

CORRESPONDENCE.

"SPURIOUS SELECTION."

To the Editor of the Journal of the Institute of Actuaries.

SIR,—In the course of the discussion on Mr. Elderton's paper recently read before the Institute, I ventured to express the opinion that the measure of "Spurious Selection" $(q' - q)K_{xt}$, deduced by the author *on the assumption that at duration t medical selection had ceased to operate*, could not properly be applied, without modification, to the example chosen, namely, to durations 5–9 in the O^[M] Table, where it is not generally admitted that the effect of such selection has vanished, and where, for at least durations 5–7, it is admitted by the author that medical selection is still effective. As the opinion was expressed in the subsequent discussion that the proof given applied to all cases, I shall be glad if you will allow me space for the following demonstration and remarks.

Let E'_t be the exposed to risk, and q'_t the rate of mortality, at duration t in one mortality experience, and E'_{t+1} and q'_{t+1} be the exposed to risk and rate of mortality in that experience at duration

($t+1$), and let the unaccented symbols E_t , q_t , E_{t+1} and q_{t+1} refer to similar functions in a different experience, and also let Q_t and Q_{t+1} be the rates of mortality, at durations t and $(t+1)$ respectively, in the experience formed by the amalgamation of the two separate experiences, the whole of the above functions applying to the same attained age.

Now,

$$Q_t = \frac{E'_t q'_t + E_t q_t}{E'_t + E_t}$$

and

$$Q_{t+1} = \frac{E'_{t+1} q'_{t+1} + E_{t+1} q_{t+1}}{E'_{t+1} + E_{t+1}}$$

$$Q_{t+1} - Q_t = \frac{E'_{t+1} E'_t q'_{t+1} + E_{t+1} E'_t q'_{t+1} + E'_{t+1} E_t q'_t - E'_{t+1} E'_t q'_t - E_t E'_{t+1} q_t - E'_t E_{t+1} q_t - E_{t+1} E_t q_t}{(E'_{t+1} + E_{t+1})(E'_t + E_t)}$$

$$= \frac{(q'_{t+1} - q'_t) E'_{t+1} E'_t + (q_{t+1} - q_t) E_{t+1} E_t + (q'_{t+1} - q_t) E'_{t+1} E_t - (q'_t - q_{t+1}) E'_t E_{t+1}}{(E'_{t+1} + E_{t+1})(E'_t + E_t)}$$

If the effect of medical selection has not vanished, let

$$q'_{t+1} - q'_t = \Delta'_t \text{ and } q_{t+1} - q_t = \Delta_t.$$

Then,

$${}_{t+1} - Q_t = \frac{E'_{t+1} E'_t \Delta'_t + E_{t+1} E_t \Delta_t + (q'_t - q_t + \Delta'_t) E'_{t+1} E_t - (q'_t - q_t - \Delta_t) E'_t E_{t+1}}{(E'_{t+1} + E_{t+1})(E'_t + E_t)}$$

$$= \frac{(q'_t - q_t)(E'_{t+1} E_t - E'_t E_{t+1}) + E'_{t+1} E'_t \Delta'_t + E_{t+1} E_t \Delta_t + E'_{t+1} E'_t \Delta'_t + E'_t E_{t+1} \Delta_t}{(E'_{t+1} + E_{t+1})(E'_t + E_t)} \tag{A}$$

If for the given attained age and durations the effect of selection is the same in both tables, *i.e.*, if $\Delta' = \Delta$, expression (A) reduces to

$$Q_{t+1} - Q_t = \frac{(q'_t - q_t)(E'_{t+1} E_t - E'_t E_{t+1})}{(E'_{t+1} + E_{t+1})(E'_t + E_t)} + \Delta_t = (q'_t - q_t) K_{xt} + \Delta_t$$

and of this, Δ_t may be assumed to be the true measure of selection under the “Combined” Table also; and in this case, as also in the case where selection has ceased to operate and $\Delta_t = \Delta'_t = 0$, we have $(q'_t - q_t) K_{xt}$ as the measure of spurious selection arising from the amalgamation, thus agreeing with Mr. Elderton’s result.

If, however, the effect of medical selection is not equal in the two experiences, and the true measure of medical selection in the Combined Table (including the "Spurious Selection" arising from causes other than the amalgamation of the separate experiences) is denoted by Δ''_t , where Δ''_t may be expected to lie somewhere between Δ'_t and Δ_t , expression (A) becomes

$$Q_{t+1} - Q_t = \Delta''_t + \frac{(q'_t - q_t)(E'_{t+1}E_t - E'_tE_{t+1}) + E'_{t+1}E'_t(\Delta'_t - \Delta''_t) + E_{t+1}E_t(\Delta_t - \Delta''_t) + E'_{t+1}E_t(\Delta'_t - \Delta''_t) + E'_tE_{t+1}(\Delta_t - \Delta''_t)}{(E'_{t+1} + E_{t+1})(E'_t + E_t)} \tag{B}$$

$$= \Delta''_t + \alpha$$

and the measure of "Spurious Selection" arising from amalgamation is α .

I must confess that I do not at present see any exact method of measuring Δ''_t , and consequently α .

If, however, we turn from the general question of the effect of amalgamating two experiences, with different rates of mortality, and showing different results of medical selection, and consider the particular case of the O^(M) Tables it would seem unnecessary to determine either α or Δ''_t .

The mortality table is of greater importance as a measure of what may be expected in the future than as a record of what has happened in the past, and the question suggested by the latter part of Mr. Elderton's paper is as to how far the inclusion of the Old experience detracts from the value of the "Combined" Table as such a measure.

In the New experience we have the best available standard of the effect of medical selection at the present time, since it comprises the very latest data, and is most free from those causes which would produce a spurious appearance of selection, and I think that a comparison of $(Q_{t+1} - Q_t)$ by the "Combined" Table, with $(q_{t+1} - q_t)$, or Δ_t , by the New Table, would give a fair idea of the error caused by the inclusion of the Old experience. If, therefore, in expression (B) we make $\Delta_t = \Delta''_t$, we obtain, as a measure of the divergence of the Combined experience, as to medical selection, from the true experience, on the above assumption,

$$(q'_t - q_t)K_{xt} + \frac{(\Delta'_t - \Delta_t)(E'_{t+1}E'_t + E'_{t+1}E_t)}{(E'_{t+1} + E_{t+1})(E'_t + E_t)}$$

$$= (q'_t - q_t)K_{xt} + \frac{(\Delta'_t - \Delta_t)E'_{t+1}}{(E'_{t+1} + E_{t+1})}$$

$$= (q'_t - q_t)K_{xt} + (\Delta'_t - \Delta_t)K'_{xt}.$$

In this connection the following Table may be of interest :

Grouped attained Ages	Duration t	$100 \frac{E_t}{E_t}$	K_{xt}	$100(q't - q)$	$(q't - q)K_{xt}$	K'_{xt}	$100(\Delta't - \Delta)$	$(\Delta't - \Delta)K'_{xt}$	(6) + (9)	$\frac{Q_{t+1} - Q_t}{\text{Combined } Q_{(M)}}$ Ungraduated
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
28-32	5	14	.0203	.071	.00001	1430	.139	.00020	.00021	.00114
...	6	17	.0197	.210	.00004	1628	.230	.00037	.00041	.00000
...	7	19	.0176	.440	.00008	1804	-.263	-.00017	-.00039	.00024
...	8	22	.0198	.177	.00004	2002	-.142	-.00028	-.00024	.00078
...	9	25	.0175	.035	.00001	2177	-.040	-.00009	-.00008	-.00115
...	10	28
...	+ Total00018369	.00059	.00062	.00216
...	- Total445	.00084	.00071	.00115
...	Average00004	...	-.015	-.00005	-.00002	.00020
38-42	5	17	.0247	.231	.00006	1697	-.327	-.00055	-.00049	.00086
...	6	20	.0271	-.096	-.00003	1968	.099	.00019	.00016	-.00055
...	7	24	.0256	.003	.00000	2223	.074	.00016	.00016	.00135
...	8	29	.0243	.077	.00002	2466	.118	.00029	.00031	-.00008
...	9	33	.0262	.195	.00005	2728	-.105	-.00029	-.00024	-.00104
...	10	38
...	+ Total00013291	.00064	.00063	.00221
...	- Total00003432	.00084	.00073	.00167
...	Average00002	...	-.028	-.00004	-.00002	.00011
48-52	5	19	.0282	.325	.00009	1893	-.032	-.00006	.00003	-.00048
...	6	23	.0342	.293	.00010	2235	-.083	-.00019	-.00009	.00254
...	7	29	.0324	.210	.00007	2559	.023	.00006	.00013	-.00058
...	8	34	.0267	.233	.00006	2826	.062	.00018	.00024	-.00060
...	9	39	.0292	.295	.00009	3118	-.307	-.00096	-.00087	-.00020
...	10	45
...	+ Total00041085	.00024	.00040	.00254
...	- Total422	.00121	.00096	.00186
...	Average00008	...	-.067	-.00019	-.00011	.00014
58-62	5	21	.0296	-.011	-.00000	2051	-.092	-.00019	-.00019	.00035
...	6	26	.0336	-.103	-.00003	2387	.221	.00053	.00050	-.00104
...	7	31	.0299	.118	.00004	2686	.305	.00082	.00086	.00060
...	8	37	.0324	.423	.00014	3010	-.518	-.00156	-.00142	.00520
...	9	43	.0341	-.095	-.00003	3351	-.148	-.00050	-.00053	-.00369
...	10	50
...	+ Total00018526	.00135	.00136	.00615
...	- Total00006758	.00225	.00214	.00473
...	Average00002	...	-.046	-.00018	-.00016	.00028

An inspection of col. (6) and col. (11) of the above table will show that, when the facts are steadied by grouping the data at five consecutive attained ages, and the actual value of $(q'_t - q_t)K_{xt}$ is compared with the actual ungraduated value of $(Q_{t+1} - Q_t)$, the ratio of "Spurious Selection" is much less on the whole than would be supposed from Mr. Elderton's Table II; whilst col. (10), which could of course have been deduced directly by taking the difference of the q 's under the New Table from that under the Combined Table, would appear to suggest that the influence of medical selection in the combined $O^{[M]}$ Tables is *under-estimated*, as compared with that likely to prevail.

It appears to me, however, very unsafe to draw general deductions from the figures in either col. (6) or col. (10), for the following reasons:

- (1) The data, though those of five ages have been employed, are scanty, and the resulting rates of mortality and selection extremely irregular.
- (2) The rate of selection under the New assurances is not altogether reliable as a standard, as the data employed in the construction of that table is itself not homogeneous.
- (3) There may, as pointed out by Mr. Elderton, be a further error caused by the "existing" in the New assurances which are exposed to risk at one duration but not at the next, without having passed out by death; and a similar error may be caused by the inclusion in the Old experience of those cases existing in 1863 which came in at one duration without having been exposed to risk at the previous duration. The former error would be lessened to some extent by the inclusion of the "existing" in both the New and Combined experiences.

The above objections appear to me to apply with much greater force to the figures in Mr. Elderton's Table II, since in that table the value of K_{xt} is deduced from the data at one attained age only, and an average value of $(q' - q)$ is taken, which, whether it be correct or not, certainly differs from the average of the values in either col. (6) or col. (7) of that table, and differs very materially from the ungraduated values of $(q' - q)$ at individual durations. Although, by taking such an average value, a certain amount of smoothness is secured, it would appear to be at the expense of accuracy, and in this connection it is interesting to note that, on comparing cols. (5), (6) and (7) of the author's Table II, the "Spurious Selection" at age 40 varies between '00005 and '00007, whilst the difference between the average graduated and ungraduated values $(q_{[x-1]+t+1} - q_{[x]+t})$ is '00010, also that at age 50 for a "Spurious Selection" of from '00006 to '00009 we have an average difference of no less than '00027 between the graduated and ungraduated figures. In this particular case graduation is certainly of more importance than the inherent error, which it is one of the functions of graduation to remove.

Whatever other result the above demonstration and illustrations may have, they will at least, I think, illustrate the very complex nature of the problem the author has attacked. They will

also, I trust, show still more clearly the debt of gratitude we are under to Mr. Elderton for his valuable and suggestive paper, and emphasize what I take to be its real message, that, failing some more scientific method of allowing for changes during the period of observations, such as is hinted at on page 226,

- (1) Only such data should be used for select tables, or tables of a similar nature, as can be traced from entry till the close of the select period, unless going out by the ordinary methods of office experience—death, withdrawal, or termination of the contract.
- (2) The period of observations should be as recent as possible; and
- (3) Subject to inclusion of sufficiently extensive data, the period of observation should be as short as possible.

I am, SIR,

Yours faithfully,

JAMES BACON.

5 & 6, Clement's Inn,
Strand, W.C.
5 March 1906.

To the Editor of the Journal of the Institute of Actuaries.

SIR,—I have carefully re-considered the discussion on my paper on "Spurious Selection", and it seems advisable before leaving the subject to bring out a few points with which I was only able to deal partially in my reply. The main point of my paper was to show the danger of amalgamating experiences or using heterogeneous material when dealing with select mortality, and the tone of the discussion gives me the satisfaction of feeling that my attempt to prove that such a danger exists was successful, although I confess that I was sorry some speakers considered it necessary to defend certain tables which I had no intention of attacking.

In the course of his speech Mr. Green asked if the "population method" at present in use is not exactly what we want. My answer is that its use leads to select rates that *may* tell us absolutely nothing about selection as defined in my paper, and the result would not give ideal premiums for office rates. The method would, I think, be more defensible for valuation purposes; and my remarks on this point, though they were, I fear, indistinctly expressed, were meant to suggest that the subject needed more consideration before additional work was incurred in office practice. The "population method" does not give an exact measure of selection, though it can give a measure which approximates closely to the truth when the term of the investigation is short; and the practical conclusion to be drawn from this is, I think, that in studying selection our statistics should only be used for a small number of years, as there is then a better chance of securing homogeneity.

Mr. Ackland asked me to give a proof of the second footnote on p. 226. The problem referred to is: "If we add together a number of groups having different rates of mortality, will there be any spurious selection"? Since the exposed to risk at nearest age x means the

summation of the exposed for all ages from $x - \frac{1}{2}$ to $x + \frac{1}{2}$, the problem in the footnote is a particular case of the one just enunciated. The proof follows from that on p. 222; but the following hypothetical example is probably the simplest way of showing the result:—

Attained Age	DURATION 5			DURATION 10		
	Exposed	Deaths	Rate of Mortality	Exposed	Deaths	Rate of Mortality
40	1,190,100	11,913	·01001	1,064,400	10,655	·01001
41	1,104,900	11,469	·01038	998,400	10,363	·01038
42	1,006,300	10,868	·01080	963,500	10,406	·01080
43	951,400	10,694	·01124	922,200	10,366	·01124
44	903,600	10,590	·01172	863,100	10,116	·01172
40-44	5,156,300	55,534	·010770	4,811,600	51,906	·010788

This shows that the results can be distorted, but, as I have already remarked, the difference is slight and may be neglected.

I may take this opportunity of giving an example illustrating my remark on p. 226, that I cannot see the *a priori* impossibility of finding two Makeham curves having different constants, so that the rates of mortality deduced from them differ widely at the start, and then converge sufficiently to give identical values for the expected deaths at the end of the tables.

Age	RATES OF MORTALITY DEDUCED FROM Colog $p_x = \cdot0025575 + \cdot0000472(1\cdot0939564)^x$	RATES OF MORTALITY DEDUCED FROM Colog $p_x = \cdot0005575 + \cdot000053293(1\cdot092645)^x$
	20	·00752
40	·00978	·00551
60	·02922	·02591
80	·1385	·1377
85	·2056	·2056
90	·3004	·3008
95	·4268	·4268

I certainly did not mean to imply that I think Makeham's formula universally true; but one reason given for its failure in the particular case seemed rather unsatisfactory, and was therefore mentioned. My remarks on the subject certainly do not "tacitly assume that there is a definite law of mortality", if that expression means that q or μ can be expressed in a simple empirical form; but I may remind Mr. Green that it seems possible that "deaths" are more or less complex frequency-curves, and, after all is said and done, the use of a formula of approximate summation is making something suspiciously like the assumption of a mathematical law of mortality. This is, however, a side issue.

My conclusion (No. (4) of p. 233) gives my personal opinion, but it is impossible to dogmatize with the particulars available, and

my estimate may easily be too high or too low. Though I have no intention of going further into this matter, I should like to point out that, so far as practical work is concerned, the sooner select tables can be run into an aggregate table the better, and I hoped I was simplifying matters for the future by suggesting that on some future occasion we might be able to dispense with a few of the ten years of selection shown by the present tables,

I may now turn to Mr. Bacon's letter, which I have had an opportunity of perusing. He deduces a formula and then shows that, in the two cases in which he finds it possible to assign a definite value to the function he uses, he has no difficulty in reaching the expression I adopted, and it is not until he is unable to see "any exact method of measuring" his function that he questions my result. It seems to me that even in this last case the simple formula I used has a meaning, for it holds if we assume that a possible measure of the selection in an amalgamation is obtained by using the same proportions of exposed at duration $t+1$ as those at duration t , and interpret the formula in which $q'-q$ was constant as $(q'_{t+1}-q_{t+1})K_x t$. I have used the term "possible measure", because I do not see how any measure in such cases can be strictly called correct, and I think Mr. Bacon is probably in as complete agreement with this remark as I am with the three final conclusions of his letter. I think it a pity, therefore, that Mr. Bacon should proceed to apply his formula to the case in which he admits he cannot measure his function exactly, and the application is so unfortunate, that he concludes that it is "very unsafe" to draw general conclusions from his figures. It is perhaps partly because he has had to obtain the result by working on the difference between the differences of two rates, both of which are based on ungraduated heterogeneous material, that he is led to conclude that he has reached an unreliable result.

With regard to my own figures in Table II, I attempted to avoid the roughnesses of data to which Mr. Bacon refers by using a constant difference; anyone who prefers a smaller difference has merely to reduce my figures in proportion to see the effect. The roughnesses do not affect me, even in using one age only, because they occur in the deaths, which are few in number, rather than in the exposed, which are many.

So much for past events; may I turn for a moment to the future, and suggest some possible developments of the study? (1) Under what conditions will a table like the "old" assurances contain negative spurious selection? (2) Heterogeneity may be introduced by subtracting part of a homogeneous body of facts. If, therefore, all with-profit policies were formerly for the whole of life, the starting of a new class like endowment assurances *might* make the whole-life material over a long period heterogeneous. This is a very difficult point, and needs investigation. (3) If we could prove that selection takes the form $A + \phi(xt)$, as suggested on p. 233, a test for homogeneity might be formed by examining changes in $\phi(xt)$. I may here mention that of course $A + \phi(xt)$ is negative and $\phi(xt)$ may probably take either the form $-\kappa t^{-m}(a-t)^n$ or $-\kappa t^{-m}e^{pt}$ where m is a fraction and κ, m, a, n and p depend on the age at entry. (4) A further study in the effects of gradual change in rates of mortality on selection showing, for instance, the effects of a decrease followed

by an increase in rates would be useful. Such an investigation might show that variations in the exposed and rates of mortality can counterbalance when selection has worn off though this would not prove that the experience was homogeneous, or that the rates when selection was in operation gave a true measure of select mortality.

I feel strongly that at present we are far more likely to get information about the effects of heterogeneous material by the construction of hypothetical examples than by the examination of existing material, and are less likely to be misunderstood.

It is admitted by Mr. Bacon in his letter, and by the speakers in the discussion, that it is unsafe to rely on the selection shown by heterogeneous material. This was what I set out to prove, and the admission makes it unnecessary for me to go further into the subject at present. Perhaps the many other problems connected with the subject may be solved by some other student, who has more time to devote to their solution than I seem likely to be able to give at present.

I am, Sir,

Yours faithfully,

W. PALIN ELDERTON.

11, *Lombard Street, E.C.*

31 *May* 1906.
