JIA 77 (1951) 0015-0074

GRADUATION TESTS AND EXPERIMENTS

By H. A. R. BARNETT, F.I.A., A.S.A.
Actuary, British-American Tobacco Company, Ltd.

[Submitted to the Institute, 27 November 1950]

I. INTRODUCTORY

LITTLE space in the *Journal* has been devoted in the past specifically to graduation tests, and nowhere have the tests generally applied been concisely set out and fully discussed. The only papers during the last forty years on the subject have been by Seal (3) and Daw (2); both these papers were submitted and discussed during the war years, with the result that many actuaries were unable to be present and state their views; indeed, many of us did not even know of them until the numbers of the *Journal* in question were published, and an opportunity to study them did not arise until after the war.

2. Neither of these papers sets out to state or discuss all the tests. The purpose of Seal's paper is to describe two tests in particular, the P_Q and χ^2 tests, and after reading the first few pages one is left with the impression that the only tests commonly employed by actuaries are the mean deviation test and the examination of deviations in five-year groups! Daw's paper does state briefly some of the customary tests, but he does not pursue the matter, as his paper is concerned primarily with the validity of the assumption that the sampling variations of q_x follow the binomial distribution.

3. Since work was started on the preparation of this paper, a further paper by L. Solomon (4) has appeared in the *Journal*, but here again the subject is specialized, and does not deal with the general question of graduation tests.

4. Frequent reference will be made to these papers, and to the discussions on the first two; this is partly to collect together the relevant considerations, and partly because the war-time discussions may not have been very widely read. On the other hand, it is felt that Seal's now well-known paper and closing remarks (as opposed to the discussion itself) have been too widely read, especially by students; it is alarming to observe the number of them who are given the impression that no graduation tests are necessary except the χ^2 test and a sign-change test. The mischief may be traced to *J.I.A.* LXXI, 61, where Seal stated:

> any actuary, and indeed anyone who had passed Part I, could carry out the χ^2 graduation test....It was then necessary to test for sequences of positive and negative signs and the work was almost finished.

The reading of Seal's paper by a student is often like trying to run before he can walk, and one of the purposes of the present paper is to state concisely just how to 'walk'; for this reason, it might be considered that this should be a Students' Society paper, but the author feels that, since the pages of the *Journal* lack a complete paper on graduation tests, its correct place is as an Institute paper. It is hoped that an instructive discussion will be provoked, especially on the more controversial questions arising.

II. THE PURPOSE OF GRADUATION

5. Before turning to the problem of which tests should be carried out, we must consider what a graduation sets out to do, and why we need graduated rates. Our investigation of a body of data will have given us crude rates (of mortality, or of whatever decrement or happening we are investigating). Why, then, need we go any further? What will graduated rates do that crude rates will not?

6. We require tables of rates to enable us to calculate premiums, contribution rates, policy values, actuarial liabilities, values of reversions, etc., in fact, values of any functions dependent on the probability of the particular event under investigation. The crude rates will exhibit irregularities which would be inconvenient in practice, and which would not be consistent with the conception that the 'true underlying rates' should progress smoothly from age to age. Several examples of the possible inconveniences could be given, the most obvious one being whole-life premium rates decreasing at certain ages. There is no reason why the graduated mortality rate should not show a regular decrease to a minimum, followed by a regular increase. What our conception of underlying rates will not admit, however, is a jerky series of increases and decreases from age to age, thereby causing further irregular series in premium rates, policy values, and any other functions calculated from the original series of crude rates.

7. The purpose of graduation may therefore be stated to be to obtain a smooth series of rates exhibiting the same general features as the jerky series given by the crude rates.

8. The calculations for which such rates might be used would assume either

(*a*) that the proportionate frequency of the event in question (e.g. death) will be that of a sample drawn from the same universe as that from which the sample giving rise to the crude rates was drawn; or

(*b*) that the proportionate frequency will be that of a sample drawn from a universe whose properties have been estimated by means of a forecast based on projection into the future of trends observed in the past.

9. Assumption (*b*) above involves some process of extrapolation and will not be considered in detail for the purpose of this paper; the problem therefore reduces itself to the estimate of the universe required for assumption (*a*). We require a smooth series of rates from which the crude rates do not differ to an extent which is statistically significant; unfortunately, there might be an infinite number of such series, of varying degrees of smoothness, and we have to choose the best of these, bearing in mind their relative smoothness and the significance of the resulting deviations between the crude and graduated rates. It is these considerations which give rise to the necessity for graduation tests.

10. Now, even when an actuary makes assumption (*a*), he will not have assumption (*b*) far out of mind, because he knows that his estimate of the universe from which his data have been drawn will be used as a forecast of expected future experience. It might therefore be argued that assumption (*a*) should be dropped completely; but the idea of a purely hypothetical forecast table has not been generally accepted, and even if we had sufficient mortality tables from past experiences to enable us to project into the future (as was done for the $a(f)$ and $a(m)$ tables) we could get so many different series by slightly varying the method of extrapolation that the forecast table would contain a large element of guesswork.

11. For this reason, actuaries generally prefer to be able to use graduated rates which have been based on some definite experience, even if it involves, when such rates become out of date, making some estimate of the changes which have since taken place or which might be expected to take place in the future. (An exception to this is, of course, annuitants' mortality tables constructed in the present century, which have been produced with assumption (*b*) in mind in view of the adverse effect of the improving trend of mortality on a company granting annuities. A similar consideration might be applicable to tables for use in pension fund valuations.)

12. If it is borne in mind, however, that although the actuary's tool-kit can only contain past experiences (it being impossible to foresee the future) what he really needs is a good forecast, it is small wonder that from time to time a high degree of adherence to data has been dispensed with in favour of a graduated series convenient for use in practice. It is not suggested that the data should be deliberately distorted beyond the limits of statistical significance merely to facilitate the fitting of a convenient curve, but in the case of graduations by a formula the general rule should be to make the simplest possible universe hypothesis consistent with the data. In other words, we would not want a curve with six parameters if we could fit one with four; we would not want one with four if three would do; we would not want a series of three blended Makeham curves when one such curve could be fitted to the whole of the data.

13. The expressions 'limits of statistical significance' and 'consistent with the data' have been used in the preceding paragraph; this presupposes the setting down of some limit of acceptance or rejection, and the 5% probability limit is the line frequently drawn. There is no particular magic in 5%, and some actuaries may prefer to draw the line at some other limit, but, for the purpose of this paper, it will be the limit taken; the principles involved would be no different were it required to use any other rejection limits.

III. TESTS OF A GRADUATION

14. Having made a graduation, as suggested in the previous section, either by fitting a curve of the simplest possible type consistent with the data or by some other method, we need to test whether the resulting series is sufficiently smooth, and whether the deviations shown between crude and graduated rates are within the acceptable limits. An infinite number of tests could conceivably be devised, and we would scarcely expect our graduation to satisfy every one; indeed, some actuary of the future might well discover a series showing the distribution of the number of tests passed by a successful graduation! The possible tests, however, may be grouped into the following broad categories which we will consider one at a time in the sections which follow:

Smoothness,
Adherence, individual deviations,
Adherence, groups of deviations regarding sign,
Adherence, groups of deviations disregarding sign,
Signs and sign-changes of deviations and accumulated deviations.

15. To illustrate the tests under these five main headings, a graduation of the duration 0 select data of the A 1924–29 experience over age 19½ has been made by a Makeham minimum-χ^2 method. Appendix 1 shows the graduated

values of q_x and the deviations required for the illustration of the tests. An outline of the theory and method of the Makeham minimum-χ^2 fit is given in Appendix 2, together with a short account of the experiments which were made in the derivation of this method, and some comments on Cramér and Wold's (1) method.

16. It will be assumed throughout the consideration of these tests, except where otherwise stated, that the exposed to risk is the sample, that the sampling variations of q_x follow the binomial distribution, and that this distribution approximates sufficiently closely to the normal; the data should therefore be grouped so that for no group would the 'expected' fall below, at the very least, 7. The validity of these assumptions has been fully dealt with in Daw's paper, and will not be considered further; it is worth mentioning here the r_x test, described in that paper, which does in fact test the validity of the binomial assumption.

17. To avoid use of the symbol 'E', which sometimes stands for 'Exposed' and sometimes for 'Expected', n_x will be used to denote the exposed to risk from age x to $(x+1)$; the number of ages or groups will be denoted by the symbol m.

18. A general word of warning may be uttered here, in connexion with the assumptions in paragraph 16; the presence of a large proportion of duplicates in the data will upset the assumptions, and unless appropriate adjustments (which are difficult to assess) are made the basis on which we define our rejection limits will break down. The exclusion of duplicates is therefore strongly advocated wherever possible. It is too late to do anything with the A 1924–29 data, and presumably it is too late to do anything about the data which will give rise to the next standard table; but it is suggested that suitable modifications should be made in the method of collecting data to enable the next-but-one standard mortality table to be graduated in the full knowledge that the variance of the number of deaths will really approximate to npq.

19. Daw and Solomon have both made suggestions as to the adjustments which might be made to allow for the effect of duplicates, and an attempt has been made to apply these adjustments to the ultimate data of the A 1924–29 experience. A regraduation of these data by a Makeham formula over the range of ages from 22 to 65 is given in Appendix 3; the author has not yet applied the Makeham minimum-χ^2 method to this section of the data, but the graduation shown is the best Makeham graduation he has so far obtained. Appendix 3 also shows the application to this graduation of the suggested tests which follow.

IV. SMOOTHNESS

Test 1. Inspection of differences.

20. Smoothness is a property which actuaries and others believe they can recognize in a somewhat undefined manner. Indeed, the recognition of such a quality is implicit in the necessity for graduating a crude series. The criteria frequently accepted in the past have been that the first and second differences should progress smoothly and the third differences should be small; but this is tantamount to saying that the curve must approximate to a polynomial of the second or third degree, and places far too much restriction on the shape of the curve to be chosen. Tetley (6) has amplified this by stating that smoothness in the successive orders of differences is more important than smallness, but

this brings us back to our undefined quality of 'smoothness'. If we say that a series is smooth if its third differences are smooth, are we not then saying that the sixth differences must be smooth, and so on *ad infinitum*? If this were to be accepted, then, to graduate, for example, a series of forty-nine values, it would be possible to fit a forty-eighth difference curve which would reproduce the original series, and to state that it must be ideally smooth. Clearly, this would not fulfil our conception, still undefined, of smoothness. Further, we should be no better off in defining a smooth curve as one approximating to the curve of any mathematical function; the series $\sin x$, at intervals of $3\pi/5$, though apparently far from smooth, follows a definite mathematical law.

21. Taking all these points into consideration, the following rather vague definition may be made of the quality implied by the word 'smooth':

A series is smooth if it displays a tendency to follow a course similar to that of a simple mathematical function.

The inclusion of the word 'simple' is intended to cut out polynomials of a very high degree (such as the forty-eighth difference curve already suggested) and the irregular series which can be produced by complicating a mathematical formula with trigonometrical functions. It would not necessarily exclude exponential functions.

22. This definition of smoothness may be said to be fulfilled if successive differences of the graduated rates tend to diminish; this would include any Makeham graduation where the function c is less than 2, and, indeed, all the usual shapes of curves which are likely to be met in practice in the analysis of those events usually investigated by actuaries.

A more comprehensive criterion is that changes in sign should be rare within each column of differences. Such a criterion as the basis of a test, however, suffers from the disadvantage that it is difficult to lay down limits of acceptable frequencies of sign changes.

23. The first test of graduation, the test for smoothness, may therefore be stated to be to take out the differences, up to a certain order, to ascertain whether there is a tendency for differences up to that order to diminish, or alternatively whether each column of differences up to that order is nearly free of sign changes (beyond that order, they might show signs of tending to increase again or to change sign frequently merely because of the restriction on the number of figures to which the original series has been expressed). A graduation would not, however, be rejected if it fails the smoothness test merely because it has retained an irregularity known or believed to be a feature of the universe, such as might occur at the end of the child-bearing range of ages in a table of mortality rates of females.

24. If the proportionate frequency of the event under investigation is known to contain properties such that this test would not be applicable (e.g. if it may be expected to approximate to a sinusoidal function) then the tester would merely have to satisfy himself as to the shape of the graduated curve and the usual smoothness test would not be applied.

25. If the differences of the series shown in Appendix 1 are taken out it will be found that they tend to diminish up to the fourth order, and therefore the series passes the smoothness test, as indeed it is bound to do, being a Makeham curve with c equal to 1·1045704. The fourth differences will be found to be somewhat irregular, due to the cutting down of the graduated values to eight places of decimals.

V. ADHERENCE, INDIVIDUAL DEVIATIONS

Test 2. Approximately half of the actual deviations disregarding sign should be less than the probable error of the expected number of events

26. On the assumptions stated, the chance of an observed deviation $|(\theta - nq)|$ being less than the probable error ($\cdot 6745\sqrt{npq}$) of the expected number of events (e.g. deaths) is one-half. Therefore, if the investigation covers m ages or groups (the grouping having been made as described in paragraph 16 to permit the binomial-normal approximation) and we regard 'less than probable error' as a 'success' and 'greater than probable error' as a 'failure', we have another binomial distribution, this time of the form $(\cdot 5 + \cdot 5)^m$, the mean of which is $m/2$ and the standard deviation $\sqrt{(m/4)} = \cdot 5\sqrt{m}$. At the 5% probability level of acceptance the graduation would pass the test if the observed number of 'successes' were within the range $m/2 \pm \sqrt{m}$; Table 1 shows these limits of acceptance for quinary values of m from 30 to 80. This paragraph assumes, for the second time, the approximation of the binomial to the normal distribution, but as the '50% success' binomial is symmetrical, the assumption is reasonable without specifying any minimum value for m.

Table 1. Limits of acceptance for Test 2 at 5% probability level

Number of graduated values m	Lower limit $m/2 - \sqrt{m}$	Upper limit $m/2 + \sqrt{m}$
30	9	21
35	11	24
40	13	27
45	15	30
50	17	33
55	20	35
60	22	38
65	24	41
70	26	44
75	28	47
80	31	49

Note. The lower and upper limits in this table are expressed to the next lower and next higher integers respectively, so that not more than one-twentieth of true hypotheses would be rejected.

27. If the number of 'successes' falls well within or well outside the acceptance limits we would respectively accept or reject the graduation on this score without hesitation. Where, however, the number of 'successes' is near the borderline it is suggested that we should test the frequencies with which the function $|(\theta - nq)|/\sqrt{npq}$ exceeds values other than $\cdot 6745$. After all, we only take $\cdot 6745$ as the basis of this test because of the simplicity of a 50% chance of 'success', and, if it exceeds $\cdot 50$, $\cdot 75$ and $1 \cdot 0$ an acceptable number of times each but because of a number of near misses narrowly fails to satisfy the $\cdot 6745$ test outlined above, it is unlikely that the graduation would be rejected provided it satisfies all the other tests described in this paper.

28. In the graduation shown in Appendix 1, it is seen from column (7) of Table 6 that only 16 of the 49 values are less than the probable error, whereas the

acceptable limits are 17–32. If we then make similar tests with reference to $\sqrt{npq}$ (32% expected to exceed), $\cdot 75\sqrt{npq}$ (45% expected to exceed) and $\cdot 5\sqrt{npq}$ (38% expected to fall below), the acceptable frequencies of these happenings would respectively be 9–23, 15–30 and 12–25; the observed frequencies, 22, 29 and 12, are all just acceptable, and the graduation should therefore not be rejected on this test of the actual deviations, although it is undoubtedly a borderline case.

29. If the number of 'successes' is greater than the upper acceptance limit, then the conclusion is that the ungraduated series has been too closely adhered to, either by a poor graphic graduation, by a weak summation graduation, or by the fitting of a curve with too many parameters.

Test 3. Approximately 95% of the actual deviations disregarding sign should be less than twice the standard error of the expected number of events

30. This is similar to Test 2, and the data must be grouped in the same way to enable the figure of 95% to hold. If we regard a 'success' as 'less than twice the standard error' the binomial distribution underlying the test is $(\cdot 95+\cdot 05)^m$; the mean of this distribution is $\cdot 95m$ and the standard deviation $\sqrt{(19m/400)}=\cdot 05\sqrt{19m}$. Clearly, the assumption that the binomial approximates to the normal is open to criticism in this case. Whereas Test 2 and its suggested subsidiaries could not be criticized on these grounds, because in each case the chance of 'success' was not very different from the chance of 'failure', in this case the binomial distribution is too skew to make the assumption without further investigation.

31. Table 2 shows the probabilities of different numbers of failures when m is equal to 20, 40, 60 and 80, and compares these probabilities with the limits of acceptance on the basis of the binomial-normal assumption. An examination of this table will show that if we take the narrow instead of the broad view of the fractional portions of these limits of acceptance, so that the lower limit is taken to the higher integer and the upper limit to the lower integer, we will not accept any hypotheses which we would reject on the true binomial probabilities. For example, column (4) of the table shows that if $m=60$, on this basis we would accept any value of r between 0 and 6, and from column (3) we see that $\cdot 970$ of true hypotheses would thus be accepted; but if we do not accept $r=6$ we would be reducing our proportion accepted to $\cdot 921$.

32. On the other hand, taking the narrow limits would, in fact, reject a number of borderline cases, as an examination of the figures for $m=20$ and 80 will show. We may therefore devise the following rule for the application of Test 3: find the acceptance limits by means of the binomial-normal approximation, expressing the lower limit to the higher integer and the upper limit to the lower integer; if the observations are within these limits of acceptance the graduation passes the test; if the observations are near but just outside these limits, a more accurate calculation based on binomial probabilities must be made.

33. For borderline cases we might also apply a subsidiary test based on the number of deviations (if any) exceeding two and a half or three times the standard error.

34. The figures in column (7) of Table 6 exceed 2 in only two cases out of 49, and the graduation easily passes this test.

Table 2. Probability of r failures when chance of any observation showing a failure is ·05

Number of observations (i.e. values to be graduated) m (1)	Number of failures r (2)	Probability of r failures $m_r \cdot 95^{(m-r)} \cdot 05^r$ (3)	Limits of acceptance if binomial assumed to approximate to normal (4)
20	0	·358	0 to 2·95
	1	·377	
	2	·189	
	3	·060	
	Over 3	·016	
40	0	·128	0 to 4·76
	1	·270	
	2	·278	
	3	·185	
	4	·090	
	Over 4	·049	
60	0	·046	0 to 6·38
	1	·145	
	2	·226	
	3	·230	
	4	·172	
	5	·102	
	6	·049	
	Over 6	·030	
80	0	·017	·10 to 7·90
	1	·069	
	2	·144	
	3	·198	
	4	·200	
	5	·160	
	6	·105	
	7	·059	
	8	·028	
	Over 8	·020	

VI. ADHERENCE, GROUPS OF DEVIATIONS REGARDING SIGN

Test 4. Examination of consecutive runs of deviations of the same sign

35. The extraction of the figures necessary for the previous two tests will show up the signs of the observed deviations. A later test will deal with the question whether the signs change sufficiently frequently, but without anticipating that test it can quite clearly be stated that we would scarcely expect the deviations to be alternately positive and negative; we shall therefore be confronted with runs of two, three or more consecutive deviations of the same sign, *and it is necessary to determine whether the total deviation of* such a run is within the acceptable limits. The square of the standard error of a run of deviations is equal to the sum of the squares of the individual

standard errors, provided there is no correlation between the deviations, and we may therefore compare the total deviation of a group with its standard error in a similar way to the previous two tests.

36. Some comment is necessary on the proviso just made. Admittedly, a summation graduation would cause a systematic connexion between the smoothed-out errors at successive ages, and therefore a constraint on the size of the deviations. If this is considerable the appropriate tests based on the assumption of no such connexion should indicate under-graduation. It is also appreciated that correlation between successive deviations is quite likely when a graphic graduation is made, but this would depend on the skill of the graduator, and again suitable tests should disclose too close adherence to the data.

37. In the example, column (6) of Table 6 shows that there are three suspect groups where the total deviation exceeds twice its standard error, namely, the groups 34–37, 51–53 and 57–58; in two of these the excess is very small, but for age-group 34–37 the total deviation is about two and a half times its standard error.

38. If we repeat Test 3, but substitute these three groups in place of the nine individual ages constituting them, we are left with 43 samples or 'cells', instead of 49, in four of which the deviation exceeds twice its standard error; this is within acceptable limits.

39. If we then make the very severe test of treating every run of deviations of the same sign as a separate cell, so that 23–26, 27–28, 31–33, etc., would each form a cell, we find there are 23 cells, and three of the deviations exceed twice their standard error; this is also within acceptable limits. It appears, therefore, that the graduation would not be rejected on the consideration of suspect groups.

Test 5. Sum of deviations regarding sign should not be large

40. This is really only a special case of Test 4, the group examined being the whole of the data. In the past, the description of this test has usually been that the sum should be approximately equal to zero, but the modification made by the above heading is deliberate. It has been usual for actuaries to aim at a zero value for this sum in the case of a formula graduation and an approximately zero value in a graphic graduation. In the discussion on Seal's paper, Longley-Cook stated that 'in almost every graduation the difference between the total actual and expected deaths should be approximately zero, and not the amount expected on the assumption of random sampling'. Seal himself, however, was of the opinion that the equalization of the actual and expected deaths (and of their accumulated deviations) was quite an arbitrary procedure, and few will disagree with him. It is felt that, in the past, actuaries have been far too stringent over this test, and possibly a broader outlook would have avoided much difficulty in graduation.

41. On the basis of the proviso discussed in paragraph 36, the sum of the deviations between actual and expected events having regard to sign would be distributed with a mean of zero and a standard error of $\sqrt{\Sigma npq}$, and if we accept the 5% probability level as our yardstick, there is no reason why any graduation should be rejected on this test if the sum differs from zero by less than twice its standard error. In practice, it may be found that if we let it differ by anything approaching the borderline the graduation will fail several of the other tests, but this is not necessarily the case, and if this test is to have any value at all it must surely be made on the same statistical basis as the other tests.

42. One reason why the zero value has been adhered to so frequently in the past for formula graduations has been the employment of the method of moments and similar methods; a possible alternative to these methods is examined in the appendices to this paper.

43. In the example in Appendix 1, the sum of the deviations is $-29{\cdot}53$ against a standard error of $44{\cdot}8$, and is therefore well within the acceptable limits of $\pm 89{\cdot}6$.

Test 6. Examination of second sum of deviations regarding sign

44. It is doubtful whether this test has any great value, but it is included for the sake of completeness. Just as it has been customary to stipulate that the first sum should be approximately zero, it has been usual to say that the second sum must be 'small', without any definition of just how small it need be. It is probable that this test has also grown up with the method of moments, and that its too stringent application has tended to encourage under-graduation. If we permit the first sum its full scope on a statistical basis, it is consistent that the second sum should be treated likewise.

45. The test is best described by illustration from our example. The sum of column (9) of Table 6 is $-681{\cdot}64$. This is a combination of the 49 values shown in column (5), the first value being weighted 49 times, the second value 48 times, and so on. The standard deviation of this function is equal to

$$\sqrt{\{49^2 \times (npq)_{19\frac{1}{2}} + 48^2 \times (npq)_{20\frac{1}{2}} + \dots + 2^2 \times (npq)_{66\frac{1}{2}} + 1^2 \times (npq)_{69\frac{1}{2}}\}}$$
$$= \sqrt{1{,}638{,}501} = 1280.$$

The value is therefore within the limits of acceptance.

VII. ADHERENCE, GROUPS OF DEVIATIONS DISREGARDING SIGN

Test 7. The mean deviation test; sum of deviations disregarding sign should approximately equal $\cdot 8\Sigma\sqrt{npq}$

46. For any particular age or sample, the mean deviation disregarding sign is approximately $\cdot 8\sqrt{npq}$, since by the grouping outlined in paragraph 16 conditions have been produced under which the distribution of the deviations approximates to the normal. The sum of the deviations should therefore approximate to $\cdot 8\Sigma\sqrt{npq}$, and Perks has suggested (see *J.I.A.* LXXII, 199) that the standard error of this sum would be approximately $\cdot 6\sqrt{\Sigma npq}$.

47. This test is of limited value as it attempts to judge the whole graduation by a single-value function; it has been considered a useful summary test when merely regarded as supplementary to the other tests, but it has now been somewhat superseded by the χ^2 test.

48. Perks also suggested in the discussion on Daw's paper a possible modification of this test by standardizing the deviations; if each deviation is divided by $\sqrt{npq}$, the average of these standardized deviations may be $\cdot 8$. This expected average should be adjusted for constraints and an approximate adjustment would be to reduce it in the proportion $\sqrt{f}/\sqrt{m}$, where f is the number of degrees of freedom; this adjustment would not be worth while making if $(m-f)$ is small compared with m. The standard error of this expected average approximates to $\cdot 6/\sqrt{m}$.

49. In the example, the sum of column (5) disregarding sign is 298·5, compared with a mean value of $\cdot 8 \times 310\cdot 63 = 248\cdot 5$ and an approximate standard error of $\cdot 6 \times 44\cdot 8 = 26\cdot 9$. This is just within the acceptable limits, at the 5% level, of 194·7–302·3. For the modified test, the sum of the standardized deviations irrespective of sign (column (7)) is 46·04, giving an average of ·94 as compared with $\cdot 8 \pm \cdot 17$.

Test 8. χ^2 test

50. Much has been written on this test, and there seems no point in describing in detail the theory and method which have been so admirably expounded in Seal's paper. It will be sufficient to state that the function χ^2, which is equal to $\Sigma\{(\theta - nq)^2/npq\}$, is one whose distribution is known, and for which the probabilities of obtaining a value at least as great as any given figure are available.

51. The data required for this test are the observed value of χ^2 and the number of groups or samples observed, again on the basis of the grouping mentioned in paragraph 16; the only qualification to this is that the number of groups, which we have termed m, must be reduced by $(m-f)$, the number of constraints. The function f is known as the number of 'degrees of freedom', and a linear constraint may be imposed, for example, if it is stipulated in the graduation that the total number of actual deaths must exactly fit the total number of expected deaths, since after $(m-1)$ of the values have been fixed there is no freedom of choice left in the fixing of the mth value; similarly, if it is also stipulated that the second sum of the deviations must be zero, a second constraint would be imposed. If a 48th difference curve were fitted to a series of 49 values (see paragraph (20)), there would be no degrees of freedom, since the original series would be reproduced and there would be no scope for any value of χ^2; the first 49 summations would all have to agree, and there would therefore be 49 constraints imposed.

52. In the example in Appendix 1, so far as the author is aware, no constraints have been imposed, and with 49 degrees of freedom the chance of obtaining a value of χ^2 equal to or greater than the observed value of 59·80 is, by interpolation in the table on p. 45 of Seal's paper, about 21%. Since, at the 5% probability level, we would accept any value of this probability between $2\frac{1}{2}$% and $97\frac{1}{2}$%, the graduation passes this single-value test.

53. A very valuable use of this test is to estimate whether a small experience can be regarded as having been drawn at random from the universe giving rise to the rates of a certain standard table.

54. A frequent disadvantage is that it is not always certain what deduction, if any, should be made from the number of 'cells' to obtain the number of degrees of freedom. For example, when a graphic graduation is made, probably a partial constraint is imposed in trying to make the actual deaths approximately equal to the expected; possibly a fractional constraint may be imposed when the graph is hand-polished to make the graduation satisfy more fully one of the other tests, although if so, it is impossible to tell the extent of this constraint. Similarly, the number of constraints is doubtful when a summation graduation is made; the different opinions expressed in the discussion on Seal's paper with regard to the number of constraints for the Kenchington formula are evidence of the difficulties. For this reason alone it would be dangerous to follow Seal and hitch our wagon to the χ^2 test to the exclusion of all other adherence tests.

55. No one test by itself can be regarded as conclusive, and this applies as well to the χ^2 as to any other test. In the discussion on Seal's paper, Haycocks described this test not as a complete test but 'a piece of evidence, the importance of which varied considerably from problem to problem'. In the discussion on Daw's paper, Prof. M. G. Kendall expressed the opinion that the four tests mentioned by Daw (which amount to a very condensed summary of a number of the tests described in this paper) were 'at best a very poor substitute for the χ^2 test'; by the same token, the χ^2 test is not a complete substitute for the other tests. In fact, it should be regarded as one method of testing a graduation; a valuable one, but not to the exclusion of all others; perhaps the best summary test, but by no means the be-all and end-all of graduation testing.

56. The limitation of this test through being based on a single value is dealt with in paragraphs 60–64, introducing Test 9.

57. It has been stated that the test would be passed if the probability found by entering the table in Seal's paper were between $2\frac{1}{2}\%$ and $97\frac{1}{2}\%$. Clearly, *if the probability of the observed* χ^2, *or of a greater value*, were less than $2\frac{1}{2}\%$, then at the 5% level the graduation would be rejected; but it is not so clear what should be the interpretation if the probability is greater than $97\frac{1}{2}\%$, i.e. if the result of this test is 'too good to be true'. This brings us to a difference in outlook between the actuary and the statistician, due to a difference in purpose rather than to any real underlying difference. If a statistician obtained a 'too good to be true' result, it might, for example, mean that there was something wrong with his method of sampling or that his instructions had not been carried out correctly. Similarly, an actuary would place a definite interpretation on a 'too good to be true' result if the graduation had been by a graphic method, the conclusion being that the graduation had been made to adhere too closely to the original data; but in the case of a formula graduation such a result can have no meaning whatever, except perhaps that a formula has been chosen with too many parameters. It seems that a graduation would not be rejected merely because of a 'too good to be true' value of χ^2, provided it passed all the other tests satisfactorily, and provided the graduated rates were ideally smooth (e.g. a Makeham graduation satisfying all the other tests); in other words, it would not be rejected wholly on account of being too probable to be probable.

58. The foregoing paragraph might be summarized by saying that whereas a 'too good to be true' result of this test may indicate too close an adherence in the case of a graphic graduation, or the choice of too complicated a formula in the case of a formula graduation, it should not be taken as a ground for the rejection of a graduation by a simple formula.

Test 9. χ^2 sectional test

59. This test is exactly similar to Test 8, but is applied only to suspect sections which may be detected under Test 2 by taking out runs of consecutive deviations exceeding their probable error.

60. Before illustrating this test it will be as well to consider the necessity for it. It has already been pointed out that the χ^2 test measures the adherence of the whole graduation by means of a single value, incurring the danger that over-graduation in one part of the table may be balanced by under-graduation in another part. In the discussion on Daw's paper, Perks compares the criteria of both Tests 7 and 8 with such single-value functions as the

expectation of life and the ratio of total actual deaths to total expected deaths; the net reproduction rate, the crude death-rate and the cost-of-living index might well be added to this 'rogues gallery'. As an extreme example of the dangers, a graduation could be envisaged in which all the ages up to 50 show deviations less than their probable errors, while at all the older ages they are greater; such a graduation could conceivably satisfy the χ^2 test without giving any grounds for suspicion. Admittedly, it would also satisfy Test 2 if applied blindly, but even a novice applying Test 2 should realize that the younger portion of the table had been under-graduated. A similar example would be a graduation in which all the negative deviations were less than their probable errors and all the positive deviations greater.

61. To demonstrate these dangers of the χ^2 test, we have no need to look further than the figures in Seal's own paper; in fairness, it should be pointed out that Seal emphasized that it was not a test that fools could use—'a certain amount of *le bon sens* was needed in applying any test'—nevertheless, it is suggested that one single probability value can give no more 'firm judgment concerning the success of a graduation' than the figures '29 below, 21 above' which Seal criticizes so heartily.

62. On pp. 12 and 13 of his paper are shown figures taken from Kenchington's graduation of the O^{JF} table. The value of χ^2 is 41·425, with an estimated number of degrees of freedom of 44, and the table of probabilities, according to Seal, 'indicates that the graduation is excellent'. If the figures in column (6) of the table are compared with their probable errors it will be found that 29 are greater ('G') and 21 less ('L'); but if we examine the order of the 'G's' and 'L's' it will be found that there is a run of eleven successive 'G's' from ages 37 to 47, suggesting that this section may have been over-graduated. If we assume that the six constraints estimated by Seal to be imposed by the formula are equally spread over the whole range of ages, it will be found that the value of χ^2 over this eleven-age section is 13·804 with f equal to 9·7, not an unsatisfactory result; however, if we examine the narrower range from ages 39 to 43 we find a χ^2 of 9·689 with f equal to 4·4, and this is rather near the limits of 'doubtful improbability' (Seal took 5% and 95% as his critical limits). Certainly the graduation would not be rejected on this account, but this sectional examination indicates that 'excellent' is scarcely the right word, and that perhaps the Kenchington formula was too powerful a wave-cutter.

63. Now, since the summary χ^2 test gave such an excellent result, it seems likely that the tendency to over-graduation in the early forties might have been counterbalanced somewhere by an under-graduation. It will be found that 10 of the last 15 values in column (6) are 'L's', and further examination will show that the value of χ^2 for the range of seven ages from 65 to 71 is only 1·863, the probability table on p. 45 telling us that this is 'rather too probable'. Again, it is not suggested that the graduation should be rejected, but merely that the χ^2 test applied to the whole range does not necessarily give the whole story, and that the graduator must collect together the other pieces of evidence before he can be satisfied that he has arrived at the best possible graduation for his purpose.

64. Similar weaknesses can be found in the figures in Table 10 on p. 35 of Seal's paper. Test 3 would show that, out of the thirty deviations, four are greater than twice their standard error, and this is very much a borderline case for acceptance; a sectional test applied to the range of twelve ages from 70 to 81 shows a value of χ^2 of 21·632, indicating 'doubtful improbability', although

Seal states that the test as a whole gives a good fit. As the figures tested were those of a graphic graduation it seems not unlikely that a better fit might have been obtained by a little further hand-polishing.

65. In the example in Appendix 1, there is a predominance of 'G's', the worst run being from ages 45 to 58 inclusive; the value of χ^2 for these 14 values is 21·85, which is not unsatisfactory. Over the smaller range from 52 to 57 the value of 13·31 indicates a possible weak feature in the graduation; again, it is suggested that this is an indication of weakness rather than a positive ground for rejection, but the lesson to be learnt is that should this test demonstrate several such weaknesses it is likely that a more satisfactory graduation could be found.

VIII. SIGN AND SIGN-CHANGE TESTS

Test 10. *Sign-change test for deviations*

66. In an ideal graduation the deviations should change sign fairly frequently. Makeham suggested a method making use of the expected distribution of the number of runs of so many of the same sign. A far simpler, and just as satisfactory, test is that suggested by Haycocks in the discussion on Seal's paper; this is based on the assumption that a sign-change is as likely as a non-change, giving rise to a binomial distribution exactly similar to that for Test 2, the number of observations being $(m-1)$, and the acceptance limits being those shown in Table 1 for this number of observations.

67. One difficulty in this test is what to do with 'zeros', where the deviation is neither positive nor negative; one solution would be to treat a change to or from zero (as opposed to a change through zero) as half a sign-change, but probably the simplest way would be to ignore all zeros and to reduce $(m-1)$ by the number of zeros before applying the test. Alternatively, they can usually be eliminated by calculating the relevant items to more decimal places.

68. It is felt that a graduation would not be rejected merely because there were too many sign-changes, though such a result might indicate a bias on the part of the graduator; for example, if a graphic graduation had been made following strictly the principle 'one up, one down' over the whole range of ages, there would be no non-changes and the graduator would have chosen a wavy curve running in and out of the spaces between the points representing the ungraduated values. It is difficult to see, however, what bias could be indicated by too many sign-changes in a formula graduation, but it seems that such a situation is most unlikely to arise in practice.

69. The application of the test can be summarized by the statement that the graduation fails the test if the number of sign-changes is less than the lower limit, while if it is greater than the upper limit it is desirable to examine for possible bias but not necessarily to reject the graduation.

70. Column (5) of Table 6 shows 22 sign-changes out of a possible 48, and the graduation therefore satisfies this test.

Test 11. *Sign test for deviations*

71. In an observed sample of n years of exposure, the expected number of events would be nq and we should expect to have an equal chance of the actual number being above or below this figure; then over a series of m ages or groups

the number of positives would be distributed with a mean of $m/2$ and a standard error of $\cdot 5\sqrt{m}$, in other words, a distribution exactly similar to that in Test 2 with acceptance limits as given in Table 1.

72. Column (5) of Table 6 has 26 positives and 23 negatives, well within the acceptance limits of 17–32.

Test 12. Sign-change test for accumulated deviations

73. This test has usually been carried out on exactly similar lines to Test 10, but certain modifications are desirable. In the first place, it has been stated under Test 5 that the sum of the deviations regarding sign need not necessarily be zero. Clearly, it is possible for this sum to be well within twice its standard error but still to be considerably in excess of the acceptable limits of the individual deviations; for example, in Table 6, the total of column (5) is about -30, and we would have accepted any value up to about ± 90, whereas the individual standard errors of the last few deviations have values round about 4 or 5. We should therefore expect a longish run of accumulated deviations of the same sign at the end of our column of summations, reaching their climax, in the example, with a value of nearly -30.

74. It is suggested that this test would therefore be better if the accumulated deviations, instead of being measured from zero in each case, were measured from successive points on a 'balance line', the first point on the balance line being (total deviations regarding sign)$/m$, the second point $2\times$(total deviations regarding sign)$/m$, and so on, so that the adjusted series of the accumulated deviations is bound to end with a final value of zero, thus removing the bias in favour of non-changes at the end of the series.

75. In Table 6, the balance line is given in column (11), and the adjusted series in column (13); the run of positives at the end of column (13) is not now due to any bias inherent in the test but to the fact that the last value in column (5) happens to be large.

76. It will also be observed that the device of the balance line has ensured that the same adjusted series is obtained, whether the summation is made from the top or the bottom of the series. This is demonstrated in columns (10), (12) and (13).

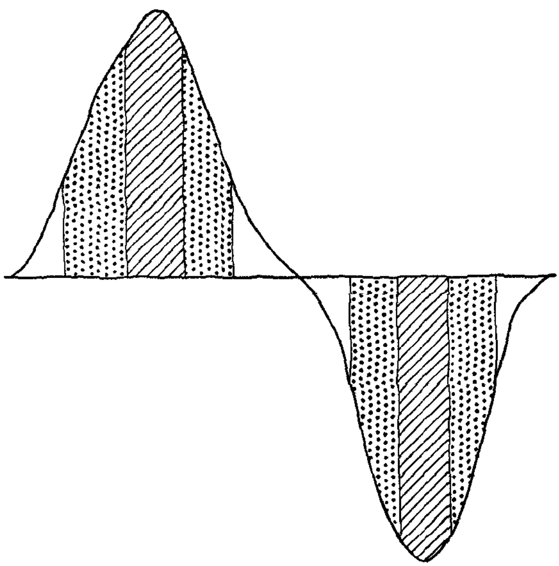
77. A far more serious defect of the test than the bias in favour of non-changes at the end of the series and the fact that we can carry out the summation in two ways is the fact that the expected number of sign-changes represents a difficult problem. The probability of a sign-change is certainly not one-half.

78. The point is best demonstrated by regarding the graph of the accumulated deviations as a series of cycles; the cycles will be of varying heights and depths, but in view of the fact that an accumulated deviation has a much higher variance than a deviation (see Test 6) we should expect the height or depth of the longest cycles to be considerably greater than any individual deviation.

79. At the top or bottom of a cycle a non-change is almost impossible; where the cycle passes through zero it would appear that a change or a non-change is equally likely; about half-way between the maximum or minimum and zero there will be an intermediate chance, perhaps a one in four chance of a change. If these three 'regions' cover an approximately equal number of values, then the average chance of a non-change is $\cdot 25$. It seems more likely that the cycles would rise and fall steeply in the regions of the maxima and minima, and if it is assumed that the 'top-and-bottom-

region' of each cycle only covers half the number covered by each of the other regions, the average chance of a change is $(1 \times 0 + 2 \times \cdot 25 + 2 \times \cdot 5)/5 = \cdot 3$. Opinions may differ as to the average shape of a cycle, but it seems reasonable to state that the areas around the maxima and minima will always be sufficiently considerable to bring the chance of a change down to about one-third or one-quarter.

80. The considerations of paragraphs 78 and 79 would appear to be applicable to the adjusted series of accumulated deviations using the balance line, rather than the unadjusted series. The only difference is that the centre of the cycle would be the point where it crosses the balance line instead of the zero line, but the unadjusted series is further complicated by the bias at the end of the series with the result that the 'no chance of a change' area is grossly enlarged.



Possible average trend of accumulated deviation cycle.

Chance of sign-change approximately ·5. Chance approximately ·25.

No chance of sign-change.

81. The diagram illustrates the conception of the three regions of the cycle, and Table 3 gives the acceptance limits, at the 5% probability level, on the assumption that the average chance of a sign-change is ·3.

82. There seems no point in hazarding a guess at the average chance of a sign-change in the unadjusted series, caused by the assumed three-tenths chance reduced by the bias at the end of the run (for what it is worth, columns (9) and (10) of Table 6 each show 6 changes out of a possible 48). The modified test suggested in paragraph 79 should be based on column (13), which shows 9 changes out of a possible 47, the acceptable frequencies being 7 to 21.

83. Even on the adjusted basis, this is probably the least valuable of all the tests; it should merely be regarded as a possible means of indicating distortion. In our example, although the number of changes in column (13) passes the test, it may be considered that there is some lack of balance in the graduation inasmuch as the early and late values in column (13) tend to be on the positive side of the balance line, while the middle values tend to be on the negative side.

84. In this test, too, a graduation could scarcely be rejected on the grounds that the accumulated deviations change sign too often, but such a case is most unlikely to be met in practice.

85. Because of the doubtful value of this test, and the fact that it can do no more than give an indication of distortion, it is felt that it is not worth while to complicate it by weighting the successive points on the balance line proportionately to the exposed to risk.

Table 3. Acceptance limits of number of sign-changes for Test 12 on the assumption of an average chance of ·3

Number of graduated values m	Possible number of sign-changes in adjusted series of accumulated deviations $m-2$	Lower limit $\cdot 3\,(m-2) - 2\sqrt{\{\cdot 21\,(m-2)\}}$	Upper limit $\cdot 3\,(m-2) + 2\sqrt{\{\cdot 21\,(m-2)\}}$
30	28	3	14
35	33	4	16
40	38	5	18
45	43	6	19
50	48	8	21
55	53	9	23
60	58	10	25
65	63	11	27
70	68	12	28
75	73	14	30
80	78	15	32

Test 13. *Sign test for accumulated deviations*

86. This would be on similar lines to Test 11, but, following a similar reasoning to that developed in the comments on Test 12, we should expect the run of non-changes at the end of the series to reduce the chance of the sign distribution being half and half; the mean proportion of positives would still be one-half over a large number of graduations, but the standard error would probably be incapable of evaluation. In order to use a series to which can be applied a normal test, this test should be applied to the adjusted series already produced for Test 12. The series in column (13) of Table 6 has 19 negatives and 29 positives, the acceptance limits along the lines of Test 2 being 17 to 31.

IX. CONSTRAINTS

87. In the discussion on Daw's paper, one of Prof. Kendall's complaints against the usual tests applied by actuaries was that they make no allowance for any constraints which may be imposed. Each of the tests outlined, with

the exception of the smoothness and χ^2 tests, will now be considered in this connexion.

88. It has been indicated in paragraph 48 that the expected average size of the standardized deviations is reduced in the proportion $\sqrt{f}/\sqrt{m}$. The constraints on the data have a similar effect on Tests 2 and 3, where we are testing the number of standardized deviations which exceed certain values (usually ·6745 and 2·0), and these values should also be reduced in the proportion $\sqrt{f}/\sqrt{m}$.

89. In Test 4, if any systematic limitation has occurred in a graphic graduation we should nevertheless judge our acceptance as though no such constraint had occurred or we may be condoning under-graduation. It is, however, certain that constraints will have been imposed on combinations of groups by a summation graduation (see under Test 9). Also if constraints have been imposed in the course of a formula graduation there will be some effect on the grouped deviations.

90. The data for Tests 5 and 6 are frequently constrained by the requirement that the sums must be zero, in which case the tests would not be made; it is difficult to see any other systematic constraint which might affect these tests.

91. The constraints for Tests 7 and 8 have been dealt with under their respective sections, but it might be repeated that the number of constraints can only be estimated approximately, if at all, in the cases of graphic and summation graduations. Similar considerations apply to Test 9; there will be partial constraints on the groups of deviations being tested if the graduation is by formula and constraints have been imposed.

92. No allowance for constraints is likely to be necessary for Tests 10 and 11 unless graduated rates have been produced which tend systematically to be generally above (or generally below) the observed rates, but such a table would be for a special purpose and would be a forecast table rather than a graduation.

93. For Tests 12 and 13 a device has been suggested which allows for the artificial restriction on the figures.

X. SUMMARY OF TESTS

94. When the thirteen tests have been completed, an assessment of the graduation may be made by a general consideration of whether, on the whole, the tests have been satisfactory. Superstitious actuaries might like to regard this general appreciation as the fourteenth test. In the example we have been considering, the only unsatisfactory features are (*a*) that it is a borderline case for Test 2, (*b*) that Test 9 indicates a weakness over ages 52–57, and (*c*) that Test 12, though the graduation passes the test, possibly indicates some lack of balance in the deviations. On the whole, therefore, it is felt that the graduation is acceptable.

95. Table 4 summarizes the tests which have been described and shows when each would be applicable.

XI. DEGREE OF ACCURACY

96. It has been fashionable to express the expected deaths to one or even two decimal places, but little consideration has been given in the past to whether the number of places could be justified. Admittedly, it is a small point and probably makes very little difference to the final results of a graduation, but i

Table 4. Summary of tests

Test no.	Description	Type of graduation: Graphic	Summation	Formula	Reference to standard table
1	Smoothness (subject to the remarks of paragraph 24)	Yes	Yes	Not necessary if simple formula, except as independent check on arithmetic	Yes
2	Individual deviations (probable error)	Yes	Yes	Yes	Yes
3	Individual deviations (twice standard error)	Yes	Yes	Yes	Yes
4	Suspect groups of deviations	Yes	Yes	Yes	Yes
5	Sum of deviations *regarding* sign	Yes	Yes	Yes, but only as arithmetical check if method of moments used	Yes
6	Sum of accumulated deviations regarding sign	Yes	Yes	do.	Yes
7	Mean deviation	Yes, provided number of constraints can be estimated	Yes, provided number of constraints can be estimated	Yes	Yes, provided number of constraints can be estimated
8	χ^2	do.	do.	Yes (*Note*: will automatically have been applied if minimum method used)	do.
9	χ^2 sectional	Yes	Yes	Yes	Yes
10	*Sign-changes* of deviations	Yes	Yes	Yes	Yes
11	Signs of deviations	Yes	Yes	Yes	Yes
12	Sign-changes of adjusted accumulated deviations	Yes	Yes	Yes	Yes
13	Signs of adjusted accumulated deviations	Yes	Yes	Yes	Yes

is felt that this is a matter which should not be overlooked, since the profession should not lay itself open to criticism from without as a stickler for precision which it cannot attain.

97. We might consider this question with regard to the example in Appendix 1. For age $51\frac{1}{2}$, the exposed to risk is about 10,000, and the graduated rate of mortality about ·005. It will therefore be seen that the first decimal place can be justified only if the exposed to risk is accurate within 10 (accuracy 1 in 1000), and the second place can be justified only if the exposed is accurate to the nearer unit (1 in 10,000). Although it is beyond the scope of this paper to consider the question of relative accuracies achieved by the census method, it seems quite possible that the approximations made would not justify even the first decimal place in the expected deaths; this means that the second significant figure of the deviations becomes suspect, and brings us to the question of how accurately we can claim to find the value of χ^2, and how many places we can take credit for in the estimate of the Makeham constants.

98. The line taken in the experiments described in the appendices has been that the problem is to find the best possible fit to the observations as given, and that, especially in the case of the Makeham constant c, it is necessary to retain a considerable number of figures. This section is, however, included as a matter worthy of consideration, in the hope that a certain amount of unnecessary work and superfluous published figures may be avoided in the future.

XII. ACTUARIES AND STATISTICIANS

99. It has been suggested in paragraph 57 that a possible difference in outlook between actuaries and statisticians is the fact that a statistician can always assign a meaning to a 'too good to be true' result of a test, while such a result may not always have any meaning to an actuary.

100. Perhaps, also, a statistician is more interested in historical facts than an actuary, while the latter, even if he does not in fact construct a forecast table, is always looking to the future; this is probably only a spurious difference between the two, since the majority of investigations made both by actuaries and by statisticians would be with the purpose of estimating something about the future.

101. A further difference should be mentioned, which was referred to by Perks in the discussion on Daw's paper; this is that statisticians usually deal with a comparatively small number of groups, whereas actuaries tend to deal with more groups and are thus more often able to make use of the normal distribution.

102. This section may appear to be irrelevant to a paper purporting to deal with graduation tests. Unfortunately there has, of recent years, been a gulf between statisticians and actuaries, which can only have been widened by Prof. Kendall's remarks on Daw's paper. The present paper seems to give a suitable opportunity for airing the question and for showing that the differences between the two do not really go very deep; it is hoped, therefore, that this section and any discussion which may take place on it will help to bridge the gulf.

XIII. CLOSING REMARKS AND ACKNOWLEDGMENTS

103. No originality is claimed for much of the subject-matter of this paper, but no apologies are offered for the elementary terms in which the arguments are couched.

104. Acknowledgements are due to the papers by Seal and Daw, and to those taking part in the discussions on these papers. A special acknowledgment should also be made to Mr R. G. Barley, F.I.A., F.S.S., without whose suggestions, criticisms, helpful comments, and encouragement this paper could never have materialized, and to Mr J. G. Day, M.A., F.I.A., F.S.S., who carried out the unenviable task of checking the algebra of Appendix 2.

APPENDIX 1

Regraduation of A 1924–29 *select data (duration* 0, *ages* $19\frac{1}{2}$ *and over) by Makeham's formula*

105. Paragraph 12 refers to a graduated series convenient for use in practice; when we consider a mortality table convenient for use in practice our thoughts immediately turn to Makeham's formula which, apart from the convenience resulting from mathematical relationships between certain functions, has a certain appeal to the actuary with a tidy mind. More complicated formulae, such as those suggested by Perks, have the same appeal but not the convenience. The same paragraph sets out the general rule that a graduation should aim at making the simplest possible hypothesis of the universe consistent with the data, and the logical outcome of this is a preference for curve-fitting methods of graduation; a graphic graduation is really only a short-cut attempt to fit a series of curves to the data, and it is felt that a summation graduation is a somewhat makeshift method which, while admittedly giving a quick answer, tends to retain certain irregularities which would be better ironed out. A combination of these considerations leads to the conclusion that wherever possible the graduation of mortality data should be by the fitting of a Makeham formula, in preference to other methods of graduation or more complicated formulae; further, the 'simplest possible hypothesis' rule leads to the preference for, say, a Makeham curve just within the limits of acceptance rather than a Perks curve well within the limits.

106. It was recalled, in dealing with Test 5, that in the past it has been usual for actuaries to employ the method of moments for curve-fitting graduations, giving rise to the custom of equating the actual and expected deaths. If the sum of the deviations regarding sign is permitted any acceptable value within the limits of statistical significance, the question arises as to how some limit might be placed on the number of curves which can be fitted to certain data. A possible answer is the employment of a minimum-χ^2 method, and provided a suitable type of curve can be selected this seems to be the ideal solution. For example, if a Makeham graduation is being attempted, quite possibly a certain value of c combined with the fixed values of A and B resulting from the method of moments would not be satisfactory; all three constants could probably be improved by a method of hand-polishing, similar to the method used in a graphic graduation, but there would be no feeling of satisfaction that the graduation so obtained was the best possible in any one respect. If, however, the value of c is combined with those values of A and B giving the minimum value of χ^2 it is quite likely to give a better graduation than is obtainable by the method of moments, and this process can be carried further by also improving c until the Makeham curve is found which, when applied to the data in question, gives a lower value of χ^2 than any other Makeham curve.

Similarly, if a Makeham curve is found not to be suitable, it would be possible to use a Perks minimum-χ^2 method.

107. *I have never been convinced that all the results of the A 1924–29* experience proved that Makeham's formula no longer represents the approximate shape of the curve of mortality, and this seemed to be suitable data on which to experiment with the minimum-χ^2 method. The duration 0 data was chosen since it is practically free from the complication of duplicates, on account of the practice of excluding duplicate policies effected concurrently

Table 5. Values of $\text{colog}_e\, p_{[x-\frac{1}{2}]}$ and $q_{[x-\frac{1}{2}]}$; A 1924–29 regraduated by Makeham's formula

Age $[x]$	$\text{colog}_e\, p_{[x-\frac{1}{2}]}$	$q_{[x-\frac{1}{2}]}$	Age $[x]$	$\text{colog}_e\, p_{[x-\frac{1}{2}]}$	$q_{[x-\frac{1}{2}]}$
20	·00125849	·00125770	45	·00307375	·00306903
21	·00127572	·00127490	46	·00328080	·00327542
22	·00129475	·00129391	47	·00350950	·00350335
23	·00131577	·00131490	48	·00376212	·00375505
24	·00133898	·00133809	49	·00404115	·00403300
25	·00136463	·00136370	50	·00434936	·00433992
26	·00139296	·00139199	51	·00468980	·00467882
27	·00142425	·00142323	52	·00506585	·00505304
28	·00145881	·00145774	53	·00548121	·00546622
29	·00149698	·00149586	54	·00594001	·00592241
30	·00153915	·00153797	55	·00644679	·00642605
31	·00158573	·00158447	56	·00700656	·00698207
32	·00163718	·00163584	57	·00762487	·00759587
33	·00169401	·00169257	58	·00830783	·00827342
34	·00175678	·00175523	59	·00906221	·00902127
35	·00182611	·00182444	60	·00989548	·00984668
36	·00190269	·00190089	61	·01081588	·01075760
37	·00198729	·00198531	62	·01183253	·01176280
38	·00208073	·00207856	63	·01295549	·01287193
39	·00218394	·00218155	64	·01419588	·01409559
40	·00229794	·00229530	65	·01556597	·01544545
41	·00242386	·00242093	66	·01707934	·01693431
42	·00256296	·00255968	67	·01875096	·01857626
43	·00271659	·00271291	70	·02488966	·02458247
44	·00288630	·00288214			

in the same office; it must be comparatively rare for an assurer to complete two or more policies at different times in the same year, or through different offices in the same year, and the small number of duplicates resulting is unlikely to have any great effect on the validity of the usual assumptions.

108. The Makeham fit giving approximately the smallest possible value of χ^2 is

$$\text{colog}_e\, p_{[x-\frac{1}{2}]} = \text{A} + \text{B}c^x,$$

where $10^3\,\text{A} = 1{\cdot}0937349$, $10^5\,\text{B} = 2{\cdot}2540558$, and $c = 1{\cdot}1045704$. The graduated values of $\text{colog}_e\, p_{[x-\frac{1}{2}]}$ and $q_{[x-\frac{1}{2}]}$ are given in Table 5.

109. The tests of this graduation are shown in Table 6, which for convenience is divided into two sections. The required figures for Tests 2–11 inclusive are shown in the first section, and the second section deals with the

accumulated deviation tests 12 and 13. The conclusion of these tests is that a Makeham graduation is quite suitable for the data in question, and it seems not unlikely that if only we could remove the duplicates from the ultimate data a considerable part of the resulting ungraduated rates could be graduated by a similar curve.

APPENDIX 2

Makeham true minimum-χ^2 method

110. Hitherto, so far as I am aware, no method of fitting a Makeham curve to give the absolute minimum χ^2 has yet been devised. The method of Cramér and Wold (1) has been referred to from time to time in the *Journal*, but that method does not, in fact, give the minimum possible value. The imperfections of their method may be summarized as follows.

(*a*) It assumes that the expected deaths are equal to $E^c_x\mu_x$, and that p_x may be taken as unity, so that χ^2 becomes

$$\Sigma\,\{(\theta_x - E^c_x\mu_x)^2/E^c_x\mu_x\}.$$

(*b*) This expression is, however, discarded in favour of the function χ_1^2, where $\chi_1^2 = \Sigma\,\{(\theta_x - E^c_x\mu_x)^2/\theta_x\}$, and it is the fit giving a minimum value to this function which the method obtains.

(*c*) The method finds the fit giving the minimum χ_1^2 consistent with the first and second summations of $(\theta_x - E^c_x\mu_x)$ equalling zero. It may be noted that it would be quite easy to make an absolute minimum χ_1^2 fit, although Cramér and Wold express the opinion that it is too complicated for practical work; it is worth mentioning that, in one of the experiments I made, such a fit gave a higher value of χ^2 than the first trial Makeham curve which initiated the experiment!

(*d*) Cramér and Wold show that, for certain Makeham curves obtained, χ^2 is not very different from χ_1^2 (out of six examples they give, χ^2 is the smaller in five cases and the greater in one); this is, however, no conclusive proof that the values of the Makeham constants giving the minimum χ_1^2 will also give the minimum χ^2.

111. I have endeavoured to derive a method of reducing χ^2 by trial and error, in which each successive trial improves all three Makeham constants, and which obtains the values giving an absolute minimum χ^2 without too large a number of trials.

112. Such a method necessitates the finding of the first three differential coefficients of χ^2 with regard to the various possible combinations of Makeham constants; the algebraical derivation of these coefficients is as follows:

$$\text{colog}_e\, p_{x-\frac{1}{2}} = A + Bc^x,$$

where A and c are the same constants as for μ_x, but $B = -(c^{\frac{1}{2}} - c^{-\frac{1}{2}})\log_e g$ and differs slightly from the corresponding constant for μ_x, which is $-\log_e c \log_e g$. When B has been found for $\text{colog}_e\, p_{x-\frac{1}{2}}$ the corresponding constant for μ_x can easily be calculated.

Now, if $\text{colog}_e\, p_{x-\frac{1}{2}} = A + Bc^x$, then

$$-(1/p_{x-\frac{1}{2}})\,\partial p_{x-\frac{1}{2}}/\partial A = 1,$$

i.e. $$\partial q_{x-\frac{1}{2}}/\partial A = p_{x-\frac{1}{2}} = 1 - q_{x-\frac{1}{2}}. \qquad (1)$$

Similarly $$\partial q_{x-\frac{1}{2}}/\partial B = c^x(1 - q_{x-\frac{1}{2}}), \qquad (2)$$

$$\partial q_{x-\frac{1}{2}}/\partial c = Bxc^x(1 - q_{x-\frac{1}{2}})/c. \qquad (3)$$

Table 6. Tests of the graduation

Age $[x]$	$\theta_{[x-\frac{1}{2}]}$ (1)	$(nq)_{[x-\frac{1}{2}]}$ (2)	$(npq)_{[x-\frac{1}{2}]}$ (3)	$\sigma_{\theta_{[x-\frac{1}{2}]}}$ $\sqrt{\text{col. (3)}}$ (4)	$(\theta-nq)_{[x-\frac{1}{2}]}$ (5) +	(5) −	Suspect groups (6)	Standardized deviation (5)÷(4) (7)	Square of col. (7) (8)
20	24	24·63	24·60	4·96		·63		·13	·02
21	39	32·51	32·47	5·70	6·49			1·14	1·30
22	31	39·03	38·98	6·24		8·03		1·29	1·65
23	39	38·96	38·91	6·24	·04			·01	·00
24	42	40·44	40·38	6·35	1·56		(16·00;	·25	·06
25	49	42·66	42·60	6·53	6·34		σ_θ=12·8)	·97	·94
26	51	42·94	42·88	6·55	8·06			1·23	1·52
27	39	41·09	41·03	6·41		2·09		·33	·11
28	30	40·77	40·71	6·38		10·77		1·69	2·85
29	47	39·77	39·71	6·30	7·23			1·15	1·32
30	32	40·89	40·82	6·39		8·89		1·39	1·94
31	40	39·89	39·83	6·31	·11			·02	·00
32	48	38·75	38·68	6·22	9·25		(12·66;	1·49	2·21
33	42	38·70	38·63	6·22	3·30		σ_θ=11·8)	·53	·28
34	28	38·82	38·75	6·22		10·82		1·74	3·02
35	32	40·72	40·65	6·38		8·72	(32·93;	1·37	1·87
36	31	41·18	41·10	6·41		10·18	σ_θ=12·7)	1·59	2·52
37	37	40·21	40·13	6·33		3·21		·51	·26
38	41	40·84	40·75	6·38	·16			·03	·00
39	46	41·41	41·32	6·43	4·59		(14·74;	·71	·51
40	47	46·50	46·40	6·81	·50		σ_θ=13·2)	·07	·01
41	56	46·51	46·40	6·81	9·49			1·39	1·94
42	30	44·35	44·24	6·65		14·35	(21·59;	2·16	4·65
43	41	45·34	45·21	6·72		4·34	σ_θ=11·6)	·65	·42
44	43	45·90	45·77	6·77		2·90		·43	·18
45	55	48·73	48·58	6·97	6·27			·90	·81
46	36	47·46	47·31	6·88		11·46		1·67	2·78
47	51	45·24	45·09	6·71	5·76			·86	·74
48	49	44·32	44·15	6·64	4·68			·70	·50
49	39	48·19	48·00	6·93		9·19		1·33	1·76
50	53	58·68	58·43	7·64		5·68		·74	·55
51	69	62·40	62·10	7·88	6·60			·84	·70
52	69	54·84	54·56	7·39	14·16		(27·37;	1·92	3·67
53	60	53·39	53·10	7·29	6·61		σ_θ=13·0)	·91	·82
54	44	54·23	53·91	7·34		10·23	(24·46;	1·39	1·94
55	51	59·55	59·17	7·69		8·55	σ_θ=12·9)	1·11	1·24
56	47	52·68	52·31	7·23		5·68		·79	·62
57	61	45·90	45·55	6·75	15·10		(20·56;	2·24	5·01
58	48	42·54	42·18	6·49	5·46		σ_θ=9·4)	·84	·71
59	40	42·78	42·39	6·51		2·78		·43	·18
60	45	52·75	52·23	7·23		7·75		1·07	1·15
61	36	34·12	33·75	5·81	1·88			·32	·10
62	32	25·68	25·37	5·04	6·32			1·25	1·57
63	19	24·03	23·72	4·87		5·03		1·03	1·07
64	20	22·61	22·29	4·72		2·61		·55	·31
65	26	24·66	24·28	4·93	1·34			·27	·07
66	21	17·93	17·63	4·20	3·07			·73	·53
67	11	10·87	10·67	3·27	·13			·04	·00
68 and over	21	31·14	30·37	5·51		10·14		1·84	3·39
Totals	1988	2017·53	2008·09*	310·63	+134·50	−164·03 = −29·53		46·04	59·80

* The sum of the series $1^2 \times 30{\cdot}37 + 2^2 \times 10{\cdot}67 + \ldots + 49^2 \times 24{\cdot}60 = 1638501$.

Table 6 (cont.)

Age [x]	Σ col. (5) from top (9)		Σ col. (5) from bottom (10)		Balance line (11)	Inverted balance line (12)	Adjusted series (9)−(11) or (12)−(10) (13)	
	+	−	+	−			+	−
				29·53		−29·53		
20		·63		28·90	− ·60	−28·93		·03
21	5·86			35·39	− 1·21	−28·32	7·07	
22		2·17		27·36	− 1·81	−27·72		·36
23		2·13		27·40	− 2·41	−27·12	·28	
24		·57		28·96	− 3·01	−26·52	2·44	
25	5·77			35·30	− 3·62	−25·91	9·39	
26	13·83			43·36	− 4·22	−25·31	18·05	
27	11·74			41·27	− 4·82	−24·71	16·56	
28	·97			30·50	− 5·42	−24·11	6·39	
29	8·20			37·73	− 6·03	−23·50	14·23	
30		·69		28·84	− 6·63	−22·90	5·94	
31		·58		28·95	− 7·23	−22·30	6·65	
32	8·67			38·20	− 7·83	−21·70	16·50	
33	11·97			41·50	− 8·44	−21·09	20·41	
34	1·15			30·68	− 9·04	−20·49	10·19	
35		7·57		21·96	− 9·64	−19·89	2·07	
36		17·75		11·78	−10·25	−19·28		7·50
37		20·96		8·57	−10·85	−18·68		10·11
38		20·80		8·73	−11·45	−18·08		9·35
39		16·21		13·32	−12·05	−17·48		4·16
40		15·71		13·82	−12·66	−16·87		3·05
41		6·22		23·31	−13·26	−16·27	7·04	
42		20·57		8·96	−13·86	−15·67		6·71
43		24·91		4·62	−14·46	−15·07		10·45
44		27·81		1·72	−15·07	−14·46		12·74
45		21·54		7·99	−15·67	−13·86		5·87
46		33·00	3·47		−16·27	−13·26		16·73
47		27·24		2·29	−16·87	−12·66		10·37
48		22·56		6·97	−17·48	−12·05		5·08
49		31·75	2·22		−18·08	−11·45		13·67
50		37·43	7·90		−18·68	−10·85		18·75
51		30·83	1·30		−19·28	−10·25		11·55
52		16·67		12·86	−19·89	− 9·64	3·22	
53		10·06		19·47	−20·49	− 9·04	10·43	
54		20·29		9·24	−21·09	− 8·44	·80	
55		28·84		·69	−21·70	− 7·83		7·14
56		34·52	4·99		−22·30	− 7·23		12·22
57		19·42		10·11	−22·90	− 6·63	3·48	
58		13·96		15·57	−23·50	− 6·03	9·54	
59		16·74		12·79	−24·11	− 5·42	7·37	
60		24·49		5·04	−24·71	− 4·82	·22	
61		22·61		6·92	−25·31	− 4·22	2·70	
62		16·29		13·24	−25·91	− 3·62	9·62	
63		21·32		8·21	−26·52	− 3·01	5·20	
64		23·93		5·60	−27·12	− 2·41	3·19	
65		22·59		6·94	−27·72	− 1·81	5·13	
66		19·52		10·01	−28·32	− 1·21	8·80	
67		19·39		10·14	−28·93	− ·60	9·54	
68 and over		29·53			−29·53			
Totals	+68·16 −749·80 = −681·64		+19·88 −814·74 = −794·86		−738·25	−738·25	+222·45 −165·84 = +56·61	

Let $X^2_{x-\frac{1}{2}}$ be the contribution to χ^2 from the cell made up of the exposed to risk in the year of life $(x-\frac{1}{2})$ to $(x+\frac{1}{2})$.

Then, dropping the suffix $(x-\frac{1}{2})$ from $X^2_{x-\frac{1}{2}}$, $\theta_{x-\frac{1}{2}}$, $n_{x-\frac{1}{2}}$, $p_{x-\frac{1}{2}}$ and $q_{x-\frac{1}{2}}$,

$$X^2=(\theta-nq)^2/npq=(\theta-nq)^2/nq(1-q), \tag{4}$$

$$\begin{aligned}\partial X^2/\partial A &= (\partial X^2/\partial q)(\partial q/\partial A)=(\partial X^2/\partial q)(1-q)\\ &= -2n(\theta-nq)(1-q)/nq(1-q)+(\theta-nq)^2(1-q)/nq(1-q)^2\\ &\quad -(\theta-nq)^2(1-q)/nq^2(1-q)\\ &= X^2-2(\theta-nq)/q-(\theta-nq)^2/nq^2\\ &= X^2+n-\theta^2/nq^2.\end{aligned} \tag{5}$$

Now $$\partial/\partial q\,(\partial X^2/\partial A)=\partial X^2/\partial q+2\theta^2/nq^3,$$

whence, from formula (1),

$$\begin{aligned}\partial^2X^2/\partial A^2 &= (1-q)\,\partial/\partial q\,(\partial X^2/\partial A)\\ &= \partial X^2/\partial A+2\theta^2(1-q)/nq^3.\end{aligned} \tag{6}$$

Also

$$\begin{aligned}\partial/\partial q\,(\partial^2X^2/\partial A^2) &= \partial/\partial q\,(\partial X^2/\partial A)-2\theta^2/nq^3-6\theta^2(1-q)/nq^4\\ &= \partial/\partial q\,(\partial X^2/\partial A)+4\theta^2/nq^3-6\theta^2/nq^4,\end{aligned}$$

whence

$$\begin{aligned}\partial^3X^2/\partial A^3 &= (1-q)\,\partial/\partial q\,(\partial^2X^2/\partial A^2)\\ &= \partial^2X^2/\partial A^2+4\theta^2(1-q)/nq^3-6\theta^2(1-q)/nq^4\\ &= \partial X^2/\partial A+6\theta^2(1-q)/nq^3-6\theta^2(1-q)/nq^4\\ &= \partial X^2/\partial A-6\theta^2(1-q)^2/nq^4.\end{aligned} \tag{7}$$

Combining formulae (2) and (3) with formulae (5)–(7) it can also be shown that

$$\partial X^2/\partial B = c^x\,(\partial X^2/\partial A), \tag{8}$$

$$\partial X^2/\partial c = (B/c)\,xc^x\,(\partial X^2/\partial A), \tag{9}$$

$$\partial^2X^2/\partial A\partial B = c^x\,(\partial^2X^2/\partial A^2), \tag{10}$$

$$\partial^2X^2/\partial A\partial c = (B/c)\,xc^x\,(\partial^2X^2/\partial A^2), \tag{11}$$

$$\partial^2X^2/\partial B^2 = c^{2x}\,(\partial^2X^2/\partial A^2), \tag{12}$$

$$\partial^2X^2/\partial B\partial c = (B/c)\,xc^{2x}\,(\partial^2X^2/\partial A^2)+(1/c)\,xc^x\,(\partial X^2/\partial A), \tag{13}$$

$$\partial^2X^2/\partial c^2 = (B/c)^2\,x^2c^{2x}\,(\partial^2X^2/\partial A^2)+(B/c^2)\,(x^2-x)\,c^x\,(\partial X^2/\partial A), \tag{14}$$

$$\partial^3X^2/\partial A^2\partial B = c^x\,(\partial^3X^2/\partial A^3), \tag{15}$$

$$\partial^3X^2/\partial A^2\partial c = (B/c)\,xc^x\,(\partial^3X^2/\partial A^3), \tag{16}$$

$$\partial^3X^2/\partial A\partial B^2 = c^{2x}\,(\partial^3X^2/\partial A^3), \tag{17}$$

$$\partial^3X^2/\partial A\partial B\partial c = (B/c)\,xc^{2x}\,(\partial^3X^2/\partial A^3)+(1/c)\,xc^x\,(\partial^2X^2/\partial A^2), \tag{18}$$

$$\partial^3X^2/\partial A\partial c^2 = (B/c)^2\,x^2c^{2x}\,(\partial^3X^2/\partial A^3)+(B/c^2)\,(x^2-x)\,c^x\,(\partial^2X^2/\partial A^2), \tag{19}$$

$$\partial^3X^2/\partial B^3 = c^{3x}\,(\partial^3X^2/\partial A^3), \tag{20}$$

$$\partial^3X^2/\partial B^2\partial c = (B/c)\,xc^{3x}\,(\partial^3X^2/\partial A^3)+(2/c)\,xc^{2x}\,(\partial^2X^2/\partial A^2), \tag{21}$$

$$\partial^3 X^2/\partial B\partial c^2 = (B/c)^2\, x^2 c^{3x}\,(\partial^3 X^2/\partial A^3) + (B/c^2)\,(3x^2 - x)\, c^{2x}\,(\partial^2 X^2/\partial A^2) + (1/c^2)\,(x^2 - x)\, c^x\,(\partial X^2/\partial A), \quad (22)$$

$$\partial^3 X^2/\partial c^3 = (B/c)^3\, x^3 c^{3x}\,(\partial^3 X^2/\partial A^3) + (B^2/c^3)\,(3x^3 - 3x^2)\, c^{2x}\,(\partial^2 X^2/\partial A^2) + (B/c^3)\,(x^3 - 3x^2 + 2x)\, c^x\,(\partial X^2/\partial A). \quad (23)$$

By definition, over any range of ages, $\chi^2 = \Sigma X^2$, and if we write $\Sigma^{(k)} f(x)$ in place of $\Sigma\,\{x^{k-1} f(x)\}$, then

$$\partial\chi^2/\partial A = \Sigma\,(\partial X^2/\partial A), \quad (24)$$

$$\partial^2\chi^2/\partial A^2 = \Sigma\,(\partial^2 X^2/\partial A^2), \quad (25)$$

$$\partial^3\chi^2/\partial A^3 = \Sigma\,(\partial^3 X^2/\partial A^3), \quad (26)$$

$$\partial\chi^2/\partial B = \Sigma\,\{c^x\,(\partial X^2/\partial A)\}, \quad (27)$$

$$\partial\chi^2/\partial c = (B/c)\,\Sigma^{(2)}\,\{c^x\,(\partial X^2/\partial A)\}, \quad (28)$$

$$\partial^2\chi^2/\partial A\partial B = \Sigma\,\{c^x\,(\partial^2 X^2/\partial A^2)\}, \quad (29)$$

$$\partial^2\chi^2/\partial A\partial c = (B/c)\,\Sigma^{(2)}\,\{c^x\,(\partial^2 X^2/\partial A^2)\}, \quad (30)$$

$$\partial^2\chi^2/\partial B^2 = \Sigma\,\{c^{2x}\,(\partial^2 X^2/\partial A^2)\}, \quad (31)$$

$$\partial^2\chi^2/\partial B\partial c = (B/c)\,\Sigma^{(2)}\,\{c^{2x}\,(\partial^2 X^2/\partial A^2)\} + (1/c)\,\Sigma^{(2)}\,\{c^x\,(\partial X^2/\partial A)\}, \quad (32)$$

$$\partial^2\chi^2/\partial c^2 = (B/c)^2\,\Sigma^{(3)}\,\{c^{2x}\,(\partial^2 X^2/\partial A^2)\} + (B/c^2)\,(\Sigma^{(3)} - \Sigma^{(2)})\,\{c^x\,(\partial X^2/\partial A)\}, \quad (33)$$

$$\partial^3\chi^2/\partial A^2\partial B = \Sigma\,\{c^x\,(\partial^3 X^2/\partial A^3)\}, \quad (34)$$

$$\partial^3\chi^2/\partial A^2\partial c = (B/c)\,\Sigma^{(2)}\,\{c^x\,(\partial^3 X^2/\partial A^3)\}, \quad (35)$$

$$\partial^3\chi^2/\partial A\partial B^2 = \Sigma\,\{c^{2x}\,(\partial^3 X^2/\partial A^3)\}, \quad (36)$$

$$\partial^3\chi^2/\partial A\partial B\partial c = (B/c)\,\Sigma^{(2)}\,\{c^{2x}\,(\partial^3 X^2/\partial A^3)\} + (1/c)\,\Sigma^{(2)}\,\{c^x\,(\partial^2 X^2/\partial A^2)\}, \quad (37)$$

$$\partial^3\chi^2/\partial A\partial c^2 = (B/c)^2\,\Sigma^{(3)}\,\{c^{2x}\,(\partial^3 X^2/\partial A^3)\} + (B/c^2)\,(\Sigma^{(3)} - \Sigma^{(2)})\,\{c^x\,(\partial^2 X^2/\partial A^2)\}, \quad (38)$$

$$\partial^3\chi^2/\partial B^3 = \Sigma\,\{c^{3x}\,(\partial^3 X^2/\partial A^3)\}, \quad (39)$$

$$\partial^3\chi^2/\partial B^2\partial c = (B/c)\,\Sigma^{(2)}\,\{c^{3x}\,(\partial^3 X^2/\partial A^3)\} + (2/c)\,\Sigma^{(2)}\,\{c^{2x}\,(\partial^2 X^2/\partial A^2)\}, \quad (40)$$

$$\partial^3\chi^2/\partial B\partial c^2 = (B/c)^2\,\Sigma^{(3)}\,\{c^{3x}\,(\partial^3 X^2/\partial A^3)\} + (B/c^2)\,(3\Sigma^{(3)} - \Sigma^{(2)})\,\{c^{2x}\,(\partial^2 X^2/\partial A^2)\} + (1/c^2)\,(\Sigma^{(3)} - \Sigma^{(2)})\,\{c^x\,(\partial X^2/\partial A)\}, \quad (41)$$

$$\partial^3\chi^2/\partial c^3 = (B/c)^3\,\Sigma^{(4)}\,\{c^{3x}\,(\partial^3 X^2/\partial A^3)\} + (B^2/c^3)\,(3\Sigma^{(4)} - 3\Sigma^{(3)})\,\{c^{2x}\,(\partial^2 X^2/\partial A^2)\} + (B/c^3)\,(\Sigma^{(4)} - 3\Sigma^{(3)} + 2\Sigma^{(2)})\,\{c^x\,(\partial X^2/\partial A)\}. \quad (42)$$

Table 7 shows the headings of the twenty-four columns necessary for the evaluation of the differential coefficients.

113. I originally experimented in some approximate methods aimed at cutting down the number of working columns; one of these was to use the

function χ_1^2 instead of the function χ^2, but this did not give the desired result since, compared with a true minimum-χ^2 fit, the approximate function χ_1^2 tends to favour values of expected deaths lower than the actual. The reason for this is not hard to seek; where the actual deaths are greater than the expected deaths, the denominator of χ_1^2 is greater than that of χ^2, and a biased graduation in favour of low expected deaths would therefore give a lower χ_1^2 than χ^2.

Table 7. Columns required for the calculation of the first three differential coefficients of χ^2 with regard to the Makeham constants A, B and c for $\text{colog}_e\, p_{x-\frac{1}{2}}$

Age x (1)	$n_{x-\frac{1}{2}}$ (2)	$\theta_{x-\frac{1}{2}}$ (3)	c^x (4)	$\text{colog}_e\, p_{x-\frac{1}{2}}$ $=A+Bc^x$ (5)	$p_{x-\frac{1}{2}}$ (6)	$q_{x-\frac{1}{2}}$ (7)	$p_{x-\frac{1}{2}}/q_{x-\frac{1}{2}}$ $=(6)/(7)$ (8)
	⋮	⋮					
	Σ	Σ					

$nq_{x-\frac{1}{2}}$ $=(2)\times(7)$ (9)	$(\theta-nq)_{x-\frac{1}{2}}$ $=(3)-(9)$ (10)	$(npq)_{x-\frac{1}{2}}$ $=(6)\times(9)$ (11)	$X^2_{x-\frac{1}{2}}$ $=(10)\times(10)/(11)$ (12)	$(nq^2)_{x-\frac{1}{2}}$ $=(7)\times(9)$ (13)	$(\theta^2/nq^2)_{x-\frac{1}{2}}$ $=(3)\times(3)/(13)$ (14)
⋮	⋮		⋮		⋮
Σ	Σ		Σ		Σ

$\partial X^2/\partial A$ $=(12)+(2)-(14)$ (15)	$2\times(14)$ $\times(8)$ (16)	$\partial^2X^2/\partial A^2$ $=(15)+(16)$ (17)	$-\partial^3X^2/\partial A^3$ $=3\times(16)\times(8)-(15)$ (18)	$c^x\partial X^2/\partial A$ $=(4)\times(15)$ (19)
⋮	⋮	⋮	⋮	⋮
Σ	Σ	Σ	Σ	Σ $\Sigma^{(2)}$ $\Sigma^{(3)}$ $\Sigma^{(4)}$

$c^x(\partial^2X^2/\partial A^2)$ $=(4)\times(17)$ (20)	$c^{2x}(\partial^2X^2/\partial A^2)$ $=(4)\times(20)$ (21)	$-c^x(\partial^3X^2/\partial A^3)$ $=(4)\times(18)$ (22)	$-c^{2x}(\partial^3X^2/\partial A^3)$ $=(4)\times(22)$ (23)	$-c^{3x}(\partial^3X^2/\partial A^3)$ $=(4)\times(23)$ (24)
⋮	⋮	⋮	⋮	⋮
Σ $\Sigma^{(2)}$ $\Sigma^{(3)}$	Σ $\Sigma^{(2)}$ $\Sigma^{(3)}$ $\Sigma^{(4)}$	Σ $\Sigma^{(2)}$	Σ $\Sigma^{(2)}$ $\Sigma^{(3)}$	Σ $\Sigma^{(2)}$ $\Sigma^{(3)}$ $\Sigma^{(4)}$

Notes. (1) In general eight significant figures were retained throughout the experiments, except where the figures of any column were only required for addition to other figures of a much higher order. (2) The table indicates which summations should be obtained. Most of these are required for the evaluation of the differential coefficients, but certain of the columns are summed merely to give an independent check on the summations of other columns, e.g. Σ column (14) is only required as a check on Σ column (15). (3) If tables of logarithms expressed to sufficient figures are not available, $q_{x-\frac{1}{2}}$ can be quite easily obtained by machine from the expression

$$q_{x-\frac{1}{2}}=\text{colog}_e\, p_{x-\frac{1}{2}}-\tfrac{1}{2}\,\text{colog}_e^2\, p_{x-\frac{1}{2}}+\tfrac{1}{6}\,\text{colog}_e^3\, p_{x-\frac{1}{2}}-\tfrac{1}{24}\,\text{colog}_e^4\, p_{x-\frac{1}{2}}+\dots.$$

114. Several attempts were then made to improve the Makeham constants by reference to the first and second differential coefficients only, in view of the large number of third differential coefficients which would otherwise have to be evaluated. It was found that it was possible to improve the constants up to a point, but that, as soon as the function χ^2 was within about ·5 of what subsequently proved to be its minimum value, no adjustments seemed possible which would finally reduce the first differential coefficients to zero; in fact, I found that I was going round in circles, obtaining a number of different Makeham curves all of which gave approximately the same χ^2, but never the minimum value.

Table 8. Differential coefficients of χ^2 found by the first three trial Makeham curves

Coefficient	First trial value	Second trial value	Third trial value
(1)	(2)	(3)	(4)
$\partial\chi^2/\partial\alpha$	− ·965439	− 1·02747	+ 1·65863
$\partial\chi^2/\partial\beta$	− 1·61622	− 4·46243	+ ·07209
$\partial\chi^2/\partial c$	− 100·280	− 152·246	+ 9·648
$\partial^2\chi^2/\partial\alpha^2$	+ 809·555	+ 802·273	+ 806·792
$\partial^2\chi^2/\partial\alpha\partial\beta$	+ 305·948	+ 360·007	+ 292·198
$\partial^2\chi^2/\partial\alpha\partial c$	+ 24150·4	+ 23293·9	+ 24302·4
$\partial^2\chi^2/\partial\beta^2$	+ 374·056	+ 545·505	+ 338·394
$\partial^2\chi^2/\partial\beta\partial c$	+ 37098·7	+ 44024·3	+ 35464·4
$10^{-1}\,\partial^2\chi^2/\partial c^2$	+380014	+367324	+383227
$\partial^3\chi^2/\partial\alpha^3$	− 1485·17	− 1452·22	− 1481·74
$\partial^3\chi^2/\partial\alpha^2\partial\beta$	− 372·546	− 428·483	− 357·074
$\partial^3\chi^2/\partial\alpha^2\partial c$	− 25344·2	− 24024·7	− 25561·8
$\partial^3\chi^2/\partial\alpha\partial\beta^2$	− 241·001	− 344·113	− 218·883
$\partial^3\chi^2/\partial\alpha\partial\beta\partial c$	− 9590·32	− 11020·5	− 9037·77
$\partial^3\chi^2/\partial\alpha\partial c^2$	−986429	−928031	−999271
$\partial^3\chi^2/\partial\beta^3$	− 407·980	− 727·910	− 350·111
$\partial^3\chi^2/\partial\beta^2\partial c$	− 6315·39	− 8802·08	− 5774·05
$10^{-1}\,\partial^3\chi^2/\partial\beta\partial c^2$	+102894	+126397	+ 97602·0
$10^{-2}\,\partial^3\chi^2/\partial c^3$	+918878	+925856	+918075
χ^2	59·852	60·449	59·808
Trial α	1·1109286	1·1561956	1·1016704
Trial β	2·0841591	1·6930957	2·2003320
Trial c	1·1060303	1·1099514	1·1050138

Note. It was not considered that more than six significant figures could be retained in these values.

115. The particular problem was taken to be to find the Makeham curve which gave the minimum χ^2 when applied to the A 1924–29 duration 0 data for all ages from $19\frac{1}{2}$ upwards; at the end of the preliminary experiments mentioned in paragraph 114, several different Makeham curves were available as possible first trials for the final experiment, and the curve was chosen which gave the lowest value of χ^2 yet obtained. The values were $10^3\text{A}=1{\cdot}1109286$; $10^5\text{B}=2{\cdot}0841591$; $c=1{\cdot}1060303$ and $\chi^2=59{\cdot}85$. It is not considered worth while to give the details of the figures of the whole twenty-four columns; column (2) of Table 8 shows the differential coefficients obtained, writing α for

10^3A and β for 10^5B. It was assumed in the calculation that the group $67\frac{1}{2}$ and over could be taken, on the average, to refer to the year of life $69\frac{1}{2}$–$70\frac{1}{2}$; it is now realized (by reference to the expected deaths according to the final regraduated values) that the weighted mean q should refer to age $69\frac{3}{4}$, and that the method would have been more scientific had the weighted mean age for this 'cell' been recalculated at the end of each trial.

116. When the differential coefficients for the first trial had been obtained, it was necessary to determine what should be the second trial values of α, β and c. The ideal values, clearly, are those giving $\partial\chi^2/\partial\alpha$, $\partial\chi^2/\partial\beta$ and $\partial\chi^2/\partial c$ all equal to zero, and it was hoped that the second and third differential coefficients would give a sufficient indication of the rates of change of the first differential coefficients.

117. Let $\delta\alpha$, $\delta\beta$ and δc be the changes in α, β and c which will provide the required changes in $\partial\chi^2/\partial\alpha$, $\partial\chi^2/\partial\beta$ and $\partial\chi^2/\partial c$ (in this case equal to $+\cdot965439$, $+1\cdot61622$ and $+100\cdot280$ respectively); then if second differential coefficients of χ^2 were all constant we could say:

$$\delta\alpha\,(\partial^2\chi^2/\partial\alpha^2)+\delta\beta\,(\partial^2\chi^2/\partial\alpha\,\partial\beta)+\delta c\,(\partial^2\chi^2/\partial\alpha\,\partial c)=-\text{observed value of }\partial\chi^2/\partial\alpha, \tag{43a}$$

with two similar equations (43b) and (43c) for the negative of the observed values of $\partial\chi^2/\partial\beta$ and $\partial\chi^2/\partial c$. Had these three simultaneous equations been employed to find $\delta\alpha$, $\delta\beta$ and δc, and hence the second trial values of the constants, the second trial would have been far better than the one actually used!

118. The assumption that second differential coefficients were all constant seemed most unlikely to be fulfilled, and the employment of formulae (43) therefore appeared to be a waste of time. But it was hoped that a similar assumption with regard to third differential coefficients would prove to be justified, and the following approximate formulae were accordingly devised:

$$\begin{aligned}&\delta\alpha\,\{\partial^2\chi^2/\partial\alpha^2+\tfrac{1}{2}\delta\alpha\,(\partial^3\chi^2/\partial\alpha^3)+\tfrac{1}{2}\delta\beta\,(\partial^3\chi^2/\partial\alpha^2\partial\beta)+\tfrac{1}{2}\delta c\,(\partial^3\chi^2/\partial\alpha^2\partial c)\}\\&\quad+\delta\beta\,\{\partial^2\chi^2/\partial\alpha\partial\beta+\tfrac{1}{2}\delta\alpha\,(\partial^3\chi^2/\partial\alpha^2\partial\beta)+\tfrac{1}{2}\delta\beta\,(\partial^3\chi^2/\partial\alpha\partial\beta^2)+\tfrac{1}{2}\delta c\,(\partial^3\chi^2/\partial\alpha\partial\beta\partial c)\}\\&\quad+\delta c\,\{\partial^2\chi^2/\partial\alpha\partial c+\tfrac{1}{2}\delta\alpha\,(\partial^3\chi^2/\partial\alpha^2\partial c)+\tfrac{1}{2}\delta\beta\,(\partial^3\chi^2/\partial\alpha\partial\beta\partial c)+\tfrac{1}{2}\delta c\,(\partial^3\chi^2/\partial\alpha\partial c^2)\}\\&\qquad\qquad=-\text{observed value of }\partial\chi^2/\partial\alpha,\end{aligned} \tag{44a}$$

with two similar equations (44b) and (44c) for the negative of the observed values of $\partial\chi^2/\partial\beta$ and $\partial\chi^2/\partial c$. I will not attempt to describe the laborious (but ultimately successful) work in solving these three simultaneous equations, giving the second trial values of the Makeham constants actually employed, since it subsequently transpired that the assumption that the third differential coefficients are constant is so far from the truth that the use of formulae (44) gives corrections to α, β and c too large and in the wrong direction! In fact, one of the solutions of the simultaneous equations (44) appeared to be turning out to be in the same directions as the solution of equations (43), but, as this proved to be one of the solutions with imaginary roots, it was discarded; as already stated, formulae (43) would have given quite a good second trial curve.

119. A comparison of the first and second trial values of the third differential coefficients shows that, whereas those dependent on α and c only are fairly constant, varying in all cases by well under 10%, all those dependent on β change rapidly. This suggests that, although formula (44b) is inaccurate, it should be possible through formulae (44a) and (44c) to find close

approximations to the best values of $\delta\alpha$ and δc if $\delta\beta$ is given the value zero, and thus to find the best α and c corresponding to a given value of β. Putting $\delta\beta = 0$ in formula (44a) and collecting up, we obtain

$$\begin{aligned}\delta\alpha\,(\partial^2\chi^2/\partial\alpha^2) + \delta c\,(\partial^2\chi^2/\partial\alpha\partial c) + \delta\alpha\delta c\,(\partial^3\chi^2/\partial\alpha^2\partial c)\\ + \tfrac{1}{2}\,(\delta\alpha)^2\,(\partial^3\chi^2/\partial\alpha^3) + \tfrac{1}{2}\,(\delta c)^2\,(\partial^3\chi^2/\partial\alpha\partial c^2)\\ = -\text{observed value of } \partial\chi^2/\partial\alpha,\end{aligned} \qquad (45\,a)$$

and we can find a similar equation (45c) for the negative of the observed value of $\partial\chi^2/\partial c$. The evaluation of simultaneous equations (45a) and (45c) will give those changes in α and c necessary, for the same value of β, to produce values of $\partial\chi^2/\partial\alpha$ and $\partial\chi^2/\partial c$ approximately equal to zero. It should, perhaps, be mentioned that the simultaneous equations may be solved by eliminating $(\delta\alpha)^2$, finding an expression of the form

$$\delta\alpha = \{h + k\delta c + l\,(\delta c)^2\} \div (m + n\delta c),$$

and substituting this in formula (45a) to obtain a quartic in δc, which may be solved quite quickly by trial.

120. Substituting the values so found of $\delta\alpha$ and δc in a similar formula involving the observed value of $\partial\chi^2/\partial\beta$, and keeping $\delta\beta$ equal to zero, we can evaluate the expression

$$\begin{aligned}\text{Observed } \partial\chi^2/\partial\beta + \delta\alpha\,(\partial^2\chi^2/\partial\alpha\partial\beta) + \delta c\,(\partial^2\chi^2/\partial\beta\partial c) + \delta\alpha\delta c\,(\partial^3\chi^2/\partial\alpha\partial\beta\partial c)\\ + \tfrac{1}{2}\,(\delta\alpha)^2\,(\partial^3\chi^2/\partial\alpha^2\partial\beta) + \tfrac{1}{2}\,(\delta c)^2\,(\partial^3\chi^2/\partial\beta\partial c^2).\end{aligned} \qquad (45\,b)$$

This function is comparable with $\partial\chi^2/\partial\beta$ for a curve with β equal to the original β, α equal to the original α plus $\delta\alpha$ as found in paragraph 119 and c equal to the original c plus δc as found in paragraph 119; it is not identical with $\partial\chi^2/\partial\beta$ for this curve, since third differential coefficients depending on β are by no means constant. If we denote expression (45b) by the symbol $(\partial\chi^2/\partial\beta)'$, we can say that, if our value of β were the ideal, then, but for the fact that third differential coefficients with regard to α and c are not quite constant, the value of $(\partial\chi^2/\partial\beta)'$ would be identical with $\partial\chi^2/\partial\beta$ and equal to zero.

121. Applying formulae (45) to the results of the first and second trials, we obtain the following values:

Trial no.	$\delta\alpha$	$\alpha+\delta\alpha$	δc	$c+\delta c$	$(\partial\chi^2/\partial\beta)'$	β
1	+·0005012	1·1114298	·0000232	1·1060535	− ·60207	2·0841591
2	+·0001013	1·1562969	·0000408	1·1099922	−2·62877	1·6930957
2 − 1					−2·02670	− ·3910634

122. For lack of any information as to how the function $(\partial\chi^2/\partial\beta)'$ might vary with β, the third trial value of β was found by a straight-line extrapolation, giving $2{\cdot}0841591 + {\cdot}3910634 \times ({\cdot}60207/2{\cdot}02670) = 2{\cdot}2003320$.

123. The problem was now to find third trial values of α and c. I tried first difference extrapolation on the values obtained for $(\alpha+\delta\alpha)$ and $(c+\delta c)$, but the first 15 columns described in Table 7 showed that such a curve, though as good as the first trial, was no better. It did, however, indicate that the best values of α and c for the value $\beta = 2{\cdot}2003320$ might be in the region of 1·1016704 and 1·1050138 respectively; it would be a waste of time to describe just how these values were arrived at, for reasons which will soon be obvious. These values were taken as modified third trial values of α and c, and formulae (45)

gave $(\alpha+\delta\alpha)=1{\cdot}0992281$, $(c+\delta c)=1{\cdot}1050268$, and $(\partial\chi^2/\partial\beta)'=-{\cdot}18121$. A third trial value of c, better than that actually used, and close to this value of $(c+\delta c)$, would have been obtained by the following device, appending a suffix to each function to indicate the number of the trial:

$$\log\beta_1={\cdot}3189309,\qquad \log(c+\delta c)_1={\cdot}0437761,$$
$$\log\beta_2={\cdot}2286815,\qquad \log(c+\delta c)_2={\cdot}0453200,$$
$$\log\beta_3={\cdot}3424882.$$

Let z be the age at which $\beta_1(c+\delta c)_1^z=\beta_2(c+\delta c)_2^z$, then

$$\log\beta_1+z\log(c+\delta c)_1=\log\beta_2+z\log(c+\delta c)_2, \tag{46}$$

whence $$z=58{\cdot}45547,$$

and the L.H.S. of formula (46) = R.H.S. $=2{\cdot}8778834$; c_3 may then be given by

$$\log\beta_3+z\log c_3=2{\cdot}8778834,$$

whence $$\log c_3={\cdot}0433731,$$

and $$c_3=1{\cdot}1050275.$$

I would have saved much time had I discovered this device and gone straight to this third trial value of c, instead of going on a meandering route leading, eventually, to a value differing from this by only 7 in the eighth significant figure.

124. It then seemed desirable to discover a similar means of finding what would have been an α_3 close to the $(\alpha+\delta\alpha)_3$ actually obtained. I have found no formulae giving such a close agreement as just obtained for c, but the following formula has proved as satisfactory as any:

$$\left.\begin{aligned}\log\beta_1+h\log(\alpha+\delta\alpha)_1&=k\log(c+\delta c)_1,\\ \log\beta_2+h\log(\alpha+\delta\alpha)_2&=k\log(c+\delta c)_2,\end{aligned}\right\}$$

$${\cdot}3189309+{\cdot}0458821h={\cdot}0437761k,$$
$${\cdot}2286815+{\cdot}0630693h={\cdot}0453200k,$$

whence $$h=6{\cdot}519193,$$

and $$k=14{\cdot}11832.$$

α_3 may then be given by

$$\log\beta_3+h\log\alpha_3=k\log c_3,$$

whence $$\log\alpha_3={\cdot}0413958,$$

and $$\alpha_3=1{\cdot}1000078.$$

Had this value been used, together with $c_3=1{\cdot}1050275$, a comparatively small $\delta\alpha$ and δc would have been needed to give the best α and c corresponding to the value, already obtained, of β_3.

125. I am not at all happy about this matter of finding successive trial values of α, since I had hoped for something closer. An attempt to find an age y, such that $A_t=B_tc^y$, t being the number of the trial, proved to give a value of α_3 further out than ever, and eventually the method suggested in paragraph 124 was decided on merely because it gave a better value than any other rough and ready method, without there being any particular reason why it should do so.

126. To revert to the experiment, the position had now become:

Trial no.	$\alpha+\delta\alpha$	$c+\delta c$	β	$(\partial\chi^2/\partial\beta)'$
1	1·1114298	1·1060535	2·0841591	− ·60207
2	1·1562969	1·1099922	1·6930957	−2·62877
3	1·0992281	1·1050268	2·2003320	− ·18121

A second difference extrapolation gave $\beta_4=2{\cdot}2540558$; the function

$$\log\beta_3+z\log(c+\delta c)_3,$$

using the same value of z as found in paragraph 123, was slightly diminished to 2·8778656, and equating this to $\log\beta_4+z\log c_4$ a value for c_4 was obtained of 1·1045710. As I was still dissatisfied with previous methods with regard to α, I tried three unknowns in a formula

$$\log\beta+u\log(c+\delta c)=v\log(\alpha+\delta\alpha)+w;$$

solving for u, v and w by substituting the above values, and putting

$$\log\beta_4+u\log c_4=v\log\alpha_4+w,$$

gave a value of α_4 of 1·0942398.

127. Columns (1)–(15) and (19) of Table 7 were completed for these preliminary fourth trial values; χ^2 was down to 59·8008, $\partial\chi^2/\partial\alpha$ to $+{\cdot}42375$, $\partial\chi^2/\partial\beta$ to ·15928 and $\partial\chi^2/\partial c$ up to 14·790. Without evaluating the second and third differentials, a consideration of those obtained at previous trials led me to correct α by $-{\cdot}0005049$ and c by $-{\cdot}0000006$, giving a modified fourth trial of $\alpha=1{\cdot}0937349$, $\beta=2{\cdot}2540558$ and $c=1{\cdot}1045704$. Again completing columns (1)–(15) and (19), it was found that for these values

$$\chi^2=59{\cdot}8007,\ \partial\chi^2/\partial\alpha=-{\cdot}00002,\ \partial\chi^2/\partial\beta=-{\cdot}00677,\ \text{and}\ \partial\chi^2/\partial c=+0{\cdot}129.$$

Bearing in mind my remarks of section XI of this paper I felt that to all intents and purposes this modified fourth trial was the ideal Makeham curve, and I therefore accepted it for the purpose of obtaining a graduation which could illustrate the main part of the paper. The total of column (15) was $+107889{\cdot}015-107889{\cdot}213$, so that $\partial\chi^2/\partial\alpha$ was, to six significant figures, zero. Similarly, $\partial\chi^2/\partial c$ was equal to $(\mathrm{B}/c)\times\Sigma^{(2)}$ col. (19) $=7994{\cdot}23-7994{\cdot}10$, and only a very slight improvement would be possible here. But Σ col. (19) came out to $+7843704-7844381=-677$, from which it seems that only the first three or four significant figures in B can be justified; the results of this graduation would, however, not be substantially changed if the ideal value of B were found to more significant figures.

128. The comparative smallness of the last set of first differential coefficients suggests that, to carry the experiment to its complete conclusion, we would need a fifth trial value of β. I decided not to waste time on this as I had already obtained a satisfactory graduation for the purpose of the paper; it seems likely, however, that four trial values would have given us the absolute ideal value had our second trial not been a long step in the wrong direction.

129. The methods of obtaining c_4 and α_4 in paragraph 126 are somewhat clumsy, and it will be instructive to consider what would have been the alternative results had we used the suggested methods of paragraphs 123 and 124 and used only the values from the two best trials to date (the first and third) in order to find the fourth trial values:

$$\log\beta_1={\cdot}3189309,\quad \log(c+\delta c)_1={\cdot}0437761,$$
$$\log\beta_3={\cdot}3424882,\quad \log(c+\delta c)_3={\cdot}0433728,$$
$$\log\beta_4={\cdot}3529647.$$

Using formula (46):

$$\cdot 3189309 + \cdot 0437761z = \cdot 3424882 + \cdot 0433728z,$$
$$z = 58\cdot 41136,$$
$$\text{L.H.S.} = \text{R.H.S.} = 2\cdot 8759524,$$
$$\log c_4 = (2\cdot 8759524 - \cdot 3529647)/58\cdot 41136 = \cdot 0431934,$$
$$c_4 = 1\cdot 1045704 \text{ (cf. modified } c_4 \text{ actually used).}$$

Using formula (47):

$$\cdot 3189309 + \cdot 0458821h = \cdot 0437761k,$$
$$\cdot 3424882 + \cdot 0410878h = \cdot 0433728k,$$
$$h = 6\cdot 060840 \text{ and } k = 13\cdot 63792,$$

therefore
$$\cdot 3529647 + 6\cdot 060840 \log \alpha_4 = 13\cdot 63792 \times \cdot 0431934,$$
$$\log \alpha_4 = \cdot 0389556,$$
$$\alpha_4 = 1\cdot 0938446$$
(cf. 1·0942398 4th trial,
1·0937349 modified 4th trial).

130. Consideration of the experiments just described suggests that the following are the operations likely to lead most rapidly to the minimum-χ^2 fit of a Makeham curve.

1. Find first trial values α, β and c by the 'flexible method' outlined in Appendix 3 and evaluate the first, second and third differential coefficients of χ^2 using headings of Table 7.
2. Find $\delta\alpha$, $\delta\beta$ and δc by means of three simultaneous equations (43) (and hence second trial values α_2, β_2 and c_2) and evaluate the first, second and third differential coefficients of χ^2 using headings of Table 7.
3. (*a*) Find $\delta\alpha_1$, δc_1, $(\partial\chi^2/\partial\beta)'_1$, $\delta\alpha_2$, δc_2, and $(\partial\chi^2/\partial\beta)'_2$ by formulae (45);
(*b*) using β_1, β_2, $(\partial\chi^2/\partial\beta)'_1$ and $(\partial\chi^2/\partial\beta)'_2$ find β_3 by straight-line interpolation or extrapolation;
(*c*) using formulae (46) and (47) respectively, find c_3 and α_3;
(*d*) evaluate first, second and third differential coefficients of χ^2 using headings of Table 7.
4. (*a*) Find $\delta\alpha_3$, δc_3 and $(\partial\chi^2/\partial\beta)'_3$ by formulae (45);
(*b*) using three previous values of β and $(\partial\chi^2/\partial\beta)'$ find β_4 by second difference interpolation or extrapolation;
(*c*) using second and third trial figures only, find c_4 and α_4 by formulae (46) and (47) respectively;
(*d*) complete columns (1)–(15) and (19) of Table 7, evaluating first differential coefficients of χ^2. By inspection of these, improve α_4 (and if necessary c_4) and by again completing columns (1)–(15) and (19) confirm that first differential coefficients are now zero.

Before each successive trial, the weighted mean age of any cell containing more than one age should be recalculated by reference to the rates of mortality according to the latest trial curve.

131. The suggested method of paragraph 130 may be suitably modified if it is found that the required number of trial β's in any particular case is three or five.

132. I have been through the algebra of the differential coefficients of χ^2 for a Perks curve; in view of the extensive number of coefficients these have not been included in the paper, but it may be said that they are not difficult to obtain. I have not yet devised a method of improving the four Perks constants to give a minimum value of χ^2, and it will probably prove to be a lengthy process, but it is a possibility worthy of consideration.

APPENDIX 3

Makeham fit by 'flexible method'; some experiments with the A 1924–29 ultimate data

133. The possibility of fitting a Makeham curve to the A 1924–29 duration 0 data led to the experiments described in Appendix 2; the results of these were so encouraging that I decided to make a similar attempt on as large a section as possible of the ultimate data. I was fully aware of the fact that the presence of a large number of duplicates would cause difficulties, but Daw had suggested on p. 181 of his paper that the standard deviation of q_x for this experience was likely to be increased by about 50%. Even before the publication of Solomon's paper I had resolved not to go as far as this, but to make the more moderate assumption that the standard deviation was increased by 40%.

134. The function estimated by Solomon was the rate by which the variance, as opposed to the standard deviation, of the number of deaths was increased, and was therefore the square of the function estimated by Daw. He examined the continuous mortality investigation data over the period 1924–38 (medical, ages 46–55, durations 5 and over), and concluded that a value of 1·6 was the best estimate, with possible values lying between 1·3 and 2·0. Taking the square root of these values, it appears that the rate of increase of the standard deviation lay between 1·14 and 1·41 for those particular data. This measure of the increased variance was approximately the same for each of the twelve groups making up Solomon's data; the 1924–29 data consisted, roughly, of four of these twelve classes combined, with the non-medical data added, and it seems reasonable to suppose that the effect of duplicates would be increased by combining the different classes. In any case, the effect of duplicates would probably be increased by combining durations 3 and over, as compared with the 5 and over taken by Solomon, so that any estimate of the increased variance below Solomon's upper limit appears to be not unreasonable. I therefore made no alteration in my decision to assume that the standard deviation of the number of deaths was $1{\cdot}4\sqrt{npq}$.

135. In the discussion on Starke's recent paper (5) I suggested that for a fit of Makeham's curve we might find trial values of all three constants rather than merely a trial value of c combined with the method of moments. Table 9 shows a possible way of arriving at first trial values of these constants. It should, at this stage, be mentioned that I originally hoped to be able to fit a curve to the data from ages $19\frac{1}{2}$ to $65\frac{1}{2}$, in order to embrace the vast majority of Endowment Assurance data. It soon became apparent that the abnormally high crude rate of mortality for the year of age $20\frac{1}{2}$–$21\frac{1}{2}$ was quite out of line, and I attributed this to the non-exclusion of children's deferred assurances

which had just passed the option date; the data to be graduated were then amended to the range of ages from 21½ to 65½.

136. It would have been possible to show one further value of A in column (14) of Table 9; this would have been negative, and it was therefore quite clear that the first trial values would not give a curve suitable for the higher ages in the range. However, in order to improve the curve it was desirable to examine

Table 9. A method of deriving trial values of the Makeham constants

Age-group (1)	Central age x (2)	Central exposed to risk ${}_5E^c_x$ (3)	Deaths ${}_5\theta_x$ (4)	Crude μ_x (5)	$\mu_{x+5}-\mu_x$ $=Bc^x(c^5-1)$ (6)	$\frac{\mu_{x+10}-\mu_{x+5}}{\mu_{x+5}-\mu_x}$ =crude c^5 (7)
25½–	28	647491·50	1505	·0023244	·0002918	2·3986
30½–	33	889071·50	2326	·0026162	·0006999	1·9380
35½–	38	1123903·75	3727	·0033161	·0013564	1·1628
40½–	43	1227620·25	5737	·0046725	·0015772	2·1167
45½–	48	1221344·50	7633	·0062497	·0033411	1·7409
50½–	53	1040377·00	9978	·0095908	·0058174	1·7677
55½–	58	769459·00	11856	·0154082	·0102832	
60½–	63	490202·75	12594	·0256914		
		7409470·25	55356			

Weight (proportionate to ${}_5E^c_x+2\,{}_5E^c_{x+5}+{}_5E^c_{x+10}$) (8)	c^x (from value of c^5 at foot of column (8)) (9)	$c^x(c^5-1)$ (10)	(6)÷(10) =crude B (11)	Weight (proportionate to ${}_5E^c_x+{}_5E^c_{x+5}$) (12)	Bc^x (13)	(5)−(13) =crude A (14)
355	29·8007	24·8356	·000011749	154	·0003780	·001946
436	54·6363	45·5333	·000015371	201	·0006931	·001923
480	100·1696	83·4803	·000016248	235	·0012707	·002045
471	183·6500	153·0521	·000010305	245	·0023296	·002342
407	336·7021	280·6042	·000011907	226	·0042711	·001978
307	617·3063	514·4569	·000011308	181	·0078305	·001760
	1131·7632	943·2001	·000010902	126	·0143564	·001051
2456 whence weighted mean $c^5=1{\cdot}83339$				1368 whence weighted mean 10^5B $=1{\cdot}2685$		Weighted mean 10^3A $=1{\cdot}908$ using col. (3) $\times 10^{-4}$ as weights

Note. The data for the above were taken from Mortality of Assured Lives 1924 (Extracts and Discussions), 662–3, as amended in *J.I.A.* LXVIII, 83.

the deviations, and these are shown in Table 10. A consideration of these led to the obvious conclusion that the curve was too steep, and that what was wanted was a lower value of c and a higher value of B. Further trials were then made giving c^5 various values between 1·7 and 1·8 and again deducing B and A and finding the deviations (at this stage it was not considered worth while going to the trouble of evaluating $q_{x-\frac{1}{2}}$ for all ages, and the expected deaths were taken as $E^c_{x-\frac{1}{2}}\mu_x$). All these trials still gave a run of positive deviations between 40 and 46 and unfortunately a negative 'bulge' also appeared between 32 and 37. The problem appeared to develop into one of reducing one bulge without enlarging others, and it appeared necessary to derive a curve giving $\mu_{34} \doteqdot \cdot 0028$, $\mu_{43} \doteqdot \cdot 0044$ and $\mu_{52} \doteqdot \cdot 0088$ without causing distortion at the extremes of the table. I found that if these values were respectively amended to ·00279, ·00441 and ·00884 the extension of the Makeham curve gave $\mu_{25} = \cdot 00220$ and $\mu_{61} = \cdot 02095$, and from the resulting value of c^9 it was possible to fill in the intervening values. The approximate deviations indicated that this curve was

Table 10. A 1924–29 durations 3 and over; actual minus expected deaths according to trial values found in Table 9

Age	(Actual − Expected)		Age	(Actual − Expected)		Age	(Actual − Expected)	
	+	−		+	−		+	−
22	6		35		22	50	104	
23	24		36	1		51	93	
24	8		37	4		52		86
			38	81		53	34	
25	6		39		6	54		90
26		—						
27	7		40	41		55		150
28	26		41	79		56		134
29	10		42	131		57		76
			43	84		58		67
30		26	44	99		59		167
31	32							
32	5		45	92		60		186
33		28	46	15		61		263
34	8		47	25		62		55
			48		30	63		250
			49		85	64		250
						65		376

quite a good fit, and it was decided to evaluate $q_{x-\frac{1}{2}}$ and test in full as described in the paper, with the exception of the smoothness test which the curve is bound to satisfy. The values of the constants for $\operatorname{colog}_e p_{x-\frac{1}{2}}$ are:

$$10^3 A = 1 \cdot 8581, \quad 10^5 B = 2 \cdot 08276, \quad c = 1 \cdot 118313.$$

137. The figures required for testing this graduation are shown in Table 11. The results of the tests are as follows.

Test 2. *Eighteen of the values in column* (8) are less than their probable error and 26 greater, compared with acceptable limits of 15–29.

Test 3. Only two of the deviations exceed twice their standard error, and this is quite satisfactory.

Test 4. Column (9) shows that for only two groups of ages do the combined total deviations exceed twice their standard errors, viz. 40–45 and 46–49; there are, however, three other groups where the combined total approaches twice the standard error, and it is agreed that the graduation can be criticized on this score but not necessarily rejected. If we repeat Test 3 with each of these five groups combined as a composite cell, we are left with 28 cells in only three of which does the deviation exceed twice its standard error, and this is *within the acceptable limits of the test.*

Table 11. A 1924–29, durations 3 and over, ages $21\frac{1}{2}$–$65\frac{1}{2}$; regraduation by Makeham's formula

Age x	$\text{colog}_e\, p_{x-\frac{1}{2}}$ (1)	$\theta_{x-\frac{1}{2}}$ (2)	$n_{x-\frac{1}{2}}$ (3)	$q_{x-\frac{1}{2}}$ (4)	$(nq)_{x-\frac{1}{2}}$ (5)	$(npq)_{x-\frac{1}{2}}$ (6)	$\sigma_\theta = 1{\cdot}4\sqrt{npq}$ (7)
22	·002102	85	37888·75	·002100	79·6	79·4	12·5
23	·002131	138	53799·5	·002129	114·5	114·3	15·0
24	·002163	162	72170·25	·002161	156·0	155·7	17·5
25	·002199	204	91195	·002197	200·4	200·0	19·8
26	·002239	235	106508	·002236	238·2	237·7	21·6
27	·002285	275	119813	·002282	273·4	272·8	23·1
28	·002335	326	131512·75	·002332	306·7	306·0	24·5
29	·002391	339	141192·5	·002388	337·2	336·4	25·7
30	·002455	330	149217·75	·002452	365·9	365·0	26·7
31	·002525	418	157507	·002522	397·2	396·2	27·9
32	·002604	426	167118·25	·002601	434·7	433·6	29·2
33	·002692	434	177870·5	·002688	478·1	476·8	30·6
34	·002791	515	188782·75	·002787	526·1	524·6	32·1
35	·002901	533	198956	·002897	576·4	574·7	33·6
36	·003025	607	208803·75	·003020	630·6	628·7	35·1
37	·003163	664	217862	·003158	688·0	685·8	36·7
38	·003317	800	226441	·003311	749·7	747·2	38·3
39	·003490	778	234839	·003484	818·2	815·3	40·0
40	·003683	878	237821·5	·003676	874·2	871·0	41·3
41	·003899	976	240528·75	·003891	935·9	932·3	42·7
42	·004140	1093	242781	·004131	1002·9	998·8	44·2
43	·004410	1129	247137	·004400	1087·4	1082·6	46·1
44	·004712	1233	250560·5	·004701	1177·9	1172·4	47·9
45	·005050	1306	249481·5	·005037	1256·6	1250·3	49·5
46	·005427	1320	248699·75	·005412	1346·0	1338·7	51·2
47	·005849	1437	248776·25	·005832	1450·9	1442·4	53·2
48	·006322	1503	248899·75	·006302	1568·6	1558·7	55·3
49	·006850	1569	246567	·006827	1683·3	1671·8	57·2
50	·007440	1804	232218·25	·007412	1721·2	1708·4	57·9
51	·008101	1867	221270·5	·008068	1785·2	1770·8	58·9
52	·008839	1821	216476·75	·008800	1905·0	1888·2	60·8
53	·009665	2100	213176·25	·009618	2050·3	2030·6	63·1
54	·010589	2119	206567	·010533	2175·8	2152·9	65·0
55	·011622	2071	187875·5	·011554	2170·7	2145·6	64·8
56	·012777	2139	173613·75	·012695	2204·0	2176·0	65·3
57	·014069	2330	165689·75	·013970	2314·7	2282·4	66·9
58	·015514	2492	158572	·015395	2441·2	2403·6	68·6
59	·017130	2517	149423·75	·016984	2537·8	2494·7	69·9
60	·018936	2378	128087·75	·018758	2402·7	2357·6	68·0
61	·020957	2262	113042·75	·020739	2344·4	2295·8	67·1
62	·023217	2595	106377·75	·022949	2441·3	2385·3	68·4
63	·025744	2553	100572·25	·025416	2556·1	2476·6	69·7
64	·028570	2682	94078	·028166	2649·8	2575·2	71·0
65	·031730	2502	82429	·031232	2574·4	2494·0	69·9
		55945	7692201·75		56029·2	55306·9	2033·8

Table 11 (cont.)

Age x	$\theta - nq$ (8) +	$\theta - nq$ (8) −	Suspect groups (9)	$\frac{(\theta - nq)^2}{\sigma_\theta^2}$ (10)	$\Sigma(\theta - nq)$ (11) +	$\Sigma(\theta - nq)$ (11) −	Balance line (12) −	(11)−(12) (13) +	(11)−(12) (13) −
22	5·4			·19	5·4		1·9	7·3	
23	23·5			2·45	28·9		3·8	32·7	
24	6·0			·12	34·9		5·7	40·6	
25	3·6			·03	38·5		7·7	46·2	
26		3·2		·02	35·3		9·6	44·9	
27	1·6			·00	36·9		11·5	48·4	
28	19·3			·62	56·2		13·4	69·6	
29	1·8			·00	58·0		15·3	73·3	
30		35·9		1·81	22·1		17·2	39·3	
31	20·8			·56	42·9		19·1	62·0	
32		8·7		·09	34·2		21·1	55·3	
33		44·1		2·08		9·9	23·0	13·1	
34		11·1	154·9;	·12		21·0	24·9	3·9	
35		43·4	$\sigma_\theta = 80·7$	1·67		64·4	26·8		37·6
36		23·6		·45		88·0	28·7		59·3
37		24·0		·43		112·0	30·6		81·4
38	50·3			1·72		61·7	32·5		29·2
39		40·2		1·01		101·9	34·4		67·5
40	3·8			·01		98·1	36·4		61·7
41	40·1			·88		58·0	38·3		19·7
42	90·1		280·1;	4·16	32·1		40·2	72·3	
43	41·6		$\sigma_\theta = 111·2$	·81	73·7		42·1	115·8	
44	55·1			1·32	128·8		44·0	172·8	
45	49·4			1·00	178·2		45·9	224·1	
46		26·0		·26	152·2		47·8	200·0	
47		13·9	219·8;	·07	138·3		49·8	188·1	
48		65·6	$\sigma_\theta = 108·5$	1·41	72·7		51·7	124·4	
49		114·3		3·99		41·6	53·6	12·0	
50	82·8		164·6;	2·05	41·2		55·5	96·7	
51	81·8		$\sigma_\theta = 82·6$	1·93	123·0		57·4	180·4	
52		84·0		1·91	39·0		59·3	98·3	
53	49·7			·62	88·7		61·2	149·9	
54		56·8	221·5;	·76	31·9		63·2	95·1	
55		99·7	$\sigma_\theta = 112·6$	2·37		67·8	65·1		2·7
56		65·0		·99		132·8	67·0		65·8
57	15·3			·05		117·5	68·9		48·6
58	50·8			·55		66·7	70·8	4·1	
59		20·8		·09		87·5	72·7		14·8
60		24·7		·13		112·2	74·6		37·6
61		82·4		1·51		194·6	76·5		118·1
62	153·7			5·05		40·9	78·5	37·6	
63		3·1		·00		44·0	80·4	36·4	
64	32·2			·21		11·8	82·3	70·5	
65		72·4		1·07		84·2	84·2		
	+878·7	−962·9 = −84·2		46·57	1493·1 − 1616·6 = −123·5		−1894·6	2415·1 − 644·0 = +1771·1	

Test 5. The sum of the deviations regarding sign is $-84{\cdot}2$ compared with a standard error of 329.

Test 6. The second sum of the deviations regarding sign (column 11) is $-123{\cdot}5$ which, without evaluating a complicated series similar to that in paragraph 45, is seen to be well within its acceptable limits.

Test 7. The sum of the deviations disregarding sign is 1841·6, compared with permitted limits on the basis of paragraph 46 of 1627 ± 395. It will be found that the sum of the standardized deviations disregarding sign is 37·83 and is also well within the limits.

Test 8. The value of χ^2 of 46·57 for 44 degrees of freedom is excellent. Even if we bias this test by combining the five suspect groups into composite cells as suggested under Test 4, we arrive at a value, 41·61, for 28 cells, which is only just outside Seal's limits.

Test 9. Possible adverse runs are ages 41–45 and 48–56. The sectional values of χ^2 for these groups are respectively 8·17 (5 degrees of freedom) and 16·03 (9 degrees of freedom), both of which are acceptable. The smaller run from 48 to 52 has a value of 11·29 for 5 degrees of freedom, which is just outside the usual limits, but the graduation would certainly not be rejected merely because for one group of five ages over a range of 44 the value of P just falls below ·05.

Test 10. The deviations change sign 19 times out of a possible 43, well within the acceptable limits.

Test 11. There are 22 positive deviations and 22 negative.

Test 12. Column (13) shows six sign-changes, compared with acceptable values *on the basis of Table 3 of 6–19.*

Test 13. Only 13 of the values in column (13) are negative as against 30 positive, while the acceptable limits on the normal basis would be 15–28. This is an unsatisfactory feature, but not sufficient to reject the graduation, especially since the unadjusted series in column (11) has 21 negatives. In this case the run of positives at the early ages combined with a negative balance line appears to have caused some bias in the adjusted series and gives further evidence that the accumulated deviation tests are not very satisfactory. Support of this statement may be quoted from Seal's closing remarks on p. 65 of the discussion on his paper, where he states: 'I...can only express my personal dissatisfaction with the accumulated deviation test in any shape or form.'

Summary. The most unsatisfactory feature appears to be that the graduation has done a certain amount of wave cutting, but it is felt that the sizes of the waves are no larger than can be justified by the statistical tests.

REFERENCES

(1) Cramér, H. and Wold, H. Mortality variations in Sweden. *Skand. Akt.* (1935), 162.

(2) Daw, R. H. (1944). On the validity of statistical tests of the graduation of a mortality table. *J.I.A.* LXXII, 174.

(3) Seal, H. L. (1940). Tests of a mortality table graduation. *J.I.A.* LXXI, 5.

(4) Solomon, L. (1948). The analysis of heterogeneous mortality data. *J.I.A.* LXXIV, 94.

(5) Starke, L. G. K. (1949). Some thoughts on the analysis of numerical data. *J.I.A.* LXXV, 183.

(6) Tetley, H. (1950). *Actuarial Statistics*, Vol. I (2nd ed.). Cambridge University Press.

ABSTRACT OF THE DISCUSSION

Mr M. C. Polman, in opening the discussion, thought that he had a considerable amount in common with the author. There seemed to have been a tendency to regard the testing of graduations as solely a problem in mathematical statistics, and public discussion took place on a very high plane. He believed that it was one of the objects of the paper under discussion to try to restore a proper balance between theory and practice. Even so, he doubted whether the paper went far enough.

The approach of the actuary to the question of graduation was strictly utilitarian. The available information had to be obtained and put into the form in which it could be used—some system of ratios, mortality rates, withdrawal rates, etc. The results might be needed for either of two main purposes, comparisons or financial calculations. For purposes of comparisons, actuaries were usually concerned with the general trend of the results; the observed ratios might be departed from only to the extent necessary to discover the trend. The National Life tables were an example of what he meant.

Tables for financial purposes, however, required an additional quality called smoothness, and to obtain it some departure from the crude results was permissible. Many methods had been evolved, and many more might yet be developed, to obtain a suitable compromise between fit and smoothness. When a table was needed for financial purposes, actuaries had their own simple and searching tests—with variations according to individual tastes—for selecting which was the most suitable graduation for their purposes. It did not matter how straightforward or how complicated the process might have been to arrive at the graduated results; the graduation stood or fell by those tests.

The graduated table was wanted for the calculation of premium rates, reserves, valuation functions and so on. Would the whole-life premiums increase progressively with age? Was it possible to interpolate with confidence between the annual premiums for reversionary annuities at pivotal ages? Could approximate methods of valuation be used with reasonable safety? The table could be said to be smooth if that kind of thing were possible.

The transition from those considerations to the properties of the graduated values was not an easy one to express in terms of mathematics, though it was generally possible to recognize whether the graduated table was what was wanted by an examination of the first two or three orders of differences. That was what the author meant, presumably, by 'recognizing smoothness in a somewhat indefinite manner'. He preferred to say that the recognition was the result of practical experience—not so much their own, but that of generations of actuaries before them—which gave them a prior knowledge of the way in which many of the tables that they used would behave.

On p. 19 of the paper the author gave his own definition of smoothness but, very wisely, he did not expect too much from it. The definition, with reservations, would probably do for the commoner actuarial functions, but might well fail altogether with the rarer ones. A table of remarriage rates, for instance, might have features which took it quite outside the suggested definition.

So far, with some differences of emphasis, he agreed with the author, but on the question of fit they parted company. It seemed to him fundamental that, after any graduation for financial purposes, the financial effect of the graduated table should be as close as possible to that of the ungraduated values. The crude experience usually represented cash, whether received or paid out—generally speaking paid out. It might be necessary to keep margins, either because future experience might not reproduce the past or purely as a precaution, but the margin should be taken deliberately and not as a by-product of the method of graduation. The actual and expected deaths, therefore, should be the same.

Again, the use of a test based on accumulated deviations merely showed the extent to which the actual and expected deaths were equal over short ranges of ages, and therefore showed how far the graduated values would reproduce the crude experience within the table. Should certain values, for instance at the extreme ends of the mortality experience, be considered unreliable, then some system of weighting could be used, but the weighted *values should still answer to the tests. In general, all graduations for financial purposes* should produce equality between actual and expected, and the number of changes of sign in accumulated deviations then showed how closely the graduated values adhered to the data.

The author's attempt to determine how many changes in sign there should be seemed to be irrelevant. That test, together with the test for smoothness, was relative; they were simply means of comparing one graduation with another. Personally, he did not believe that there was any way of arriving at the best graduation other than by trial and error, and, because the final decision was a compromise, there would always be room for difference of opinion, beauty being in the eye of the beholder.

When an exceptional deviation stuck out like a sore thumb, the data should be reviewed. But it was convenient to have some measuring rod for the size of deviations, and for that purpose there was a mathematical model based on some form of the binomial distribution. The only justification for the model was that it gave results useful in practice which could not be obtained otherwise. That was undoubtedly true, but if the model were developed too far the number of variations would become embarrassing. There were standard deviations and mean deviations of varying sizes, and perhaps one day the probable error might be heard of again. They were all based on their own sets of assumptions. To use more than one such test seemed mere duplication, and that one should be used merely as an adjunct to the main test, based on the changes in sign of the accumulated deviations.

He thought that actuaries should beware of making too many assumptions in devising their tests, and for that reason he saw no purpose in the additional assumptions required for the χ^2 test. The test had its place in general statistics, but seemed to him to be out of place in work of the kind under discussion. The test bristled with difficulties. There was a difference of opinion over the proper size of the standard deviations. There were further differences of opinion over the number of degrees of freedom and he believed that, at the end, the test gave no more information than could be obtained by other means.

Perhaps he might illustrate what he had said by referring to the sample graduation. Curve-fitting by least squares was well known for producing unpleasant sets of equations, and in the instance in question it certainly did. No doubt the author enjoyed himself immensely but, in spite of all the work that he had put into it, probably most actuaries would look for a graduation which gave a rather better fit. From Table 6, columns (1) and (2) or column (5), it would be seen that the expected differed from the actual by 29·53; in other words, the expected was greater than the actual by roughly $1\frac{1}{2}\%$ of the actual deaths. Column (13) showed how those excess deaths could be best distributed; but a glance at that column showed that the graduated values were too low at the ends and too high in the middle. Not only had the number of deaths been overstated, but the curve had been straightened out. Such a result often followed from the fitting of a Makeham curve. He wondered—he had no means of telling with certainty—whether that kind of difficulty was the reason why the Committee responsible for the A 1924–29 table rejected a Makeham graduation.

Features of that kind might not be vital in the first year of selection, but if similar features appeared in the ultimate curve both premiums and reserves would be affected. The premiums would be too low for certain ages and too high for others. The reserves on the whole-life plan, the level of which depended on the steepness of the curve of μ_x, would tend to be too low. Perhaps further investigation would be needed to determine the extent of the particular distortion. Whether the convenience of the Makeham formula was sufficient compensation was a matter of opinion; personally he doubted it.

Mr H. L. Seal, in a written contribution which was read at the meeting, mentioned that his paper—criticized by the author as having been 'too widely read'—was printed and circulated to interested members in mid-October, 1939; as might be expected, he had somewhat modified his ideas during the intervening eleven years. He was no longer happy with the suggested *two* tests of a graduation (namely, a χ^2 test of 'fit', and a test for improbable sign-sequences in successive deviations of actual from expected deaths) and wished to see them replaced by a single truly efficient test. Since that view was diametrically opposed to the author's he would try to explain his reasons.

It had been a feature of statistical theory during the past quarter of a century that the one-time multiplicity of tests had tended to be replaced by a single test when the hypothesis to be tested had been formulated in precise terms and the class of alternative hypotheses had been specified. And, furthermore, it had sometimes been possible to prove that no other test devised could add to the information so provided; in other cases, the single test could be shown to have desirable properties which were not duplicated by any other test.

Those remarks were relevant to the graduation tests under discussion because their very diversity pointed to an imprecisely-posed problem. Could not actuaries put into exact mathematical terms the statistical hypothesis and its alternatives to be tested in a unique graduation test? Was an array of hypotheses essential? Could they not be all subsumed under one head?

The following comments on the tests advocated in the paper would illustrate his arguments:

(*a*) Tests 2, 3, 7 and 8 appeared to be directed to testing whether the individual deviations were distributed binomially, but without regard to the fact that the deviations were ordered according to age.

(*b*) Tests 10, 11, 12 and 13 were all—as the author indicated—tests of the improbability of groups of positive and negative deviations, without reference to the sizes of such deviations or their distribution.

(*c*) Tests 4 and 9 marked an attempt to test, simultaneously, the binomiality of the deviations and their ordering according to age.

(*d*) Tests 5 and 6 were more in the nature of checks on the method of fitting than of statistical tests. The (unknown) constraints set on those tests by almost any conceivable method of graduation seemed to invalidate a probability judgment of success or failure. The hypotheses thus tested were not mutually exclusive and most of the tests were correlated with the others. In consequence, it was difficult to obtain a clear picture from the dozen or so individual, correlated and conflicting probability judgments.

Tests 4 and 9 were incorrectly applied in the paper and both for the same kind of reason. Thus, in describing Test 4, the author overlooked that the distribution of n unit normal variates known to be positive was not normal with variance n, but was of complicated form (Triconi, *Gior. ist. ital. attuar.* 1937) and had a larger variance.

Though he was keenly aware that the statistical theory of graduation testing was still fragmentary with regard to the 'laws' most popular with actuaries, he would indicate one case where existing theory was adequate. Though only a particular case could be treated in that manner, the graduation procedure and its subsequent testing was not far removed from practice. It had the supreme advantage of leading to a single statistical test which was known to be 'good' and might well be the 'best' possible.

Statistical tests for the adequacy of a given mortality law in the representation of a series of observations encountered the difficulty that the (approximate) normal distribution of deaths at any age involved the parametric (universal) value of q_x in both its mean and variance. In order to avoid the consequent mathematical difficulties a transformation of variable was desirable.

If y was a binomial random variable with parameters q and E the transformed random variable

$$z=\sqrt{(\mathrm{E}+\tfrac{1}{2})}\ \mathrm{arc\,sin}\left(\frac{y+\frac{3}{8}}{\mathrm{E}+\frac{3}{4}}\right)^{\frac{1}{2}}$$

had, for large E, a mean value of $\mathrm{E}q+\frac{1}{2}q-\frac{1}{4}$ and a variance $\frac{1}{4}$ (Anscombe, *Biom.* XXXV, 1948) and was, in the limit, distributed normally (Curtiss, *Ann. Math. Statist.* XIV, 1943).

He then supposed that a series of values of z_x had been observed corresponding to the successive attained ages of life ($x=\alpha, \alpha+1, \ldots, \beta$, n values in all), and that the graduation formula

$$z_x=\sum_{j=1}^{k}\alpha_j x_j,$$

where x_j was a determinable function of x (e.g. $x_j = c_j^x$, c_j known), had been fitted to the observations by means of the method of maximum likelihood with the result that the parameter α_j was estimated by the value $\hat{\alpha}_j$ ($j = 1, 2, \ldots, k$). It would be desired to test the adequacy of the graduation by investigating the necessity of the last, most complicated, term of the formula. In other words the hypothesis to be tested was that $\alpha_k = 0$.

His description of a graduation procedure and its subsequent testing was not far removed from what might happen in practice. Admittedly, it was q_x and not z_x that was graduated and the formula chosen was seldom applied by means of maximum likelihood (*which, however, was closely allied to the minimum-*χ^2 *used by the author*). Furthermore, should expressions of the form c^x occur, c had to be estimated—it was not known *a priori*. However, it could not be seriously argued that the choice of z_x instead of q_x entirely altered the theoretical content of the problem, though an inefficient method of fitting invalidated the subsequent distribution theory. It should also be remembered that a complex graduation formula might often be approximated to by a polynomial of sufficiently high degree in x.

Should the above formulation be adopted it could be shown (e.g. Mood, *Introduction to the theory of statistics*, New York, 1950) that the likelihood ratio test of the hypothesis mentioned was a quotient of two random variables, each distributed as χ^2, with 1 and $n-k$ degrees of freedom respectively, of which the denominator was

$$\sum_{x=\alpha}^{\beta} \left(z_x - \sum_{j=1}^{k} \hat{\alpha}_j x_j \right)^2$$

That expression was recognizable as being closely analogous to the χ^2 value used in mortality table graduations.

In some ways that was the 'best possible' solution. As Kendall remarked (*The advanced theory of statistics*, Vol. II, London, 1946): 'These [viz. tests based on the likelihood ratio] will give uniformly most powerful tests if such exist, and in the contrary case will do their best, so to speak, by finding the greatest common denominator among the best critical regions.' In less technical terms, he might say that the statistician had devised a certain type of test of a hypothesis subject to specified alternatives and called it 'uniformly most powerful' because any other test or combination of tests of that hypothesis would be less powerful in detecting deviations. Failing such a 'uniformly most powerful' test, that provided by the likelihood ratio was a useful and reliable test.

The purpose of the preceding outline was to indicate that, contrary to what the author believed, there was usually one test of a given statistical hypothesis which was distinctly preferable to any other, and that an important preliminary to any statistical test was the specification of the hypothesis to be tested and the permissible alternatives. Unquestionably, there was there a fruitful field of study for the graduate student of actuarial mathematics.

Mr G. A. Brown congratulated the author on a monumental work. Like the author, he had been out of touch with things actuarial for the six years of the war, the end of which found him in Oslo, where he had had the opportunity of meeting some Norwegian colleagues. The then secretary of their association paid British actuaries a tribute when he said 'We always admire British actuaries very much, because you are so practical. We Norwegians are hopelessly academic'. He had replied that that was part of the tradition; Norwegian actuaries were trained in the University of Oslo while British actuaries were trained in the insurance offices, and so there was a difference in their point of view. After being demobilized, he began to peruse the numbers of the *Journal* issued during the war years. But when he had waded through a few, he began to wonder what his Norwegian friends would think when they received them, because it seemed to him that since the beginning of the war there had been quite a trend away from the practical to the academic.

The author had been rather clever, because he had evidently set out to suit both tastes. There was enough of the academic in the paper to satisfy the most devout 'x-chaser', and there was enough of the practical for the rest of them—he made no apologies for including himself among the latter group. From his own point of view, the significance of the paper was in the fresh hope it gave for the Makeham graduation, with its very considerable advantages.

One interesting subject on which the author touched was in paragraph 97—on the limitations of the data in the exposed to risk. That started a train of thought which was worthy of much more serious further consideration.

Mr M. D. W. Elphinstone referred to the paper which he had read before the Faculty a week earlier. At first sight, beyond discussions on roughness and smoothness, the paper under discussion appeared to have little in common with his own paper. There was, however, one other question of major importance with which both had been concerned. He referred particularly to section II, *The Purpose of Graduation*.

The making of a graduation implied some assumption about the nature of the progression from age to age. Without such an assumption the crude rate had to be left alone or, at most, replaced by some function of itself. There were two fundamental considerations: the first was that the assumption ought to be determined (though it often was not) by the purpose of the graduation: the second was that it ought to be possible to state the assumption with mathematical precision when a mathematical method was used. It was a curious fact that, although many mathematical methods of graduation had been devised, it had hitherto been possible in respect of only two of them to say precisely what assumption had been made. One of the methods was curve-fitting; the other was Whittaker's method. In section II, the author had not examined the underlying assumption quite so deeply as he might usefully have done.

When, for example, a Makeham curve was fitted to crude data the assumption could be expressed in either of two ways: it might be said that the true rates followed some Makeham curve with three unknown constants; however, it was better to discuss the equivalent, unique, assumption that μ_x satisfied the difference equation

$$\Delta\mu_x . \Delta^3\mu_x = (\Delta^2\mu_x)^2$$

—that was quite definite with no unknown element. The author's assumption was not that a particular Makeham curve was true, but that the difference equation was true; having postulated the truth of the difference equation, he proceeded to find its most probable solution by an extremely ingenious application of the minimum-χ^2 method. But the object of significance tests was to check the validity of the assumption actually made—namely the truth of the difference equation, not of the actual curve.

If it were supposed that, in fact, no Makeham curve properly represented the true rates, there was a greater chance that the crude rates could be represented by accident by *some* Makeham curve (i.e. by the difference equation) than by a specified Makeham curve. In applying the χ^2 test that greater chance had to be allowed for by reducing the degrees of freedom by the order of the difference equation; the Makeham graduation required a reduction by three. The author should, he thought, have applied the χ^2 test with 46 degrees of freedom, not 49.

The practical importance of the adjustment was small. It would not affect the judgment of the success of the author's graduation because the ·05 significance level for such a judgment was arbitrary, and it was impossible to distinguish between any but the very broadest classes of significance levels.

However, the argument illustrated a general question of the utmost theoretical importance. It was necessary to be absolutely clear about the meaning of the particular mathematical process used before significance tests could be properly applied to a graduation. The tests had then to be designed to show whether the assumptions made in using that process were justified.

In his own paper, the speaker had shown just what was being assumed about the progression of successive rates when a summation formula was used. The assumption was rather different from what was assumed in curve-fitting. Nobody had considered the implications of the difference, but the difference was such that it was illogical to apply significance tests of the traditional type to summation graduations. Even more so was it wrong to apply them to graphic graduations. The traditional tests were only justifiable as being approximations to the suitable tests—and even so, the approximation could not be quantified.

There was a possible objection to the argument. If it were supposed that the author's graduation was proffered without any indication of how it had been obtained, what should be the reply to an enquiry whether the crude A24–29 data could have arisen from the universe characterized by that graduation? Most people would use the χ^2 test with 49 degrees of freedom. If that were so, it was equally correct to apply the χ^2 test to a summation graduation, after allowing for any obvious constraints. Indeed, that procedure would, he thought, have the general support of those who were accustomed to modern statistical theory. But modern statistical theory also demanded the use of 46 degrees of freedom if it were known how the universe was formulated. He thought that the difficulty could be resolved by going back to his original argument—that the purpose must be formulated with precision. The objection would then be seen to be a 'catch' question. No answer could be given until it had been discovered how the universe was defined, and the discovery led to the use of the χ^2 test with 46 degrees of freedom for the Makeham graduation, and to some yet unknown test for the summation graduation.

If the use of 49 degrees of freedom could be justified for the whole table, then he would concede that, in applying the test over a section of, say, 10 ages, the author was correct in using 10 degrees of freedom. But if 46 degrees of freedom were appropriate for the whole table, then the 10 should be scaled down slightly. It might be that the correct distribution was not one of the usual distributions of χ^2 at all, but something rather different. He did not know; the necessary analysis had not been carried out.

A χ^2 sectional test need not be applied to consecutive groups. It could be applied, for example, to those ages which were prime numbers or, with perhaps more sense, to every even age. Failure to pass the χ^2 test in even ages might suggest that there was inherent in the data a wave of period 2, and it should be considered on general grounds, not in the light of conventional sampling theory, whether the wave of period 2 should be retained. Such a situation might arise in graduating a census table. Were every possible χ^2 sectional test to be considered, the graduation would be expected to fail to pass a proportion of them. It would be of real interest to know whether those sectional tests which the graduation failed to pass represented anything of practical importance in the light of the purpose of the graduation.

Mr Seal's contribution to the discussion emphasized that different significance tests might overlap each other, and he doubted whether Tests 2 and 3 contained much information which could not be derived from the χ^2 test. For example, out of a hundred miscellaneous graduations it might be found that, say, twenty passed Test 2, and twenty the χ^2 test. Would there be any degree of association between the graduations which passed the two tests? He considered that there would be; perhaps fifteen would pass both tests. Different tests should be independent, and reveal different features of the experience. Alternatively, it should be recognized that the tests were dependent and that the graduations might to some extent be expected to pass or fail them together. He doubted whether, if the χ^2 test was to be used, there was any point in putting forward Tests 2 and 3.

Finally, the χ^2 test depended on the squares of deviations and could not therefore test deviations according to sign.

He might summarize his arguments under three heads:

1. Work such as the author's was invaluable: the theory of today might well become the practice of tomorrow, but before the tests he described could come into habitual use much more had to be known about the logic underlying them and their associations and differences.

2. Meanwhile, before a graduation was tested it should be made clear what was being attempted. One thing was being attempted in a minimum-χ^2 Makeham graduation, something else in a forecast table, and yet something different in the use of summation formulae: nobody had shown what was being attempted in graphic graduation or osculatory interpolation.

3. So long as graduations were made with vague ideas about what was being attempted, the tests, in his opinion, need not be too intensive nor formidable in application.

The last time he had occasion to make a graduation he did the same as everyone else—he took a pencil and drew a line between some points, and that was all. There was nothing wrong in that; it was sufficient for the purposes for which he was making the particular graduation. Had a test been necessary, one of the crudest and simplest sort would have served.

Mr F. H. Spratling was especially interested in paragraph 6 of the paper where the reasons for graduation were summarized. He thought that that paragraph might not unfairly be paraphrased by saying that the principal purpose of graduation was the introduction of smoothness so that financial anomalies might be avoided.

It could be said that gross premium rates were net premium rates distorted by loading. In the same way, valuation reserves were theoretical values distorted by assumptions made for practical convenience in grouping and also by the laudable desire, in the circumstances of commercial life assurance, to be on the safe side. Those considerations supported the view which the author expressed.

But it was important to remember that comparisons of mortality and sickness rates, and resignation and withdrawal rates had sometimes to be made for administrative purposes into which financial considerations either did not enter or entered only indirectly. He could think of no better way in which the actuary could destroy his credit than by saying that a particular difference between two comparable sets of rates was not a difference revealed by the data but a difference which the actuary thought ought to have been revealed by the data if it had been more extensive and if the imperfections of an imperfect world could have been ruled out. For those reasons, he personally preferred to work from ungraduated data wherever it was possible and reasonable to do so, provided, of course, financial considerations were not directly involved.

It was known that mortality was influenced by age, sex, marital status, geographical distribution of the population, occupation, class selection, and possibly other factors as well. He suggested, therefore, that no mathematical expression of mortality could be adequate unless it reflected each of those various influences as an independent variable and that that approach would carry them into much more elaborate mathematical analyses than had yet been attempted.

He came back to the argument, therefore, that the primary purpose of graduation was, quite simply, financial convenience. When a set of mortality rates had been graduated, considerations of financial prudence and practical convenience came into play. The finished product in the form of gross premium rates or valuation reserves, or whatever it might be, reflected the general level of the experience and also the imported smoothness. In those circumstances, might not a reactionary plea be entered in defence of the ancient and, he suspected, still most frequently used method of graduation, namely, graphical graduation? A curve was drawn through the irregular ungraduated data plotted on squared paper. The shape of the curve had, of course, to be restrained by tests of fidelity to data—the simpler the better provided that they were reasonably adequate—and also by the actuary's professional sense of fitness for the purpose to be served.

Mr R. H. Daw wondered whether statistical tests applied to the rate of mortality really tested what the actuary wanted to test. In perhaps nine graduations out of ten he was interested in producing premium rates, policy values and the like to use in his business activities; i.e., he was interested in the financial effects of using the mortality table. Instead, however, of investigating the difference between the financial effects of the graduated and ungraduated tables, he investigated whether the deviations between the actual and expected deaths were statistically significant. The procedure might indicate whether the graduation was acceptable from the financial point of view, but personally he saw little reason why it should always do so or why it should be the best method to use. Might not a comparison between graduated and ungraduated annuity values at a suitable rate of interest be a test more in conformity with the reasons for making the graduation? If that idea was accepted, it seemed that the process of graduation should be applied to the annuity value.

With that in mind, he had drawn a graph of the whole-life annuity values on the A 1924–29 ultimate table at 3 % interest. The shape of the curve appeared to be that of a logistic function, but for ages between 35 and 75 it differed only slightly from a straight line, and in fact looked quite a promising curve for fitting by a mathematical function. He thought that some experiments on those lines might be of interest.

In perhaps one out of ten graduations where the actuary was interested in comparing rates of mortality or in the effect of certain types of selection (e.g. by type of policy), the statistical graduation tests considered in the paper had their legitimate and proper application. Unfortunately, many mortality experiences, and in particular the Continuous Mortality Investigation, contained duplicate policies, and that fact alone rendered many of the tests described in the paper inapplicable without some adjustment. That would not be very important if the size of the adjustment was known, but the proportion and distribution of the duplicates was unknown, and only an arbitrary adjustment could be made. The inclusion of duplicates had been a serious hindrance to research, as could be seen from a study of the papers by L. Solomon and S. Vajda in the *Journal*, to name only two. Appendix 3 of the paper was yet another instance. He wished to reinforce the plea in paragraph 18 for the exclusion of duplicates from future experiences, or alternatively for some course of action which would enable exact statistical tests to be employed. Until that was done, much serious statistical work on the great body of data collected by the Life Offices was hardly possible.

He gave a warning about test 9—the χ^2 sectional test used to check the detailed agreement of portions of the curve. The test was described as being applied to suspect sections of the graduation, but if that were done at the 5 % significance level the actual chance of finding a significant deviation on the null hypothesis would be greater than 5 %, and might be considerably greater. The reason was that the section over which the test was applied had been chosen because it appeared to show deviations which were either too large or too small (i.e. chosen *because* it appeared to be abnormal) and thus the chance of finding significant deviation was increased. The uncertainty about the significance level of the test limited its use, and he doubted whether it would bring to light any feature not shown by other tests.

Mr N. L. Johnson made, first of all, a few technical remarks on Test 12—the sign-change test for accumulated deviations. While agreeing with the author that in general, as he said, 'the expected number of sign-changes represents a difficult problem' it was possible to throw some light on the problem by theoretical investigation. On certain simplifying assumptions, that the deviations were independent, Normal, and of constant standard deviation, the expected number of sign changes was

$$(2\pi)^{-1}\sum_{j=1}^{n-1}\cos^{-1}\left(\frac{j-1}{j+1}\right)$$

in the unadjusted case and

$$(2\pi)^{-1}\sum_{j=1}^{n-2}\cos^{-1}\left(\frac{j-1-j\,(j+1)/n}{j+1-j\,(j+1)/n}\right)$$

in the adjusted case. For $n=40$ the expressions gave, on the unadjusted basis, 3·35 and on the adjusted basis 4·95. For $n=50$ the figures were 3·82 for the unadjusted and 5·70 for the adjusted basis. Those figures were very considerably lower than those given in the paper; the author's figures, both on the adjusted basis, would be 11·4 and 14·4.

It was true that, in the situation which would usually arise, on going through the series of data the standard deviations of each deviate of actual from expected would increase to a maximum and then fall away. The effect would be to increase the expected numbers of sign changes, but he thought that the increase would be something of the order of 1 at the most, and would not bring the figures up to 11·4 and 14·4 which the author had given. The standard deviation of the number of sign changes was much more difficult to evaluate, even in the simplified case; but theory plus a controlled guess gave, for $n=50$, estimates of 3·1 and 3·0 as standard deviations on the unadjusted and adjusted bases respectively.

The standard deviation was fairly large, considering what the average number was, and that was to be expected, because the distribution of sign changes was probably very skew. It was possible to get sign changes as big as 45, 46 or 47, though not very often, while the mean was 5·7. Allowing for that, he suggested a lower limit of 2, or perhaps 3, and an upper limit of 11, or perhaps 12, for the adjusted basis when $n=50$, in place of the limits of 8 and 21 given in the paper in Table 3. It would be noticed that the observed values obtained in Table 6, of 8 sign changes where $n=49$, and of 6 for $n=44$ in Table 11, would be well within the limits based on those calculations, whereas they were at the lower extremities of the limits given by the author. It could be justly said that a very low number of sign changes would be required to detect departures from graduations if the test were used rigorously, because the test was not very sensitive.

On a more general question, a graduation by minimum-χ^2, such as the author had carried out, always called forth very great admiration; but, since the theoretical justification for using the minimum-χ^2 method for the author's graduation was that it might be an approximation to the maximum-likelihood method and *might* be easier to work out than that method, and since it did not seem probable that the maximum-likelihood method would be any worse to apply than the χ^2 method in the particular case, he thought that the maximum-likelihood method might as well be used. The formula for the function to be maximized was expressed simply in terms of expected and actual deaths, and the actual computation would not, he thought, be any more complicated.

He agreed with the author, and disagreed with Mr Seal, on the question whether a single test or a number of tests should be applied. It was only under artificially restricted conditions that one unique test was likely to be available and that no others were likely to be suitable. Whilst what looked to be the best test could often be found for a particular class of alternative hypotheses, it was also true that a better test could probably be found for a specified sub-class of those hypotheses. Why should a test which was pretty good with respect to a whole wide range of alternative hypotheses be preferred to a much better test for a particular sub-set of alternatives—unless a quick single test was required. Ideally, as many tests should be carried out as ingenuity, patience and time permitted. Some tests might be so time-consuming, or be useful for such rare forms of deviation from the graduation, that they might be dispensed with; also, a chance effect should not be elevated to the status of significance because of an excessive proliferation of tests; but the investigation of a graduation, or of any statistical problem, from as many aspects as possible could, in general, only increase the actuary's appreciation of the properties of the graduation or the problem.

Mr R. E. Beard referred to the use of the χ^2 test for mortality data. He was concerned at the central position into which that test was being jockeyed. The essential point of his criticism lay in the very nature of mortality data. Generally speaking, the exposed to risk in the various groups from which the contributions to χ^2 were calculated would vary in magnitude over the range of the graduation, and it was pertinent to ask whether it was proper in those conditions that equal weight should be given to the various contributions to χ^2.

A similar remark was made by G. F. Hardy in his lectures, when discussing least-square methods, and quite recently the problem had been considered by Patnaik in *Biometrika*. It was ironical to find that support was being given by actuaries to the χ^2 test at a time when there was a swing of opinion amongst theoretical statisticians to the position which actuaries occupied some forty years ago.

Unfortunately, the author had left the χ^2 test in a somewhat elevated position, and other speakers supported that view. Clearly, there were circumstances in which it was the appropriate test, but for mortality data it seemed to him that, if χ^2 were calculated without weighting, then all the information available was not being used, and thus, in the language of theoretical statistics, it was not an efficient test.

Once it was recognized that the χ^2 test was merely a special case of a weighted-χ^2 test in which the weights were equal, Perks's remarks in the discussion on Daw's paper on the mean-deviation test assumed an even greater significance, and drew attention to four distinct summary tests of a graduation, namely the weighted and unweighted mean-deviation test and the weighted and unweighted-χ^2 test. Each measured different characteristics of a graduation. In that context it was of interest that Cramér, when discussing the χ^2 test in *Mathematical Methods of Statistics*, suggested that other tests should also be applied.

He had submitted a paper for discussion at a later sessional meeting of the Institute in which the application of such tests to a number of graduations was considered. In particular, he had dealt with the problems of the sampling distribution and of constraints. The problem of constraints was, to his mind, important. Mr Elphinstone might be happy to ignore the uncertainty of 2 or 3 degrees of freedom when dealing with 30 or 40 groups, but in practical work he liked to group the data—it made the work so much less. When that was done, and by the time the data had been weighted for the application of a weighted test, it would be found that the equivalent number of separate groups was only 9 or 10. If allowance was then made for constraints, the number of degrees of freedom might be reduced to 5, and it was then necessary to know how many constraints were imposed by the particular graduation process. It was thus important to consider the subject theoretically as well as practically. It was not sufficient to say 'That is theory; it does not help us'. As Mr Elphinstone had already remarked, the theory of today might well become the practice of tomorrow.

He congratulated the author on his minimum-χ^2 fit of a Makeham curve; it was something which he had been thinking of for twenty years. He was glad that the author had done it because it showed those who were interested in graduation what it would look like.

When, as with other physical problems, a complicated solution had to be found, it often paid to go back and to reconsider the mathematical formulation of the problem. If the emphasis were moved from the χ^2 test to the mean-deviation test, it would be possible to devise an alternative criterion for fitting in which the sum of the deviations without regard to sign would be minimized; the arithmetic was relatively simple. The A and B constants were calculated using a few suitably chosen values of c and a simple interpolation led rapidly to values for which the sum of the deviations without regard to sign was a minimum. Using that method on the select data in the appendix, and working in quinquennial groups, he had, in under two hours, obtained a χ^2 which was only 1 more than that obtained by the author, so that there was no significant difference between the results, but an appreciable one in the times taken to fit the data.

The flexible method described in Appendix 3 was merely a systematic application of the technique which would be used in the preliminary stages of a graduation, and he thought that it would have been better to adjust the data by deduction of 1/24th of the second central differences of the Exposed to Risk and Deaths before calculating the values of μ_x. Had that been done, and the value at age 28 ignored as being out of line, the resulting values of the constants would have been found to be much closer to the values finally adopted.

The author had not emphasized that he had restricted himself to a limited range of ages in the application of the Makeham formula to the A1924–29 data. The Makeham formula would not fit over about age 65, and therefore the author's statements about the use of a Makeham curve required some qualification. When the 1924–29 statistics were published, he (the speaker) had made a graduation of the ultimate data by Perks's modification of the Makeham curve. On referring to his file of papers it was interesting to find that, taking the same range as the author's Table 11 and using a duplicates factor of 1·4, the χ^2 determined from the Perks graduation was 45·97, as compared with the author's value of 46·57. The fit by the Perks curve was satisfactory up to the highest ages, whereas the Makeham graduation would depart from the data as soon as the range covered in the paper was left. The extra parameter in the curve produced a reasonable fit over the whole range. It should, however, be mentioned that neither graduation would be described as the best possible.

Mr W. Perks thought that Mr Beard had rightly thrown serious doubts on whether the χ^2 test was in fact the best single measure test so far as mortality was concerned. To take an extreme case, if there were a million exposed to risk at each of the even ages, and a thousand exposed to risk at each of the odd ages, would anybody in their senses let the odd ages contribute in the same degree as the even ages to the χ^2 test? If any theory required that to be done, the theory was inadequate. It needed to be remembered that the problem was that of a set of binomial distributions—not, usually, a single *multinomial distribution.*

Professor Jeffreys had pointed out that the χ^2 test asked all the questions at once and provided only one answer. It mixed up the questions according to its own recipe, and claimed that the proof of the pudding was in the cooking. He was quite sure that the author was right in wanting to test his results from many different points of view. The questions should be asked one at a time, even if, as Mr Seal said, and he thought rightly, they were not all independent of one another. Perhaps, however, he was a little peculiar, because he did not find any difficulty in thinking of more than one number at once; he had got beyond the baby class, when everything had to be brought down to a one-dimensional measure. When all the testing had been done, an educated judgment had to be made whether the graduation was useful for the purpose in hand, whether theoretical or practical.

Mr Seal and Mr Daw had criticized the author's test of suspect groups. Personally, he thought that they were asking a different question from the one put by the author—he did not include in the data of his probability the fact that all the deviations in the group were known to be of the same sign. Mr Seal wanted to include that knowledge in the data of his probability, i.e. in the H of the probability Pr. $\{X \mid H\}$. The inclusion of that fact in H completely stultified the test. The author was asking the question 'How reasonable is it to expect a group deviation of this size?'—a legitimate and sensible question.

With regard to fitting processes, he, like Mr Beard, was a devoted follower of G. F. Hardy, and he thought that much of the modern work in that field was just hair-splitting when it came to practical work with large quantities of data. As a theoretical example of the strict application of the minimum-χ^2 method the author's work was valuable for educational purposes, but it was important, he thought, to realize that the method did not increase the statistical efficiency, even from the χ^2 point of view, by more than the equivalent of a fraction of one, or at the most two, degrees of freedom over the whole range of ages.

He thought it was perfectly clear that for minimum-χ^2 the total of the standardized deviations had to be equal to zero. He understood that the author had tried that suggestion on his graduation and had found that that was so.

Mr Seal referred to a circular transformation of q_x. It might be remembered that something of that kind appeared in his own probability paper. He agreed with Mr Seal that the transformation made the standardizing factors independent of the unknown parameter q_x, and should therefore ease the work of minimizing χ^2 or of minimizing the total of the standardized deviations taken without regard to sign. However, the transformed variable z included a multiplier, $\sqrt{(E+\frac{1}{2})}$. To his mind, that made z completely hopeless as a function for graduation, because it was irregular.

In the discussion on Daw's paper he had questioned the propriety of adjusting for constraints imposed by the fitting process. He still remained of the opinion that in mortality graduations the adjustment ought not to be made. The graduated curve should be taken as characterizing the hypothetical universe without regard to how it had been formulated, and the test should be whether the data could reasonably have been obtained by random sampling from such a universe. An adjustment for the constraints of fitting seemed to him to make too much concession to under-graduation.

It seemed desirable to distinguish between constraints on the data (such as $\Sigma m_i = N$ in a multinomial distribution) and constraints imposed on the hypothesis in fitting. There were occasions and repetition processes calling for proper allowance for constraints on the hypothesis but mortality graduation was not usually one of them.

Mr H. Tetley, in closing the discussion, welcomed the paper as a good piece of work and an excellent contribution to the proceedings of the Institute. He did so for several reasons, among which were the following. In the first place, it had made him, and probably a good many others, think carefully about some of their fundamental ideas, about what they were trying to do when they graduated data, and therefore what they were trying to do when they tested the results. Moreover, he thought that it was an excellent counter-balance to Seal's paper, which was deservedly famous, but which did put one point of view—very cogently and forcefully. The present paper put an alternative point of view. In the third place, he thought it was quite useful to break away from the tyranny of making the first and second summations of the deviations equal to zero. He admitted that the equality had in fact been honoured more in the breach than in the observance, but the author had at least given a more precise idea of what was meant by saying that the sum of the deviations should be 'reasonably small'.

In the rest of his remarks he would be dealing with what might be regarded as details. He did not apologize for that; he agreed so much with the author on the general outline of the paper that it was only on some of the details that he found anything to criticize.

When so much was given in the paper it was ungracious to ask for more, but he was surprised that no mention was made of a test devised by David in *A χ^2 Smooth Test for Goodness of Fit* (*Biometrika*, 34, 297). That was an attempt—and, as far as he could judge, a very successful attempt—to modify the χ^2 test to make allowance for the changes and runs of signs, a piece of information which was completely lost sight of in the ordinary χ^2 test. He made no apology for suggesting an addition to the list of tests, because, as he saw it, few people would ever be likely to apply all the 13 or 14 tests to any one graduation. From his point of view, he would regard them all as weapons which could be extremely useful in appropriate conditions, and he would make a suitable selection of two or three which he thought would be most helpful in testing a particular graduation. From that point of view, the fact that there was a good deal of overlapping of the tests was largely immaterial.

In paragraph 52, the author said that as far as he was aware no constraints had been imposed. On that question he, the speaker, found himself somewhat at variance with Mr Perks. As he saw it, the only occasion on which no constraints were imposed was in the type of problem which arose in practice of which he would give an example. An actuary might be about to value a pension fund and might wish to find out whether a standard table such as Scottish Bankers' Mortality was suitable for that purpose. To do so he would test the data of the fund, using the standard table. In those circumstances there would be no constraints imposed, because the table was not in any sense based on the data; in other circumstances the table would be, to some extent, 'pegged down' to the data. It was usually impossible to find to what extent constraints had been imposed, and it was only in rather artificial circumstances that a definite measure could be given; but he thought that some small deduction—it might even be a large one—should be made from the total number of cells in arriving at the degrees of freedom, or the test was not being applied impartially.

Another question, which was referred to on pp. 26 and 34, was rather more difficult and went rather deeper than the others he had made. The author followed Seal in saying that the χ^2 test, as it was usually applied, gave a warning if a fit was 'too good to be true', and implied that in that sense actuaries were at variance with statisticians. Personally, he believed that that was a mistake, and that the fundamental question was not merely the hypothesis being tested but the alternative hypotheses against which it was being judged. A test of a mortality table was not concerned with fits which were too good to be true. The test was a one-ended test, and there was no difference, as he saw it, between the point of view of actuaries and of statisticians. The question was dealt with, he hoped clearly, in *Statistics: an intermediate text-book*, Vol. II.

In Appendix 1, the author referred to the minimum-χ^2 fit as being in certain circumstances the 'ideal' solution. Again he thought that that was going too far. It was obviously a solution which had great intuitive appeal as being eminently reasonable, and one which should give very good results, but there were others, such as one which had been mentioned—fitting by the method of maximum likelihood—which had equal intuitive appeal and would probably give slightly better results.

At the top of p. 40 reference was made to the contribution to χ^2 from the cell made up of the Exposed to Risk at a particular age or age-group. The formula which was used was not correct for one cell. What actually happened was that at each age or age-group there were two cells, those who died and those who survived. Each of the cells contributed a term to χ^2, though for practical purposes it was easier to amalgamate those two contributions into a single term.

Although he had a tremendous admiration for the doggedness and thoroughness of the author's fitting by minimum-χ^2, he agreed with others that it was unlikely to be widely used. It was likely to be worth while only for a big standard table, and in such a table, particularly for assured lives, there were so many imperfections in the data that it was probably an undue refinement. The number of duplicates of unknown extent and, what was much more serious, of an extent which probably varied systematically with age, and the fact that the data were based to a very large extent on endowment assurances up to age 65, and above age 65 predominantly on whole-life experience, raised difficult problems. One solution would be to split the data and have two complete tables, but a little consideration would show that that would probably raise more problems than it would solve.

The President, Mr F. A. A. Menzler, C.B.E., B.Sc., in proposing a vote of thanks to the author, said that the paper took his mind back to his examination days. He was among those who, while not being mathematicians, were called 'good at maths' at school, and he remembered what an attractive relief it was from the study of life contingencies to go to Part III(*a*) and to study problems of graduation. It seemed to offer some systematic means of putting constraint on disobedient data. At the time, his studies made him look down on the graphical method as the lowest form of actuarial expertise, and he derived much intellectual sustenance from the thought that a suitable mathematical formula would do the work for him, particularly as he had no capacity for freehand drawing.

Not long previously, all the reading on the subject, apart from papers, was the red book of G. F. Hardy's lectures, to which Mr Perks had referred—lectures delivered, incidentally, forty-five years earlier. It was interesting to look back at that book, as he had done recently, and to see what progress had been made since then. He was inclined to agree with what Mr Perks had said on the subject. There was one other book (whose author was present that evening), namely Sir William Elderton's *Frequency Curves and Correlation*, which he tackled when he was safely through the examinations. Since those days, much more was expected of the student, but at least the students had the advantage of excellent text-books written by the closer of the debate and another member of the Institute who had spoken that evening.

Of recent years, there had been a great deal of research into the testing of graduations, and the paper would be of service to the younger generation of actuaries in particular, but not excluding the vocal group of older post-graduate actuaries, in bringing together the studies of the last decade since Seal's paper of 1940. He felt, however, that the general practitioners of the actuarial profession, who worked out premiums and did valuations, should have a word of comfort after hearing the discussion that evening, because, after all, it was rare to be concerned with a standard table, with massive data behind it, which demanded the full apparatus of tests described in the paper. For every major graduation of the type which had been discussed there must be thousands of graduations done in the offices of consultants in the course of the valuation of pension funds—when all too frequently there were no true underlying rates at all, and if there were it would be hazardous to use them. In fairness to the author, it should be mentioned that he had stated clearly, in paragraph 12, that what the actuary needed was a good forecast.

He had not to sum up the debate; that had already been done more ably than he could do it, but he rather supported Mr Spratling's remarks, though he would like to make a friendly debating point—what would he do about the preference for certain digits exhibited in the age-distribution of the census population? He noticed with interest that Mr Elphinstone in the end came round to the graphical method of graduation, which was very comforting to the older generation. Personally, as one with some experience of pension funds, he would have liked an appendix on salary scales, but perhaps that was asking too much. It was of importance to reconsider from time to time the fundamental ideas underlying graduation, which they were so likely to take for granted. In helping them to do that, the author had rendered them a great service.

Mr H. A. R. Barnett, in reply, thanked Mr Tetley for accompanying him so far in the paper, but suggested that there was bound to be some correlation between their views, since Mr Tetley had been his tutor in the subject.

In the discussion of Section II of the paper, both the opener and Mr Spratling seemed to him to be too retrospective in their outlook, and he thought that the President agreed with him. Should it be desired to find what table would have been best to use over the period of the experience, it would be as well to have equality between the total actual and total expected deaths; but, since the experience would never be repeated exactly, it was harking back too much to the past to insist on equality. Further, if the opener was to be completely guided by financial considerations, he should weight his experience by the amounts at risk. When the first and second summations of the standardized deviations were made equal to zero, the result was almost exactly the minimum-χ^2 fit which he had produced.

On Section III, there might be a lot in what Mr Daw said about comparing values other than θ and nq. It was not possible to generalize, and it might be a different problem every time a graduation was made, but he thought that q would still be the basic function. He was glad that Mr Daw had mentioned duplicates and has supported his plea for their exclusion. The data for assured lives were produced in a form which actuaries largely shaped for themselves. Sometimes, they had to do their best with data over which they had no control, but when they shaped the data themselves they could surely so arrange that the problem of duplicates did not arise.

Mr Elphinstone, in his paper to the Faculty, put forward the interesting conception that smoothness was 'absence of the positive quality, roughness', which might well point the way to a possible statistical basis for the smoothness test, complete with significance limits. Just as the sizes of deviations were examined to determine whether they amounted to non-adherence, could not a test be devised whereby the sizes of the roughnesses would be examined to determine whether they amounted to non-smoothness? While on that subject, he would ask whether Mr Seal's all-embracing test also included a test for smoothness.

In 1940, Mr Seal wanted two tests; in 1950, he reduced them to one; and by extrapolation, by 1960 he would be relying on intuition—effectively on Tests 2 and 3 of the paper, because probably nine times out of ten the intuitive test of a graduation would be subconsciously guided by Tests 2 and 3. When Mr Seal and Mr Elphinstone tested a graduation, did they not have a surreptitious look to see whether one or two of the deviations were or were not in excess of twice their standard error? The tests might be rough and ready; he had tried to make them less rough; it was because they were ready to hand that they always had been, and always would be, applied.

He agreed with Mr Seal that the items examined by Test 4 had a variance differing from the normal. He had avoided reference to the question lest he should appear to overdo the justification of his own graduations. He was grateful to Mr Perks for showing that the test could do two different things. One was to see whether a particular group of deviations could have arisen, and the other was what he had done in paragraph 39, where the test gave a summary of all the different groups that could be tested that way.

A good deal had been said about Test 9, but, like so many other tests, it was merely a means of indicating distortion. No one or two or even three tests should be taken as an absolute test for a graduation; it was desirable to get all the different answers to as many questions as possible. In that connexion he quoted from Jeffreys's *Theory of Probability*, p. 91:

The trouble is that with regard to a large number of data we may want to ask several questions. To some of them the answer will be 'yes', to others 'no'. But if we try to sum up all the information in one number we shall not know what question we have answered. It is desirable to arrange the work, when several questions arise simultaneously, so as to provide answers to each of them separately. When this is done it is still found that the χ^2 form persists, but it is now broken up into separate parts each of which has its own message.

With regard to the number of constraints, he admitted that the χ^2 test should be applied in the theoretically correct way if the graduation were by a Makeham curve. But there were dangers if more complicated formulae were used and Kendall's *Advanced Theory* was taken too literally; for example, if the select data used in the paper were fitted by a curve with 40 constants, would it be correct to say that there were only 9 degrees of freedom, and then to find from the χ^2 test that the graduation was satisfactory? He would say certainly not. Such a curve could be fitted passing exactly through 40 particular values, and without applying any test it could be seen that it would be an under-graduation; yet the χ^2 test as recommended by Professor Kendall would accept the graduation if the remaining 9 values displayed reasonable deviations.

As he saw it, it was necessary to apply the χ^2 test in two ways; first, it should be applied without reducing the number of degrees of freedom to test whether the type of curve was appropriate; it could then be applied, in the theoretical way, after reduction of the number of degrees of freedom, to test for over-graduation.

He had completed the minimum-χ^2 fit for the ultimate data (see Table 12); it made very little difference, except that the test of signs of the adjusted accumulated deviations was now satisfactory. He was quite prepared to believe that, as Mr Johnson had stated, the probability underlying the test was less than ·3. He had avoided putting it too low, again for fear of accepting his own graduation unjustifiably.

Mr K. J. Burton submitted the following written contribution.

Whittaker and Robinson state in *Calculus of Observations*, 4th ed. p. 303:

> ...we must remember that the problem of graduation belongs essentially to the mathematical theory of probability; we have the given observations, and they would constitute the 'most probable' values of u for the corresponding values of the argument, were it not that we have *a priori* grounds for believing that the true values of u form a smooth sequence, the irregularities being due to accidental causes which it is desirable to eliminate. The problem is to combine all the materials of judgment—the observed values and the *a priori* considerations—in order to obtain the 'most probable' values of u.

In a footnote they draw attention to the remarks of George King in the discussion on Sprague's well-known paper on graduation by the graphical method (*J.I.A.* XXVI, 114):

> ...what was the real object of graduation? Probably the reply would be, To get a smooth curve; but he did not think that quite correct. To his mind, the reply should be, To get the most probable deaths.

The author of the paper under discussion makes clear in paragraphs 6 and 7 what he considers to be the objects of graduation, and I think it is well that these practical considerations should be emphasized. There is no important difference between what the author says and the remarks which I have just quoted. In making a probability judgment we should take account of all the evidence and, as Whittaker and Robinson indicate, part of the evidence is constituted by our own knowledge of other mortality experiences.

There is a tendency at the present time amongst statisticians to over-elaborate techniques without sufficient consideration whether the material to which the techniques are to be applied will bear the assumptions on which the theory rests. Professor Kendall, in his recent inaugural lecture, *The Statistical Approach* (*Economica*, May 1950), warned his audience that 'we have reached a critical phase in the development of statistical method when pure mathematicians looking for something to research on are throwing up such a dust that the practical nature of the subject is being obscured and theoretical statistics is in danger of being discredited in the eyes of the more practical man'. As practical men, actuaries will no doubt be grateful to Professor Kendall for what he has said. Heterogeneity imposes practical limitations on the extent to which it is possible to analyse mortality data.

In particular, I am not at all sure that it is legitimate to apply the χ^2 test to most of the graduations with which we have to deal. As I understand its application for this purpose, the assumption is tacitly made that the 'actual deaths' at successive ages are themselves uncorrelated. There is no reason to expect that fiddling with the number of degrees of freedom will make proper allowance for such correlation, if it is present. Since most mortality tables are constructed on the basis of an experience stretching over several years the exposed to risk at age $x+1$ includes persons who were exposed to risk at age x. The deaths which occur amongst the exposed to risk at age x affect the constitution of the group of persons exposed at age $x+1$. Any aberrations in the experience of these persons at age x, which are of a selective character, will be reflected in the constitution of the exposed to risk at age $x+1$. For example, an epidemic may operate to remove the weaker members of the group, or a war may operate to remove the stronger. Older exponents of graduation often tended to regard it as a process of redistributing the actual deaths and I think this idea played some part in the assumption —which the author criticizes—that the expected deaths must be made equal to the actual deaths.

I am not clear what the author has in mind in paragraph 57 where he states that, when the χ^2 test gives a 'too good to be true' result for a graduation by a mathematical formula, the result may indicate that the formula has too many parameters. If the right number of degrees of freedom has been employed, the χ^2 test can never give an indication that the number of parameters is too great. That is a matter purely for the judgment of the person who is graduating the table.

In paragraph 97 the author draws attention to the somewhat specious practice of retaining too many places of decimals in the expected deaths, and he goes on to question the number of places that should be used in the estimation of the Makeham constants. For this latter purpose I think the only criterion that needs to be applied is how many significant figures it is desired to retain in the commutation columns eventually calculated. I do not think the limitations of the crude data condition the number of decimal places in the Makeham constants, which are supposed, of course, to relate to a hypothetical universe.

Mr H. A. R. Barnett has subsequently written:

Mr Elphinstone complains that the section on *The purpose of graduation* does not go far enough. I am bound to agree with him but, as the paper is primarily on tests and has already multiplied the allotted span of pages nearly threefold, it is impossible for it to go further. For that reason I am grateful to Mr Polman for having described in very clear terms the practical purposes of graduation. I agree that my definition of smoothness would probably not be applicable to a table of remarriage rates, and this is another example of the type of event I had in mind in paragraph 24.

I agree that the fitting of Makeham to the A 1924–29 data is open to objection, but so, apparently, are any graduations of this unruly experience. I cannot, however, agree that column (13) of Table 6 proves that the graduated values are too low at the ends and too high in the middle; the deviations themselves (column (5)) show that this is not so. The trouble with any consideration of the column of accumulated deviations is that, even after adjusting, it perpetuates certain values, and that is why I feel that it can do no more than give the merest hint of distortion; I think, also, that Mr Polman's criticism takes no account of the fact that the largest figure in column (13) is 20·41. Similar remarks apply to column (13) of Table 11, where, with a much larger exposed to risk, the greatest figure is 224·1. Column (8) of that table certainly shows that some waves have been severely cut, but that may be a desirable feature of the graduation of the 'ultimate' data. Consider the age-group 40 to 45; the average date of exposure would be about 12 years after the outbreak of the 1914 war, and the persons exposed to risk in this group would have reached the most popular ages for life assurance before 1914; by 1924–29, they would have become a somewhat impaired group by reason of the war, and would therefore be expected to experience actual deaths higher than those expected in a hypothetical universe free from the effects of war. Now consider the group 32 to 37; many of these persons would not have reached the usual ages for assurance by 1914; the war would have affected lives of these ages as much as those a little older, but they would have been impaired first and selected by the assurance companies later. In other words, the age-group 32–37 is a more select group, and it is therefore reasonable that the hypothetical universe for 'ultimate' lives should show higher rates of mortality. Perhaps it should be interpolated here that, especially so far as effects of war are concerned, selection never really wears off. The only other group to comment on is 22–25, which is probably affected by two influences; the first is the presence of childrens' deferred assurances after the vesting date, already referred to in paragraph 135, and the second is the fact that rates of mortality tend to decrease to a minimum at age 25. However, at these ages mortality is relatively unimportant, and it might be inconvenient to start off with decreasing rates.

If it be a fact that the graduated curve is not sufficiently steep for use for whole-life reserves, that is a criticism of the use of the same tables for whole-life and endowment assurances—a gnat which has surely been strained at often enough. In many offices the whole-life business is now a comparatively small section, and a combined table fitting all types of business up to age 65 might well be adequate. When, however, the whole-life business is substantial, the steepness of the curve must not be underestimated. I therefore suggest that, for the next standard tables, instead of having the choice of normal, heavy, and light as for the A1924–29 tables, the three published standards should be normal, steep and flat.

The sum of a run of deviations of the same sign has a variance differing from the normal, but Mr Seal is incorrect in saying that the variance is larger; the use of the word 'variance' implies that the values are measured from the mean, and it can be shown that the variance of the run (as measured from its mean value) is smaller than the normal in the ratio $(\pi-2):\pi$. What Mr Seal undoubtedly means is that the expectation of the square of the run deviation as measured from zero is greater than the sum of the individual variances.

The following approximation would probably suffice for the Test 4 distribution. Consider first a run of two deviations of the same sign. If σ_θ is approximately constant for the two individual ages, the mean run deviation will be $1{\cdot}6\sigma_\theta$. The 'run variance' will be approximately

$$2\sigma_\theta^2\,(\pi-2)/\pi \simeq {\cdot}7268\sigma_\theta^2,$$

whence the 'run standard error' is approximately ${\cdot}85\sigma_\theta$. The testing of a run of two by reference to twice its standard error would appear, then, to be applicable if the size of the run is compared with $3{\cdot}3$ times the individual σ_θ. Similarly, for a run of three the criterion would be

$$2{\cdot}4+2\sqrt{\{3\,(\pi-2)/\pi\}}=2{\cdot}4+2{\cdot}1=4{\cdot}5$$

times the individual σ_θ; for a run of four, $3{\cdot}2+2{\cdot}4=5{\cdot}6$; for a run of five, $4{\cdot}0+2{\cdot}7=6{\cdot}7$; and for a run of six, $4{\cdot}8+2{\cdot}9=7{\cdot}7$. A sufficient approximation seems to be that, if there are t values in the run, the total run should be compared with $1{\cdot}1\,(t+1)$ times the standard error of one value—say the central value—instead of twice the root of the sum of the individual variances.

At the same time, I still agree with Mr Perks that if we ask the question 'Is a group deviation of this size too large if we have no particular reason to expect the whole group to be of the same sign?', then Test 4, as I have applied it in the paper, will give the answer 'such and such a group may be suspect'. It can indicate where there *might* be room for improvement, but cannot reject outright.

Most of Mr Seal's criticisms have already been answered in the course of the discussion, but I would say that, even if he were correct that there is usually one test preferable to any other, unless he can say 'always' instead of 'usually', it would not be safe to rely on one test only. What Mr Seal calls the multiplicity of tests is aimed at assisting, but not supplanting, the use of personal judgment.

I am interested in Mr Johnson's remarks on Test 12. I suspected that my estimate of

30 % for the probability was too high and that my diagram of a possible average cycle was too flat. I am not sure whether it would be safe to go as low as his limits, particularly since his estimate of the standard deviation of the number of sign changes depends on a 'controlled guess'. However, what he has said indicates that 15 % would be a closer guess than 30 %, and that if the probability were 20 % the test would still be on the stringent side. The lesson is, of course, that the test itself is somewhat unsatisfactory, and that Mr Polman should not be too dogmatic in drawing his conclusions.

Mr Beard and Mr Burton have cast doubts on the use of the χ^2 test for mortality data. I think it has its use as a summary of all the adherence tests, but it should be regarded as that and nothing more. Mr Perks's hypothetical example, of an experience containing exposed to risk of alternately one thousand and one million at successive ages, certainly points to the danger of standardizing. The minimum-χ^2 only gives a 'best' fit from one particular point of view. I agree that 'ideal' is too strong a word. Mr Perks, in this discussion, has said 'why standardize?', and in the discussion on Daw's paper 'why square?' and I quite agree that a curve fit giving the minimum value to the total deviations disregarding sign would be as ideal as any. Unfortunately, although good approximate methods have been devised for this sort of fit, so far as I am aware the difficulties of a true minimum method have not yet been overcome. I believe they can be overcome by a method similar to that described in Appendix 2, but with simpler differential coefficients. If so, it would be more satisfactory—and I hope not much slower—to find the true minimum than to choose a few values of c and to interpolate.

I am interested in Mr Beard's remark to the effect that if, in my illustration of the flexible method, I had deducted the second central difference term of Hardy's formula, I should have saved a certain amount of hand-polishing. This is a good lesson for students (and others) to learn.

There is little difference between my own and Mr Beard's points of view with regard to the best type of curve to fit the A 1924–29 ultimate data. Perks's curve is the simplest hypothesis over the whole range of the data, and I should have liked to see such a curve as the basis of the standard table. The simplest hypothesis over the range 22–65 is Makeham, and I should also have liked to see this curve published. These alternative tables would have been useful in view of their differing gradients.

I am grateful to Mr Tetley for supporting my suggestion that the first and second summations of the deviations need not be zero. Surely, if a curve has been fitted making the total deviations irrespective of sign approximately a minimum, it does not matter whether the positives and negatives exactly balance, provided that the adherence tests are satisfied.

The possibility, mentioned by Mr Burton, of small inverse correlation at adjacent ages was fully dealt with in section 7 of Daw's paper, where he concludes that such correlation would be swamped by the random errors. Epidemics are more or less recurrent and should be reflected in the mortality curve itself. Wars are a different problem; they are not as frequent as epidemics but their effects are felt for longer. Since there is a tendency for war risks to be regarded as 'extra', I feel inclined to suggest a further assumption (c) to those stated in paragraph 8, namely that the proportionate frequency will be that of a sample drawn from a universe which differs from the universe from which the observed sample was drawn only in so far as it excludes special influences (e.g. wars). My graduation of the A 1924–29 data might be regarded as an assumption (c) graduation.

Mr Burton also takes me to task with regard to the number of degrees of freedom for the χ^2 test, whilst aligning himself with those who would not use this test at all. I dealt with this in the discussion.

I cannot agree that the choice of a curve with a large number of parameters is a matter for personal judgment; this would be an easy path to the quicksand of under-graduation. I agree that the suggested 40-constant curve is an extreme case; but how would Mr Burton decide whether a 10-constant curve would be appropriate?

Table 12 shows the constants and tests for the true minimum-χ^2 fit of the data of Appendix 3, allowing for the reduction of 3 in the number of degrees of freedom. Even though the method may perhaps not be used again, I think it is worth mentioning how the operations described on p. 49 can be shortened. To arrive at my second trial I applied formulae (43) to the Appendix 3 fit, but added 20% to all the corrections. This ensured an over-correction and, as every gunner knows, it is better to have a bracket on a target than to creep up to it. The third trial, found by operations 3(*a*) to (*c*), was then sufficiently close for formulae (43) to give the exact corrections required, in other words it was in the 'inner region' where second differentials were, to all intents and

Table 12. A 1924–29, durations 3 and over, ages $21\frac{1}{2}$–$65\frac{1}{2}$. Tests of Makeham minimum-χ^2 fit ($10^3\mathrm{A} = 1{\cdot}8764747$, $10^5\mathrm{B} = 2{\cdot}0629023$, $c = 1{\cdot}1184201685$). (NOTE: $m = 44$, $f = 41$, $\sqrt{f}/\sqrt{m} = {\cdot}9653$, duplicates factor taken as 1·4.)

Test no.	Criterion	Observed	Acceptance limits
2	Number of standardized deviations less than $\cdot6745 \times \cdot9653 = \cdot6511$	16	15–29
3	Number less than $2 \times \cdot9653 = 1\cdot931$	3	0–5
4	Individual groups exceeding $1\cdot93\sigma_\theta$	Three groups just in excess. 40–45 well in excess, but better than shown in Table 11	
	Summary of groups	None exceeds $1\cdot1\,(t+1) \times$ individual σ_θ	
5	Sum of deviations regarding sign	$-\ 44\cdot73$	Approx. ± 658
6	Second sum regarding sign	$-868\cdot6$	Approx. ± 8000
7	Sum of deviations disregarding sign compared with* $\cdot8\Sigma\sqrt{npq}$	1829	Approx. 1627 ± 395
	Average standardized deviation compared with $\cdot8 \times \cdot9653 = \cdot77$	·85	$\cdot77 \pm \cdot18$
8	χ^2	46·35	$P \doteqdot 30\%$
9	Sectional χ^2 (degrees of freedom taken as number of ages $\times\, 41/44$)		
	Ages 41–45	7·84 (4·66 degrees)	$P \doteqdot 20\%$
	Ages 48–52	11·21 (4·66 degrees)	$P \doteqdot 4\frac{1}{2}\%$
	Ages 48–56	15·60 (8·39 degrees)	$P \doteqdot 7\%$
10	Sign-changes of deviations	17	14–29
11	Signs of deviations	21+, 23−	15–29
12	Sign-changes of adjusted accumulated deviations	6	6–19 on basis of 30% (see also Mr Johnson's remarks)
13	Signs of adjusted accumulated deviations	26+, 17−	14–29

* If the sum is compared with $\cdot77\Sigma\sqrt{npq}$ it is found that its size is still acceptable.

purposes, constant. I have confirmed my figures by applying the method, as described, in full. Paragraph 114 is rather misleading because continued use of formulae (43) will eventually give the correct answer (probably after about seven trials); it was only when I tried short cuts that I went round in circles, until I found the described method which, by finding approximately the best A and *c* for a given B, arrives at the 'inner region' by a direct route. I mention this because it may be that a 'minimum deviation' fit can be derived by a similar method, but with simpler functions. I should also mention, in confirmation of one of Mr Burton's contentions, that to justify six figures in the differential coefficients it has been necessary to express A and B to eight, and c to ten or eleven significant figures.

Finally, I should like to express my gratitude for the warm reception given to my paper and the many kind words spoken in the discussion. My only regrets are, first, the unavoidable absence of Mr Barley and, secondly, that only Mr Daw has voiced his support of my plea for the exclusion of duplicates wherever possible; perhaps I may also claim that Mr Tetley, who only disagrees with me on details, gives it his tacit support.