Practical Applications of the Statistics of Repeated Events, Particularly to Industrial Accidents.

By Ethel M. Newbold, B.A., M.Sc.

[Read before the Royal Statistical Society, April 26, 1927, Sir Bernard Mallet, K.C.B., Honorary Vice-President, in the Chair.]

In the study of accident prevention, attention has of late years been directed increasingly towards the personal factor. It is repeatedly pointed out in the Annual Reports of the Chief Inspector of Factories that the great bulk of accidents are not caused by machinery, and that only a relatively small proportion can be prevented by mechanical safeguards. A great deal of work has been done by various investigators on this side of accident causation.* The particular question of differing individual liability was first treated statistically by Dr. Greenwood and Miss Woods.† During the war they were able to get some interesting figures relating to small accidents to women in munition factories, which suggested the existence of a certain amount of measurable tendency for such accidents to fall unevenly on the population at risk. In 1920 a joint paper by Dr. Greenwood and Mr. Yule, on the theoretical side of the distributions of repeated events of this kind, was published in the *Journal* of this Society.‡ Arising from this, some further investigations on this subject were instituted, and are still in progress under the Industrial Fatigue Research Board and the Statistical Committee of the Medical Research Council, and it has been suggested that a summary of the statistical side of these investigations might be of interest to the Fellows of this Society, and might, perhaps, elicit useful criticism. My own connection with the

* For a summary of much of this work *see* Preface to Industrial Fatigue Research Board Report No. 19.

† Report No. 4 of Industrial Fatigue Research Board : "A Report on the Incidence of Industrial Accidents upon Individuals, with Special Reference to Multiple Accidents," by Major Greenwood and Hilda M. Woods.

‡ "An Inquiry into the Nature of Frequency Distributions representative of Multiple Happenings with particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents," *Journ. Roy. Stat. Soc.*, vol. lxxxiii, part II, March, 1920.

investigation has only been a share in part of the statistical work, which originated with Dr. Greenwood and Mr. Yule. and has since been continually under their advice and guidance. The investigation is now mainly concerned with the psychological side of the personal tendency to accidents, which is being examined by Mr. E. Farmer and Mr. E. G. Chambers. Of that side I am less competent to speak than of any other, as I am not a psychologist, so that with regard to this I shall confine myself to the statistical results obtained.

The main questions that the originators of these investigations had in mind were—

(1) Does any definite tendency exist under uniform conditions of risk for certain people to sustain more accidents than others ?
(2) If such a tendency exists, how, if at all, is it modified by the occurrence of accidents ?
(3) If such a tendency exists, to what extent can it be measured, and does it show any association with other qualities ?
(4) Is it possible to devise any sort of test by which people liable to accidents, if such exist, could be roughly sorted out, so that in the choice of persons for occupations to which special risk is attached such individuals could be avoided ?

These questions at once call up a cloud of practical difficulties and suggest more questions ; for example : What is an " accident ? " Are not the causes of, and reactions to, accidents clearly very complex, possibly too complex to show any orderly results ? May not all these things be very different in case of trivial accidents with no material consequences, and, in the case of really serious accidents, involving danger to oneself and others ? Is it possible ever to approach to " uniform conditions " in real life ? and so on. No one concerned with these investigations has been unaware of these and many other difficulties, admittedly only a very small corner of the field has been entered, and great caution must be observed not rashly to extend results obtained with a definite limited type of accidents under certain limited conditions to cases where they may fail to apply. The fact that a complicated problem is not soluble at once, and possibly not completely soluble, in any practical sense at all, is no reason for abstaining from attacking one side of it.

May I, for the moment, postpone the discussion of these very important practical points, and take first a much simpler and purely theoretical point of view ? For this purpose an accident is considered simply as an event, and we assume that we have a record of the numbers of such events happening to different people in certain periods of time; we assume also that external conditions are uniform, and confine ourselves first to the question whether the distribution of events among the individuals is a purely chance one, and then, if it is not, to what extent are the underlying peculiarities masked by chance variations, and how far are we able to strip off the mask and see the form of these peculiarities themselves. In the purely chance scheme, as Dr. Greenwood and Mr. Yule have shown, the proportion of people who might be expected to have 0, 1, 2, 3, etc., accidents in a period in which the mean number of accidents is λ, is given by the terms of the ordinary Poisson limit to the binomial $e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\ldots\right)$, the assumption made being that the time period can be looked on as consisting of an indefinitely large number of very small time intervals in each of which any person might receive an accident, and that the chance of an accident happening in any one of the time intervals is alike for each person and over the whole period. As the Poisson series turns up so continually in questions of this kind, in either time or space, and as it is involved in the modification of the simple chance scheme that we have to consider, perhaps those of you who are very familiar with it and its applications will bear with me if I go shortly over some of its points and uses for the sake of any non-statistical members of the audience. By Poisson's limit is meant simply the limit of the expanded binomial $(p + q)^n$, when p becomes very small and n very large; or perhaps it would be truer to say that it is restricted to applications of this limit to probability problems, in which, as usual, p is the probability of an event happening, q of its not happening, in any one trial, and n the number of trials, the terms of the series giving the proportion of observations in which the event may in the long run be expected to happen 0, 1, 2, 3, etc., times when each observation consists of a set of n trials. The path to the limit is very simple; as Bortkiewicz has remarked, it does not need a first-class mathematician like Poisson to obtain this limit, and it has been arrived at independently more than once by other people dealing with statistical or physical problems in time or space. "Student"* obtained it when searching for the probable error of the number of yeast cells counted in the squares of a hæmacytometer; Bateman,† in an appendix to a paper

* "On the Error of Counting with a Hæmacytometer," *Biometrika*, V, 1906–7, pp. 351–60.

† "The Probability Variations in the Distribution of α Particles," *Phil. Mag.*, 6th series, 1910, vol. xx, p. 696.

by Rutherford and Geiger, by a different method of approach obtained it as an exact formula to describe the frequency of emission of α particles per unit of time in radioactive radiation. Von Bortkiewicz* and Mortara† have used it to describe infrequent events in vital and social statistics, and it has become familar in many later applications. Incidentally, Poisson's name has stuck to the series, because it has often been stated that he was the first to apply it to probability problems, but this is, I think, incorrect. Poisson's *Recherches sur la Probabilité des Jugements* was published in 1837; but over a hundred years before, in 1718, in the first edition of his *Doctrine of Chances*, De Moivre applied the exponential limit to the following problem:—Problem V (p. 14): "To find in how many Trials an Event will Probably Happen, or how many Trials will be requisite to make it indifferent to lay on its Happening or Failing; supposing that a is the number of Chances for its Happening in any one Trial, and b the number of Chances for its Failing." De Moivre make the chances of the event happening (*i.e.* happening at least once in the total number of trials) or failing equal, by equating the first term of the binomial (or, as he calls it, "Sir Isaac Newton's theorem") to the rest; his binomial is $\left(1+\frac{1}{q}\right)^x$ where $\frac{a}{b}=\frac{1}{q}$ and $x=$ number of trials, so that his equation‡ is

$$1+\frac{x}{q}+\frac{x(x-1)}{1.2q^2}+\frac{x(x-1)(x-2)}{1.2.3.q^3}+\ldots=2.$$

He then proceeds to two limits, first he makes the chance of an event $\cdot 5$, in which case his x and q both become unity, then he makes x and q both infinite, the mean x/q remaining finite, and so the left hand of his equation becomes

$$1+\frac{x}{q}+\frac{x^2}{2!\,q^2}+\frac{x^3}{3!\,q^3}+\ldots,$$

and hence $\frac{x}{q}=\log_e 2$. For the whole range of possible original chances x/q lies between 1 and $\cdot 7$, and x can without much error be taken as $\cdot 7q$ for all values of q. This problem of De Moivre's really contains the basis of Proposition LI of Whitworth's *Choice*

* *Das Gesetz der Kleinen Zahlen*, 1898.

† "Sulle Variazioni di Frequenza di Alcuni Fenomeni Demografici Rari Annali di Statistica," serie V, vol. 4, 1912, pp. 5–61.

‡ I am quoting from the first edition; in the second edition the wording is rather different.

and Chance: "If an event happens at random on an average once in time t, the chance of its not happening in a given period T is $e^{-T/t}$," *i.e.* the first term of a Poisson series. This use of a Poisson, not as a complete series, but as a dichotomy, the first term as opposed to the sum of the rest—*i.e.* an event happening or not happening, irrespective of how many times it may happen—is a form in which it is very often applied now in practical problems, *e.g.* in testing dilutions for bacterial growths, and dividing into sterile and fertile cases; and we find it useful in the accident problem, when we come to consider the proportion of people who escape accident altogether. It is true that the limiting series only comes in incidentally in De Moivre's problem; but it certainly does come in, and I think there is as much historical reason for calling it De Moivre's series as Poisson's.

In Poisson's *Recherches* the exponential expansion occurs first in the same problem as De Moivre's, para. 8, p. 40: "To find the number of trials necessary for it to be an even chance for an event to happen at least once, or not to happen." He solves this just as De Moivre does, and then goes on to say that, however small the chance p of the event E may be other than zero, we can always choose n so that the probability that E will happen at least once is as near certainty as we like, and that when p is very small this limit is $1 - e^{-np}$, and the opposite probability e^{-np}. In a later paragraph (para. 81, p. 205, which is the one usually quoted), Poisson arrives at the series by the usual method of approach. He has been deducing the normal curve as a limit to the binomial when n is large and neither p not q are small (Bortkiewicz,* by the way, showed later that these two conditions may be replaced by the single one, that the mean nq should be large), and then proceeds to the case where q is very small. Poisson's way of expressing the problem requires the sum of a finite number of terms of the series, which we can now get by a single entry, either into Pearson's Tables of the Incomplete Γ function, or, for values of the mean above ·5, by Elderton's Tables for the P of the Goodness of Fit.†

Poisson's limit has also been called the Law of Small Numbers. This name is due to Professor Bortkiewicz, who published in 1898 a small treatise‡ on this series. But (and this point has some application to the accident problem), by the Law of Small Numbers,

* *Loc. cit.*

† Sum of the first p terms of a Poisson series with mean $m = 1 - \Gamma_m(p)/\Gamma(p)$ = the P of the Goodness of Fit test for $\chi^2 = 2m$ and $n' = 2p + 1$ (n' — one more than the number of degrees of freedom).

‡ *Das Gesetz der Kleinen Zahlen.*

Bortkiewicz did *not* mean simply or even primarily the exponential limit to the binomial, but a general law involving it.

What he stated as the Law of Small Numbers was the fact, that statistical series which yield small absolute numbers of events (even though each observation is based on a large number of trials) do in actual fact more frequently show an apparently reasonable agreement with an expected simple sampling distribution than do series of larger absolute numbers. He illustrates this by several examples, some of which would not, perhaps, come up to the standard of the more rigid test of fit that would now be used. Having shown the facts, he then searches for the theoretical reason for this apparent paradox. When the observations are not really homogeneous but are grouped about different submeans, the total variation consists of two parts, the simple sampling error about each of the submeans, and the variation of these submeans about the mean of the whole. The latter Bortkiewicz calls the excess error, and the point of his argument is that when the number of trials in a single observation is increased, the ratio of the excess error to the sampling error is also increased. In the case when the sampling errors follow Poisson's limit, we can state the result thus: if we increase our mean in the ratio k to 1, then the ratio of the excess error to the sampling error will increase in the ratio $\sqrt{k}$ to 1 (*see* para. 4 (ii) of Appendix). To put it in a simpler way, when the expected numbers are absolutely small, the sampling errors are so large that they may swamp the excess error, unless the latter is very large, while, when the expected numbers are large (*i.e.* absolutely large) the simple sampling error is relatively small and less important. In the case of a small mean we may thus get a fictitious agreement with the ordinary simple sampling distribution, which in this case happens to be Poisson's exponential limit; hence the connection of this limit with the Law of Small Numbers. An example of this is given by the sickness claims in one year (1922) of 134 miners, all working in a single pit, and all of ages 35—44; 103 had no claims, 27 had one and 4 had two. These are identical with the numbers given by the Poisson limit, but we can be confident that if these same men were observed over a longer period, some want of homogeneity would appear. We may have to reject a group of people as unsuitable for experimental work in correlating other tests with their accident records, because these records apparently follow the simple sampling law, but we are not justified in concluding that their tendency to accidents is undifferentiated unless the same result appears when they are watched over longer periods.

The property of the Poisson limit which makes the series so applicable to time and space problems, is that the *sum* of a set of numbers, each following a separate Poisson series (about different means) is itself a Poisson series. We can thus split up our time or space units into smaller ones, divide our cells, or change our periods of exposure, or sum the records of separate individuals into records of groups, and each of the single sets as well as the sum of the whole, will give a Poisson series. The logical advantage given thus to the Poisson over the binomial, as sometimes applied to this type of data, was pointed out by Dr. Greenwood and Mr. Yule in the paper referred to above.

In this most interesting paper the authors discussed the forms of the distributions that would arise from various hypotheses. They considered first what would happen on the assumption that the fact of a person receiving an accident made him more or less likely to have another. They found a general solution to this problem, but in a complicated form not very suitable for computation. For the simpler assumption, that the chance changes after the first accident has been received, but not again, they found a solution easy to apply, but concluded that there was no theoretical justification for this modification, so that this biassed scheme could only be looked on as a possibly convenient smoothing formula. Finally, they found a very simple form of distribution on the assumptions (1) that the individual chances varied from one person to another, but remained constant throughout the period, and (2) that these initial liabilities were distributed in the population in a simple curve of the form of Pearson's Type III. This curve was chosen only because it was of the requisitely skew form and led to conveniently simple equations for fitting. The expected distribution of accidents could easily be obtained from the mean and second moment of the observed accident distribution.* They tested these different schemes on the small accidents incurred by women in different processes of munition work; the nature of this work made it comparatively easy to get groups of women under similar working conditions. The simple chance scheme clearly was not in accordance with the facts; in some cases the modified biassed scheme graduated the data fairly well, but the most successful form was that assumed for differing individual

* The expected frequencies of 0, 1, 2 . . . accidents are given by the terms of the series

$$\left(\frac{c}{c+1}\right)^k \left[1 + \frac{k}{c+1} + \frac{k(k+1)}{2!\,(c+1)^2} + \frac{k(k+1)\,(k+2)}{3!\,(c+1)^3} + \ldots\right],$$

where k and c are found by equating the mean and second-moment coefficient of the accident distribution to $\frac{k}{c}$ and $\frac{k\,(c+1)}{c^2}$ respectively.

liabilities, which gave reasonable fits to 14 groups of women. Dr. Greenwood and Miss Woods, in the first report, had also found a positive correlation of the order ·4 to ·7 between the accidents of the same individual in two consecutive periods of three months each. It was then thought advisable to test these results on rather wider data. A number of factories engaged in various industries agreed to keep records of accidents, for periods varying from three months to two years, in chosen occupations, where there was opportunity for many small accidents, where the work was as homogeneous as possible under normal conditions, and also where strict reporting of all small accidents was in force. Any injury, however slight, which was recorded as treated at the ambulance-room or from ambulance boxes, was counted as an accident. The information that it was found possible to get for individual workers was: Age, sex, length of time in the factory, number and type of accidents, visits to the ambulance-room for minor ailments, time lost for sickness or accident, lateness, and other causes and, in some cases, output.

The details of this investigation are given in Report No. 34* of the Industrial Fatigue Research Board, with full numerical tables; here I can only take the main points. The defects of the data are only too obvious. Perfect uniformity of work and surroundings in any group is very difficult to find under normal factory conditions. Much trouble was taken in the selection of groups; but, even so, they must to some extent vary in homogeneity. Each group had to be treated separately, as conditions were not comparable from one group to another. There were in all 25 groups of males, and 14 of females, *i.e.* 6,938 males and 2,024 females, and the total number of accidents was 16,188. It may be objected that these small cuts, bruises, burns, etc., are too trivial to take into account. There might be some weight in this criticism, if these minor accidents were themselves the sole subject of enquiry, although, taken in the mass, they do represent a good deal of wasted time and expense, but they are taken rather as a tentative measure—inadequate though it probably is—of the underlying personal tendency to accident. If such a tendency exists, there is reason to believe that though, in, say, 99 cases out of 100, the events to which it leads may be of little or no importance, in the hundredth this may be disastrous. The possibility, however, of a different attitude of mind of workers to surroundings capable of producing serious and minor accidents

* "A Contribution to the Study of the Human Factor in the Causation of Accidents," by E. M. Newbold, 1925.

is not to be overlooked; some evidence bearing on this point was obtained later by Mr. Farmer on dockyard accidents. In the factory data, the numbers of serious accidents in any one group were too few for statistical treatment. Another possible source of error is that the data naturally cover *reported* accidents only. The practice of reporting small accidents varies very much in different factories, but in those groups finally chosen it is believed that very few accidents escaped notice. The distributions of accidents among the members of the groups were (with one small exception) found to differ considerably from pure chance distributions, in the same way as did those of the munition-workers, *i.e.* an excess of people with no accidents and also of those with several.

If, instead of being based on the observed mean, the Poisson series is based on the proportion who escape accident altogether, the excess of people observed with repeated accidents is increased. Where the numbers were too small for a good fit to be expected in any single occupational group, separate Poisson series were fitted to each small group and the results summed for comparison with the whole factory, but the same difference always appeared. Of the other two hypotheses discussed by Dr. Greenwood and Mr. Yule, that of differing individual liabilities alone had any success with these data. Some examples of these distributions are given in Table I, other groups are given in the original report; in all cases the Greenwood-Yule curve is a great improvement on the Poisson, in some it is far from a good fit, but the results on the whole are fair. The third example in Table I refers to a group of engine-fitters' apprentices, whose records were kindly given me by Mr. Farmer.

Many factories now keep records of small accidents as a matter of routine, but not so many make any use of them. Quite apart from the relevance to the possibility of discovering people potentially unsafe for dangerous jobs, it is of some practical importance for reducing accidents in a department to know if a high rate is unduly affected by a few people, or if the cause is to be sought in the general conditions which are alike for all. A quick, rough test is given by comparing the observed mean number of accidents with the mean expected on the basis of equal risk from the percentage of people with no accidents.*

It is a common experience in a factory that an influx of new workers means an increase in accident rates; this is sometimes put down to the speeding-up of output which often happens at the same

* If this percentage is $100e^{-m}$, the expected mean will be m. For a table of these values, *see* Industrial Fatigue Research Board Report No. 34, p. 26.

TABLE I.—*Accident Distributions Fitted to the Greenwood-Yule Curve.*

No. of accidents.	Factory B (Whole Factory).					
	Men.			Women.		
	Observed.	Calculated by Greenwood-Yule* Curve.	Sum of Poissons.	Observed.	Calculated by Greenwood-Yule* Curve.	Sum of Poissons.
0	239	250·5	202·0	111	113·5	80·5
1	98	92·9	130·4	66	63·4	84·3
2	57	46·1	65·6	42	36·8	53·5
3	33	24·7	29·5	17	21·7	26·4
4	9	13·8	12·4	10	12·8	12·0
5	2	7·9	4·8	9	7·6	5·5
6	2	4·6	1·6	4	4·5	2·6
7	1	2·7	·5	4	2·7	1·3
8	—	1·6	·2 } ·7	1	1·6 } 6·7	·6
9	4	·9	—	1 } 8	over 8 2·4	·2 } 1·0
10	—	·6 } 6·5	—	1	—	·2
11	1 } 7	·3	—	1	—	—
12	—	·2	—	—	—	—
13	—	over 12 ·2	—	—	—	—
14	1	—	—	—	—	—
15	—	—	—	—	—	—
16	—	—	—	—	—	—
17	—	—	—	—	—	—
18	—	—	—	—	—	—
19	—	—	—	—	—	—
20	—	—	—	—	—	—
	447	—	—	267	—	—
	8 groups χ^2 = 13·7, P = ·058		χ^2 = 75·5, P = < ·00001	8 groups χ^2 = 3·1, P = ·88		χ^2 = 40·0, P = < ·00001

* I take this opportunity to correct an unfortunate error in the calculated values of both these series as given in the Industrial Fatigue Research Board Report No. 34. They should stand as given here, and so give better fits than as originally miscalculated.

TABLE I—*contd.*

No. of Accidents.	Engine-Fitters' Apprentices. Observed.	Engine-Fitters' Apprentices. Calculated by Greenwood–Yule Curve.	Factory M. Group II. (Night-shift. Males.) Observed.	Factory M. Group II. Calculated by Greenwood–Yule Curve.	Factory M. Group II. Single Poisson Fitted to Mean.	Factory M. Group II. Single Poisson Fitted to Zero Group.
0	61	69·9	54	51·4	7·5	54·0
1	55	50·3	56	58·2	29·5	104·7
2	46	33·8	57	53·8	57·6	101·6
3	16	22·2	53	46·0	75·1	65·8
4	18	14·4	29	37·7	73·4	32·0
5	5	9·3	28	30·1	57·4	12·4
6	4	5·9	19	23·7	37·4	4·0
7	4 } 38	3·8 } 39·8	20	18·3	20·9	1·1
8	20 { 2	25·4 { 2·4	9 } 20	14·1 } 24·8	10·2 } 14·6	·3 } ·4
9	3	1·5	11	10·7	4·4	·1
10	1	1·0	11	8·1	1·7	—
11	1	over 10 1·7	7 } 28	6·1 } 18·8	·6 } 2·5	—
12	—	—	10	4·6	·2	—
13	—	—	3	3·4	·1	—
14	—	—	3	2·6	—	—
15	—	—	1	1·9 } 13·3	—	—
16	—	—	2	1·4	—	—
17	—	—	2 } 12	over 16 4·0	—	—
18	—	—	—	—	—	—
19	—	—	—	—	—	—
20	—	—	1	—	—	—
	216	—	376	—	—	—
	6 groups $\chi^2 = 9\cdot8$, P = ·084; 5 ,, $\chi^2 = 7\cdot8$, P = ·10		11 groups $\chi^2 = 10\cdot2$, P = ·42			

time, and sometimes to the inexperience of the new workers. In the present data, it was found that, on the whole, there was a distinct tendency for the younger workers to have more accidents, quite apart from their inexperience, so far as this could be judged from the length of time they had been in the factory. This held for all the groups, both male and female, with only one exception, in which the age-range was very small. Age and length of time in the factory were naturally fairly highly correlated, but when partial correlation was used, the effect of youth appeared greater than that of inexperience. The final correlation between age and accidents was of the order $-\cdot 2$. Occupational mortality statistics show a much higher accident death-rate at the older than at the younger ages, which is in apparent conflict with the greater liability of younger workers, if this holds for more serious accidents also. The difficulty of getting the true "exposed to risk" for occupational mortality is well known, but there is no need to seek for the explanation there as records of compensated accidents show that the number of accidents to one death decreases steeply with age, while the average length of incapacity per accident increases. We are concerned at the moment only with the event of an accident, not with its effect.

A more interesting association that came out very clearly, was that between the number of minor accidents and the occurrence of small ailments necessitating a visit to the ambulance-room for treatment. Records of this were only available for 6 groups of men and 6 of women, but in every case a positive association appeared, of the order of $\cdot 2$ to $\cdot 4$. An obvious criticism is that "tendencies to report" accidents and sickness go together, but as the association was highest in the factory where penalties for not reporting small accidents were most strictly enforced, it is probable that this is far from a complete explanation. The suggestion from the figures, is that so far as these ailments are a measure of lower general health—and this may be true of the "tendency to *report* sickness" just as much as of the "tendency to *be* sick"—accident liability is associated with such a state.

No relation appeared between accident and length of recorded absence from sickness, but for such comparatively short periods the latter is not a very reliable measure.

If there is any practical importance in the fact of differing accident tendency among different people it is necessary to go a little farther and see how far this quality is a stable one. The best measure of this is the correlation between accidents of the same individual in two different periods. Eleven of the factory groups were examined for this, 2 of them gave negligible correlations, the other 9 showed positive coefficients of the order ·2 to ·6, very little lower than those found by Dr. Greenwood and Miss Woods. In some groups, it is true, the possibility of some part of this correlation being due to varying risk is present, and can hardly be avoided, but in 4 groups of one factory—2 of men and 2 of women—the stability could be tested without this complication, as in these groups some of the workers were treated in the ambulance-room for accidents received at home, and the number of "home accidents" and of "factory accidents" showed in all four cases a positive correlation of the order ·2 to ·3.

Dr. Karl Marbe,* Professor of Psychology at Wurzburg, has obtained similar results at opposite ends of the scale of accidents; he examined the numbers of visible small cuts, bruises, scars, etc., on the exposed parts of the body of various sets of school children (124 in number) in two different ten-day periods, taking care that none was recorded twice. He also took the records of 3,000 officers and non-commissioned officers of the German Army from an Accident Insurance Company for a period of ten years, and compared individual records in the first two and last two years, also in the first five and last five. In all cases, those with most in the first period also showed an excess in the second. These officers had been put into three "Danger Classes" by the Insurance Company, according to the risk of their special occupation, and these were taken separately. It is not possible, however, without considerably more knowledge of their conditions, to be satisfied that this classification would remove the possibility of any correlation arising from constant unequal risk.

In the meantime, on the psychological side of the investigation, Mr. Farmer and Mr. Chambers had been trying to devise tests that would correlate with accident tendency, and the results of the factory investigation, though not very definite, seemed to justify the continuation of this work. For the experimental comparison of these tests with accident records ("accident" again includes any injury, however small, reported and treated at the surgery) they were able to use 6 groups of workers; 2 of females (17 and 23 in number, average age 17 to $17\frac{1}{2}$) employed in sweet-covering and sweet-packing, respectively; 2 of dockyard apprentices (57 and 100 in number, average age 16 to 17); and 2 of Royal Air Force apprentices (175 and 279 in number, average ages 19 and 16, respectively).

* *Praktische Psychologie der Unfälle und Betrieb-schäden*, von Dr. Karl Marbe, München o. Berlin, 1926.

Various tests were tried (described in detail in Report No. 38* of the Industrial Fatigue Research Board), and, as may be imagined, the application of these to 650 individuals involved a very large amount of detailed work.

Some of these tests showed a small, but apparently consistent, relation with the crude accident rate. The correlation, however, was only of the order ·11, which is hardly of practical use for predictive purposes. When the combined score of the best of these tests was used, each single score being weighted with the proper partial regression coefficient of accidents on tests, then, in the 4 male groups tested, all those below the average in the tests by this method of scoring, had an accident score on the average 48 per cent. worse than those above the average.

The investigators only give this as a tentative result, subject to confirmation on other data, and point out that a good deal remains to be done before the reliability of these tests can be established.

From the dockyards they were also able to obtain some data bearing on the relation of major to minor accidents. For this purpose a "major" accident was defined as one entailing at least one day's absence from work. The major and minor accidents of the same person were correlated from a year's records of 14,524 dockyard workers, taken in 10 different trade groups separately, 6 of these groups showed a significantly positive correlation of the average order ·25, the other groups all gave correlations too small to be of any meaning, and 3 of these were negative.

On the whole, there is some indication that the same people are likely to incur both small and major accidents, but it is not a very definite result. The small average number of major accidents that happen in the short period of one year probably makes it difficult for correlation to appear.

To go back now to the more theoretical side. The next step was an attempt to get back, by some simple statistical modifications, from the number of accidents observed to the underlying individual liabilities, of which the actual accidents in any limited period can only be an imperfect measure. We again make some assumptions in order to get a standard scheme for comparison. Suppose that we have N persons exposed to the same outward conditions as far as possible. Assume for the time being that differing individual liability exists, that it is constant in time for the same person but

* "A Psychological Study of Individual Differences in Accident Liability," by Eric Farmer, M.A., and E. G. Chambers, M.A., 1926.

differs from one person to another. Let $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_n$, be the average number of accidents per person per period chosen for observation, of the people with 1st, 2nd, . . . nth degree of liability, that would be expected to occur, if we could observe either one person in each of these groups over a very large number of such periods, or a very large number of persons in each of these groups over a single such period under the same conditions. The λ's are, of course, a function of the length of unit period chosen, and may be supposed to contain, in most practical cases, as well as the actual personal tendency, any constant external differential risk affecting particular persons which may exist in the same occupational group. The actual measure we have for any person is the number A of his accidents in a single period, *i.e.* $\lambda_s +$ a sampling error. We know that this sampling error may be large, and that, if it is purely random, its effect will be that the expected number of accidents in the member of a group all having the same λ, λ_s, say, would be distributed as a Poisson series with mean λ_s. We cannot, of course, by any statistical modification distinguish between that part of the λ which is personal and that which depends on constant external bias—such bias must be guarded against, as far as possible, by careful choice of conditions, and the unavoidable remainder in any particular case discovered by the study of the particular people concerned—but we can, to some extent, make allowance for the random sampling error, assuming independent trials.

The Greenwood-Yule curve assumes a particular distribution of the λ's, and this has turned out to be a happy assumption in many cases, but it is possible, in theory, to get a good deal of information about the λ's without any assumption as to the form of their distribution.

If we have a series of sample periods taken on the *same* set of N people, the numbers of accidents observed may be expected to behave in the same way as the results of taking samples of N in fixed proportions as regards the λ's from a set of Poisson distributions about different fixed means. This type of compound sampling is a particular case of one of those which have been discussed in detail by Tschuprow.* In his more general case it is not possible, as a rule, to get back from the constants of the observations to those of the submeans, but, owing to the convenient fact that a Poisson series is completely given by one constant only—the mean—we are in our particular case able to do this, approximately, by taking the observed mean number of accidents for the whole group as an estimate of

* *See* references I and II given in Appendix.

the mean λ for the whole group. From this we can get the mean values of the moments of the different Poisson series, and so substitute in the general equations given by Tschuprow. Appendix I gives the details of this and the resulting equations necessary to answer the more obvious questions that arise in our particular problem. Though some of these results will be familiar to many, I have included them in full because it seems possible this compound Poisson sampling scheme may often arise in practical questions of an entirely different kind—as we find the simple Poisson does, and it may be of use to others to have these formulæ all together in a simple and accessible form. The reason for arriving at the more familiar results indirectly through this not very attractive notation is partly for the sake of uniformity, but mainly so that this particular application may serve as a simple introduction to Tschuprow's methods and the more complicated notation that can hardly be avoided in the general case. § 3 of the Appendix gives the moments of the λ's in terms of the expected moments of the A's, and incidentally the expected moments of the A's in terms of the moments of the λ's. The product of the mean and second moment has been added, as it is wanted for the fourth moment.

As the choice of our length of period of observation is not always at our disposal, we may need to compare periods of different length. § 4 estimates the effect on the lower moments of changing the length of the period of observation, or in a space-period of the area, incidentally illustrating Bortkiewicz's Law of Small Numbers. § 5 gives the sampling errors (always assuming that we are taking repeated samples on the *same* set of people, so that the λ's keep their original values) of the mean, and second moment of the A's (whether found directly or estimated from a longer or shorter period), and of $\mu_2^{(A)}$, as found from the A's. § 6 estimates the expected correlation between the λ's and the A's; between the λ's and any other measure x, from the observed correlation between the A's and x; and also the expected correlation between the accidents of the same individual in two periods, and the effect on this of changing the period of observation. In the Reports I have quoted, these correlations have been given in their crude form, with no allowance for the sampling errors. The method of allowing for this is the same as that used by Cobb* to correct for errors of observation, and also similar to the use made of reliability coefficients by psychologists,

* "The Effect of Errors of Observation upon the Correlation Coefficient," by J. A. Cobb, *Biometrika*, vi, p. 109.

the assumption being made in both cases that the errors are uncorrelated with each other and with the true value. (I prefer to compare the observed with expected correlations, rather than raise the value of the observed by dividing by the expected; the former method draws attention to the low correlation attainable in the observed values, while the latter stresses the hypothetical higher correlation behind this, both mean the same thing, but the first is less liable to misinterpretation.)

From these formulæ the moments and β_1 and β_2 of the λ's have been calculated from those of the A's for some of the distributions of factory accidents, and from another sample relating to engine-fitters' apprentices which Mr. Farmer has kindly given me. The results are given in Table II. I do not think that they are of any practical value in our case as regards the higher moments, as the A's all give J-shaped distributions, which must have large probable errors attached to their higher moments. Three of them give for the distributions values of β_1 and β_2 which fall in the impossible area. We might interpret this as possibly due to the large probable error, or as showing that the theory of constant λ's does not hold in these cases. If the β's in these tables mean anything at all, they show that the λ distributions fall, some in the Type VI_J, some in the Type I_J area. The Type III curve chosen by Dr. Greenwood and Mr. Yule falls on the dividing line between these two areas, so that it probably describes the data as well as any one type would do. The λ distributions are very little less asymmetric than the A distributions.

Table III compares the correlation already mentioned as found between the accidents of the same person in two different periods with that expected on the constant λ theory. The agreement is in most cases good. In 9 cases out of 12 the difference is under three times the probable error of r. The effect of the chance variations in lowering the expected maximum value of the correlation attainable is very noticeable.

In choosing a group of people on whom to experiment with tests, the first step is to see whether $\sigma_\lambda{}^2$ exists, *i.e.* if $\sigma_A{}^2 - \overline{A}$ is positive. If it is not positive there is no indication of real differentiation among the members of the group, and to try tests on such a group before one had watched them long enough to get a real value for σ_λ would be idle. Increasing either the number of people in the group, or the period of observation, will give us a better estimate of the mean, but to get a better approximation to the true individual variation it is better to increase the period of observation.

TABLE II.—*Constants of the* λ *and* A *Distributions.*

	Whole Factory A. Men.		Factory A.						Whole Factory B.			
			Group I.		Group II.		Group III.		Men.		Women.	
	N = 3420.		N = 1312.		N = 1330.		N = 777.		N = 437.		N = 376.	
	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ
Mean	3·98	3·98	6·44	6·44	3·78	3·78	2·56	2·56	·98	·98	1·41	1·41
μ_2	22·85	18·88	41·88	35·45	17·16	13·38	10·56	8·0	2·58	1·60	3·55	2·14
μ_3	238·17	177·57	391·80	279·02	160·80	116·89	67·00	40·44	13·96	8·18	13·79	5·96
μ_4	5102·08	3355·24	9114·42	5568·36	3655·85	2468·20	844·73	381·65	134·48	58·09	102·30	20·12
β_1	4·75	4·69	2·09	1·75	5·12	5·71	3·82	3·19	11·40	16·26	4·24	3·62
β_2	9·77	9·42	5·20	4·43	12·42	13·79	7·58	5·96	20·25	22·64	8·11	4·39
Skewness	1·22	1·35	1·33	1·84	·78	·78	1·72	4·85	1·65	7·80	1·87	—·80
$\frac{100\,\sigma\,\lambda}{\bar{\lambda}}$	—	109·3	—	92·5	—	96·8	—	110·7	—	129·3	—	103·7
Type of curve	I_J	I_J	I_J	I_J	VI_J	VI_J	I_J	I_J	VI_J	I_J	I_J	In the impossible area $\beta_2 - \beta_1 - 1$ negative

TABLE II.—*contd.*

	Whole Factory E. Men.		Engine-Fitters' Apprentices.							
	N = 4915.		1st year. N = 216.		2nd year. N = 216.		3rd year. N = 216.		All 3 years. N = 216.	
	A	λ	A	λ	A	λ	A	λ	A	λ
Mean	1·36	1·36	·52	·52	·69	·69	·70	·70	1·92	1·92
μ_2	7·18	5·82	·77	·25	1·03	·34	1·20	·51	5·10	3·20
μ_3	89·90	71·08	1·54	·28	1·91	·23	2·78	·59	28·39	17·00
μ_4	1987·03	1459·86	6·21	·74	7·38	·21	11·59	·38	332·31	160·06
β_1	21·84	25·63	5·21	5·37	3·38	1·42	4·47	2·71	6·06	8·85
β_2	38·55	43·10	10·54	12·07	6·99	1·83	8·03	1·50	12·76	15·67
Skewness	1·84	2·23	1·24	·91	1·62	—·35	2·68	—·21	1·05	1·71
$\frac{100\,\sigma\,\lambda}{\bar{\lambda}}$	—	177·4	—	95·1	—	83·4	—	101·7	—	93·3
Type of curve	VI_J	VI_J	I_J	VI_J	VI_J	Impossible area	I_J	Impossible area	VI_J	I_J

TABLE III.—*Comparison between Observed and Estimated Correlation between the Accidents of the Same Individuals in Two Periods.*

Groups.	Observed r.	Estimated r.	Difference.	No. of people.	Length of periods. 1st.	Length of periods. 2nd.
DI Females	$\cdot 21 \pm \cdot 15$	·63	Between 2 and 3 times Probable Error	19	2 years	5 months
DII ,, ...	$\cdot 36 \pm \cdot 09$	·27	Just over P.E.....	42	2 ,,	5 ,,
EI Males	$\cdot 57 \pm \cdot 02$	·59	Under P.E.	445	1 year	1 year
EII ,, ...	$\cdot 25 \pm \cdot 04$	·32	Between 1 and 2 times P.E.	288	1 ,,	1 ,,
EIII ,, ...	$\cdot 20 \pm \cdot 04$	·32	Just over 3 times P.E.	226	1 ,,	1 ,,
EIV ,, ...	$\cdot 62 \pm \cdot 03$	·54	Between 2 and 3 times P.E.	288	1 ,,	1 ,,
GI ,, ...	$\cdot 36 \pm \cdot 09$	·50	Between 1 and 2 times P.E.	47	1 ,,	1 ,,
GII ,, ...	$\cdot 57 \pm \cdot 05$	·56	Under P.E.	82	1 ,,	1 ,,
GI Females	$\cdot 53 \pm \cdot 04$	·41	Between 2 and 3 times P.E.	120	1 ,,	1 ,,
GII ,, ...	$-\cdot 01 \pm 1 \cdot 0$	·51	Between 5 and 6 times P.E.	50	1 ,,	1 ,,
H ,, ...	$\cdot 05 \pm \cdot 05$	·40	Between 7 and 8 times P.E.	227	6 months	6 months
Engine-fitters' apprentices*	$\cdot 28 \pm \cdot 08$	·41	Between 2 and 3 times P.E.	57	1 year	1 year

* Data kindly supplied by Mr. E. Farmer for the apprentices. Data for the other groups taken from Industrial Fatigue Research Board Report No. 34.

The best criterion for the variability of a group is $\frac{100\sigma_\lambda}{\bar{\lambda}}$, *i.e.* $\frac{100\sqrt{\sigma_A{}^2 - \bar{A}}}{\bar{A}}$. For the factory workers this was of the order 100,* though it varied considerably from one group to another. In only one case here, and in one of Dr. Greenwood's munition groups, was $\sigma_A{}^2 < \bar{A}$. We may note that the constant k on which the form of the Greenwood-Yule curve depends is the square of the reciprocal of this criterion, *i.e.* $\frac{1}{k} = \frac{\sigma_A{}^2 - \bar{A}}{\bar{A}^2}$.

Mr. Chambers and Mr. Farmer, in Report No. 38, have examined their tests by comparing the mean number of accidents of persons above and below the mean in the tests. Here, too, as in the case of the correlation coefficient, unless we have a criterion for the maximum attainable ratio, it is difficult to judge the value of the test. If we assume that the Greenwood-Yule curve holds, which in any case is not a wildly wrong assumption, we can get the ratio of the accidents of those above the mean in tests and those below, if there were perfect correlation between the tests and λ, from the equation

$$\frac{\bar{A}_A}{\bar{A}_B} = \frac{I_k(k-1)\,[1 - I_k(k)]}{I_k(k)\,[1 - I_k(k-1)]}$$

where $I_x(p) = \frac{\Gamma_x(p+1)}{\Gamma(p+1)}$, when k is known.

The values of k in the distributions of these reports usually fall between ·3 and 1·5. (The smaller k is the more scope for differentiation by tests.) Table IV gives on these assumptions the expected ratio for perfect correlation for different values of k. It seems simpler, however, to deal directly with the correlation coefficient itself. If we correct the correlations of Mr. Farmer's combined tests with accidents for the various groups given in Report No. 38, we find that they are only raised from an average of ·11 to one of ·16 (*see* Table V), so that they are far from being of predictive value as yet, though the consistently positive sign suggests that the correlation is real. Further work on these tests is now in progress.

* *See* Table VI, p. 24, of Report No. 34.

TABLE IV.—*Expected Value for Perfect Correlation between a Test x and λ, of the Ratio* $\frac{\bar{A}_A}{\bar{A}_B}$ *of the Mean Number of Accidents of Persons worse than the Mean in Test x, to that of Persons better than the Mean in Test x, for different Values of k* $\left(k = \frac{\bar{A}^2}{\sigma_A^2 - \bar{A}}\right)$, *assuming that the λ's follow a Type III distribution.*

k	·2	·3	·4	·5	·6	·7	·8	..	1·0
$\frac{\bar{A}_A}{\bar{A}_B}$	24·1	14·9	10·9	8·7	7·3	6·4	5·7	..	4·8
k	1·1	..	1·3	..	1·5	..	2·0	....	9·0
$\frac{\bar{A}_A}{\bar{A}_B}$	4·5	..	4·0	..	3·6	..	3·1	....	1·7

TABLE V.—*Correlation between λ's and Neuro-muscular Tests (x) estimated from the Correlation between the A's and the x's.*

	C.	D.	E (1).	E (2).	F.
Observed r_{xA}*	·14	·16	·09	·03	·13
Estimated $r_{x\lambda}$	·20 ± ·09	·24 ± ·06	·10 ± ·07	·03 ± ·07	·21 ± ·04

* Data kindly supplied by Mr. E. Farmer. Groups as in Report No. 38.

Repeated Sickness Claims.

The correlation found between small accidents and visits to the ambulance-room for minor ailments among the factory workers (Report No. 34) was confirmed in the only one of the groups used for the experimental work on tests for which the data were available. Also, as before, when sickness was measured by the total length of time absent, no correlation appeared, but in most cases the time of observation was really too short to expect individual tendencies to appear clearly by this standard.

The two measures of sickness available—length and number of attacks—probably measure quite different things, and both should really be considered together. It is of interest to get some idea of the extent of individual variation in sickness. As we had easily available a number of National Health Insurance records, covering five years' experience, I took two samples of them out of curiosity to see how the form of the distributions of repeated claims compared with those of minor accidents. The records were those of printers only, and were kindly put at my disposal by Dr. A. B. Hill. They form part of a much larger number he has collected for a special inquiry into the health of printers that he is undertaking for the Medical Research Council. I have taken into account the number of *claims* only (sickness and disablement both included), entirely ignoring length of incapacity. I am well aware this is open to objection, as those for instance, who had long periods of incapacity had not the opportunity for so many claims as those with short, but in the time at my disposal was not able to consider the length of incapacity as well. Apart from this, all were exposed to risk for the whole five years. The distribution of claims is, I think, one that one would expect to follow the same compound Poisson scheme that we have been assuming for the accidents. Tables VI and VII show the distribution by claims in a sample of male and one of female* printers in the five years 1921–5 inclusive. They differ from the pure chance schemes (whether based on the mean or on the proportion with no claims), in the same way as did the accidents—an excess of multiple claims; and the relative variation $\frac{100\sigma_\lambda}{\bar{\lambda}}$, when freed as far as possible from sampling error, is of about the same order as that for accidents (*see* Table VIII), and is so large for the single ages, that it is not increased by taking all ages together. The Greenwood-Yule curve based on the Type III distribution of λ's gives a reasonable fit both to females (all ages) and to the largest, *i.e.* 35—39† years age-group of males, but does not describe all ages of the males as well (*see* Table IX). Table X gives the A and λ moments found as before, but here, too, the distribution is too skew for high moments to be of value and again we get a case in the impossible area. These are, of course, only very small and select samples, and without any standard of variability for comparison, one cannot say if they are abnormal or not. The report of the Departmental Actuarial Committee to the Royal Commission on National Health Insurance‡ (1926) gives the proportion of claimants for different

* The female group was too small to subdivide into married and single. The main bulk were single up to age 35; after age 45 about half were married.

† As the experience covers five years, the five-year age-groups cover ten years of age and overlap.

‡ Cmd. 2596.

TABLE VI.—*Printers.* (*Males.*) *Distribution of Claims* (*Sickness or Disablement*) *in the Five-Year Period* 1921–5.

Age last Birthday on Jan. 1, 1921 ...	16–19.			20–24.			25–29.			30–34.			35–39.			40–44.		
		Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on	
No. of Claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.
0	71	57·4	71·0	118	98·9	118·0	95	78·2	95·0	173	130·7	173·0	247	196·7	247·0	242	183·7	242·0
1	19	36·7	30·4	59	79·1	73·4	57	69·0	65·3	83	128·3	121·4	152	191·4	184·0	130	173·5	162·0
2	8	11·8	6·6	27	31·6	22·9	21	30·5	22·5	51	63·0	42·6	66	93·3	68·9	49	82·2	54·4
3	9	2·6	1·0	10	8·4	4·8	6	9·0	5·2	24	20·7	10·0	32	30·4	17·3	21	26·0	12·3
4	2	·4	·1	3	1·7	·8	3	2·0	·9	9	5·1	1·8	14	7·4	3·3	14	6·2	2·1
5	—	·1	—	1	·3	·1	6	·4	·1	6	1·0	·2	3	1·5	·5	11	1·2	·3
6	—	—	—	1	—	—	—	·1	—	2	·2	—	3	·2	·1	4	·2	—
7	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—
8	—	—	—	—	—	—	1	—	—	1	—	—	1	—	—	—	—	—
9	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—
10	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
11	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—	—	—	—
12	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
13	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
14	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
15	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—
Total	109	—	—	220	—	—	189	—	—	349	—	—	521	—	—	473	—	—

TABLE VI—*contd.*

Age last Birthday on Jan. 1, 1921.... ...	45–49.			50–54.			55–59.			60–64.			All Ages.		
		Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on	
No. of Claims.	Ob-served.	(i) Mean.	(ii) Pro-portion with no claims.	Ob-served.	(i) Mean.	(ii) Pro-portion with no claims.	Ob-served.	(i) Mean.	(ii) Pro-portion with no claims.	Ob-served.	(i) Mean.	(ii) Pro-portion with no claims.	Observed.	(i) Mean.	(ii) Pro-portion with no claims.
0	189	122·3	189·0	139	87·1	139·0	73	42·3	73·0	29	18·2	29·0	1376	999·3	1376·0
1	96	142·3	137·7	91	111·9	113·6	40	58·3	60·8	19	24·2	25·1	746	1030·3	979·3
2	43	82·9	50·3	35	71·9	46·5	24	40·2	25·4	8	16·2	11·0	332	532·2	349·2
3	28	32·3	12·3	19	30·9	12·7	12	18·5	7·1	5	7·2	3·2	166	183·6	83·2
4	19	9·4	2·3	11	9·9	2·6	12	6·4	1·5	2	2·4	·7	89	47·6	14·9
5	3	2·2	·3	7	2·6	·4	3	1·8	·3	4	·6	·1	44	9·9	2·1
6	5	·4	—	3	·6	·1	—	·4	—	1	·1	—	19	1·7	·3
7	4	·1	—	2	·1	—	1	·1	—	—	—	—	8	·3	—
8	3	—	—	4	—	—	—	—	—	1	—	—	11	—	—
9	2	—	—	1	—	—	—	—	—	—	—	—	4	—	—
10	—	—	—	1	—	—	—	—	—	—	—	—	2	—	—
11	—	—	—	—	—	—	1	—	—	—	—	—	3	—	—
12	—	—	—	1	—	—	1	—	—	—	—	—	2	—	—
13	—	—	—	1	—	—	—	—	—	—	—	—	1	—	—
14	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
15	—	—	—	—	—	—	1	—	—	—	—	—	2	—	—
Total	392	—	—	315	—	—	168	—	—	69	—	—	2805	—	—

TABLE VII.—*Printers.* (*Females.*) *Distribution of Claims* (*Sickness or Disablement*) *in the Five-Year Period* 1921–5.

Age last Birthday on Jan. 1, 1921	16–19.			20–24.			25–29.			30–34.			35–39.			40–44.		
		Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on	
No. of Claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.
0	25	18·7	25·0	79	52·9	79·0	57	43·6	57·0	49	34·8	49·0	56	39·5	56·0	43	28·4	43·0
1	23	24·1	25·0	36	62·7	61·9	38	49·9	49·9	32	38·8	37·8	19	33·4	27·8	11	27·9	24·4
2	10	15·6	12·5	33	37·1	24·3	25	28·6	21·9	6	21·6	14·6	9	14·2	6·9	12	13·8	7·0
3	3	6·7	4·2	11	14·7	6·4	9	10·9	6·4	10	8·0	3·8	2	4·0	1·1	5	4·5	1·3
4	4	2·2	1·0	3	4·4	1·3	4	3·1	1·4	5	2·2	·7	1	·9	·1	3	1·1	·2
5	—	·6	·2	8	1·0	·2	1	·7	·3	2	·5	·1	1	·2	—	—	·2	—
6	2	·1	—	3	·2	—	1	·1	—	—	·1	—	2	—	—	1	—	—
7	—	—	—	—	—	—	1	—	—	2	—	—	2	—	—	1	—	—
8	1	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—
Total	68	—	—	173	—	—	137	—	—	106	—	—	92	—	—	76	—	—

TABLE VII—*contd.*

Age last Birthday on Jan. 1, 1921 ...	45–49.			50–54.			55–59.			60–64.			All Ages.		
		Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on			Expected from chance based on	
No. of Claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.	Observed.	(i) Mean.	(ii) Proportion with no claims.
0	31	29·2	31·0	23	16·3	23·0	18	16·4	18·0	6	2·5	6·0	387	278·3	387·0
1	27	26·4	26·1	14	17·9	17·4	13	13·8	13·4	3	4·5	5·5	216	302·6	292·9
2	8	11·9	11·0	5	9·9	6·6	3	5·8	5·0	2	4·0	2·5	113	164·7	111·3
3	3	3·6	3·1	3	3·6	1·7	3	1·6	1·3	1	2·4	·8	50	59·8	28·3
4	2	·8	·7	2	1·0	·2	1	·4	·2	1	1·1	·2	26	16·3	5·4
5	1	·2	·1	—	·2	—	—	·1	—	—	·4	—	13	3·6	·8
6	—	—	—	1	—	—	—	—	—	1	·1	—	11	·7	·1
7	—	—	—	1	—	—	—	—	—	1	—	—	8	·1	—
8	—	—	—	—	—	—	—	—	—	—	—	—	2	—	—
Total....	72	—	—	49	—	—	38	—	—	15	—	—	826	—	—

TABLE VIII.—*Printers.* *Mean and Variance of Number of Claims (Sickness and Disablement) in the Five-Years* 1921–5, *and Relative Variation freed from the effect of Random Sampling.*

MALES.

Ages....	16—19.		20—24.		25—29.		30—34.		35—39.		40—44.		45—49.		50—54.		55—59.		60—64.		All Ages.	
	N = 109.		N = 220.		N = 189.		N = 349.		N = 521.		N = 473.		N = 392.		N = 315.		N = 168.		N = 69.		N = 2805.	
	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ
Mean	·6	·6	·8	·8	·9	·9	1·0	1·0	1·0	1·0	1·0	1·0	1·2	1·2	1·3	1·3	1·4	1·4	1·3	1·3	1·0	1·0
μ_2	1·1	·5	1·5	·7	1·6	·8	1·7	·7	1·9	1·0	2·1	1·2	2·8	1·6	3·8	2·5	4·3	3·0	3·0	1·6	2·4	1·4
$\frac{100\sigma_\lambda}{\bar{\lambda}}$	—	104·5	—	102·9	—	98·2	—	86·6	—	99·9	—	114·8	—	109·5	—	122·6	—	124·6	—	96·1	—	112·3

FEMALES.

Ages....	16—19.		20—24.		25—29.		30—34.		35—39.		40—44.		45—49.		50—54.		55—59.		60—64.		All Ages.	
	N = 68.		N = 173.		N = 137.		N = 106.		N = 92.		N = 76.		N = 72.		N = 49.		N = 38.		N = 15.		N = 826.	
	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ	A	λ
Mean	1·3	1·3	1·2	1·2	1·2	1·2	1·1	1·1	·9	·9	1·0	1·0	·9	·9	1·1	1·1	·8	·8	1·8	1·8	1·1	1·1
μ_2	2·6	1·3	2·2	1·0	2·0	·9	2·3	1·2	2·4	1·5	2·1	1·2	1·2	·3	2·4	1·3	1·1	·2	4·8	3·0	2·2	1·1
$\frac{100\sigma_\lambda}{\bar{\lambda}}$	—	88·0	—	84·9	—	81·7	—	97·4	—	145·4	—	109·0	—	57·4	—	104·1	—	58·0	—	96·7	—	95·8

TABLE IX.—*Printers, 1921–5. Distribution of Claims for Sickness and Disablement, Fitted to the Greenwood-Yule Curve.*

No. of Claims in 5 years.	Females. (All Ages.)		Males. (All Ages.)		Males. (Age at beginning of period 35–39 years.)	
	Observed.	Calculated.	Observed.	Calculated.	Observed.	Calculated.
0	387	388·3	1376	1447·5	247	263·7
1	216	211·5	746	649·6	152	130·3
2	113	110·2	332	329·5	66	64·3
3	50	56·7	166	173·5	32	31·8
4	26	29·0	89	93·1	14	15·7
5	13	14·7	44	50·5	3	7·7
6	34 { 11 } 60	30·3 { 7·5 } 59·3	19	27·6	3	3·8
7	8	3·8	8	15·1	24 { — } 56	30·9 { 1·9 } 62·7
8	2	4·3	11	8·3	1	·9
9	—	—	4	4·6	1	·5
10	—	—	52 { 2 } 96	61·3 { 2·6 } 111·8	—	·2
11	—	—	3	1·4	2	·2
12	—	—	2	·8	—	—
13	—	—	1	·4	—	—
14	—	—	—	·2	—	—
15	—	—	2	·3	—	—
	826	826	2805	2805	521	521
	For 5 groups	$\chi^2 = \cdot 98$, P $= \cdot 91$	For 6 groups	$\chi^2 = 20\cdot 6$, P $= \cdot 00098$	For 4 groups	$\chi^2 = 5\cdot 4$, P $= \cdot 15$
	,, 6 ,,	$\chi^2 = 1\cdot 73$, P $= \cdot 88$			,, 5 ,,	$\chi^2 = 6\cdot 3$, P $= \cdot 18$
	,, 9 ,,	$\chi^2 = 9\cdot 1$, P $= \cdot 34$				

TABLE X.—*Printers.*

Moments of the Distribution of Claims for Sickness and Disablement Benefit in the Five Years 1921–5. A, *Crude;* λ *freed from the Effect of Random Sampling.*

	Males. (All Ages.)		Females. (All Ages.)	
	A.	λ.	A.	λ.
Mean	1·0	1·0	1·1	1·1
μ_2	2·4	1·4	2·2	1·1
μ_3	10·4	5·3	6·0	1·7
μ_4	90·5	36·7	32·6	3·2
β_1	8·0	11·6	3·5	2·2
β_2	16·0	20·3	6·9	2·7
Skewness	1·2	1·7	2·1	—·5
$\frac{100\sigma_\lambda}{\bar{\lambda}}$	—	112·3	—	95·8
Type of curve	VI_J	VI_J	I_J	Impossible area
Number of observations	2,805		826	

ages for a typical sample from selected societies; but without the mean number of claims per person in the same groups, one cannot judge how much real individual variability this denotes apart from chance. That the individual variation is very large is, however, clear from the following conclusion drawn by the same Actuarial Committee after an examination of the after-history of a number of insured persons who had drawn disablement benefit*: "It is thus evident, that in respect of both men and women, the cases of frequent claims (which include many prolonged claims) must account for an appreciable amount of the total expenditure on sickness and disablement benefit. From the point of view of public health, as well as from that of administration, there is undoubtedly much that deserves attention in this feature of the working of the Health Insurance System."

The points that the distributions of minor accidents and of these sickness claims have in common, is that the extreme skewness still remains, even when the random sampling element is removed; and, so far as these very limited samples go, the underlying form of both is not very far different, and the corrected relative variability of about the same order.

* *Loc. cit.*, p. 367.

For fear of misunderstanding, may I make it quite clear that the variable denoted here by λ is not meant to represent any mysteriously vague quality called accident or sickness tendency. To speak of the distribution of such a " tendency " has, of course, no meaning at all until the measure chosen to represent it is clearly defined and scaled. Different measures and scales may give quite different forms of distribution, and the choice for such a vague quality, where not dictated by necessity, is largely arbitrary. In our present cases, the λ only means the average number of accidents or of sickness claims that may be expected *in the long run* under the given conditions. We know that in both cases it must be a very inadequate measure, either of general carefulness, or aptness, or whatever is the underlying quality that avoids accidents, or of general health, if only for the reason that it does not differentiate the varying degrees of these qualities finely enough. Just as a detailed medical examination could give a much more accurate health classification, and might possibly differentiate far more between people at one part of the scale of " claims " than at another, and thus alter the form of the distribution altogether, so it would not necessarily condemn a test as useless if it gave at first trial a very different distribution from that of observed accidents, provided that a suitable alteration of scale in the test could bring about a linear relation between the two. The low correlations observed between the tests and accidents do not, however, appear to be due to any such want of linearity, and in this case it does not seem that any reasonable adjustment of the scale of marking the tests tried would appreciably raise the correlation.

Conclusion.

That there are really individual differences in liability to sustain accidents when other things are equal might well be regarded as a matter of common knowledge. Dr. Greenwood and Mr. Yule brought this to the test of statistical measurement and suggested ways of measuring variations of liability. The work done since the publication of their paper has been directed to an improvement of the statistical technique of description and to seeking measures of individual qualities correlated with the unknown characters upon which liability to sustain accidents depends.

It has been shown that there is some evidence that a sorting out of the persons who ought not to be placed in particularly dangerous positions can be done. But it is quite clear that the results are not yet sufficiently definite to provide a basis for administrative proposals. I think, however, that enough has been done to warrant the conclusion that research in this field should be continued. The great importance of the end justifies a continuance of the inquiry even if it is sure to be a long one.

On the purely statistical side, some of the methods may, I think, be found of service in problems seemingly remote from that which was the starting-point.

APPENDIX.

COMPOUND POISSON SAMPLING SCHEME FOR CONSTANT SUBMEANS.

§ 1. *Index.*

---

* *i.e.* non-overlapping.

§ 2. *Notation and General Formulæ.*

Let N denote the total number of observations, A_i any particular observed value of A, and $\bar{A}$, $\mu_2^{(A)}$, $\mu_3^{(A)}$, $\mu_4^{(A)}$ the observed mean, 2nd, 3rd and 4th moment coefficients of the A's about the observed mean.

Let λ_i be any particular value of λ and $\bar{\lambda}$, $\mu_2^{(\lambda)}$, $\mu_3^{(\lambda)}$, $\mu_4^{(\lambda)}$ the mean, 2nd, 3rd and 4th moment coefficients of the λ's about their mean.

Let E (x) denote the mathematical expectation of x, i.e. the sum of every possible value of x multiplied by the probability of that value.

I refers to A. A. Tschuprow, " Zur Theorie der Stabilität statistischer Reihen " (*Skandinavisk Aktuarietidskrift*, 1919, pp. 80–133).

II refers to A. A. Tschuprow, " On the Mathematical Expectation of the Moments of Frequency Distributions," Part II (*Biometrika*, XIII, pp. 283–295).

The general sampling scheme discussed in the relevant sections of these two papers is as follows:—From an unhomogeneous population of N chance variables, each variable following its own law of frequency, N independent draws of single individuals are made, the first draw on the variable x_1, the second on x_2, and so on. The following notation is used:—

$$\mathrm{E}\,(x_i) = m_1^{(i)}, \qquad \mathrm{E}\,(x_i^2) = m_2^{(i)}, \qquad \mathrm{E}\,(x_i^k) = m_k^{(i)},$$

$$\mathrm{E}\,[x_i - m_1^{(i)}]^2 = m_2^{(i)} - [m_1^{(i)}]^2 = \mu_2^{(i)}, \quad \mathrm{E}\,[x_i - m_1^{(i)}]^k = \mu_k^{(i)},$$

$$\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} m_1^{(i)} = m_{[1\mathrm{N}]}, \qquad \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} m_k^{(i)} = m_{[k\mathrm{N}]}, \qquad \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} \mu_k^{(i)} = \mu_{[k\mathrm{N}]},$$

$x_i' =$ observed value drawn for the variable x_i,

$$\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} x_i' = x_{(\mathrm{N})}, \qquad \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} [x_i' - x_{(\mathrm{N})}]^k = \nu'_{[k\mathrm{N}]}, \quad \mathrm{E}\,\{\nu'_{[k\mathrm{N}]}\} = \nu_{[k\mathrm{N}]}.$$

As an example of Tschuprow's notation, I give in full (*see* vi(a) below) the working for $\nu_{[4\mathrm{N}]}$, the expectation of the 4th moment coefficient of the observed values, for the sake of those who may find it difficult to follow the general analysis without some practice on a particular case. This is the most complex of those required for my purpose.

Hence, in the notation of I and II,* $x_{(N)} = \bar{A}$, $\nu'_{[kN]} = \mu_{(k)}^{(A)} =$ observed kth moment coefficient of the A's about their observed mean.

$m_1^{(i)} = \lambda_i$, $m_{[1N]} = \bar{\lambda}$, $\frac{1}{N}\sum_{i=1}^{N} [m_1^{(i)} - m_{[1N]}]^k = \mu_k^{(\lambda)}$, and since the distributions of the separate series about their own means $m_1^{(i)}$ or λ_i, are all assumed to be Poisson series, we have, since $E(x_i - m_1^{(i)})^k = \mu_k^{(i)}$,

$\mu_2^{(i)} = \lambda_i$, and hence, since

$$\mu_{[kN]} = \frac{1}{N}\sum_{i=1}^{N} \mu_k^{(i)} \qquad \mu_{[2N]} = \bar{\lambda}$$

$$\mu_3^{(i)} = \lambda_i \qquad\qquad \mu_{[3N]} = \bar{\lambda}$$

$$\mu_4^{(i)} = \lambda_i + 3\lambda_i^2 \qquad\qquad \mu_{[4N]} = m_{[1N]} + 3\left(\mu_2^{(\lambda)} + \bar{\lambda}^2\right)$$

$$= \bar{\lambda} + 3\mu_2^{(\lambda)} + 3\bar{\lambda}^2.$$

Also, since $m_{[kN]} = \frac{1}{N}\sum_{i=1}^{N} m_k^{(i)}$, where, $m_k^{(i)} = E(x_i^k)$,

$$\therefore\ m_{[2N]} = \frac{1}{N}\sum_{i=1}^{N} (\mu_2^{(i)} + \lambda_i^2) = \mu_{[2N]} + \mu_2^{(\lambda)} + \bar{\lambda}^2 = \bar{\lambda} + \bar{\lambda}^2 + \mu_2^{(\lambda)}$$

$$m_{[3N]} = \mu_{[3N]} + \frac{3}{N}\sum_{i=1}^{N} m_1^{(i)} m_2^{(i)} - \frac{2}{N}\sum_{i=1}^{N} \{m_1^{(i)}\}^3,$$ by the last line of p. 81 of I, or equation (1) on p. 284 of II,

$$= \bar{\lambda} + \frac{3}{N}\sum_{i=1}^{N} \lambda_i(\lambda_i + \lambda_i^2) - \frac{2}{N}\sum_{i=1}^{N} \lambda_i^3$$

$$= \bar{\lambda} + 3\mu_2^{(\lambda)} + 3\bar{\lambda}^2 + \mu_3^{(\lambda)} + 3\mu_2^{(\lambda)}\bar{\lambda} + \bar{\lambda}^3$$

$$\frac{1}{N}\sum_{i=1}^{N} m_1^{(i)} m_2^{(i)} = \frac{1}{N}\sum_{i=1}^{N} \lambda_i(\lambda_i + \lambda_i^2)$$

$$= \mu_2^{(\lambda)} + \bar{\lambda}^2 + \mu_3^{(\lambda)} + 3\mu_2^{(\lambda)}\bar{\lambda} + \bar{\lambda}^3.$$

* The quantity written as $\nu'_{[kN]}$ in I is written as $\nu'_{k(N)}$ in II. The quantity written as $\nu_{[kN]}$ in I is written as $\nu_{k(N)}$ in II.

§ 3. *To get the expectation of the constants of the* A *distribution in terms of those of the* λ *distribution, and hence to estimate the moments of the distribution of* λ *from those of the observed* A *distribution.*

(i) *Mean.*

$$\mathrm{E}(\bar{\mathrm{A}}) = \sum_{i=1}^{\mathrm{N}} \frac{1}{\mathrm{N}} \mathrm{E}(\mathrm{A}_i) = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \lambda_i = \bar{\lambda},$$

$\therefore$ our best estimate of $\lambda = \bar{\mathrm{A}}$.

(ii) *Square of Mean.*

$$\mathrm{E}(\bar{\mathrm{A}}^2) = \bar{\lambda}^2 + \frac{1}{\mathrm{N}} \bar{\lambda} \text{ by 1st equation on p. 82 of I,}$$

or equation (4) on p. 285 of II;

$$\therefore \qquad \bar{\lambda}^2 = \mathrm{E}(\bar{\mathrm{A}}^2) - \frac{1}{\mathrm{N}} \mathrm{E}(\bar{\mathrm{A}}) \text{ by (i)};$$

$\therefore$ our best estimate of $\bar{\lambda}^2 = \bar{\mathrm{A}}^2 - \frac{\bar{\mathrm{A}}}{\mathrm{N}}$ or, approximately, if N is large, $\bar{\mathrm{A}}^2$.

(iii) *Second Moment Coefficient.*

$$\mathrm{E}(\mu_2^{(\mathrm{A})}) = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} [m_1^{(i)} - m_{[1\mathrm{N}]}]^2 + \frac{\mathrm{N}-1}{\mathrm{N}} \mu_{[2\mathrm{N}]}$$

by equation (1) on p. 84 of I,
or equation (1) on p. 284 of II,

$$= \mu_2^{(\lambda)} + \frac{\mathrm{N}-1}{\mathrm{N}} \bar{\lambda}.$$

$\therefore$ $\mu_2^{(\lambda)} = \mathrm{E}(\mu_2^{(\mathrm{A})}) - \frac{\mathrm{N}-1}{\mathrm{N}} \mathrm{E}(\bar{\mathrm{A}})$ from (i),

$\therefore$ our best estimate of $\mu_2^{(\lambda)} = \mu_2^{(\mathrm{A})} - \frac{\mathrm{N}-1}{\mathrm{N}} \mathrm{A}$ or, approximately, if N is large, $\mu_2^{(\mathrm{A})} - \bar{\mathrm{A}}$.

(iv) *Third Moment Coefficient.*

$$E(\mu_3^{(A)}) = \mu_3^{(\lambda)} + \frac{(N-1)(N-2)}{N^2}\bar{\lambda}$$

$$+ \frac{3(N-2)}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\lambda_i^2 - \bar{\lambda}^2\right]$$

by equation (2) on p. 84 of I,
or equation (2) on p. 294 of II,

$$= \mu_3^{(\lambda)} + \frac{3(N-2)}{N}\mu_2^{(\lambda)} + \frac{(N-1)(N-2)}{N^2}\bar{\lambda};$$

$\therefore$ substituting from $\mu_2^{(\lambda)}$ and $\bar{\lambda}$ from (i) and (iii)

$$\mu_3^{(\lambda)} = E(\mu_3^{(A)}) - \frac{3(N-2)}{N}E(\mu_2^{(A)}) + \frac{2(N-1)(N-2)}{N^2}E(\bar{A}).$$

$\therefore$ our best estimate of $\mu_3^{(\lambda)}$

$$= \mu_3^{(A)} - \frac{3(N-2)}{N}\mu_2^{(A)} + \frac{2(N-1)(N-2)}{N^2}\bar{A},$$

or, approximately, if N is large, $\mu_3^{(A)} - 3\mu_2^{(A)} + 2\bar{A}$.

(v) *Product of Mean and Second Moment Coefficient.*

$$E[\mu_2^{(A)}\bar{A}] = E\left\{\frac{1}{N}\sum_{i=1}^{N}(x_i' - x_{(N)})^2 x_{(N)}\right\}$$

$$= E\left\{x_{(N)}\frac{1}{N}\sum_{i=1}^{N}x_i^2\right\} - E(x_{(N)}^3).$$

But $E(x_{(N)}^3) = m^3{}_{[1N]} + \frac{3}{N}m_{[1N]}\mu_{[2N]} + \frac{1}{N^2}\mu_{[3N]}$, by 2nd equation on p. 82 of I, or equation 5 on p. 285 of II,

$$= \bar{\lambda}^3 + \frac{3}{N}\bar{\lambda}^2 + \frac{1}{N^2}\bar{\lambda},$$

$$\text{and } \mathrm{E}\left\{x_{(\mathrm{N})}\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} x_i^2\right\} = \mathrm{E}\left\{\frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}} x_i \sum_{i=1}^{\mathrm{N}} x_i^2\right\}$$

$$= \mathrm{E}\left(\frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}} x_i^3\right) + \mathrm{E}\left(\frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}}\sum_{j\neq i}^{\mathrm{N}} x_i x_j^2\right)$$

$$= \frac{1}{\mathrm{N}} m_{[3\mathrm{N}]} + \frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}}\sum_{j\neq i} \mathrm{E}(x_i)\,\mathrm{E}(x_j^2),$$

since there is no correlation (in the widest sense) between x_i and x_j,

$$= \frac{1}{\mathrm{N}} m_{[3\mathrm{N}]} + \frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}}\sum_{j\neq i} m_1^{(i)} m_2^{(j)}$$

$$+ \frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}} m_1^{(i)} m_2^{(i)} - \frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}} m_1^{(i)} m_2^{(i)}$$

$$= \frac{1}{\mathrm{N}} m_{[3\mathrm{N}]} + \left(\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} m_1^{(i)}\right)\left(\frac{1}{\mathrm{N}}\sum_{j=1}^{\mathrm{N}} m_2^{(j)}\right) - \frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}} m_1^{(i)} m_2^{(i)}$$

$$= \frac{1}{\mathrm{N}} m_{[3\mathrm{N}]} + m_{[1\mathrm{N}]}\, m_{[2\mathrm{N}]} - \frac{1}{\mathrm{N}^2}\sum_{i=1}^{\mathrm{N}} m_1^{(i)} m_2^{(i)}$$

$$= \frac{1}{\mathrm{N}}(\bar{\lambda} + 3\mu_2^{(\lambda)} + 3\bar{\lambda}^2 + \mu_3^{(\lambda)} + 3\mu_2^{(\lambda)}\bar{\lambda} + \bar{\lambda}^3)$$

$$+ \bar{\lambda}^2 + \bar{\lambda}^3 + \bar{\lambda}\mu_2^{(\lambda)}$$

$$- \frac{1}{\mathrm{N}}(\mu_2^{(\lambda)} + \bar{\lambda}^2 + \mu_3^{(\lambda)} + 3\mu_2^{(\lambda)}\bar{\lambda} + \bar{\lambda}^3),$$

$$\therefore\ \mathrm{E}[\mu_2^{(\mathrm{A})}\bar{\mathrm{A}}] = \frac{\mathrm{N}-1}{\mathrm{N}^2}\bar{\lambda} + \frac{2}{\mathrm{N}}\mu_2^{(\lambda)} + \bar{\lambda}\mu_2^{(\lambda)} + \frac{\mathrm{N}-1}{\mathrm{N}}\bar{\lambda}^2.$$

Substituting for $\bar{\lambda}$, $\bar{\lambda}^2$ and $\mu^{(\lambda)}$ from (i), (ii), and (iii), we get

$$\bar{\lambda}\mu_2^{(\lambda)} = \mathrm{E}[\mu_2^{(\mathrm{A})}\bar{\mathrm{A}}] + \frac{2(\mathrm{N}-1)}{\mathrm{N}^2}\mathrm{E}(\bar{\mathrm{A}}) - \frac{2}{\mathrm{N}}\mathrm{E}(\mu_2^{(\mathrm{A})}) - \frac{\mathrm{N}-1}{\mathrm{N}}\mathrm{E}(\bar{\mathrm{A}}^2);$$

$\therefore$ our best estimate of $\bar{\lambda}\mu_2^{(\lambda)}$

$$= \bar{\mathrm{A}}\mu_2^{(\mathrm{A})} + \frac{2(\mathrm{N}-1)}{\mathrm{N}^2}\bar{\mathrm{A}} - \frac{2}{\mathrm{N}}\mu_2^{(\mathrm{A})} - \frac{\mathrm{N}-1}{\mathrm{N}}\bar{\mathrm{A}}^2$$

or, approximately, if N is large, $\bar{\mathrm{A}}\mu_2^{(\mathrm{A})} - \bar{\mathrm{A}}^2$.

(vi) *Fourth Moment Coefficient.*

$$E(\mu_4^{(A)}) = \mu_4^{(\lambda)} + \frac{6}{N}\bar{\lambda}\mu_2^{(\lambda)} + \frac{6(N-2)}{N}\sum_{i=1}^{N}(\lambda_i^3 - 2\lambda_i^2\bar{\lambda} + \bar{\lambda}^3)$$

$$+ \frac{4(N^2-3N+3)}{N^2}\left[\frac{1}{N}\sum_{i=1}^{N}\lambda_i^2 - \bar{\lambda}^2\right]$$

$$+ \frac{(N-1)(N^2-3N+3)}{N^3}(\bar{\lambda} + 3\mu_2^{(\lambda)} + 3\bar{\lambda}^2)$$

$$+ \frac{3(2N-3)}{N^2}\bar{\lambda}^2 - \frac{3(2N-3)}{N^3}\frac{1}{N}\sum_{i=1}^{N}\lambda_i^2,$$

by equation (2)* on p. 294 of II,

$$= \mu_4^{(\lambda)} + \frac{6}{N}\bar{\lambda}\mu_2^{(\lambda)} + \frac{6(N-2)}{N}\{\mu_3^{(\lambda)} + 3\mu_2^{(\lambda)}\bar{\lambda} + \bar{\lambda}^3$$

$$- 2\mu_2^{(\lambda)}\bar{\lambda} - 2\bar{\lambda}^3 + \bar{\lambda}^3\} + \frac{4(N^2-3N+3)}{N^2}\mu_2^{(\lambda)}$$

$$+ \frac{(N-1)(N^2-3N+3)}{N^3}(\bar{\lambda} + 3\mu_2^{(\lambda)} + 3\bar{\lambda}^2)$$

$$+ \frac{3(2N-3)}{N^2}\bar{\lambda}^2 - \frac{3(2N-3)}{N^3}(\mu_2^{(\lambda)} + \bar{\lambda}^2)$$

$$= \mu_4^{(\lambda)} + \frac{6(N-2)}{N}\mu_3^{(\lambda)} + \frac{6(N-1)}{N}\mu_2^{(\lambda)}\bar{\lambda}$$

$$+ \frac{7N^2-24N+24}{N^2}\mu_2^{(\lambda)} + \frac{3(N-1)^2}{N^2}\bar{\lambda}^2$$

$$+ \frac{(N-1)(N^2-3N+3)}{N^3}\bar{\lambda};$$

$\therefore$ substituting for $\bar{\lambda}$, $\bar{\lambda}^2$, $\mu_2^{(\lambda)}$, $\mu_3^{(\lambda)}$ and $\mu_2^{(\lambda)}\bar{\lambda}$ from (i), (ii), (iii), (iv) and (v), we get

$$\mu_4^{(\lambda)} = E(\mu_4^{(A)}) - \frac{6(N-2)}{N}E(\mu_3^{(A)}) - \frac{6(N-1)}{N}E(\mu_2^{(A)}\bar{A})$$

$$+ \frac{(11N^2-36N+36)}{N^2}E(\mu_2^{(A)}) + \frac{3(N-1)^2}{N^2}E(\bar{A}^2)$$

$$- \frac{6(N-1)(N^2-3N+3)}{N^3}E(\bar{A});$$

* For proof of this equation *see* (vi) (*a*) below.

$\therefore$ our best estimate of $\mu_4^{(\lambda)}$

$$= \mu_4^{(A)} - \frac{6(N+2)}{N}\mu_3^{(A)} - \frac{6(N-1)}{N}\mu_2^{(A)}\overline{A}$$

$$+ \frac{(11N^2 - 36N + 36)}{N^2}\mu_2^{(A)} + \frac{3(N-1)^2}{N^2}\overline{A}^2$$

$$- \frac{6(N-1)}{N^3}(N^2 - 3N + 3)\overline{A},$$

or, approximately, if N is large,

$$\mu_4^{(A)} - 6\mu_3^{(A)} - 6\mu_2^{(A)}\overline{A} + 11\mu_2^{(A)} + 3\overline{A}^2 - 6\overline{A}.$$

As a check on 3 (iii), (iv), (v) and (vi), we note that for the particular case in which all the λ's are equal, our estimates of $\mu_2^{(\lambda)}$, $\mu_3^{(\lambda)}$, $\overline{\lambda}\mu_2^{(\lambda)}$ and $\mu_4^{(\lambda)}$ all vanish as they should for large N when the A distribution is assumed to be a Poisson series itself.

(vi) (*a*) *Proof of the equation used for* $\nu_{[4N]}$ (*equation* (2), *p.* 294 *of II*).

$$\nu_{[4N]} = E\frac{1}{N}\sum_{i=1}^{N}[x_i' - x_{(N)}]^4,$$

$$x_i' - x_{(N)} = [x_i' - m_1^{(i)}] + [m_1^{(i)} - m_{[1N]}] - \frac{1}{N}\sum_{j=1}^{N}[x_j' - m_1^{(j)}]$$

$$= \frac{N-1}{N}[x_i' - m_1^{(i)}] + [m_1^{(i)} - m_{[1N]}]$$

$$- \frac{N-1}{N}\frac{1}{N-1}\sum_{j \neq i}[x_j' - m_1^{(j)}].$$

Put $a = \frac{N-1}{N}[x_i' - m_1^{(i)}]$, $\quad b = m_1^{(i)} - m_{[1N]}$, and

$$c = -\frac{N-1}{N}\cdot\frac{1}{N-1}\sum_{j \neq i}[x_j' - m_1^{(j)}],$$

$$\therefore\ \nu_{[4N]} = E\,\frac{1}{N}\sum_{i=1}^{N}[a+b+c]^4$$

$$= E\,\frac{1}{N}\sum_{i=1}^{N}[a^4+b^4+c^4+4a^3b+4bc^3+6a^2b^2+6b^2c^2+6c^2a^2] \quad (1)$$

as, since a, b and c are independent, and $E(a) = 0$, $E(c) = 0$, the other terms in the expansion of the multinomial vanish.

Now, for $r \neq 0$, $E(a^r) = \left(\frac{N-1}{N}\right)^r \mu_r^{(i)}$,

and $E(c^2) = \frac{1}{N^2}[N\mu_{[2N]} - \mu_2^{(i)}] = \frac{1}{N}\mu_{[2N]} - \frac{1}{N^2}\mu_2^{(i)}$

$$E(c^3) = -\frac{1}{N^3}[N\mu_{[3N]} - \mu_3^{(i)}] = -\frac{1}{N^2}\mu_{[3N]} - \frac{1}{N^2}\mu_3^{(i)},$$

$$E(c^4) = \frac{1}{N^4}\Big\{[N\mu_{[4N]} - \mu_4^{(i)}]$$

$$+ 6\sum_{j\neq k\neq i}\sum E(x_j' - m_1^{(j)})^2\,E(x_k' - m_1^{(k)})^2\Big\}$$

$$= \frac{1}{N^4}\Big\{[N\mu_{[4N]} - \mu_4^{(i)}] + 6\sum_{j\neq k\neq i}\sum \mu_2^{(j)}\,\mu_2^{(k)}\Big\}.$$

But $\left[\sum_{i=1}^{N}\mu_2^{(i)}\right]^2 = \sum_{i=1}^{N}(\mu_2^{(i)})^2 + 2\mu_2^{(i)}\sum_{j\neq i}(\mu_2^{(j)}) + 2\sum_{j\neq k\neq i}\sum\mu_2^{(k)}\mu_2^{(j)}$,

$$\therefore\ 2\sum_{j\neq k\neq i}\sum\mu_2^{(j)}\,\mu_2^{(k)} = N^2\mu^2_{[2N]} - \sum_{i=1}^{N}(\mu_2^{(i)})^2 - 2\mu_2^{(i)}[N\mu_{[2N]} - \mu_2^{(i)}]$$

$$= N^2\mu^2_{[2N]} - \sum_{i=1}^{N}(\mu_2^{(i)})^2 + 2(\mu_2^{(i)})^2$$

$$- 2N\mu_2^{(i)}\mu_{[2N]},$$

$$\therefore\ E(c^4) = \frac{1}{N^3}\mu_{[4N]} - \frac{1}{N^4}\mu_4^{(i)} + \frac{3}{N^2}\mu^2_{[2N]} - \frac{3}{N^4}\sum_{i=1}^{N}(\mu_2^{(i)})^2$$

$$+ \frac{6}{N^4}(\mu_2^{(i)})^2 - \frac{6}{N^3}\mu_2^{(i)}\mu_{[2N]}.$$

$\therefore$ Substituting in (1),

$$\nu_{[4N]} = \left(\frac{N-1}{N}\right)^4 \mu_{[4N]} + \frac{1}{N}\sum_{i=1}^{N} [m_1^{(i)} - m_{[1N]}]^4$$

$$+ \frac{1}{N^3}\mu_{[4N]} - \frac{1}{N^4}\mu_{[4N]} + \frac{3}{N^2}\mu^2_{[2N]}$$

$$- \frac{3}{N^4}\sum_{i=1}^{N}(\mu_2^{(i)})^2 + \frac{6}{N^4}\frac{1}{N}\sum_{i=1}^{N}(\mu_2^{(i)})^2$$

$$- \frac{6}{N^3}\ \frac{1}{N}\sum_{i=1}^{N}\mu_2^{(i)}\mu_{[2N]}$$

$$+ 4\left(\frac{N-1}{N}\right)^3 \frac{1}{N}\sum_{i=1}^{N}\mu_3^{(i)}[m_1^{(i)} - m_{[1N]}]$$

$$- \frac{4}{N}\sum_{i=1}^{N}[m_1^{(i)} - m_{[1N]}]\left\{\frac{1}{N^2}\mu_{[3N]} - \frac{1}{N^3}\mu_3^{(i)}\right\}$$

$$+ \frac{6}{N}\sum_{i=1}^{N}\left(\frac{N-1}{N}\right)^2 \mu_2^{(i)}[m_1^{(i)} - m_{[1N]}]^2$$

$$+ \frac{6}{N}\sum_{i=1}^{N}]m_1^{(i)} - m_{[1N]}]^2\left\{\frac{1}{N}\mu_{[2N]} - \frac{1}{N^2}\mu_2^{(i)}\right\}$$

$$+ 6\left(\frac{N-1}{N}\right)^2 \frac{1}{N}\sum_{i=1}^{N}\mu_2^{(i)}\left\{\frac{1}{N}\mu_{[2N]} - \frac{1}{N^2}\mu_2^{(i)}\right\}$$

$$= \frac{(N-1)(N^2-3N+3)}{N^3}\mu_{[4N]} + \frac{1}{N}\sum_{i=1}^{N}[m_1^{(i)} - m_{[1N]}]^4$$

$$+ 4\frac{1}{N}\sum_{i=1}^{N}\mu_3^{(i)}[m_1^{(i)} - m_{[1N]}]\left(\frac{N^2-3N+3}{N^2}\right)$$

$$+ 6\frac{N-2}{N}\ \frac{1}{N}\sum_{i=1}^{N}\mu_2^{(i)}[m_1^{(i)} - m_{[1N]}]^2$$

$$+ \frac{6}{N}\ \frac{1}{N}\sum_{i=1}^{N}[m_1^{(i)} - m_{[1N]}]^2\mu_{[2N]}$$

$$+ \frac{3(2N-3)}{N^2}\mu^2_{[2N]} - \frac{3(2N-3)}{N^3}\frac{1}{N}\sum_{i=1}^{N}(\mu_2^{(i)})^2.$$

§ 4. *Estimation of the Effect on* $\bar{A}$, $\mu_2^{(A)}$ *and* $\mu_3^{(A)}$ *of Changing the Length of the Period of Observation.*

Suppose the time of observation be multiplied by n (n may be $>$ or $<$ 1).

Let undashed letters refer to the original period, and dashed letters to a period n times as long—other things being unchanged. Then, though A_i' will *not* be nA_i, it will be true of the λ's that $\lambda_i' = n\lambda_i$.

$\therefore$ $\bar{\lambda}' = n\bar{\lambda}$, $\mu_2^{(\lambda')} = n^2\mu_2^{(\lambda)}$, $\mu_3^{(\lambda')} = n^3\mu_3^{(\lambda)}$, etc., the effect being the same as changing the unit of λ. The expected distributions of A's in persons in the same λ group will also be Poisson series with the same unit and means $n\lambda_1$, $n\lambda_2$, etc.

(i) *Mean.*

$$E(\bar{A}') = nE(\bar{A}).$$

(ii) *Second Moment Coefficient.*

$$\begin{aligned} E(\mu_2^{(A')}) &= \mu_2^{(\lambda')} + \frac{N-1}{N}\bar{\lambda}' \text{ by 3 (iii)} \\ &= n^2\mu_2^{(\lambda)} + n\frac{N-1}{N}\bar{\lambda}^* \\ &= n^2E(\mu_2^{(A)}) - n(n-1)\frac{N-1}{N}E(\bar{A}). \end{aligned}$$

$\therefore$ our best estimate of $\mu_2^{(A')} = n^2\mu_2^{(A)} - n(n-1)\dfrac{N-1}{N}\bar{A}$, or, approximately, if N is large, $n^2\mu_2^{(A)} - n(n-1)\bar{A}$.

(iii) *Third Moment Coefficient.*

$$\begin{aligned} E(\mu_3^{(A')}) &= \mu_3^{(\lambda')} + \frac{3(N-2)}{N}\mu_2^{(\lambda')} + \frac{(N-1)(N-2)}{N^2}\bar{\lambda}' \text{ by § 3 (iv)} \\ &= n^3\mu_3^{(\lambda)} + \frac{3(N-2)}{N}n^2\mu_2^{(\lambda)} + \frac{(N-1)(N-2)}{N^2}n\bar{\lambda} \\ &= n^3E(\mu_3^{(A)}) - n^2(n-1)\frac{3(N-2)}{N}E(\mu_2^{(A)}) \\ &\qquad + n(2n-1)(n-1)\frac{(N-1)(N-2)}{N^2}E(\bar{A}). \end{aligned}$$

* In this notation Bortkiewicz's Law of Small Numbers depends on the fact that $E(\mu_2^{(A)}) = n^2\mu_2^{(\lambda)} + n\dfrac{N-1}{N}E(\bar{A}) = n^2$ (sq. of original "excess error") $+ n$ (mean square of original sampling errors about the λ's).

$\therefore$ our best estimate of $\mu_3^{(A')} = n^3\mu_3^{(A)} - n^2(n-1)\dfrac{3(N-2)}{N}\mu_2^{(A)}$ $+ n(2n-1)(n-1)\dfrac{(N-1)(N-2)}{N^2}\bar{A}$, or, approximately, if N is large, $n^3\mu_3^{(A)} - 3n^2(n-1)\mu_2^{(A)} + n(2n-1)(n-1)\bar{A}$.

§ 5. *Sampling Errors of* $\bar{A}$, $\mu_2^{(A)}$ (*when found directly from the observed values*), $\mu_2^{(A)}$ (*when estimated from the observed values for a period of different length*), *and of* $\mu_2^{(\lambda)}$ *as estimated from the* A'*s*.

(i) *Sampling error of* $\bar{A}$.

$$E(\sigma_{\bar{A}}^2) = \frac{1}{N}\mu_{[2N]} \text{ (by p. 82 of I, or equation 8 on p. 286 of II)}$$

$$= \frac{\lambda}{N} = \frac{E(\bar{A})}{N}.$$

$\therefore$ our best estimate of $\sigma_{\bar{A}}^2 = \dfrac{\bar{A}}{N}$. This also follows at once from the fact that since the *sum* of a number of variables, each following a Poisson series about its own mean, is itself a Poisson series (with the same class unit) about the *sum* of the submeans, the *mean* of N such variables also follows a Poisson series in which the class unit is not 1 but $\dfrac{1}{N}$, *i.e.* the terms of the Poisson whose mean is equal to the *mean* of all the submeans, gives the number of times that the mean of the N variables takes the values (in the old units)

$$0, \frac{1}{N}, \frac{2}{N}, \frac{3}{N} \ldots \text{etc.}$$

(ii) *Sampling error of* $\mu_2^{(A)}$ (when found directly)

$$E[\sigma^2_{\mu_2(A)}] = E\{\nu'_{[2N]} - \nu_{[2N]}\}^2 \text{ in the notation of I}$$

$$= \frac{(N-1)^2}{N^3}\bar{\lambda} + \frac{2(N-1)}{N^2}(\mu_2^{(\lambda)} + \bar{\lambda}^2) - \frac{2}{N^2}\mu_2^{(\lambda)}$$

$$+ \frac{4}{N}(\mu_3^{(\lambda)} + \bar{\lambda}\mu_2^{(\lambda)}) + \frac{4(N-1)}{N^2}\mu_2^{(\lambda)}$$

by equation (4), on p. 85 of I, or p. 295 of II

$$= \frac{(N-1)^2}{N^3}\bar{\lambda} + \frac{2(N-1)}{N^2}\bar{\lambda}^2 + \frac{2(3N-4)}{N^2}\mu_2^{(\lambda)}$$

$$+ \frac{4}{N}\mu_3^{(\lambda)} + \frac{4}{N}\mu_2^{(\lambda)}\bar{\lambda};$$

$\therefore$ substituting for $\bar{\lambda}$, $\bar{\lambda}^2$, $\mu_2^{(\lambda)}$ and $\bar{\lambda}\mu_2^{(\lambda)}$ from 3 (i), (ii), (iii), (iv) and (v).

$$\mathrm{E}\left[\sigma^2_{\mu_2(\mathrm{A})}\right] = \frac{3(\mathrm{N}-1)^2}{\mathrm{N}^3}\mathrm{E}(\bar{\mathrm{A}}) - \frac{2(\mathrm{N}-1)}{\mathrm{N}^2}\mathrm{E}(\bar{\mathrm{A}}^2)$$

$$- \frac{2(3\mathrm{N}-4)}{\mathrm{N}^2}\mathrm{E}(\mu_2^{(\mathrm{A})}) + \frac{4}{\mathrm{N}}\mathrm{E}(\mu_3^{(\mathrm{A})}) + \frac{4}{\mathrm{N}}\mathrm{E}(\mu_2^{(\mathrm{A})}\bar{\mathrm{A}});$$

$\therefore$ our best estimate of $\sigma^2_{\mu_2(\mathrm{A})}$

$$= \frac{3(\mathrm{N}-1)^2}{\mathrm{N}^3}\bar{\mathrm{A}} - \frac{2(\mathrm{N}-1)}{\mathrm{N}^2}\bar{\mathrm{A}}^2 - \frac{2(3\mathrm{N}-4)}{\mathrm{N}^2}\mu_2^{(\mathrm{A})}$$

$$+ \frac{4\mu_3^{(\mathrm{A})}}{\mathrm{N}} + \frac{4(\mu_2^{(\mathrm{A})}\bar{\mathrm{A}}}{\mathrm{N}},$$

or, approximately, if N is large

$$\frac{1}{\mathrm{N}}(3\bar{\mathrm{A}} - 2\bar{\mathrm{A}}^2 - 6\mu_2^{(\mathrm{A})} + 4\mu_3^{(\mathrm{A})} + 4\bar{\mathrm{A}}\mu_2^{(\mathrm{A})}).$$

(iii) *Expected mean product moment of deviations due to random sampling in* $\mu_2^{(\mathrm{A})}$ *and* $\bar{\mathrm{A}}$ *;* i.e. *in the 2nd moment coefficient and the mean of the* A*'s.*

E [product moment of deviation in $\mu_2^{(\mathrm{A})}$ and $\bar{\mathrm{A}}$]

$= \mathrm{E}\{\nu'_{[2\mathrm{N}]} - \nu_{[2\mathrm{N}]}\}(x_{(\mathrm{N})} - m_{[1\mathrm{N}]})$ in the notation of I,

$$= \mathrm{E}(x_{(\mathrm{N})}\nu'_{[2\mathrm{N}]}) - m_{[1\mathrm{N}]}\nu_{[2\mathrm{N}]}$$

$$= \mathrm{E}[\mu_{(\bar{\mathrm{A}})2}\bar{\mathrm{A}}] - m_{[1\mathrm{N}]}\nu_{[2\mathrm{N}]}$$

$$= \frac{\mathrm{N}-1}{\mathrm{N}^2}\bar{\lambda} + \frac{2}{\mathrm{N}}\mu_2^{(\lambda)} + \bar{\lambda}\mu_2^{(\lambda)} + \frac{\mathrm{N}-1}{\mathrm{N}}\bar{\lambda}^2 - \bar{\lambda}\left(\mu_2^{(\lambda)} + \frac{\mathrm{N}-1}{\mathrm{N}}\bar{\lambda}\right)$$

by 3 (v) and 3 (iii)

$$= \frac{\mathrm{N}-1}{\mathrm{N}^2}\bar{\lambda} + \frac{2}{\mathrm{N}}\mu_2^{(\lambda)};$$

$\therefore$ substituting for λ and $\mu_2^{(\lambda)}$

E [mean product moment of errors in $\mu_2^{(\mathrm{A})}$ and $\bar{\mathrm{A}}$]

$$= \frac{\mathrm{N}-1}{\mathrm{N}^2}\mathrm{E}(\bar{\mathrm{A}}) + \frac{2}{\mathrm{N}}\left\{\mathrm{E}\mu_2^{(\mathrm{A})} - \frac{\mathrm{N}-1}{\mathrm{N}}\mathrm{E}(\bar{\mathrm{A}})\right\}$$

$$= \frac{2}{\mathrm{N}}\mathrm{E}(\mu_2^{(\mathrm{A})}) - \frac{\mathrm{N}-1}{\mathrm{N}^2}\mathrm{E}(\bar{\mathrm{A}});$$

∴ our best estimate of the mean product moment of sampling errors in $\bar{A}$ and $\mu_2^{(A)}$

$$= \frac{2}{N} \mu_2^{(A)} - \frac{N-1}{N^2} \bar{A},$$

or, approximately, if N is large, $\frac{1}{N}(2\mu_2^{(A)} - \bar{A})$.

(iv) *Sampling error of $\mu_2^{(A)}$ when this is not found directly but estimated from the* A *values for a period of different length.*

Let $\Sigma^2{}_{A'}$ be the estimate of $\mu_2^{(A')}$ for a period of n times as long as that to which the observed values $\mu_2^{(A)}$ and $\bar{A}$ refer.
Then

$$\Sigma^2{}_{A'} = n^2 \mu_2^{(A)} - n(n-1) \frac{N-1}{N} A \quad \text{by 4 (ii)};$$

$$\therefore \mathrm{E}\left(\sigma^2_{\Sigma^2{}_{A'}}\right) = \mathrm{E}\left(n^4 \sigma^2_{\mu_2^{(A)}}\right) + \mathrm{E} n^2 (n-1)^2 \frac{(N-1)^2}{N^2} \sigma^2_{\bar{A}}$$

$$- 2n^3 (n-1) \frac{N-1}{N} \mathrm{E}\ [\text{mean product of deviations in } \mu_2^{(A)} \text{ and } \bar{A}]:$$

∴ substituting from 5 (i), (ii) and (iii), we get

$$\mathrm{E}\left(\sigma^2_{\Sigma^2{}_{A'}}\right) = \frac{4n^4}{N} \mathrm{E}\left(\mu_3^{(A)}\right) + \frac{4n^4}{N} \mathrm{E}\left(\mu_2^{(A)} \bar{A}\right)$$

$$+ \frac{2n^3}{N^2} \{2(N-1) - n(5N-6)\} \mathrm{E}\left(\mu_2^{(A)}\right)$$

$$- \frac{2n^4(N-1)}{N^2} \mathrm{E}\left(\bar{A}^2\right) + (6n^4 - 4n^3 + n^2) \frac{(N-1)^2}{N^3} \mathrm{E}(\bar{A}).$$

∴ our best estimate of $\left(\sigma^2_{\Sigma^2{}_{A'}}\right)$

$$= \frac{4n^4}{N} \mu_3^{(A)} + \frac{4n^4}{N} \mu_2^{(A)} \bar{A} + \frac{2n^3}{N^2} \{2(N-1) - n(5N-6)\} \mu_2^{(A)}$$

$$- \frac{2n^4(N-1)}{N^2} \bar{A}^2 + (6n^4 - 4n^3 + n^2) \frac{(N-1)^2}{N^3} \bar{A},$$

or, approximately, if N is large,

$$\frac{1}{N}\{4n^4(\mu_3^{(A)}+\bar{A}\mu_2^{(A)})+2n^3(2-5n)\mu_2^{(A)}-2n^4\bar{A}^2$$
$$+(6n^4-4n^3+n^2)\bar{A}\}.$$

When $n=1$, these values reduce to those given in 5 (ii).

(v) *To Estimate the Sampling Error of* $\mu_2^{(\lambda)}$ *as found from the* A'*s*.

Let Σ_λ^2 be our estimate of $\mu_2^{(\lambda)}$ as found from the A's.

$$\therefore\ \Sigma_\lambda^2=\mu_2^{(A)}-\frac{N-1}{N}\bar{A}.$$

$\therefore\ E\left(\sigma^2_{\Sigma^2_\lambda}\right)=E\left(\sigma^2_{\mu_2^{(A)}}\right)-\dfrac{2(N-1)}{N}E$ [mean product moment of deviations in $\mu_2^{(A)}$ and $\bar{A}$] $+\dfrac{(N-1)^2}{N^2}E(\sigma^2_{\bar{A}})$.

$\therefore$ substituting from 5 (ii), (iii) and (i) we get

$$E\left(\sigma^2_{\Sigma^2_\lambda}\right)=\frac{6(N-1)^2}{N^3}E(\bar{A})+\frac{4}{N}E(\mu_3^{(A)})+\frac{4}{N}E(\mu_2^{(A)}\bar{A})$$
$$-\frac{2(N-1)}{N^2}E(\bar{A}^2)-\frac{2}{N^2}(5N-6)E\mu_2^{(A)}.$$

$\therefore$ our best estimate of $\sigma^2_{\Sigma^2_\lambda}$

$$=\frac{6(N-1)^2}{N^3}\bar{A}+\frac{4}{N}(\mu_3^{(A)}+\bar{A}\mu_2^{(A)})-\frac{2(N-1)}{N^2}\bar{A}^2$$
$$-\frac{2}{N^2}(5N-6)\mu_2^{(A)},$$

or, approximately, if N is large,

$$\frac{1}{N}(6\bar{A}+4\mu_3^{(A)}+4\bar{A}\mu_2^{(A)}-2\bar{A}^2-10\mu_2^{(A)}).$$

§ 6. *Correlation Coefficients* (for large N only).

(i) Correlation between the λ's and any other measure x estimated from the observed correlation between the A's and x (approximately).

$A_i = \lambda_i + \varepsilon$ where ε is a sampling error. Let x be measured from its mean,

$$\therefore\ r_{xA} = \sum_{i=1}^{N} \frac{x_i(\lambda_i+\varepsilon)}{N\sigma_x\sigma_A} \quad \text{and} \quad r_{x\lambda} = \sum_{i=1}^{N} \frac{(x_i\,\lambda_i)}{N\sigma_x\,\sigma_\lambda},$$

$$\therefore\ r_{xA} = \frac{r_{x\lambda}\,\sigma_x\,\sigma_\lambda}{\sigma_x\,\sigma_A}$$

approximately (*cf.* Cobb, *Biometrika* VI, p. 109), assuming there is no correlation between x and ε

$$= \frac{r_{x\lambda}\,\sigma_\lambda}{\sigma_A}$$

$$\therefore\ r_{x\lambda} = \frac{r_{xA}\,\sigma_A}{\sigma_\lambda} = r_{xA}\frac{\sigma_A}{\sqrt{\sigma_A{}^2 - \bar{A}^*}} = \frac{r_{x\,A}}{\sqrt{1 - \dfrac{\bar{A}}{\sigma_A{}^2}}}.$$

(ii) *To estimate the correlation between the* λ*'s and the* A*'s.*

$$r_{A\lambda} = \frac{\sum\limits_{i=1}^{N}\left(\dfrac{A_i\lambda_i}{N}\right) - \bar{A}\bar{\lambda}}{\sigma_A\sigma_\lambda} = \frac{\sum\limits_{i=1}^{N}\dfrac{\lambda_i(\lambda_i+\varepsilon)}{N} - \bar{A}\bar{\lambda}}{\sigma_A\sigma_\lambda}$$

$$= \frac{\sigma_\lambda{}^2}{\sigma_A\sigma_\lambda},$$ approximately, assuming there is no correlation between λ and ε

$$= \frac{\sigma_\lambda}{\sigma_A} = \sqrt{1 - \frac{\bar{A}}{\sigma_A{}^2}}$$

* The $\dfrac{N-1}{N}$ factor has been omitted, as it is unity to the order of terms kept above.

(iii) *To estimate the expected correlation between the* A's *of the same individual in two different* (i.e. *non-overlapping*) *periods of the same length.*

Let ${}_iA_1\ {}_iA_2$ be the A's for any one individual in the λ_i group in the 1st and 2nd periods respectively.

Then $\qquad {}_iA_1 = \lambda_i + \varepsilon_1 \qquad {}_iA_2 = \lambda_i + \varepsilon_2$

$$\therefore\ r_{A_1 A_2} = \frac{\sum\limits_{i=1}^{N} \frac{(\lambda_i + \varepsilon_1)(\lambda_i + \varepsilon_2)}{N} - \bar{A}_1 \bar{A}_2}{\sigma_{A_1}\sigma_{A_2}}.$$

But $E(\bar{A}_1) = E(\bar{A}_2) = \bar{\lambda}$ and $E(\sigma_{A_1}\ \sigma_{A_2}) = E(\sigma_A{}^2) = \sigma_\lambda{}^2 + \frac{N-1}{N}\bar{\lambda}$, and we neglect as before the correlation of ε_1 with ε_2 and of either with λ.

$$\therefore\ r_{A_1 A_2} = \frac{\sigma_\lambda{}^2}{\sigma_\lambda{}^2 + \bar{\lambda}} = \frac{\sigma_{A_1}{}^2 - \bar{A}_1}{\sigma_{A_1}{}^2} = \frac{\sigma_{A_2}{}^2 - \bar{A}_2}{\sigma_{A_2}{}^2} = 1 - \frac{\bar{A}_1}{\sigma_{A_1}{}^2} = 1 - \frac{\bar{A}_2}{\sigma_{A_2}{}^2}$$

Hence the correlation between the λ's and the A's in any period is approximately equal to the square root of the correlation between the values of A for the same person in two periods, each of the same length as that to which the first correlation refers.

(iv) *Effect on the correlation between the* A's *of the same individual in two periods of changing the length of the periods.*

Without altering other conditions let $r_{A_1' A_2'}$ be the correlation between the same individuals' A's in two periods, each n times as long as those to which A_1 and A_2 refer.

Then from 6 (iii)

$$r_{A_1' A_2'} = 1 - \frac{\bar{A}_1'}{\sigma^2_{A_1'}} \text{ approximately}$$

$$= 1 - \frac{n\bar{A}_1}{n^2\sigma^2_{A_1} - n(n-1)\bar{A}_1}$$

$$= 1 - \frac{\bar{A}_1}{n\sigma^2_{A_1} - (n-1)\bar{A}_1}, \text{ or a similar expression in } A_2.$$

(v) *Estimate of expected correlation between the* A's *of the same individual over two periods differing in length, the second, say, n times the first.*

Let $\bar{\lambda}$ and σ_λ refer to periods of the same length as the first, then the corresponding values for periods of the same length as the second will be $n\bar{\lambda}$ and $n\sigma_\lambda$. Then with the same notation as before

$$_iA_1 = \lambda_i + \varepsilon_1, \qquad _iA_2 = n\lambda_i + \varepsilon_2,$$

$$\therefore \; r_{A_1A_2} = \frac{\displaystyle\frac{\overset{N}{\underset{i=1}{S}} (\lambda_i + \varepsilon_1)(n\lambda_i + \varepsilon_2)}{N} - \bar{A}_1\bar{A}_2}{\sigma_{A_1}\sigma_{A_2}}.$$

$\therefore$ with the same approximations as before

$$r_{A_1A_2} = \frac{n\sigma_\lambda^2}{\sqrt{(\bar{\lambda} + \sigma_\lambda^2)(n\bar{\lambda} + n^2\sigma_\lambda^2)}}$$

$$= \frac{n\left(1 - \dfrac{\bar{A}_1}{\sigma_{A_1}^2}\right)}{\sqrt{n\left\{n - (n-1)\dfrac{\bar{A}_1}{\sigma_{A_1}^2}\right\}}}$$

If $\bar{A} = \sigma_A^2$ the correlations in 6 (iii) 6 (iv) and 6 (v) become zero for large N as they should, for in this case all the λ's are equal and the A's follow a Poisson distribution, the only variations from $\bar{A}$ being independent sampling errors.

Discussion on Miss Newbold's Paper.

Dr. Greenwood: Before proposing the vote of thanks, I should like to read a note I have received from Mr. Yule, who is unfortunately prevented from attending by the illness of a relative. Mr. Yule writes :—

"I greatly regret that, owing to unfortunate circumstances, it is unlikely that I shall be able to be present to-morrow at Miss Newbold's reading of her paper on 'The Statistics of Repeated Events,' a most valuable paper, not merely summarizing, but greatly generalizing, and extending previous results. The same circumstances which prevent my attendance have also prevented my reading the paper with all the attention that it deserves, and I must content myself with a gentle objection to one statement as misleading: I refer to the sentence in the first paragraph: 'My own connection with the investigation has only been a share in the statistical work, which originated with Dr. Greenwood and Mr. Yule, and has since been continually under their advice and guidance.' For myself I must wholly deny this allotted share in the merits of Miss Newbold's work. As a matter of fact, matters are, if anything, the other way round. Some weeks ago I sent her a result at which I had arrived, for comment, and found that she had long ago got not merely that result, but general theory including it. Miss Newbold has made a most useful contribution to the theory of accidents, and I should like to join in the congratulations which I am sure will be offered to her."

I should like to associate myself in the disclaimer that Mr. Yule mentions; in fact, I think the association of Miss Newbold with me is an illustration of the verse:

> "To teach his grandson chess, his leisure he'd employ,
> Until at last the old man was beaten by the boy."

Miss Newbold was in a certain sense a pupil of Mr. Yule and myself, but has long passed beyond the need of our tuition. I have a special pleasure in moving a vote of thanks on this occasion, one of the reasons being that the work Miss Newbold is now doing recalls memories of a time which I should be very sorry to live through again, but memories of which are interesting.

In 1918 Mr. Yule and I were engaged on war-work, and you will recollect that the definition of war-work was any work, provided it had nothing to do with one's ordinary work. So far as Mr. Yule and I were concerned, that particular requirement was not strictly adhered to, because we were both engaged in doing sums; but they were not the kind of sums that we habitually did, and if there were any reasonably patriotic excuse for employing the minute paper of the Ministry of Food, where Mr. Yule was winning the war, or of the Ministry of Munitions, where I was impeding the plans of the Kaiser, for algebraical calculations, we welcomed it. The work arose in this way:—I was on a Committee appointed to advise on certain medical aspects of the Flying Service. *Inter alia* we had to consider the nature of tests which would be suitable to determine the aptitude of entrants for the Flying Service, and to determine whether persons who had broken down, or who had been sent back from the front for wounds or other reasons, were fit to return to duty. In the course of the investigation, the point arose whether a man who had had a small accident not sufficient to damage himself or his machine seriously might yet have sustained such a shock that it would be undesirable for him to fly again. The problem emerged whether it would be possible to distinguish by analysis of the frequency-distributions of accidents between the three possibilities: (1) That accidents are pure accidents in the sense of being "simple" chance events; (2) that the distribution of *first* accidents is that of "simple" chance events; (3) that the distribution whether of first or subsequent accidents differs in a specific way from the "simple" chance scheme.

That was the theoretical problem, and we realized that at that stage it would not be much more than a theoretical problem. It was obvious that we could not expect to get from the records of the Air Force data on which we could test any hypothesis, because the question of unequal exposure to risk would arise. We decided that it would be necessary to confine our investigation to the statistics of trivial accidents in munitions factories.

With regard to the theoretical problem, the solutions that we obtained have, on the whole, stood criticism. If I had to suggest to a young mathematical statistician a theoretical problem connected with the work which probably would repay more mathematical investigation, it would be the case of what we call the biased distribution, viz. the first accidents distributed at random; the subsequent accidents not distributed at random. The solution we obtained was formally complete but exceedingly unhandy.

Possibly a young mathematical statistician might see some other, and neater, way of formulating in statistical terms that particular problem. But whatever we may have failed to prove, we did succeed in making it probable, and all the subsequent work has made it still more probable, that the case of unequal initial liability is the practically most important form of the problem, and in extending the study of that case it seems to me that Miss Newbold has taken the right line. I should like particularly to commend Miss Newbold for utilizing the late Professor Tschuprow's treatment of the sampling problem, because it seems to me that Professor Tschuprow's notation should be familiar to all mathematical statisticians; it marks a great step in the logical advance of the treatment of these problems, and it is very desirable that we should have some practical introduction to it in our *Journal*.

Miss Newbold's own extensions of the mathematical treatment of the subject are important. Particularly her study of the correlation between certain variables, one of which is accessible to direct study, because there is the crux of the practical work which is now going on. When our work was first published, a critic in one of the technical engineering journals was very angry with us, because he thought our work was either trivial or useless. In a sense it *is* common knowledge that liability to accident varies; but the whole object of our investigation and subsequent investigations has been to get quantitative knowledge of the variability, and until that has been done, one has not got very far in a practical sense. Consequently, the problem that is now facing us is how to measure liability. I hope we shall hear later in this discussion the views of physicians and psychologists who are dealing practically with individuals and attempting to assess on various lines of evidence their liability to sustain accidents. It would be extremely foolish to suppose that by the most lavish use of algebra or arithmetic one could possibly devise a method by which the application of some rule-of-thumb test to individuals would enable anybody to assess individual liability to accident as well as a medical man or a psychologist of first-rate ability and long experience may be able to do. But we are dealing here with a question of great practical importance, viz. how is the incidence of accidents to be diminished when the people who have to deal with this subject are not all of them highly expert neurological physicians or psychologists. To such people, developments of the work on lines about which we may hear later in the discussion may be at least of some value. It seems to me that Miss Newbold is entering on work that will ultimately lead to important results. As statisticians we should try to help people by the use of our art, although we admit its limitations, and I think Miss Newbold deserves a hearty vote of thanks which I have great pleasure in moving.

Mr. D. R. Wilson: I have much pleasure in seconding the vote of thanks, and in congratulating Miss Newbold on her very able and interesting paper. I feel bound to say, however, that I should have accepted the invitation to do so with some hesitation if I had already seen the paper. It is doubtless a shameless confession to make at a meeting of this Society, but circumstances force me to admit, that whilst I now regard r as an old personal friend, and have a nodding acquaintance with η, I am not yet on speaking terms with χ^2 and P, although I know them both well by sight. In these circumstances, the meeting will probably be relieved to hear that I do not propose to embark on any statistical appreciation of Miss Newbold's paper, but to confine my remarks to the practical significance of her results.

As I see it, one can regard an accident of the kind dealt with by Miss Newbold as an event entailing a momentary conflict between inanimate and animate material, and following the usual precedent for problems involving the human factor, progress in accident prevention has been much more rapid on the inanimate side than the other.

May I illustrate this from my own experience? When I was appointed a Factory Inspector twenty years ago, we were then rather vague about what constituted a dangerous machine. Then, as now, new machines were being continually introduced into industry, and our usual test of whether one of these was dangerous or not was based on a well-known formula, namely "to wait and see." If we found by experience that a particular machine caused an undue number of accidents, then we all set to work to get proper safeguards provided. Now all that has changed; years of investigation on the part of the Engineering Inspectors of Factories and others have resulted in the discovery of mechanical actions which are inherently dangerous, and it is now quite possible to have safeguards provided for a new machine *before* it is operated in the factory and so before it has a chance of doing any damage.

As I see it, our aim is to do with persons exactly what has been done with machines, that is to say, to be able to tell beforehand which people are likely to incur accidents, and to prevent them from entering into what are to them dangerous occupations. The fulfilment of this aim must involve years of patient research, and some of the difficulties of the subject have been mentioned this afternoon. Miss Newbold implies that the results obtained by her are not sufficiently definite to form any basis for future research; with that I agree generally, although I should think that the associations between accidents and minor ailments and between accidents and youth are of interest to the psychological investigators now dealing with the subject.

But if Miss Newbold's results do not show *how* this end could be achieved, they show two other things of equal practical importance—first, that research on the subject is worth continuing, and secondly, that when the tests are ultimately worked out, their practical application will not involve undue interference with existing conditions.

Taking the first point, probably the most definite conclusion arising from these statistical inquiries is that there is unequal susceptibility to accident among individuals. It may be said—it has, in fact, been said—that that is a point so obvious as to be hardly worth proving by statistical methods, but I remember, in the case mentioned by the proposer of the vote of thanks, the writer expressed himself as very sceptical about this conclusion, and claimed that the worker who sustained an accident was, just as probably, merely unlucky, in the sense that the accident was due entirely to chance. There is therefore one person whom Miss Newbold's work may have convinced, though, from what I remember of him, I think it rather unlikely.

Anyhow, these repeated proofs of unequal individual susceptibility are, to my mind, very encouraging in regard to the ultimate possibilities of research on the lines already mentioned, for obviously, if everyone were equally susceptible there would be no specially susceptible people to exclude, and so, from a practical point of view, continuance of this kind of research would be useless.

The administrative practicability of eventually reducing industrial accidents is suggested by another conclusion in Miss Newbold's work, which seems to me even more important; that is, that the bulk of accidents occur amongst a comparatively limited number of workpeople or, to put it in her own words, that "the average number of accidents in any homogeneous group is much influenced by a comparatively small number of workers."

Let us imagine a factory with 1,000 workers and a total of 100 accidents a year. Clearly it would be much easier to reduce the number of accidents, by the exclusion of susceptible persons, if those accidents were distributed amongst 20 instead of amongst 500 persons.

The suggestion conveyed by Miss Newbold's second conclusion is that in the course of years, when the psychological tests which are now being worked upon by Mr. Farmer and Mr. Chambers have been developed to their full extent, their practical application will be a matter comparatively easy to apply, and will result, without any great administrative disturbance, in a marked diminution of the number of accidents incurred.

In conclusion, I would like to refer to a report which happened to reach me this morning, and which quotes the accidents incurred in 1925 included in the returns published under the Workmen's Compensation Act. In 1925 there were 476,000 accidents, 3,300 of which were fatal. The total compensation paid amounted to £6,600,000 for the one year. I think everyone will agree that research that possesses any possibility of reducing these figures is well worth pursuing from both the humanitarian and economic points of view. I have again much pleasure in seconding the vote of thanks.

Dr. MILLAIS CULPIN said that he must congratulate Miss Newbold upon successfully applying statistics to that most erratic phenomenon —human behaviour—for ultimately accidents were a matter of human behaviour. The subject was of great interest to him, because for several years he had been doing work that apparently started far away from this subject, but ultimately brought him into close contact with it.

Towards the end of the war he was engaged in treating men suffering from shell-shock, and he made it his business to go into their history as far as possible, and to pick out those in whom he could find a predisposition to nervous breakdown. He took as his standard whether, with his knowledge of the man's history, he would have rejected him on enlistment. On that basis he found that 64 per cent. of the men suffering from shell-shock had had a predisposition or actual nervous trouble. Sir Frederick Mott put the figure as high as 80 per cent. Those figures were unsatisfactory, however, because there were no control groups; it was impossible at the time to take the history of men without shell-shock. In fact, they had no grounds, excepting personal judgment, for saying that certain symptoms were a mark of predisposition. Amongst his notes of these patients, however, Dr. Culpin found two cases in which the men's temperament showed itself in accidents, and he would give these briefly. One man had many symptoms; he was in a bad way, and when under treatment he casually mentioned that he had had a bicycle accident before a nervous breakdown. When made to talk about this, he showed an unexpected emotional outbreak, telling how he found the machine running away with him, how he had become excited, pedalling faster and actually inviting the accident. That case was worked out, and it was found that there was an unconscious impulse to inflict self-punishment upon himself. That might sound a freak case, but similar cases had been described.

The second case was more typical of a general principle. The patient, at Dr. Culpin's instigation, wrote out his autobiography, which made up to six closely printed pages. His life was marked by pathological fear; he was always afraid from childhood onwards, and this was of great importance in the light of what followed. In his autobiography he explained how at an early age he got his first job, when he was taken to an old shed and set to keep lifting a number of little pigs up to an old sow. He felt afraid and began to tremble. His mistress saw him and asked what was wrong; he told her, and she burst out laughing, which only made him worse. He ran away, and she accompanied him back to the pigs, but still he was afraid, and she had to give him another job and look after the pigs herself. He went on with his history until he got a railway job in Scotland and learnt to work a signal-box. On his first night the signal-box had to be washed out. His mate left him after a while, and directly he was left alone fear attacked him. To drive away the fear he started to wash the floor whilst a goods train was shunting. Then he "took on" two more trains, and the next thing he heard was a crash. The guard of the train came running up and told him he had fouled both main lines. He telephoned to stop both trains, and then collapsed, "useless and fit for nothing." As a result of the accident he was "court-martialled," and the last question asked him was whether he had ever been seriously hurt about the head. If he had been led to talk about his fear of the pigs it would have thrown some light upon his temperament as related to the accident.

This case prompts speculation about a recent railway accident, where a signalman pulled the wrong lever. At the enquiry he stated that he had suffered from influenza and had only been back at work for three weeks. A person temperamentally nervous and inclined to break down often suffers an exacerbation after influenza, but we have not yet learnt to consider such a possibility in an accident enquiry.

These two cases were perhaps extreme examples, but the disturbed mental states experienced by the subjects and contributing to the accidents are typical of conditions that, in varying degrees of severity, exist quite commonly.

Dr. Culpin's work had led him to the subject of telegraphists' cramp, and in investigating a control group for psycho-neurotic symptoms—which occur in persons who might be described as highly strung or nervous—he found these symptoms in about 50 per cent. of what were supposed to be ordinary people. With the help of the Fatigue Board the matter was carried further, and he and his colleague Miss May Smith were trying to find out the percentage of psycho-neurotic persons among the general population. The standard of normality necessarily depended upon personal judgment, and Dr. Culpin said he must confess that neither he nor his colleague would pass the other as normal. Varying with the groups taken, between 50 and 60 per cent. of the subjects were passed as normal. Then one came to border-line cases, and then to those people with severe symptoms in need of treatment. In one group only 3 per cent. showed severe symptoms, whereas in another group the percentage was as high as 17.

At the same time his colleague had carried out certain tests similar to those used by Mr. Farmer, and it was found that there was a definite correlation between the results of the dotting tests and Dr. Culpin's own findings based on the results of an interview. He did not know whether the figures would turn out to be of statistical value, but they constantly pointed in one direction. If psycho-neurotics were taken, their mean was well below the mean of the average, with the exception of one small group described as obsessionals. Obsessional subjects were very highly strung, but they put up excellent records, so much so that when Dr. Culpin's colleague obtained superior results to her tests, she knew that he would, on other tests, classify the subject as obsessional or absolutely normal. The point of this communication was that Mr. Farmer had managed to find a certain group prone to accident who did in a general way give certain results to his tests of co-ordination, and when those same tests were applied to his people, Dr. Culpin found these same results given by those suffering from psycho-neurotic symptoms. From that it would appear that it was those people who provided the accident population.

It might be interesting to know the sort of finding upon which judgments were based. The following was an actual case :—

A woman aged 33, employed in a rather low grade of clerical work. Dr. Culpin opened by asking her how she liked her work, and if she would call herself nervous: to which she replied, "Yes," and said that a month or so previously she had had to go off work for "her nerves," which had been upset by the making of a single mistake. She had always been afraid of the dark and of being alone. She had a feeling of being watched by others. (In some cases this feeling of observation was so extreme that Dr. Culpin had known some people, apparently normal, who would not go into a restaurant because of a feeling of being looked at; he proposed to deal with that symptom later on.) In the present case the patient was nervous if spoken to by strangers. She had no traffic fears. She did not like picture shows. She had good and bad days; on a bad day she had no energy, although she improved as the day went on, and had a feeling of "something awful" going to happen to her. When asked about sports, she said that tennis was too much for her. She had had influenza, which had left her heart weak. Questioned about her heart, she said she felt faint in trains and in crowds. She said, "I cannot sit in small rooms, but it would not make me frightened; it is my heart." Being in small rooms can have no direct relation to the heart; such symptoms are phobias, and this particular one is claustrophobia, although the woman herself thought it was a heart symptom. Probably fear produced palpitation and she preferred to blame her heart, for she would not admit, even to herself, that she was afraid. When it came to the subject of marriage, she admitted that she was not social and did not like mixing with people. She wore very weak glasses to avoid eye-strain, and this symptom—the wearing of very weak glasses by a young adult to avoid eye-strain—might be considered to be in itself a bad sign. Her hands were blue—a sign of emotional circulatory disturbance. She had had all her teeth out for gastritis before she was 21—another suspicious symptom.

That was a case of a very nervous woman, and it could well be imagined that such a woman, placed in any position where accidents could happen, would probably have one. She was always in a state of emotion or fear about something. The case shows also the relation of psycho-neurosis to sick rates.

Dr. Culpin said he would now attempt to show how certain symptoms had a definite relation to accidents. There was the fear of observation to which Mr. Farmer had referred. Dr. Culpin once had a case of a man whose business it was to polish the inside of aluminium horns for loud speakers. This had to be done while the horn was rotated at a great speed. The foreman came and stood by him, and the man said, "If you stand and watch me there will be an accident." The foreman reported the man for impudence, but the manager had sufficient insight to say: "If he feels like that, don't watch him." The fear of being watched was possibly a common cause of accidents.

Then there were those people who must rush at something. For instance, a patient who had writer's cramp. He had to rush at everything he did. When asked to write, he tried to do so at such a pace that he could not form his letters. He said that he behaved in the same way when riding a bicycle, and felt compelled to rush. He admitted that he had had several accidents, and said, "I shall break my neck one day." It was that kind of thing that accounted for at least some speed accidents. If a man like that were put at work where accidents were possible, his rushing tendency would sooner or later cause him to be the subject of accident.

There is another type that may influence accident rates. The normal person on monotonous work can let his mind wander. An efficient shorthand-typist can transcribe her notes and think of last week-end at the same time, and yet do her work efficiently and satisfactorily. As a result of investigation it is established that that is the way the healthy person does routine work. The question might be put to the worker, "When you are at your work, must you think of it all the time, or not?" If the worker says, "I have to concentrate on my work all the time, or I shall make a mistake," it is symptomatic, and other nervous trouble is generally present. The effort of concentration upon work which should be routine and easy is a bad sign.

It is such points that emerge in the investigation of individuals and carry on to a further stage the work which Miss Newbold has been doing. Whether it would be possible to correlate them with more objective tests in order to make the latter of value, it was impossible to say, but it was a thing to be hoped for. Although all the information given above was got out of a girl in a quarter of an hour, it was voluntary, and the girl was sure that nothing she said would be used against her, and she was therefore willing to let herself go and tell all about herself. The situation might arise that, when dealing with employees, an interview of that character might not be possible, and that was why Mr. Farmer's work was so valuable. If objective tests that would pick out the same class of people could be discovered they would be of great assistance.

Dr. Culpin said he did not wish to detain the meeting longer, but he would like to add that he agreed with Miss Newbold that enough had been done in this particular work to warrant a conclusion that research in this field should be continued.

Mr. ERIC FARMER said that it gave him very great pleasure to be allowed to state publicly before the Society how greatly indebted he and his colleague were to the work of Mr. Yule and Dr. Greenwood, and how much he personally was indebted to the constant co-operation with Miss Newbold through several years of research work on accident causation. It was an exceedingly interesting example of the work that had been done by the Industrial Fatigue Research Board, and of the way in which two sciences could co-operate.

The work done by Dr. Greenwood and Mr. Yule was originally of a purely theoretical kind, and they had no particular practical end in view, but their work made possible the subsequent work which the Industrial Fatigue Research Board had carried out, the object of which was to try and diagnose still further the reason for the unequal distribution in accident causation.

Mr. Farmer said that he had been mainly concerned in giving psychological tests, and as Dr. Culpin had given an idea of the method by which the more intimate psychological side of the investigation was approached, it might be of interest to Fellows of the Society to know that the tests which had given the best results dealt with sensory motor co-ordination. As Miss Newbold had said, the correlation between performance in these tests and accident causation was rather low; Miss Newbold had said "extremely low," but Mr. Farmer would prefer to say "rather" low. He thought there were many reasons for this; one might be that the tests were not well adapted to their purpose. That was possible, and in this work, which was entirely new, one could only go on giving further tests and developing those that showed promise of fruitful results.

There was another reason why tests of this kind might give rather low correlations. It was somewhat different from getting correlations between intelligence tests and tests of scholastic ability as measured by examination. The correlation between these two forms was on the whole of a fairly high order, but when one came to deal with the question of accident causation, one was not dealing with a simple factor like educational ability, but with a very complex situation—in fact, with the whole reaction of a person during a large portion of his working life. It was obvious from a theoretical point of view that accidents were caused by a multitude of different causes, and that any one test or group of tests could at the best only hope to measure qualities which were responsible for a relatively small number of accidents. The complex nature of the problem presented special difficulties to the research worker, who could not be certain at the present stage whether a low correlation between a test and accident incidence was due to the poorness of the test, or because interfering factors were operating. Mr. Farmer, however, believed this line of research would ultimately lead to a combination of a large number of tests, which would more accurately test the whole mental state predisposing to a high accident rate than any one test.

The part played by Miss Newbold in this research was far greater than strangers listening to her paper would for a moment suppose; she gave so much credit to others, but in her criticism of the work which Mr. Chambers and Mr. Farmer so constantly brought to her—work which presented new problems—it often happened that she was able to help with notable suggestions as to how to work out the data. Her co-operations had raised the standard of psychological methods of approach, and Mr. Farmer and his colleagues were really greatly indebted to her. Those attending the meeting should not be allowed to go away with the impression that Miss Newbold had done nothing more than analyse or criticize someone else's data; she had played an important creative part in the examination of the psychological aspects of accident causation. It was a subject thoroughly worth going on with, and could never have been made possible without the work which Mr. Yule, Dr. Greenwood, and Miss Newbold had done in pointing out the inequality of accident distribution.

Dr. Isserlis said that he would not occupy time by repeating at length the very high opinion that had been expressed of the theoretical treatment in Miss Newbold's paper. It had given him particular pleasure to see the most excellent use that she had made of one of Tschuprow's papers, as it had been his duty to communicate that particular paper to an English journal. When he read Miss Newbold's paper in proof, he had some little difficulty in finding parts with which he could quarrel or criticize, but he did find one or two places, and he would like to draw attention to those, because they suggested that the very excellent work that Mr. Yule, Dr. Greenwood, and Miss Newbold had done on the subject was not final, and that mathematics had not absolutely solved the problem. He was led to that by the reference in the practical part of the paper to accidents among dockers. Most of the accidents to which the statistical treatment in the paper applied were accidents where the person responsible for the accident and the person suffering from the accident were one and the same. The interesting thing about the distribution of accidents amongst dockers was this, that the document from which Mr. Wilson was quoting earlier in the afternoon showed that amongst dockers there was that same persistence of a small number occurring with small variation year after year. The proportion of fatal accidents among dockers was fairly constant from year to year, but, in the case of dockers, one man was usually responsible for the lapse, but it was another man who was killed. It was the man working the derrick who made one of those slips, but it was someone in the hold who was killed as a result.

It seemed to Mr. Isserlis that the tests employed by Mr. Farmer and by Dr. Culpin could hardly be applied to the man who figured as one of the victims of accident; they would have to be applied to someone not mentioned in the return.

Miss Newbold had drawn attention to the fact that she could look at a problem in two ways, and reach the same result. She could take one person for a number of periods and enumerate the periods in which an accident occurred, or consider the accidents to many persons in one period. The first method suggested an application to Metchnikoff's problem. Metchnikoff, after trying very hard, failed to find an example of natural death amongst man. According to him all mankind died from accident. The chance that any speakers in the discussion would die from fatal accident within the next quarter of a second was very small, but if one took a sufficient number of these brief intervals one would meet with one of those things which Metchnikoff described as a fatal accident. The old gentlemen who, having attained the age of 100 years, put their longevity to having abstained from a particular brand of cigars, really belonged to those who had been sufficiently lucky in a given period. That was a field that would, perhaps, permit of further mathematical investigation.

There was one other point to which he wished to refer, namely, the question of different liability to accident when one realized the difference of responsibility. Dr. Isserlis was not a tennis player, but he enjoyed watching tennis, and he had learned that players were apparently permitted to serve two balls. If fairly good play were watched, it would be seen that double faults were comparatively rare, but there seemed to be an extraordinary proportion of faults in the first service. People who were allowed to play two balls played the first ball badly. The newspapers that morning were dealing with the proposal to establish at Manchester a bowling-green of the southern type, and referred to the fact that the northern players were apt to find themselves handicapped when they played the new game against southern exponents who were only allowed two bowls each. That would suggest that the feeling of responsibility did in a great many cases diminish liability to accident.

Dr. Isserlis had very great pleasure in supporting the vote of thanks that had been accorded to Miss Newbold, both for the excellence of her mathematical treatment and the general importance of her results.

Miss NEWBOLD, in reply, thanked the Society for the kind reception given to her paper; it was much kinder than the paper deserved. Some of the points that had been raised in the discussion dealt with sides of the question rather outside her province. Dr. Culpin's particular cases had been specially interesting, because this problem of accident was one in which careful consideration of detailed particular cases which did not lend themselves to statistical treatment was obviously just as necessary as the statistical consideration of mass figures in which great detail was not available.

With regard to Dr. Greenwood's point that Professor Tschuprow's notation should be familiar to all mathematical statisticians, Miss Newbold felt that there was often an unnecessarily wide gap between people who were interested in the theoretical side of statistics and those who were dealing with practical problems, and that one aim of the paper had been to give a simple practical illustration of the application of Professor Tschuprow's work.

Miss Newbold agreed with Mr. Farmer that it was more than probable that the low correlation between tests and accidents was due to the complexity of the factors concerned making a high correlation impossible.

If there were more points in the discussion to which Miss Newbold had not replied, she would like to answer them in writing, but she wished to thank all those who had joined in the discussion for their helpful comments and suggestions.

Miss Newbold's written communication is as follows:—

I do not think I have anything to add, except to agree with Dr. Isserlis as to the need for distinguishing between the cause and the victim of an accident. The constancy of small numbers might be applicable in either case, but the individual distributions naturally involve quite different points. The question of the effect of responsibility is clearly one that needs consideration, and Mr. Farmer has been going into that point, but it is difficult to get enough data to draw any statistical conclusions in the case of accidents in which serious consequences might be foreseen.

As a result of the ballot taken during the meeting, the candidates named below were elected Fellows of the Society:—

Sydney Harold Bladon.
Leonard Frank Cheyney.
Henry Arthur Churchill.
John Henry Cooper.
Ernest Spencer Gallimore.
N. F. B. Osborn, C.B., M.A.
John Henry Richardson, M.A., B.Sc.
Robert Bruce Wycherley, M.C.