THE GENERATION METHOD OF ANALYSIS APPLIED TO MORTALITY FROM RESPIRATORY TUBERCULOSIS

By C. C. SPICER

Central Public Health Laboratory, Colindale, London, N.W. 9

(With 2 Figures in the Text)

INTRODUCTION

It has been suggested by Frost (1939) and Andvord (1930) that mortality from tuberculosis can be explained on the assumption that every generation has the same pattern of mortality at various ages, while its general level of mortality is determined very early in life and remains peculiar to it thereafter. This hypothesis has been applied to English overall tuberculosis mortality by Daw (1950) and Springett (1950). A similar explanation for mortality from all causes had previously been put forward by Derrick (1927) and Kermack, McKendrick & McKinlay (1934). It is the purpose of this paper to see how far such an explanation, in a simple arithmetical form, agrees with the data on respiratory tuberculosis, as most previous work has been of a more qualitative kind. It is especially interesting to see how far the mortality in recent years can be quantitatively explained by the generation hypothesis, since, if the hypothesis is correct, the mortality at older ages should be almost completely determined by the tuberculosis mortality in previous decades. The fitting of a simple mathematical scheme is not likely to be better than a first approximation, but it will be seen that it can give a remarkably good account of the observations.

MATERIALS AND METHODS

Only mortality from respiratory tuberculosis has been considered in this study. Other authors have dealt with mortality from all forms of tuberculosis, but it was thought worth while to see how well the hypothesis applied to respiratory tuberculosis alone. The mortality rates used were obtained from the reports of the Registrar General, and the analysis was made using a decennial grouping. Thus, all those born in a 10-year period were taken to form a 'generation'. The death-rates (per 100,000) at different ages were, therefore, also grouped decennially, starting at ages 5–14 years.

Death-rates at ages greater than 75 years were not considered. No generation born before 1850 was included in the investigation.

The rates to be expected on the generation hypothesis were calculated by the method used by Kermack, McKendrick & McKinlay (1934) described below. However, as this seemed to be open to some theoretical objection, a second set of calculations was made using the method of least squares. This had the advantage of providing a general check on the work, and of confirming that there were no gross inconsistencies in the hypothesis of generation mortality.

The actual computing involved in Kermack's method is very simple. First the ratios of the mortalities at successive ages are calculated for each generation. The relative mortalities at different ages are taken as the averages of these ratios over all generations, and on the generation hypothesis they are unaffected by secular trends in mortality.

If now the mortalities at different ages in any generation are divided by their ratios to some base-line mortality then the quotients so obtained should be approximately constant and equal to the base-line rate. The average of these values is taken as the 'true' value of the mortality in the base-line age group and from it the expected mortalities at other ages are calculated using the factors of proportion previously obtained.

For example, suppose that the mortalities in successive age groups starting at 5–14 years are found, on the average, to be in the following ratios for all generations

$$1:5:7:7:6:5:3.$$

Then a generation with a calculated mortality of 50 per 100,000 at ages 5–14 would have the following calculated mortalities at other ages,

5–14	$50 \times 1 = 50$
15–24	$50 \times 5 = 250$
25–34	$50 \times 7 = 350$
	$\vdots$
65–74	$50 \times 3 = 150$

The other method used for calculating mortalities was to assume that the logarithm of the mortality at any age and period could be expressed as the sum of three factors: one for age, one for generation, and one a general factor which forms a base-line for all the observations. The values of these factors are chosen so that the sum of the squares of the differences between observed and calculated figures is a minimum. As the results do not differ significantly from those found by the simpler method no details are given here.

RESULTS

The general results of the investigation for 1856–1926 are given in Table 1 and Figs. 1 and 2.

In Table 1 the observed mortality rates at various ages in each generation are compared with those calculated. A similar comparison for the age-specific rates in different decennia is made graphically in Fig. 1. It will be seen that agreement on the whole is remarkably good. In the period immediately before the first World War the predicted rates are rather too high and during the war years too low, which indicates the presence of influences not taken into account by the hypothesis. In the years after the war mortality was lower than expectation, especially in the case of older males. There is no sign of any shift in the incidence of the peak mortality in males towards old age as it remains in the decennium 45–54 years throughout. The female peak appears to be shifting into the younger ages.

The comparison of the rates at different ages observed in 1930–40, 1940–50 with

Table 1. *Comparison of observed and calculated age-specific mortality rates (per 100,000) from respiratory tuberculosis, by generations*

Age group	1856		1866		1876		1886		1896		1906		1916		1926	
	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.
	Males. Generation (mid-year of birth)															
5–14	54	64	44	51	35	41	29	30	22	20	18	15	17	16	(10)	(10)
15–	280	299	228	234	181	177	150	141	114	112	93	100	88	91	—	—
25–	372	371	303	304	241	237	199	197	152	166	124	129	—	—	—	—
35–	376	358	306	310	243	245	201	209	153	153	—	—	—	—	—	—
45–	335	314	272	275	216	229	179	165	—	—	—	—	—	—	—	—
55–	251	238	204	202	162	135	—	—	—	—	—	—	—	—	—	—
65–	138	127	112	87	—	—	—	—	—	—	—	—	—	—	—	—
	Females. Generation (mid-year of birth)															
5–14	97	94	75	74	56	60	43	51	33	37	30	29	32	30	(19)	(19)
15–	338	353	261	276	195	205	150	144	115	111	104	107	111	118	—	—
25–	354	356	274	280	204	192	157	148	120	127	110	112	—	—	—	—
35–	280	274	217	212	162	155	124	130	95	85	—	—	—	—	—	—
45–	174	164	134	131	100	108	77	66	—	—	—	—	—	—	—	—
55–	106	105	82	86	61	54	—	—	—	—	—	—	—	—	—	—
65–	58	63	45	42	—	—	—	—	—	—	—	—	—	—	—	—

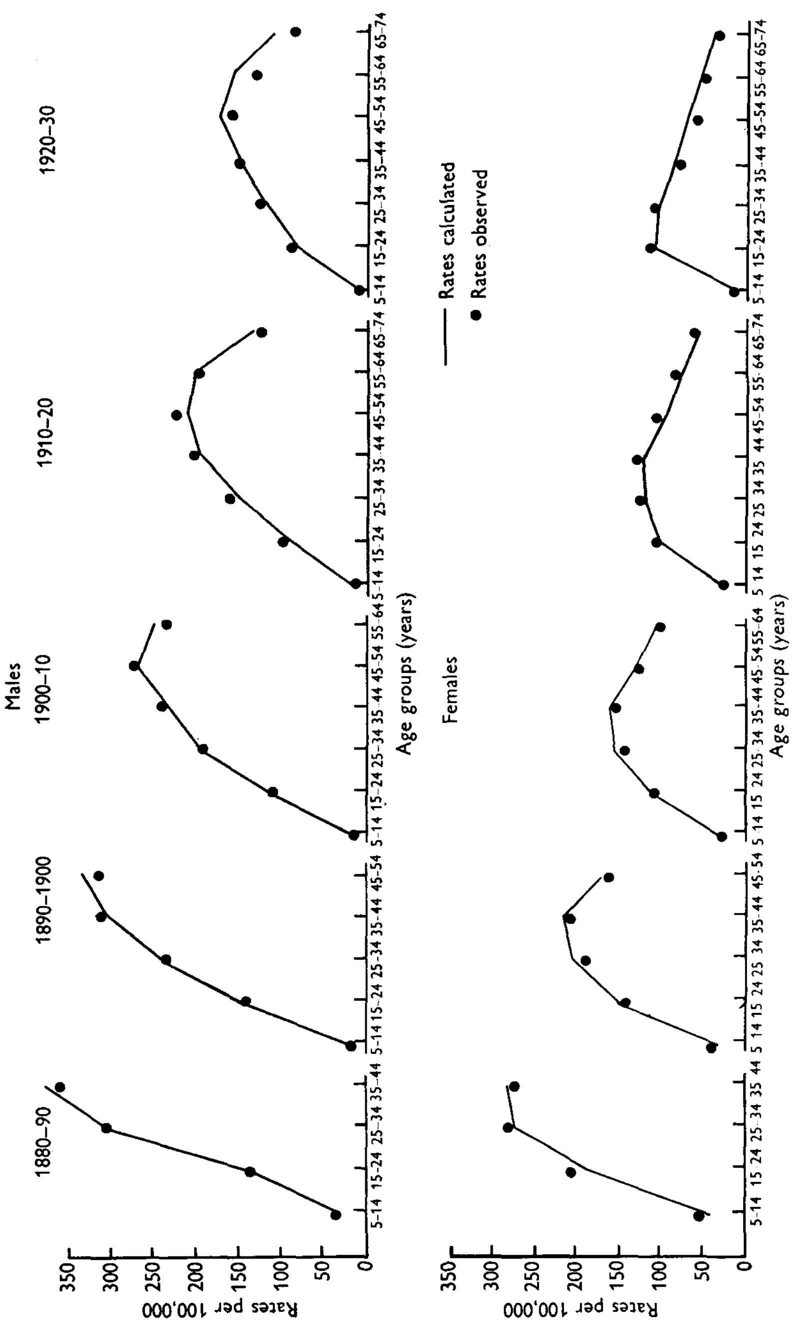

Fig. 1. Comparison of observed and calculated mortalities from respiratory tuberculosis in the period 1880–1930

https://doi.org/10.1017/S0022172400027558 Published online by Cambridge University Press

those calculated on the basis of the previous mortality are set out in Table 2 and Fig. 2. It will be seen that there are considerable discrepancies which are more serious than would be expected from the previous results. In particular, the appearance of a sharp peak of mortality among males in later life is not predicted by the generation hypothesis, and the female mortality in middle age is considerably overestimated with a corresponding underestimation in youth.

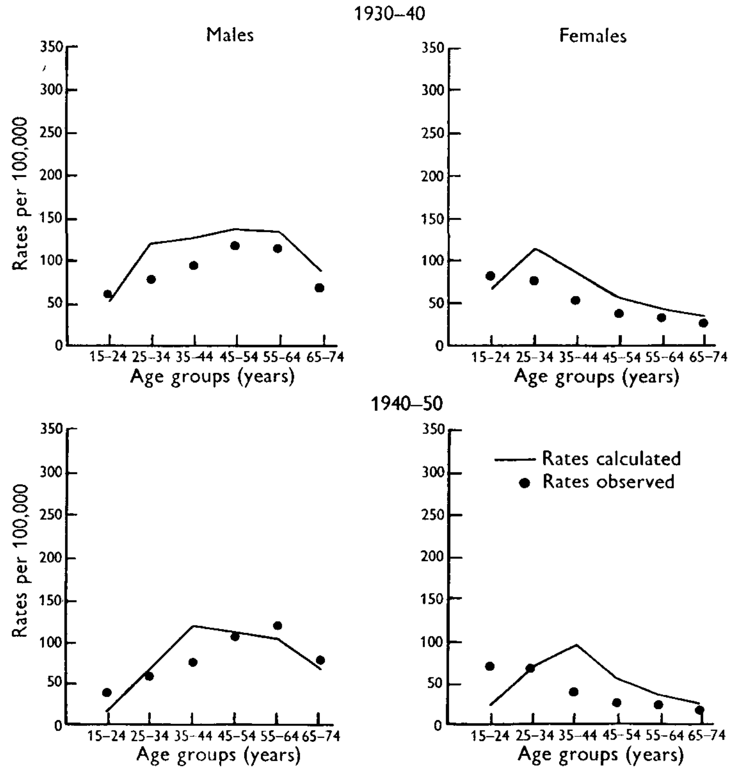

Fig. 2. Comparison of observed and calculated mortalities from respiratory tuberculosis in the decades 1930–40 and 1940–50

Table 2. *Comparison of age-specific mortality rates (per* 100,000) *from respiratory tuberculosis in the decennia* 1930–40 *and* 1940–50, *with those calculated*

	1930–40				1940–50			
	Male rates		Female rates		Male rates		Female rates	
Age group	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.	Cal.	Obs.
15–24	52	59	66	85	16	42	24	72
25–	117	81	117	78	69	60	69	69
35–	125	96	87	53	118	76	92	42
45–	136	122	59	39	111	107	54	29
55–	134	115	47	34	102	120	36	26
65–	89	68	34	26	74	78	26	22

DISCUSSION

It appears from this analysis that the generation hypothesis is capable of giving a very good description of the mortality from respiratory tuberculosis until about 1930. After that time it seems to break down and no longer gives a satisfactory agreement with the facts, but this conclusion needs further consideration.

In the first place it may be argued that the comparison of the rates in 1930–50 with those predicted on previous experience is too stringent a test of the generation hypothesis; that the calculations are admittedly an over-simplification and it would be fairer to base the predictions on the whole experience from 1850 to 1950. In fact, it makes no difference to the general conclusions if this is done. The agreement is still very poor in the two last decades and the appearance of a peak mortality in old age remains unexplained. Furthermore, the agreement between theory and fact in earlier decades is good enough to justify some confidence in any predictions made.

How good the agreement should be is a statistical problem to which it is difficult to give a satisfactory answer. A general comparison between the observed and predicted values would certainly show that statistically significant differences existed. This is because the observed rates have low sampling errors, being based on large numbers; in addition the deviations are usually systematic, and the observed sequences of positive and negative discrepancies are not likely to have arisen by chance. On the other hand, the coefficients of correlation between observed and calculated figures are about 0·99, which implies that $(0{\cdot}99)^2$ or 98% of the variation in mortality rates can be explained on the generation hypothesis, so that although other influences are at work they are comparatively unimportant. A general estimate of the errors to be expected in making predictions is provided by the mean of the squares of the differences between observed and calculated rates, and this estimate was made from the logarithmic data used in the least squares fit. The corresponding mean squared differences between observed and predicted figures in 1930–50 should be of about the same size as this estimated error, but in fact it is found that they are between 7 and 40 times as large and are far greater than could reasonably be expected. Unfortunately no statistical test is available for assessing the degree of significance, but as judged by the usual variance ratio test all the differences are significant at well below the 0·1% probability level.

One further point should be mentioned. The figures for tuberculosis mortality after 1940 are not strictly comparable with those before, since there was a change in the system of classifying the causes of death in that year. The effect of such a change would be, generally, that tuberculosis was rather less likely than before to be entered as a cause of death if other possible causes were also present. Such a change might interfere with the predictions of the generation hypothesis but would not explain the main discrepancy, which is a deficiency in the mortality during early life. On the whole, multiple causes are more probable in later life, and it would be expected that the change in certification procedure would depress the relative mortality in old age rather than accentuate it.

The preceding analysis shows quite clearly that movements of the peak of mortality do not necessarily result from the existence of a generation effect. For although the generation hypothesis agrees very well with the observed mortalities from respiratory tuberculosis there is, nevertheless, no significant movement of the peak mortality in males, which remained from 1906 to 1926 in the decade 45–54 years. The mortality at 5–14 years declined from 1856 to 1926 in an approximately logarithmic manner, and the calculated rates at other ages are obtained from these on multiplying by a factor which depends only on the age, so the rate at age θ in year t can be put in the mathematical form

$$M\ (t, \theta) = Ae^{-\kappa(t-\theta)} f(\theta),$$

where A and κ are constants, and $f(\theta)$ is a function of θ alone. The peak mortality in year t is the maximum of $M\ (t, \theta)$ with respect to θ and is independent of t. It seems dangerous, therefore, to place too much reliance on qualitative interpretations of movements of the peak mortality as evidence of a generation effect.

No detailed explanations of the failure of the generation hypothesis can be made from the present investigation, but some general remarks are permissible. It can be seen from the comparison of the predicted and observed figures that, in the decade 1930–40, there was a large overall decline in tuberculosis mortality affecting males and females from 30 years onwards. In the subsequent decade this decline was not maintained in males except in middle life; there was also an excessive mortality in youth and the peak in mortality at 55–65 years is much higher in relation to the earlier ages than could be explained on the generation hypothesis. This peak value is also absolutely higher than that predicted by theory. Taken altogether the figures for the last decade suggest that there are causes quite apart from the generation effect which have arrested the decline of male mortality in old age, or even reversed it. In youth there is also an unexplained excess both among males and females.

It might be argued that the generation effect is still operative in that the older groups have been heavily exposed when young and are more seriously affected by increases in tuberculosis mortality. The emphasis here, however, is more on the changing conditions than on the generation effect. It seems possible that the disease is epidemiologically different in later life, since its incidence in the county boroughs at ages greater than 45 years was shown to be correlated with indices of poverty but not with those of substandard housing while exactly the reverse held at younger ages (see Hart & Wright, 1939, p. 44). This suggests that, while the disease in early life is largely determined by overcrowding, in later years factors causing breakdown of old lesions are more important.

SUMMARY

1. A quantitative comparison has been made of the observed mortalities from respiratory tuberculosis with those expected from a simple scheme of generation mortality.

2. This scheme gives a good account of the observed rates until about 1930, but is unsatisfactory thereafter.

3. It is concluded that these discrepancies cannot be due to chance, and that the generation effect has become obscured by other factors.

4. In particular the shift of maximum mortality into later life is not directly explained by the hypothesis, nor is such a shift a necessary consequence of the generation hypothesis of mortality.

I am greatly indebted to Dr J. C. McDonald for advice and constructive criticism during the writing of this paper; my thanks are also due to Dr G. S. Wilson for his comments.

REFERENCES

Andvord, K. F. (1930). What can we learn by studying tuberculosis by generations? *Norsk Mag. Laegevidensk.* **91**, 642.

Daw, R. H. (1950). The trend of mortality from tuberculosis. *J. Inst. Actu.* **76**, 143.

Derrick, V. P. A. (1927). Observations on (i) Errors in age in the population statistics of England and Wales, and (ii) The changes in mortality indicated by the national records. *J. Inst. Actu.* **58**, 117.

Frost, W. H. (1939). Age selection of mortality from tuberculosis in successive decades. *Amer. J. Hyg.* (Sect. A), **30**, 91.

Hart, P. D'A. & Wright, G. P. (1939). *Tuberculosis and Social Conditions.* London.

Kermack, W. O., McKendrick, A. G. & McKinlay, P. L. (1934). Death rates in Great Britain and Sweden: Expression of specific mortality rates as products of two factors, and some consequences thereof. *J. Hyg., Camb.*, **34**, 433.

Springett, V. H. (1950). A comparative study of tuberculosis mortality rates. *J. Hyg., Camb.*, **48**, 361.

(*MS. received for publication* 17. II. 54)