

APPENDIX F

DUPLICATES

The theoretical effect on variance of the inclusion of duplicates in a mortality investigation was examined by Daw (1944). The variance of the number of death claims from policies for lives aged x is increased by a factor

$$\frac{\sum_{k=1}^{\infty} k^2 n_k}{\sum_{k=1}^{\infty} k n_k} \quad (1)$$

where n_k ($k = 1, 2, 3, \dots$), is the number of lives aged x who hold k policies.

The distribution of k may be unknown, even though an estimate of the average value of k may be available. Daw illustrated the application of this formula by examples of conjectured distributions of the numbers of duplicates and suggested that, in the particular circumstances being considered, it would be suitable to assume that k followed a geometric progression.

As a by-product of other investigations into the 1975-78 Aggregate experience, some indications were obtained of the numbers of duplicates occurring in sickness claims inceptions. These took the form of estimates of the number of policies, among the total numbers of inceptions under each deferred period table, which were first policies, the number which were second policies and the number which were third or subsequent policies, as shown in the following table:

	Deferred 1 week	Deferred 4 weeks	Deferred 13 weeks	Deferred 26 weeks
Numbers of:				
1st policies	8,135	3,431	828	395
2nd policies	3,272	361	59	41
3rd and subsequent policies	1,960	77	11	10
Total inceptions	13,367	3,869	898	446
Average number of policies per life	1.643	1.128	1.085	1.129

The average number of policies per life is calculated as the total number of inceptions divided by the number of first policies.

If a geometric distribution of k is postulated, then from the average number of policies per life, \bar{k} , the geometric factor may be estimated as $1 - 1/\bar{k}$. On this basis,

theoretical frequencies corresponding to the previous table were found as follows:

	Deferred 1 week	Deferred 4 weeks	Deferred 13 weeks	Deferred 26 weeks
Numbers of:				
1st policies	8,135	3,431	828	395
2nd policies	3,184	388	65	45
3rd and subsequent policies	2,048	50	5	6
	<hr/>	<hr/>	<hr/>	<hr/>
	13,367	3,869	898	446
	<hr/>	<hr/>	<hr/>	<hr/>

There is a reasonable correspondence between the two tables.

If the frequencies n_k are assumed to follow a geometric progression with a factor $1 - 1/\bar{k}$, then expression (1) reduces to $2\bar{k} - 1$. Ignoring any variation of \bar{k} with age, this leads to estimates that the variance of sickness claim inceptions may have been raised, by the presence of duplicates, by factors of 2.29, 1.26, 1.17 and 1.26 for the deferred 1, 4, 13 and 26 weeks tables respectively. A scaling down of the values of χ_f^2 stated in Table 6.6 by these factors results in the following adjusted values:

	χ_f^2	f
Deferred 1 week	39	36
Deferred 4 weeks	56	36
Deferred 13 weeks	37	33
Deferred 26 weeks	28	32

On this adjusted basis, the test results become far more acceptable. However, the result for the deferred 4 weeks table remains significantly high.

It is concluded that the very limited amount of information to hand on the presence of duplicates does point to this as a major contributing factor to the high values of χ^2 often arising in the attempted graduations.