only of the membership is neither profitable nor safe, even for funds where the valuation groups contain as many as 8000 lives. These comments apply equally if salary particulars are available only for a sample (full age distribution and particulars of past service being available), provided that the determination of the valuation salary ratio scale and the estimation of actual salaries on the valuation date depend on the sample. Where, however, the salary ratio scale is determined from other considerations, apart from the sample data, sampling methods can be used satisfactorily for estimating actual salaries at each age, providing the particular circumstances of the fund are not unfavourable, and the valuation group totals 8000 lives. For funds of this size sampling can also be used for estimating average duration at each age, or both durations and actual salaries. In the experiments made, the sampling errors corresponding to these two characteristics were of the same order of magnitude and sampling for them may be satisfactory for funds with fewer than 8000 lives. For example, with the valuation group of 2482 lives, sampling for duration alone, the standard error in the net liability expressed as a percentage was estimated as $(2{\cdot}6 \pm {\cdot}6)\%$. Circumstances may be such that errors of this order can be accepted.

It was to be expected that errors due to sampling for duration would be smaller than those due to other characteristics. Duration is a stable characteristic increasing steadily with time and unlike salary is insensitive to changes in the cost of living. The errors introduced by sampling for duration do not affect the valuation basis, nor the value of contributions, nor the rate of contribution for a new entrant. The errors are mainly in the value of past service pensions, and for recently established funds this would further limit their effect on the valuation results.

If sampling is to be used, there are advantages in taking the same sample of the data from valuation to valuation, with appropriate application to new entrants, not only for administrative reasons, but because it enables a fairly precise picture to be drawn of how at least the sample of the fund is progressing.

It is possible that sampling for all characteristics, i.e. using data for only a sample of the cases, could be considered for making rough estimates of the financial effects of changes in large funds, especially as other approximate calculations can often be made against which the estimates can be measured. The results obtained would not be accurate and a full valuation of the fund should follow as soon as possible so that the effect of any errors could be limited or corrected. In the special circumstances envisaged, if an analysis of the experience of the fund over recent years is required, the exposed to risk

corresponding to the 'in service' might also be estimated from a sample, but full particulars of movements should be obtained.

For checking the results of a normal valuation the work involved in sampling methods would be too great in relation to the valuation itself and better methods are available.

Summing up, it appears that valuation results derived from sample data may be subject to serious errors and the use of sampling in practice is likely to be limited. It can, however, be used more freely for data which have a relatively minor effect on the valuation, such as the totals of accumulated contributions. The general comment may fairly be made that the actuary needs to be fully alive to the possibilities of sampling in pension fund valuations.

23. Finally, I want to make it clear that, although I am a member of a Government Department, I write with no official authority, and that the views expressed are purely my own opinions. Ever since a remark by F. H. Spratling some five years ago, I have felt prompted to look into the question whether more could be done to use sampling methods in pension fund work, but I am conscious of the many aspects of the subject I have hardly touched upon or have omitted altogether. I wish to thank R. C. Gilder of the Government Actuary's Department for his invaluable help and am grateful for the encouragement and valuable advice given by a number of other members of the Institute.

APPENDIX I

TABLES SHOWING THE STANDARD DEVIATION OF SALARY, AND ESTIMATES OF THE STANDARD ERROR OF MEAN SALARY WHEN THE MEAN IS ESTIMATED FROM A SAMPLE FOR VARIOUS GROUPS OF PENSION FUND DATA

Group A of 80,073 male lives

Age last birthday (1)	No. of cases in 10% sample, 8053 lives (2)	Mean salary (3)	Standard deviation of salary calculated from the 10% sample (4)	Standard error of the mean of the 10% sample (5)	No. of cases in 1% sample, 823 lives (6)	Standard error of mean of the 1% sample (7)
		£	£ %	£ %		£ %
27	199	362	65 (18)	4·6 (1·3)	22	14 (3·9)
32	304	442	60 (14)	3·4 (0·8)	27	12 (2·7)
39	278	557	98 (18)	5·9 (1·1)	26	20 (3·5)
47	136	605	113 (19)	9·7 (1·6)	11	36 (5·9)
55	164	636	132 (21)	10·3 (1·6)	14	37 (5·8)

Group B of 2482 male lives

Age last birthday (1)	No. of cases: full data (2)	Mean salary (3)	Standard deviation of salary calculated from full data (4)	No. of cases in 10% sample, 225 lives (5)	Standard error of mean of 10% sample (6)
		£	£ %		£ %
25	42	374	43 (12)	6	20 (5·3)
35	99	512	66 (13)	7	27 (5·3)
45	74	631	165 (26)	6	74 (11·7)
55	44	633	140 (22)	5	70 (11·0)

Group C of 3036 male lives

Age last birthday	No. of cases: full data	Mean salary	Standard deviation of salary calculated from full data
		£	£ %
25	258	64	27 (42)
35	94	73	43 (59)
45	25	75	36 (48)

Notes. (i) The figures in brackets express the standard deviation or error as a percentage of the mean. (ii) The standard error of the sample mean has been estimated by the formula $S/(n-1)^{\frac{1}{2}}$, where S is the standard deviation of salary and n is the number of cases in the sample.

APPENDIX II

VALUATION RATES FOR MALE LIVES USED BY THE GOVERNMENT ACTUARY IN THE 1948 VALUATION OF THE TEACHERS' SUPERANNUATION SCHEME IN ENGLAND AND WALES

Age last birthday at beginning of year	Probability of death in service	Probability of withdrawal from service with refund of contributions	Probability of retirement on ground of ill-health or age
22	·0008	·007	—
27	·0009	·006	·0001
32	·0010	·004	·0002
37	·0011	·002	·0003
42	·0016	·001	·0005
45	·0022	—	·0006
6	·0025	—	·0007
7	·0028	—	·0008
8	·0031	—	·0010
9	·0035	—	·0013
50	·0039	—	·0017
1	·0043	—	·0023
2	·0047	—	·0030
3	·0052	—	·0037
4	·0057	—	·0045
55	·0063	—	·0055
6	·0069	—	·0070
7	·0074	—	·0089
8	·0081	—	·0124
9	·0082	—	·156
60	·0088	—	·193
1	·0099	—	·160
2	·0108	—	·185
3	·0118	—	·221
4	·0075*	—	·993

Probability of death of ill-health pensioners	
(1) During the first three years after retirement	
Year of duration	Probability
0	·140
1	·080
2	·055
(2) In the fourth and subsequent years after retirement	
Age last birthday at beginning of year	Probability
All ages to 68	·035
For ages over 68 the rates for age pensioners were adopted	

Age last birthday at beginning of year	Probability of death of age pensioners
60	·0209
1	·0214
2	·0220
3	·0228
4	·0239
65	·0254
6	·0273
7	·0297
8	·0327
9	·0363
70	·0405
1	·0453
2	·0506
3	·0563
4	·0624
75	·0689
6	·0758
7	·0831
8	·0908
9	·0990
80	·1078
1	·1172
2	·1273
3	·1382
4	·1499
85	·1624
6	·1757
7	·1898
8	·2046
9	·2202
90	·2366
For ages over 90 the mortality rates of Life Offices Annuitants, 1900–20, were adopted	

* Exposure for half a year

Note. Most of the above rates are extracted from H.C. 128/1950–51 and are reproduced here by kind permission of the Controller of H.M. Stationery Office.

APPENDIX III

VALUATION RESULTS WHEN A 10% RANDOM SAMPLE IS USED TO OBTAIN VARIOUS VALUATION CHARACTERISTICS IN THE CASE OF TWO GROUPS OF PENSION FUND DATA

Characteristics for which sampling is employed	Valuation results					
	Group of 80,073 lives			Group of 8053 lives		
	Value of pensions in £000's	Value of contributions in £000's	Net liability in £000's	Value of pensions in £000's	Value of contributions in £000's	Net liability in £000's
(*a*) Full valuation: no sampling	164,043	63,424	100,619	16,284	6,382	9,902
(*b*) All characteristics	162,490 (−0·9)	63,950 (+0·8)	98,540 (−2·1)	16,584 (+1·8)	6,624 (+3·8)	9,960 (+0·6)
(*c*) All characteristics but adjusting numbers	161,561 (−1·5)	63,569 (+0·2)	97,992 (−2·6)	16,227 (−0·4)	6,482 (+1·6)	9,745 (−1·6)
(*d*) Mean actual salaries, duration and age distribution	162,332 (−1·0)	63,839 (+0·7)	98,493 (−2·1)	16,665 (+2·3)	6,584 (+3·2)	10,081 (+1·8)
(*e*) Mean actual salaries and duration	162,679 (−0·8)	63,186 (−0·4)	99,493 (−1·1)	16,343 (+0·4)	6,427 (+0·7)	9,916 (+0·1)
(*f*) Mean actual salaries, salary ratio scale and duration	162,797 (−0·8)	63,269 (−0·2)	99,528 (−1·1)	16,277 (−0·0)	6,486 (+1·6)	9,791 (−1·1)

Notes. (i) The figures in brackets represent the error expressed as a percentage of the value found from the full valuation. (ii) In (*c*) the sample numbers at each age are adjusted by a constant factor so that they amount in total to 10% of the full data.

APPENDIX IV

TABLES SHOWING THE STANDARD DEVIATION OF DURATION AND ESTIMATES OF THE STANDARD ERROR OF MEAN DURATION WHEN THE MEAN IS ESTIMATED FROM A SAMPLE FOR VARIOUS GROUPS OF PENSION FUND DATA

Group A of 80,073 male lives

Age last birthday (1)	No. of cases in 10% sample, 8053 lives (2)	Mean duration in years (3)	Standard deviation of duration calculated from the 10% sample in years (4)	Standard error of the mean of the 10% sample in years (5)	No. of cases in 1% sample, 823 lives (6)	Standard error of mean of the 1% sample in years (7)
25	169	2·4	2·1 (88)	·16 (7)	16	·54 (23)
35	331	9·4	5·9 (63)	·33 (4)	38	·97 (10)
45	173	20·2	7·0 (35)	·53 (3)	16	1·81 (9)
55	164	32·0	6·3 (20)	·49 (2)	14	1·75 (5)

Group B of 2482 male lives

Age last birthday (1)	No. of cases: full data (2)	Mean duration in years (3)	Standard deviation of duration calculated from full data in years (4)	No. of cases in 10% sample, 225 lives (5)	Standard error of mean of 10% sample in years (6)
25	42	2·0	1·8 (90)	6	·8 (40)
35	99	8·8	4·7 (53)	7	1·9 (22)
45	74	18·2	7·1 (39)	6	3·2 (18)
55	44	26·2	9·0 (34)	5	4·5 (17)

Group C of 3036 male lives

Age last birthday	No. of cases: full data	Mean duration in years	Standard deviation of duration calculated from full data in years
25	258	2·6	1·9 (74)
35	94	8·6	4·4 (51)
45	25	13·4	7·8 (58)

Notes. (i) The figures in brackets express the standard deviation or error as a percentage of the mean. (ii) The standard error of the sample mean has been estimated by the formula $S/(n-1)^{\frac{1}{2}}$, where S is the standard deviation of duration and n is the number of cases in the sample.

ABSTRACT OF THE DISCUSSION

Mr F. Gordon Smith, in introducing the paper, said that when sampling was used in pension fund valuations a difficulty was that the valuation result—the surplus or deficiency—was sensitive to relatively slight changes in the net liability. G. C. Campbell had found a difficulty of a similar nature in sample reserve valuations of a large life assurance company.* Although the error expressed as a percentage of the total reserves might be small, it might represent a significant part of the disposable surplus for that year.

Mr J. P. Holbrook, in opening the discussion, said that it was a pleasure to welcome to the Institute a paper on a statistical subject which involved no more difficult mathematical idea than the standard deviation. The reason was easy to find: the author had chosen as his subject the valuation of a pension fund, and that was a field where so far the actuary's judgment and experience had proved more valuable to him than advanced mathematical techniques. It was only necessary to consider, for example, the item 'Total of the annual rates of salary at age x' to realize the type of initial difficulty which was encountered. To obtain that, some assumption had to be made about the probability distribution of the salary at each age; if the fund were sufficiently large, and could be investigated over a period which had seen fairly stable conditions, it would probably be possible to obtain some idea of the shape of those probability distributions, but whether they would be applicable to other similar funds would remain a matter of mere guesswork. That lack of homogeneity existed in almost all the items that entered into a pension fund valuation. The author had realized the difficulty clearly. He had wisely avoided what might be called the higher statistical obscurities and, as a result, had written a paper which his readers could understand.

The object of the paper was not in any sense to 'make out a case' for sampling. In the author's words, the object was

> to investigate the possibilities of its use in the actuarial field.

The case for sampling would only be made

> if it were possible to reduce the degree of uncertainty introduced by sampling sufficiently to accomplish all that is really required of a valuation,...providing that there was a real saving of labour.

That seemed to be an admirable criterion and it was interesting in the first place to consider the theoretical question 'what degree of uncertainty is acceptable?'

That was a general question equally important whether sampling for life offices or for pension funds was being considered. The reputation of actuaries as financial advisers derived largely from recognition of their capacity to make realistic judgments from statistical material, judgments in which uncertainty was limited as far as possible. It was therefore to be expected that individual actuaries would differ widely in their views on the degree of uncertainty which was permissible.

He thought it relevant to relate the experience of an American statistician who had a reputation for extreme caution. The statistician was travelling with a friend by train in the early part of the summer. His friend looked out of the window and remarked 'Look at all these sheep! They have been sheared already.'

* *Journal of the American Statistical Association*, 1948, p. 413.

The statistician looked and, after a moment, said 'They appear to have been sheared—on this side at least'. It was fairly clear what that gentleman would have thought of the idea of valuing by reference to a 10% sample. Probably few actuaries would go so far as that, but there would be a number who, when asked to sign a report including a valuation result based on a 10% sample, would be tormented with doubts about what the other 90% would show.

In considering that difficulty, it was necessary to remember that the valuation result, in statistical terms, was the mean of a probability distribution, and assuming that all the basic assumptions were correct, there would still be deviations between the actual and expected experience owing to purely random variations. When a valuation result was quoted that degree of uncertainty was tacitly accepted. It seemed fair to say that sampling techniques should be accepted also if it could be established that the extra uncertainty introduced thereby was small in relation to that which was already inherent in the result, remembering that it might be possible to use stratified sampling instead of random sampling. The question was how the matter could be investigated, as the probability distributions of the basic elements were unknown and probably unknowable. It was clear that precise methods could not be used, such as were adopted by Lidstone in investigating the errors due to valuing endowment assurances by the Z-method. The most that could be done was to make a number of experiments with different types of pension fund and to see whether any general working rules could be deduced from the results. That, for the most part, had been the author's method.

In paragraphs 7–11 of the paper the problem of sampling for experience rates was investigated by reference to a hypothetical service population of about 35,000 lives, supported by 1000 new entrants a year. The author considered separately sampling for decrements and for the exposed to risk. Large as that population was, it gave rise to only 1040 expected deaths and 900 expected ill-health retirements during the quinquennium and, although those decrements were not of great financial importance, the sampling errors would have been large and the saving in labour negligible; the author rightly did not pursue the matter further. In sampling the exposed to risk the errors were more acceptable, but that conclusion was of theoretical rather than of practical interest, because in practice the census method would produce satisfactory results with less labour.

The second part of the paper dealt with the question of sampling the valuation data. Samples drawn from an actual fund were used, and the errors due to sampling for the characteristics salary, duration and age distribution were calculated separately and also in combination. In his view, the important elements were the mean actual salaries, the durations, and the 'spread' of age distribution. The other two items—the number in the sample and the salary-ratio scale—arose only as a result of the methods adopted in the paper. The error due to the number in the sample not being exactly 10% of the number in the universe was a result of the method of sampling, whereby the sample of the existing was drawn from data which included exits, and was due to random variations in the proportion of exits to existing in different blocks of data. It could be avoided by extracting the exit cards before the sampling took place and renumbering the existing cards, and although extra labour would be involved, he suggested that such a course would be advisable, in view of the magnitude of the errors brought out in Table 7.

The author recognized the fact that, owing to recent inflation, the 'classic' technique of deducing a salary scale from increase ratios obtained from past

experience had largely broken down. At the time of writing, as he said, average salaries at the valuation date, current scales of remuneration, and promotion policy were among the materials from which the scale was evolved. It was unfortunate, however, that in the subsequent investigation the author should have taken into account only the first of those items, and should have deduced valuation salary scales by the purely mechanical process of applying summation formulae to the average salaries at each age. The object of that procedure was to eliminate bias in the process of graduation; unfortunately it also eliminated professional judgment, as was evident from the last column of Table 6. That column showed the graduated salary scale calculated by summation formulae from Sample XI from the group of 2482 lives. It gave a value of 213 at age 47, which dropped to 187 at age 52, rising to 205 at age 57, and only at age 62 did it rise above the figure attained at age 47. It was fair to say that in no conceivable circumstances would such a scale be appropriate, and it was hardly surprising that its use gave rise to errors of 15% in the net liability and 10% in the contribution rates. Those errors might well have been greater if a more realistic method had been adopted for deducing the salary scale from the full data, but even there the progression was most irregular and the salary at 62 was less than at age 57. It was wrong to ascribe those errors of 10% and 15% to the sampling process. Current average salaries were only a guide, and an inherently imperfect one, to the valuation salary scale, and errors in those averages would not result in similar errors in the valuation salary scale, provided professional judgment were given a chance to operate and the other relevant facts were taken into account.

Concerning the results conveniently summarized in paragraph 21 of the paper there were four theoretical points which seemed to the opener to be of some interest. First, the valuation group was a large one, containing over eight thousand lives, and the benefit was, by modern standards at any rate, a simple one. Secondly, the method of sampling by means of reference numbers allocated in order of entry gave, at least as regards the element of duration, a stratified rather than a random sample—although the extent of the stratification depended on when the Central Record Office was set up and on how the reference numbers of the then existing members had been allocated. Stratified sampling was likely to lead to better results than random sampling in practice, but it might not always be possible to adopt it. Thirdly, in calculating the liability for past service pensions there was required at each age the sum of the product (salary) × (duration) for the individual members. The approximation (salary) × (mean duration for the age group as a whole), which had been used in the paper, had been used consistently for the universe and the samples, but the sampling errors in the former case were likely to be larger than those of the approximation, particularly if there was correlation between salary and duration at certain ages. There was no correlation in the example under consideration, but there might well be in other cases. Finally, owing to ignorance of the form of the probability distribution of the valuation results, it was not safe to assume that twice the standard error gave a sound estimate of the maximum sampling error.

Because of the size of the errors brought out by the author's calculations, and of the difficulties of generalizing from them, it appeared that the outlook for sampling methods in estimating the capital liability of a pension fund was not very promising. It was essential, however, to relate those considerations to the practical commercial background of pension fund finance. The result of the pension fund valuation was merely a guide to the employer in financing

his pension commitments. If a deficiency was revealed or an additional liability imposed as the result of an improvement in benefits, it was almost invariably financed by payments over a period of years, and it was frequently possible to adjust those payments upwards or downwards in the light of experience. In those circumstances, the valuation result had not the same immediate significance in capital values as had the result of a life-office valuation, on which depended the extent to which definite additional liabilities were immediately undertaken in the form of bonuses. In that perspective it could be seen that, when considering the errors introduced by sampling processes, it was justifiable to allow more latitude than that permissible if it were necessary to think exclusively in terms of capital values as in life-office work.

There might, therefore, be differences of opinion about the practical importance of the errors brought out by the author's calculations, but it seemed clear that it was not yet possible to evolve rules of general application in sampling pension funds. But however much the sampling errors might be reduced or limited as a result of further research, there remained the second part of the author's criterion—the question of how much time was saved.

Any concern which had a pension fund had to keep records for each individual member to determine when that member was entitled to benefit and to enable the amount to be calculated, as well as the usual staff records of annual rates of pay and pension fund deductions. In concerns which were large enough for sampling to be a possibility those records were usually well organized, so that the extra information required by the actuary could be assembled at a relatively small cost. The saving in clerical costs due to sampling was therefore likely to be small, and might well be outweighed by the extra professional costs owing to the need for close actuarial supervision at all stages in the valuation of a sample. Moreover, in sampling many valuable checks on the data were lost, it became difficult or impossible to trace valuation profits and losses, and a host of other practical difficulties arose. In particular, if all the data were available there was very little point in sampling, and if they were not available it was questionable how it could be decided whether that part which was available would contain a satisfactory sample.

With those facts in mind, it might be considered that the conclusions reached in the paper merely confirmed what was apparent at the outset, namely that only in the most exceptional circumstances would there be scope for sampling in a pension fund valuation. The author's conclusions were admittedly negative, but by following him in his exploration much was learned about the fundamental problems of pension fund valuations. It became necessary to re-examine ideas about the importance of the various elements entering into the basis and to mark out afresh the frontiers between practice and theory. There was much material in the paper for thought and discussion, which he thought might well lead to useful results in the field of life assurance. But those members of the profession who were perturbed by the recent invasion of mathematical statistics into actuarial territory might take heart from the fact that in pension funds they had a fortress which seemed likely to hold out against further assaults.

Mr P. R. Cox said that it was suggested in section 3 of the paper that sampling might save labour to the authorities of a fund but would not help the actuary much. There were circumstances, however, in which judicious sampling applied in the right manner could save work not only to the fund but to the actuary as well. There was a tendency to set up large and complicated

schemes, or to reconstruct or add to schemes and so make the benefits very intricate. That could occur where members were assimilated from former schemes, in an effort to ensure fairness. In such an instance, a small preliminary sample might give the actuary a good idea of which of the rules could safely be ignored or treated in an approximate manner, and which had to be attended to in considerable detail. That meant cutting down the arithmetical processes of valuation, and also reduced the size and scope of the data required.

An example arose from a fund of which he had had some experience a short time previously. In that fund the benefit on the death of a member actively in service was computed as the best of four different amounts, the annual salary, accumulated personal contributions, a function depending on the length of service, and a function depending on age and length of service. It was difficult to value a benefit of that kind accurately without some prior knowledge, perhaps based on a small sample, of how it worked out in practice.

Reference had been made to a possible correlation between duration of service and amount of salary at a given attained age that was found in some pension schemes. To multiply the average salary by the average duration would give too low a liability if the correlation were positive; a small preliminary sample might give an indication of the necessary correction.

The opener had referred to the form of the valuation records. The paper provided one more reason why an actuary should come in from the beginning when the system of records for a pension fund was being set up, because he might be able to increase the possibility of getting a properly stratified, and therefore more efficient, sample when it was needed.

If the author or anyone else felt inclined to do some more research along the lines followed in the paper, there were two subjects which he would suggest for attention: the errors in estimated emerging costs in sample valuations, and the extent to which changes from time to time in the valuation results of funds of various sizes might arise from purely statistical fluctuations. Knowledge derived from enquiry upon the second of those subjects should be of assistance in considering what action was necessary in dealing with any surplus or deficiency.

Mr F. J. Lloyd recalled the story of the sergeant-major who selected a group of volunteers by saying 'You, you and you!', so obtaining a sample which was almost certainly biased, although that was not due to the easy fault of asking for volunteers. In practice it was often uncommonly difficult to select a sample without introducing bias of one sort or another. It was, of course, common practice to test a particular attribute of the sample against the known distribution of that attribute in the universe. That test might be satisfied, but another attribute of the sample might still be biased.

Another danger might arise even if the sample had been drawn to avoid the introduction of bias: it was possible in a few cases that the random sample would be unrepresentative of the universe for the attribute under consideration. Such difficulties would be less likely to arise and the overall standard error would be reduced if, as the opener suggested, use were made of stratified random sampling. As a simple example of what he meant by stratified random sampling, let it be assumed that he required a 10% sample from a body of 10,000 workers of whom 3000 were women; he would prefer to select his sample by aggregating a 10% random sample drawn from the women with another 10% random sample drawn from the men. He would use separate adjusting factors for the men and the women, and so ensure that the ratio between the sexes was representative.

The principle of stratification could be extended so that each broad category of members was satisfactorily represented. For example, it might be desirable to stratify hourly paid, weekly paid and monthly paid workers.

The author described clearly how he had selected his samples. In one respect he had been fortunate, as each member had an index number and it had been possible to extract a 10% sample by working on the last digit. The speaker had selected similar 10% samples for other purposes using punched card data, which were peculiarly appropriate for drawing such a sample, as it would be certain that the sample included every required case and no other. He would have less faith in such a sample if extracted clerically from drawers of manuscript cards, because in those circumstances it was possible to include a wrong card or to exclude a card which should be in the sample. It was necessary, as the author said, to check the work of extracting the sample.

It was not always that employees would have an index number the last digit of which could be taken. It might therefore be necessary to devise another method of selecting the sample, and he would suggest that one possible method was to use the initial letter of the employee's surname with an adjusting factor to bring the sample up to the size of the universe. The initial letter of the surname was usually independent of such things as age, length of service and salary. It was, however, possible for bias to be present. In the first place, certain nationalities favoured certain initials; for example, there were more names starting with the letter M north of the Border than there were south of it. That would be important if comparing geographical distributions, and it would also apply if the distribution by type of employment were being examined as there were certain occupations, such as the medical profession, which were favoured by the Scots.

Secondly, a sample based on the initial letter of the surname would include more persons of the same family than a pure random sample. For pension fund purposes those limitations would probably be irrelevant, but they should be borne in mind. A sample covering workers in England and Wales and based on the letter B included generally about 10% of the universe.

Though he agreed with what had been said about actual pension fund valuations, he considered that the use of sampling might be justified when it was proposed to set up a new pension fund and the employer, who might be a public authority, required an estimate of the cost of setting up the fund. In those circumstances the employer required a broad indication of the cost of the back service benefits which the fund was to bear before he could come to the proper policy decision. He wanted to know whether the cost would be one million, five million or ten million pounds. In those circumstances, the margins of error given by the author for a 10% sample would appear to be acceptable.

He had studied the paper with great interest. There was no doubt that the size of pension funds was growing and many lower-paid workers not provided for were asking that pension funds should be set up. The managements of large organizations required sound estimates of cost on which to base their policy decisions. He was sure that, as progress took place, more use would be made of sampling in making such estimates, and he considered that the paper was a valuable contribution to the greater understanding of a subject requiring much practical skill.

Mr R. C. B. Lane referred to a practical experience which he had had in valuing a fund using a sample. It was not, he remarked, a pension fund, and

he thought that in selecting a pension fund the author had made things as difficult as he possibly could, especially in regard to pension funds in which salary scales were really important, as it was probably fair to say that no two members of a scheme were ever exactly alike. The case with which he himself had had to deal had been completely different. It was a large organization with some hundreds of thousands of members spread all over the United Kingdom in many branches.

The fund which had to be valued was operated in three sections. One of those could be ignored, because it was so small. There were two important sections, one providing accident-pay benefits of a more or less standard kind, and the other provident-fund benefits, quite a small capital sum payable at age 65 or earlier on disablement down to age 60, the amount varying to some extent with the duration of membership.

He had come into it at a late stage. There had nominally been an actuary previously, but so far as he could see they had made almost no reference to him at all, and they had certainly given him no opportunity of doing any really useful work in connexion with the fund. Four or five years before he himself had become connected with it, however, they discovered that their provident benefits at age 65 had been running right away from the contributions coming into the fund, and that had been a sign to them. No doubt they had been warned many times but had taken no notice, as often happened. As a result of the trend they increased the contribution in the ratio of between two and two and a half to one and found a large sum from another source to bring into the fund.

The valuation which he had been asked to do was the first which could be called in any sense a reasonable one, and they wanted it done on a sample. He had not been keen, because he had had no experience of that sort of thing, and, like the actuary asked to sign a report of that kind for the first time to whom the opener had referred, he wondered what the other 90% would show. It was not, however, a fund which had to be valued under any statutory provisions; whether it was valued or not depended entirely on the management of the fund. It was clear to him from what they said, and also from the history as he knew it, that there was no hope of getting them to have the fund valued regularly every five or three years—they contemplated a three-yearly valuation—if they had to upset all their records to get the complete data. They had record cards to cover everyone, but they were not good cards, and the practical work involved in getting them into a really good state would have been quite intolerable to the management.

He had wondered whether to use a random sample of members or a selection of branches and, if he used selected branches, how to pick them. For a first sample he told the management to pick about 10% of the branches, feeling that if they themselves picked them they would at least believe that they formed a fair sample—and that was something. The data were extracted. He had absolutely nothing to go on as far as the accident experience was concerned except the totals in the accounts. He took the data from what was supposed to be a 10% sample and got out some accident-pay rates which seemed absurdly low, and so he added a 50% margin but still found a good deal more money than was needed on that side of the fund. The provident benefit was not difficult to deal with and experience there mattered little, because it was only a question of finding the reasonable probability of numbers going early, and the greater part of them did in fact go at 65.

He thought that it was a 10% sample, though he did not really know, but he

was able to obtain a great deal of useful information by comparing as many things as he possibly could with the figures in the accounts. He had the contributions under several heads. He had the claims under quite a number of heads both for the main and subsidiary benefits (not all of which he had mentioned). By seeing how the claims which he expected from the sample and the contributions which he expected from the sample, and so on, matched up, he was able to get a reasonable degree of certainty about how big the sample really was, and he finally decided that in fact it was the equivalent of about one-eighth of the effective operative part of the fund. It seemed clear to him that they had made a reconstruction which was probably sound, and that there was nothing much to be done unless it was to increase the accident-pay, but they had had enough of increased benefits and would have nothing to do with that.

Three years later he used the same sample again, and that was when some of the more interesting points began to come out. The reason he used the same sample again was to form some judgment on how stable the experience was likely to be. If he had used a different sample the final results would have been impossible to interpret, because he would never have known whether differences were due to the different sample or to instability in the experience.

The first point which emerged was that they were not marking off their retirements properly. He found that a very substantial number indeed had passed the age of 65 but were still left in. He mentioned that particularly because that came out from the 10% sample. That was by far, he felt, the greatest defect in the data with which he had had to deal. If he had gone to all the trouble of taking out the whole data he would simply have had ten times as many of those retirements which had not been properly recorded.

Another point which came out was that they did not seem to collect all their contributions and seemed to have a number of void members still kept in the records. He had not found any way of getting rid of them.

He was again doing a valuation, and in the current one he was taking a random sample, he thought a truly random sample. He had fixed a random number for each branch, which was to be the last digit of the membership number. It might be stratified in some curious way which he had not thought of, but it was as near a random sample as he could get and he was going through every branch to get it. What the results would be he could not say.

When arranging that with the organization, he found that they said that it would be awkward, because they would not be able to make sure that the data of those branches were right. That was usually the difficulty in getting a sample from any administration. It was not without its advantages, because it did make them try to get their records right. Even then they were not out of trouble, and the branches, though perhaps sending in voids earlier than usual, sent them in too late to be much good. He did not see the point of waiting for all that to be done, however, because it would mean being another year behind, and he did not think that the data would be really right even then.

What had been the final upshot of it all? It had certainly saved a tremendous amount of labour. Probably it had been possible to do the work reasonably for about one-fifth of the cost, because the cost of taking out the data was tremendous, quite apart from the disturbance in the office itself. It therefore fulfilled the requirement that a substantial amount of labour should be saved. Whether it did all that could be expected of a full valuation he did not know, but he had some confidence that it showed that the funds were currently all

operating on a sound and stable basis, and that they were running at a profit. There were various *prima facie* reasons for that; for example, the contributions seemed to be thoroughly adequate, and that was a good beginning. The process of looking at the profits and attempting to analyse the samples and tie them back at as many places as possible to the accountant's figures, gave confidence in the method. If the fund was run on a sound basis, and the management were not anxious to spend every pound of surplus disclosed, it was possible that a sample valuation would always be satisfactory. It might not distribute surplus right up to the hilt, but it would enable control to be exercised over the fund and prevent serious trouble.

The author had done for his fellow actuaries, who might use samples in practice, something which they could not do themselves. When they used a sample they could not know what the error was. In doing his experiments the author had drawn attention to the important factors which arose outside life-office organization or the valuation of any other fund where the organization was more or less under the control of the actuary, and he drew attention to the fact that the errors inherent in the valuation where benefits depended on sickness, accident, salary scale and so on were much greater than in a life office. There was always an inherent probable error which was not found elsewhere. The author also drew attention to the difficulty which existed in many cases of being sure that the right data had been collected: it was a real problem.

Mr M. T. L. Bizley said that the method of valuing by random sample consisted essentially in selecting a 10% (say) random sample, valuing those members in the usual way, and multiplying the result by 10. To appreciate what was being done, it might be helpful to imagine that a card had been written for every member of the fund, and that on each card there had been entered the liability in respect of that member. The object was to find the sum of the entries on all the cards. The method adopted was to draw at random 10% of the cards and in effect to add up the entries on those 10% and to multiply by 10.

The problem, therefore, amounted to this. If there were a large number of cards, say $10n$, and on each card there was a figure which might be anything from 0 to perhaps £20,000 or so, and n of the cards were drawn at random, what was the probability that the sum of the entries on the n sample cards would differ from one-tenth of the sum of the entries on all the cards by more than an acceptable predetermined margin of error? That was a problem of a familiar type, and it was tempting to apply familiar methods to tackle it; there was, however, a serious obstacle, namely that the liability-distribution (as it might be called) of the fund was not known. In other words it was necessary to know for each value of x how many cards there were with a liability entry of £x. If the problem could not be solved, the probability of a bad result was not known and, if that was not known, not only was it not possible to say that a bad result would not occur but, what was worse, it was not possible even to say that such a result was unlikely to occur.

The author had provided much arithmetical information, but, interesting though his results were, it seemed to the speaker that what he had really done was to pose a problem in probability, which demanded a solution before the method of valuing by sampling could be used safely. The liability-distribution would vary from one fund to another, and even within one fund it would vary from time to time. There seemed, therefore, little hope that a theoretical

investigation based on the results from one fund at one time could be used in another fund at another time, or even in the same fund at a later time. It might be that if an extreme set of data were employed the results derived might be of use, provided the results of the investigation showed that the probability of an unacceptably large error was satisfactorily small, even for the extreme data, and hence, if the deduction was legitimate, for other data *a fortiori*. If any such theoretical investigation could not be made it seemed to him that the actuary could never really have confidence in his results.

Mr J. K. Scholey remarked that when considering the valuation of a pension fund it was necessary to bear in mind exactly what it was hoped to gain from the valuation. There had been talk that evening of errors in the results as though there were a definite result of the valuation which stood, but he did not know that there was any definite result of that kind. After all, there might be a fund which was solvent to the extent of 105%, but which might still be regarded as in difficulties. There had been life offices which were invariably solvent on valuation, and yet the actuary had felt it prudent to say that no bonus should be declared for the current year or for the 5-year period. Alternatively, there might be a fund which was in deficiency, the valuation showing it to be in deficiency, but there might still be a case for some increase in benefits.

It seemed to him, in a way, that the more important part of the valuation was to trace what had happened since the previous valuation was made. It was not a question of stopping something to examine it, but of examining the flow of the fund, and it was necessary to examine the profits and losses which had arisen since the last valuation had been made. If samples had been taken with an interval of five years between them, there might be a result which was 2% too high in the first case and 2% too low in the second, and when the time came to analyse the profits and losses it would not be possible to check the result or to know which valuation was wrong, if either was, or whether something had gone wrong with the data. It seemed to him that if valuations were being done over a period of years it would be better to take the sample to start with and continue to examine the same sample. A sample of, for example, one in ten of the people who entered the fund would be taken and that sample would be kept on its own. It might be necessary to keep a separate fund for that 10% sample; otherwise, there might be difficulties.

Another purpose of a valuation was to tell the members of the fund what a certain additional benefit might cost. In that case it was perhaps legitimate to take a sample. There would still be the difficulty that the sample taken might not be a true one. For example, there might be a big fund of 10,000 members, but there might be only half a dozen earning more than £3000 a year, and if a 10% sample were taken they might all be cut out.

On the question of what was a proper sample, he said that no one had actually come down to the numbers that were needed in it. His own experience had been that if there were, say, 50,000 members in a fund, a sample of 5000 would often give reasonable decrement rates. The actuary did not start from scratch in those matters; he started with a general knowledge of the trends of rates of mortality and ill-health retirement, and could check his ideas against the experience of the fund itself.

On the question of time it was often a question not of how much time there was absolutely but of how much time there was relatively. If someone asked a question and wanted the answer in three months, two and a half months could

not be spent in producing data He himself had had a fund with 50,000 members, and even with punching machines, etc., it had taken 6 weeks to produce the data after setting the machinery in operation. First it was necessary to make sure that all the data had been got together, then all the punching had to be done and even then there was all the clerical valuation work. It was not reasonable to dismiss it and to say that in any case the clerical work did not take so much time. He agreed with Mr Lane that it was something to be borne in mind.

On the question of salary scales, he felt that the author had given a wrong impression by seeming to countenance the method of deriving salary scales by taking the average salaries and smoothing them out. He did not think that smoothing them out gave any kind of scale. As the opener had pointed out, in some samples given in the paper people at age 50 might be getting a higher salary on average than those at 65, but that did not mean that they were going to get a lower salary at age 65 than they were then receiving. If a sample were taken the actuary had still to look at the facts of the case and say to himself 'What sort of salary are the people who are now 50 likely to get at 65?'

Then there was the question of the omission of various decrements. If a particular decrement did not affect the result very much it might be argued that it could be left out, but in his view if there were a particular decrement it was much better to bring it in. It might be proposed subsequently to alter the benefits in certain ways which would make it unreasonable to have left that decrement out; then, if it had been left out, it would be necessary to start all over again, whereas it was not very difficult to bring in another decrement at the beginning.

Mr W. Perks pointed out that there was a solution of the theoretical problem that Mr Bizley had formulated, because the liability distribution could be estimated from the sample itself. Mr Bizley had assumed that the service table and the salary scale to be used in the valuation were given, so that for each member there was a fixed liability amount which could be put on his card. If, then, a sample was taken of those cards the distribution of liabilities could be obtained from that sample or even from quite a small sub-sample and the standard deviation of that distribution would provide an adequate estimate of the standard deviation of the universe, and therefore would provide the actuary with sufficiently accurate estimates of the probabilities of the sampling errors.

In the paper under discussion, however, it was not assumed that the service table and the salary scales were in fact given; it was assumed that they were obtained from the samples themselves, and therefore the liability values obtained from the sample would be slightly different from the liability values obtained if the whole process were carried through for the whole membership. The estimate of the standard deviation from the sample would be reliable enough for the sampling error but there would remain the error due to the discrepancy in the valuation basis. However, there was nothing sacrosanct about any service table or salary scale, and in practice the service table and salary scales often had a certain element of arbitrariness about them; he thought, therefore, that with the idea of estimating the liability distribution from the sample there was something of practical value in Mr Bizley's theoretical formulation of the problem.

Mr C. A. Poyser said that the different ways of drawing the samples had been discussed, and it seemed to him that the feature which was being investigated dictated the kind of sample to be drawn. The author had attempted to investigate virtually everything about a pension fund, and had therefore quite

rightly taken a 10% sample in more or less random fashion; but if, for example, it was desired to estimate the valuation liabilities alone he thought that, as the liabilities started at nil for the very youngest members and reached a very high figure as the retiring age was reached, it would be better to weight the sample so as to have perhaps 1% or 2% in the age-group 20–30, 5% or 6% in the age-group 30–40, and so on, increasing the percentage towards the higher ages, at which the liabilities were largest and most important.

Mr H. Tetley confessed that in his rather limited experience in the field of pension funds the problem of whether or not to sample did not usually arise; as a rule sampling was essential, there being no other practical way of finding the necessary information. In fact, he would say not only that sampling had a future but that it had a very definite present.

It had been emphasized that in life-office valuations the size of the surplus was important, both relatively and absolutely, because it would lead normally to a distribution of surplus, i.e. to an immediate increase in the liability. The same point did not arise in a pension scheme and much larger errors could be accepted; but he confessed to a completely illogical hesitation about altering the sign of the surplus. In other words, he would not mind very much an error of £3000 in a surplus, but he felt a little uneasy about showing a deficiency of £1000 which should in fact have been a surplus of £2000. From that point of view he had misgivings about doing the entire valuation of a pension fund by sampling. As far as he was concerned, therefore, the problem posed by Mr Bizley did not arise, but what was often of great importance was to investigate some of the barnacles which had grown so thickly on the clean lines of a pension scheme. Pension schemes were becoming more and more complicated and the benefits were becoming greater in number and far more complex. It was often essential to treat many of them by approximate methods which were still sufficiently accurate.

There were several points which arose in that connexion. It was surprising how little the managers of some funds knew about the actual importance of some of the excrescences on their funds—for instance, how many members had taken up a particular option. They were usually completely in the dark, particularly if their records were not centralized. It was often extremely valuable to investigate many of the frills by a sampling process, but it was essential in doing so to get particulars of one or two items of basic data such as age and salary. That point had been made earlier. It enabled the actuary to test whether his sample seemed to be truly representative in those two important points, and then he could work on it with a good deal of confidence with regard to smaller items. It would almost certainly be found that the financial effect was so modest that quite roughly approximate methods would be satisfactory and that the data thrown up by the sample would be all that anybody could reasonably require.

That meant that the calculation of those rather recondite factors could often be started well ahead of the main valuation. While the basic data were being collected—a process which consumed much time—the actuary could be analysing his sample, preparing his factors and producing figures for many of the ancillary benefits which he could have confidence were sufficiently accurate for practical purposes, and he would not be held up waiting for the main data to come in.

With regard to the actual methods of drawing a sample, in his experience the

most important point of all was for the actuary who was going to be in charge of the work to know exactly who was going to draw the sample, who was going to supervise it, and what was the calibre of the people who would be doing the work for him. Those factors could vary enormously, and depended greatly on whether the data were centralized at head office, where there would be an efficient staff fully conversant with them, or scattered throughout the length and breadth of the country and obtainable only by asking the individuals involved to fill in a form or to supply the information orally. A typical example was a wayside station with a total staff of 12. Random numbers in such a case would be useless, because the data when obtained would not have been correctly drawn. In other words, the sampling process must be fitted to the capabilities of the people who would carry out the work. It was far better to have a less satisfactory method of sampling which was sufficiently simple for the people involved to be able to carry it out efficiently and with scrupulous accuracy than to have a method which was much more satisfactory from a theoretical point of view but so complicated and seemingly so remote from reality that there could be little confidence that reliable data would, in fact, be produced.

There were several snags which had come to light from the little work which he had seen which might be of interest to members of the Institute. The first concerned the method of sampling by final digit, or final two digits, which had been used to a large extent in the U.S.A. and had produced some extremely good results, but had shown some rather dangerous warning lights. In the first place, if it was a question of tracing a sample from one valuation to another the utmost vigilance was essential to make sure that all the decrements or increments were faithfully recorded. If one sampled every number ending in 8, it was important, if the records were not centralized, to see that each office or sub-group faithfully reported every case coming in—every death, withdrawal and retirement—which ended in 8. That did not always happen, and it could throw out the results badly.

There was another curious feature: if every number ending, say, in 7 were taken, it would seem that that should give 10% of the total; but in actual practice it never did. The reasons for that varied, and were not without interest. On one occasion the clerks who were allotting numbers were told to fill in a book, each page starting with 0 and ending with 9. They went through the book taking each case as it came, and the process should have been perfectly random, but when the numbers were analysed it was found that those ending in 7, 8 and 9 were seriously deficient, 7 being worse than 6 and so on. An investigation was made to find what was happening, and it was discovered that when a clerk went to lunch or stopped for a cup of tea or a chat he often, when he started work again, said to himself 'There are only a couple of lines left on that page; I will start a new page'.

There was also another and rather more subtle reason, which was particularly noticeable in the American experience. There were hundreds of points throughout the U.S.A. which were issuing numbers to people who were taking out a certain benefit, and each point was allotted a certain range of numbers and told that when those numbers were exhausted they were to ask head office for more, when they would be given another range. Every range started with a number ending in 0. If a census were taken at any moment there would be a complete set of 0's, but at that moment an office might just have issued a number ending in 7. It would automatically follow that every number below 7 was fully represented, but the 8's and 9's would be one short. That, by itself, would not matter but when it happened in hundreds of cases it would be found

that there was a serious deficiency of numbers ending in the higher digits; there would in fact be a progressive shortage from 0 upwards, because every range started with a 0 and throughout the country there would be so many incomplete series. The only way to avoid that difficulty was, not to give the points a group which always started with 0, but to give to one a series starting with a number ending in 3, to another a series starting with a number ending in 9 and so on. If they were sufficiently random throughout the country the difficulty would not arise.

Then there was the use of the initials of surnames referred to by Mr Lloyd, which brought in the difficulties of families, of Welsh names and so on. That presented serious problems, because some industries were concentrated in Wales, and unless care was taken there would be a deficiency. To take Jones, Evans and Williams would give a heavy sample, but other names would produce one seriously deficient. A method of correcting for that had been suggested and was satisfactory for most purposes, namely to ask for a statement of the total number at the particular point where the sample had been drawn and to correct for any deficiencies. That gave a good correction for the mean expected value, but considerable care needed to be taken with any measures of dispersion based on data some of which had been grossed up while the rest had had no correcting factor applied to them. Such an amalgamation would upset the dispersion, and care would be needed in interpreting the results.

Another method which he had seen used, and which was particularly appropriate if all the data were in one central place and were numbered, was for the actuary or statistician who would be in charge to prepare a list of random numbers and to ask for full particulars of all the corresponding people and no others. That, however, had two snags. The first was that the numbers might not be allotted continuously throughout the whole range. It might be found, as he himself had found on one occasion, that a particular range of numbers was reserved for the managerial staff, that another range represented some other stratum, and that there were odd gaps. (Occasionally, for no good reason, numbers suddenly appeared in the middle of the gap!) The best plan then was to stratify the sample, but it was necessary to know exactly where the numbers were and where the gaps occurred before constructing the list of random numbers.

The results obtained in that way were often extremely valuable, but it was essential that someone who was really reliable and conscientious should be on the spot to check the work because, when a clerk was given a list of numbers and a schedule of particulars to be filled in, there was a great temptation for him to put in 'Retired' or 'Withdrawn' if he failed to find a particular file, even after repeated trials. It was often extremely important, however, that such files should be found, because they usually related to the sort of person about whom one would particularly want to have information. It was therefore necessary to stipulate that every one of the cases on the list should be produced. If alive and active the fact should be recorded; if it was alleged to be a death, withdrawal or retirement the file should be found and produced, or else the sample would be unreliable and might be dangerously biased.

He had tended to concentrate on the practical aspects because he felt that they were extremely important. Sampling was a practical weapon which, whether they liked it or not, they had to be prepared to use; though it was not ideal in many ways, it could be perfected. They should preserve an open mind about it and be prepared to consider when it was useful and when it was likely to be dangerous. It was only by experience that they would find out exactly in which category to put a particular problem, but unless they were prepared to investigate

the uses of sampling it seemed to him that they would never get very far with the complicated investigations which arose.

Mr W. H. Clough, in closing the discussion, emphasized the different approach of the actuary when he was valuing an internal fund from that which he adopted when he was valuing the liabilities of an organization in which the distribution of profits was the first essential. When a private pension fund was set up, the technique was to provide as good an estimate as possible of the liability of the employer year by year in making up that deficiency of the fund which was left after the contribution of the employees and interest earnings had been taken into account. There was a good deal of significance in the author's remark that the first valuation of a fund was of much more importance than the subsequent valuations. While he would not be satisfied, at least with a number of clients, to make an estimate of the cost of the fund on a 10% sample at the outset, and while he would be anxious if at all possible to make a full valuation and to obtain full experience of a fund at an early date, there was much to be said, particularly in times when changes were so rapid, for considering the possibility sometimes of making a sample valuation, involving less total labour and possibly less expense, at more frequent intervals than those at which in the ordinary way a full valuation would be conducted.

He thought that the conclusions drawn by the author in the paper left that as a practical point to be considered, as should his remarks with regard to the estimation of subsidiary benefits by sampling methods in order to avoid labour.

If the significance of the valuation of a pension fund was accepted as the speaker had expressed it, there was not the same necessity to pursue the point of discussion between Mr Bizley and Mr Perks with regard to the significance of the liability attaching to each member and the importance of the valuation balance sheet to which Mr Bizley had referred in disclosing a deficiency or a surplus. What the actuary was most interested in for a pension fund was first of all the surrounding circumstances, to which reference had already been made, the 'feel' he had of the fund and the trends which were taking place. Whether in fact at any time, on the assumptions which were being made, those trends revealed a small deficiency or a small surplus was not, in itself, important. There should be borne in mind the difference of treatment of depreciation of assets in summing up the position of a pension fund as compared with what might be regarded as the right criteria when dealing with a life assurance fund, because in a pension fund benefits would be allowed to flow as they had been flowing if, with a well-spread portfolio of investments, the situation at the time did not cause too much worry, and much more latitude in that respect was allowed than was allowed in a life office valuation.

Mr Lane had touched on the significance of sampling methods in other connexions, in those cases when it was much better for the actuary to know that he had a sample of the data than to fear that only a sample was being presented when he was supposed to have the whole of the data. There were many instances of the necessity to employ sampling methods. One had come across his path recently where an organization had suddenly been called upon to make a valuation because the type of benefit which it had been offering had changed and, owing to a small contribution and a large spread of membership, no reliable facts as to data had been kept. It was possible, as Mr Lane pointed out, by a sampling method at least to get the 'feel' of such a fund and achieve over a period of years what Mr Lane described as control of the situation. That was more important in many instances than arriving at what looked like a good arithmetical answer

but one based on flimsy data, although, as Mr Tetley had said, great significance attached to being certain that the sample which was being taken had been given with every emphasis on the surrounding circumstances.

He admired the author for the full way in which he had worked out the significance of sampling in the handling of valuations, and thought that he had placed much valuable information in the hands of consultants who had from time to time to face such problems.

The President (Mr W. F. Gardner), in proposing a vote of thanks to the author, said that what had pleased him was that, among other things, the author set out not so much to show what sampling would do as to inquire whether sampling was appropriate or not, and when he found that it was not he did not hesitate to say so.

He had also been glad to find that the contributors to the discussion had been able to exhibit, as usual, not only a knowledge of the technique of the subject under discussion, but the possession of a gentle fund of humour and a full understanding of human nature. Those qualities were very important to actuaries, as indeed to others.

He thought that the author would be encouraged by some of the remarks which had been made about the possibilities of sampling even in the most difficult field which he had chosen. He might also be heartened by some of the practical examples which had been given.

Mr F. Gordon Smith, in reply, said that he was prepared to admit that he could have evolved more appropriate salary scales from the data. He had used a method of graduation that was a purely mechanical process because he might have been accused, if he had obtained better results, of selecting a suitable graduation of the salary scale to produce small errors in the valuation. It was difficult to avoid bias if other methods were used.

He had been interested in the use of sampling for the valuation of complicated benefits, a subject which had been referred to by several speakers. The question of emerging costs, especially with large funds, was an important one. Accordingly, he thought that an investigation into the effect of sampling on emerging-cost estimates would be valuable. He had also been interested in the reference to the actual use of sampling in the case of large funds, even although they were not pension funds.

One speaker had stated that actuaries were more interested in the flow of the fund than in the position at a particular point of time, and with that he was in complete agreement. That had been in his mind in mentioning that a sample selected in a particular way should be taken for successive valuations; it made it easier to see how the fund was progressing. So he was led to the point that there was no bonus distribution in a pension fund and often—not always—the employer was behind it financially. That made the problem a little easier than it was for a life office and might make it possible to accept a larger margin of error. If the results from the sample were striking enough to suggest that something was going wrong, and that a change in benefits or contributions was necessary, a full valuation could always be made. In that case a certain amount of time might have been wasted but, fortunately, such a situation was not the usual one disclosed by a valuation.

He felt that some mistrust of sampling was due to lack of familiarity with its use. Improvement in technique and greater experience might well mean that it would be used more in pension fund valuations in the future.