JSS 17 (1) (1962) 42-47

NOTES ON FORMULA GRADUATION OF SOUTH AFRICAN 56/58 ULTIMATE MORTALITY

by

R. C. LLOYD

(*The form of these notes was dictated by the Editors. The usual form of a report such as this would no doubt start at paragraph* 13 *or thereabouts. The result would be more gratifying to the author but, we believe, less useful to 'students'—especially those qualified ones like ourselves who have never attempted a graduation by the full-scale fitting of a mathematical formula.*—Eds.)

1. The S.A. 56/58 Ultimate data (see *J.I.A.* Vol. **87**, p. 371) was originally graduated by summation formula and arising out of the discussions it was decided to attempt to fit a mathematical formula of the Perks type.

2. To start with, data below age 30 were excluded.

3. The generalized Perks formula

$$q_x = \frac{A+Bc^x}{Kc^{-x}+1+Dc^x}$$

was chosen—the thought being that if some of the parameters were not needed the process of curve fitting would result in these parameters assuming small values.

4. In order to reduce the arithmetic it was decided to make use of the Spencer graduated values of q_x at quinquennial points, rather than to graduate with reference to the exposed to risk and the actual deaths. Starting with an assumed value of c it was necessary to have four equations, and two summations, each on half the data, were used (in preference to four summations of all the data—which would have exaggerated the importance of the high ages).

5. We have

$$Kc^{-x}q_x+q_x+Dc^xq_x = A+Bc^x.$$

First and second summations were done on

$$(a)\ x = 30, 40, \ldots, 90,$$
$$(b)\ x = 35, 45, \ldots, 95.$$

From the resultant four equations the values of *K*, *D*, *A* and *B* were obtained.

6. The first trial was with $c^5 = 1.7$. The results were too high below age 60, too low from 60 to 85 and thereafter too high. Before trying other values of c, some variations to the formula were tried. The term Kc^{-x} was replaced by Ec^{-2x} and the results were worse, i.e. the deviations were in the same direction, but larger. Kc^{-x} was then replaced by $Fc^{-x/5}$. The results were disastrous as *A* and *F* became negative and there was a resulting discontinuity in the curve of q_x at about age 44. Kc^{-x} was then replaced by $Gc^{-x/2}$ and again a discontinuity was found.

7. The original formula was then reverted to and the results were worked out with the following values of c^5

1·1, 1·25, 1·3, 1·35, 1·4, 1·5, and 1·8

although not in that order. All these trials revealed the same pattern as the first trial except that as would be expected some values of *c* gave smaller deviations. None of the results could, however, be regarded as a suitable graduation, although that with $c^5 = 1{\cdot}4$ was the best.

8. The trial with $c^5 = 1{\cdot}4$ was then reworked leaving out the four highest ages i.e. 80 to 95. This improved the fit up to age 75, but thereafter the values became impossibly high. An attempt was then made to adjust this latest trial more or less arbitrarily to improve it, but this attempt was not successful.

9. Next the term Kc^{-x} was dropped (and one second summation disregarded), $c^5 = 1{\cdot}5$ was used, but the results were poor.

10. The next two trials left the term Kc^{-x} in and added an additional term Ec^{-2x}; in the denominator. To get the additional equation q_{35} was taken at the Spencer value (·00196). c^5 was taken as 1·7 and 1·4. The trial with $c^5 = 1{\cdot}4$ proved to be the best of all the trials made, without, however, producing a really satisfactory graduation.

II. At this stage it seemed clear that on the method used and with the formulae tried, no values of c would give a satisfactory graduation. Some correspondence passed with Mr Perks in which, ***inter alia***, he made the following remarks:

Now with regard to the graduation of the data the comparisons with the A24/29 and A49/52 help to show how far the $(A + Bc^x)/(1+Dc^x)$ formula will work. We know that the A 24/29 can be well fitted by this formula and that the A49/52 cannot. An adjustment in the middle of the A49/52 table had to be made to pull the graduated rates up. The percentages in relation to the A 24/29 rates show a hump between 45 and 65 and this may lead to difficulties. The ***A49/52*** percentages show a hump between 35 and 45. This may be offsetting the adjustment to the A49/52.

It is clear that the data show a hump in the early 20's and a dip at about age 30. Now the c^{-x} and the c^{-2x} terms that you have tried may be made to reflect this when there is an *A* constant in the formula but experience has shown that it is extraordinarily difficult to get the thing right by a full-dress fitting process throughout the age-range. The practical course that has worked in the past is to do the fitting without these terms from about age 35 and then put the extra terms in arbitrarily—it is a bit of a tricky operation.

12. Following the lines suggested by Mr Perks the main portion of the data was tackled. From a study of the ratios of q_x/q_{x-5} it seemed clear that a Gompertz curve would fit from age 30 to 65 and that a second Gompertz curve (with a lower value of c) would fit from 65 to 80. If the constants B and c were determined from the values of q_x from 35 to 65 the curve would be too high thereafter, but division by $1 + Dc^x$ together with an increase in B would probably give a good fit up to age 80 and the rest of the curve above that age would not matter so much.

13. Taking $c = 1{\cdot}1$ (determined from the ratios of q_x/q_{x-5}) and using q_x from 35 to 80, two summations were performed to fit $q_x = Bc^x/(1+Dc^x)$. The fit was acceptable from 35 to 80 and the deviations for higher ages were tolerable in view of the lack of importance of that section of the curve.

14. While better results would no doubt be possible by trying a slightly higher and a slightly lower value of c, this was not done as the fit already obtained was good enough. The purpose of the experiments was not to produce a 'best fit', but rather to establish method.

15. The values of q_x were then calculated back to age 15 and as was to be expected were in general much too small. An adjusting term or terms thus had to be added. According to Mr Perks terms involving c^{-x} and c^{-2x} in the denominator will work, provided the numerator contains a substantial *A* term, but not otherwise. This approach was, therefore, abandoned as there was no *A* term in the numerator, and some other method of adjustment had to be sought.

16. As it was realized that to reproduce the 'hump' in the early 20's might be a little difficult, the first aim was to produce the more customary shape, i.e. the aim was to have rates of mortality substantially flat up to age 25 or so, thereafter increasing to blend into the main curve at age 35, while at the same time obtaining a reasonable correspondence between actual and expected deaths for the age-range 15 to 35.

17. The first step was to settle on a rough progression of mortality rates from age 15 to 35 that would satisfy these criteria and then to subtract from them the values obtained from the main formula. A function then had to be fitted to the differences, bearing in mind that this function should be negligible after age 35. A polynomial would have worked for the ages 15 to 35, but would then have become increasingly negative and would have ruined the main part of the graduation. It did not seem right to use a polynomial and to add a note to the effect that it should be disregarded over age 35—this seemed t6 be cheating. In the end it seemed that, whereas $Ec^{-x} + FC^{-2X}$ in the denominator would not work, the same terms in the numerator, with *E* positive and *F* negative, would be a fruitful line of approach. Clearly here, taking the origin at 35, there would be little difficulty in getting the function to have little effect above age 35. A little experimenting soon settled the constants and a check of actual and expected deaths over the whole range of ages led to the introduction of a small negative *A* factor to pull the whole curve down slightly. The final formula was

$$q_x = \frac{A + Bc^{x-35} + Ec^{-(x-35)} + Fc^{-2(x-35)}}{1 + Dc^{x-35}},$$

where

$A = \cdot 00015,\quad D = \cdot 00411,$
$B = \cdot 00188,\quad E = \cdot 000344,$
$C = 1 \cdot 1,\quad F = -\cdot 0000264.$

and the results could be considered acceptable.

18. Having found a successful formula it seemed worth while to see if the method used in the early trials would have worked had the correct formula been used, and also to see the effect of using the exposed to risk and actual deaths. In fact the trial using the final formula (and q_x from 30 upwards) produced hopeless results, and the use of q_x for the earlier ages would probably not have improved matters. The trial using the exposed to risk and actual deaths was equally hopeless. In both cases a substantial *A* factor was obtained, and there cannot be a good fit to this set of data at the middle ages with a substantial *A* factor. Also the constants *E* and *F* in both cases took the wrong sign. Finally, the *A* term was left out of the formula, but the trial was not successful as some values of q_x became negative.

19. The lesson that was learned is that it is no use choosing a formula and trying by mathematical means to force the data to conform. Rather one must recognize that each term in the formula serves a specific purpose and must be fitted with that purpose in mind—and in the main the empirical approach seems the best way to do this.

20. Mr Perks commented on the graduation as follows:

Your new graduation is quite good on the basis of flat rates up to about age 25. My comments in my previous letter regarding terms in c^{-x} and c^{-2x} related, of course, to terms in the denominator. You have now introduced them in the numerator. This is, of course, quite legitimate and the success arises out of the fact that the data do not require a significant value for the *A* constant. I do not much like negative constants—particularly for *A*—but to get the results you have this seems inevitable. Fortunately you do not run through zero for *q* in the practical range. Negative constants in the denominator are also troublesome sometimes. Your graduation formula has six parameters and we must accept (1) that the formula is just a fitting instrument—it has no 'meaning', (2) that other forms with six parameters could be found to produce as good a fit. However, such a fit means that the *q's* are mathematically 'smooth' and interpolation and approximate integration work on a table based on these would not run into difficulties.

21. Having found one function that could adjust the main curve at the young ages it seemed a good idea to see what could be done with different functions, and in particular to see how many parameters would be needed. The two functions that seemed worth trying were

(a) the reciprocal of a polynomial, and

(b) the normal curve of error.

It also seemed worth trying to use them not only to reproduce the conventional shape of curve back to age 15, but also to reproduce the 'hump' and the 'trough' at the two ends of the 20's.

22. It was found that both the 'level' and the 'humped' curve could be reproduced by using either form of adjustment. However, the adjustment by means of the reciprocal of a polynomial led to final formulae with at least seven parameters. The normal curve of error proved to be a very powerful and flexible means of adjustment and the final formulae each had only six parameters. Both for the 'level' and 'humped' curve the adjustment was of the form $Ae^{-h^2/2}$ where h equalled $(x-15)/10$ and $(x-19)/5$ respectively, 15 and 19 being the origins chosen for the two normal curves. The adjustment is very simply transformed into the form $Ac_1^{-(x-a)^2}$, where a is the origin of the normal curve. The constant B in the main curve required a small reduction in the case of the 'level' curve, otherwise B, c and D were not altered. The final (six parameter) formula used to reproduce either the 'level' or the 'humped' curve was of the form

$$q_x = Ac_1^{-(x-a)^2} + \frac{Bc^{x-35}}{1+Dc^{x-35}}.$$