ACCIDENT PRONENESS

By MAJOR GREENWOOD, F.R.S.

[*Editorial note.* This paper was probably Major Greenwood's last completed work. It was submitted for publication on the day of his death, 5 October 1949. It is hoped to include some account of his life and work in the next issue of *Biometrika*.]

An eminent psychologist has written: 'When during the last war two statisticians had their curiosity aroused by some accident figures in munition works, history was made' (Smith, 1943, p. 183). Dr May Smith did not mean that these two statisticians were the first to remark that some people had clumsier fingers and (or) less alert brains than others, a fact no doubt commented on by the overlookers in manufactories of flint arrows in the old stone age; she meant that they were the first (so far as at present known) to propose an arithmetically simple test, based on the accident frequency-distribution which would result if those exposed to the risk of accident, in an environment which did not change, had some human characteristic, later to be christened Accident Proneness, which was a continuous variable adequately represented by a Pearsonian curve of type III. Were this true, then the distribution of 0, 1, 2, etc., accidents would be a negative binomial of the form

$$\left(\frac{c}{c+1}\right)^r\left[1+\frac{r}{c+1}+\frac{r(r+1)}{2!(c+1)^2}+\frac{r(r+1)(r+2)}{3!(c+1)^2}+\dots\right], \tag{1}$$

where r and c can be deduced from the first and second moments of the distribution (Greenwood & Yule, 1920).

This simple test interested psychologists and suggested to them the possibility of detecting persons prone to accidents *before* they had had any accidents. The first thing to do was to apply suitable tests to persons who had been or were about to be exposed to the risk of accidents, and to learn whether the persons with high (or low) marks in the tests had low (or high) accident rates.

The first results of such an inquiry appeared in 1926 (Farmer & Chambers 1926), and several other reports have since been published by the same and other investigators; indeed, there is a large 'literature' of the subject. I do not propose to epitomize the literature, but believe that the following paragraphs fairly estimate what is now established.

A large number of tests have been tried. Some of these are primarily measures of corporeal dexterity, others—at the other end of the scale—explore, mainly, the conative or the cognitive sides of human nature or, if the reader prefers, of the psyche. Results show conclusively enough that there is a statistically significant correlation (measured, of course, in various ways) between performance in some of the tests and accident scores. We may therefore hold not only that 'accident proneness' is a human characteristic, but that it is measurable and that it is reasonable to believe that if persons whose scores in tests found to correlate with accident scores in these researches were excluded from occupations involving special risk of accidents, the accident rate in such occupations would be reduced. But so far no test, or battery of tests, has been found which would exclude the highly accident-prone *without* excluding a considerable proportion of entrants who are not specially 'bad risks'.

This difficulty is not peculiar to the study of accident proneness. It is found in all attempts at *a priori* selection of persons suitable for occupations. But it is a *serious* difficulty because, until it is overcome, it means that we cannot by purely arithmetical methods discover the true frequency distribution of the variable we call accident proneness, because we only have the distribution of accidents. We can only guess. We know empirically that a very large proportion of accident distributions give *some* of the criteria of a negative binomial, viz. the appropriate first and second moments of such a distribution. We know nothing more. E. M. Newbold (1927, pp. 504–5) in her classical memoir found in eleven instances that the proneness distribution was either in the type VI_J or type I_J area for ten cases while for one case it was in the 'impossible area', when the first four moments of the accident distributions were used. She justly remarks that, owing to the large probable errors of the higher moments and the fact that type III is at the dividing line of two areas, all one could say was that type III was a reasonable choice. Readers of *Biometrika* do not need to be told that because an aggregation of frequency distributions, made by summing the zeroes, the 1's, the 2's, etc., of the separate distributions, give a negative binomial, the finding is no proof that proneness was responsible. We should necessarily reach a negative binomial if the items so aggregated were Poissons—unless, of course, they were identical Poissons (Pearson, 1917, p. 139). Some beginners have overlooked this. It may, perhaps, be worth noting that if the components aggregated—all 0's together, all 1's together, etc.—were Poissons, the variance of the resulting negative binomial will be smaller than it would have been if the several components were themselves negative binomials with the same means.

There is an oddity about equation (1) as practically applied which has not been noticed. As the constants of the proneness distribution are deduced from those of the accident distribution, they must vary from factory to factory if the external risk varies. But, by hypothesis, proneness is a character of the individual, a psycho-physiological property which exists whether he is subjected to the hazard of accidents or not. This did not affect the use of the method made by Greenwood & Yule, who assumed that in each factory the risk was constant, but it does make comparison of factory with factory difficult. Suppose then that we put the fundamental Poisson in the form

$$e^{-kx}(1+kx/1!+k^2x^2/2!+k^3x^3/3!\ldots), \tag{2}$$

where k is constant for any one factory or department of a factory, but different for different factories or departments. Then our accident distribution will become

$$c^r/(c+k)^r\,(1+kr/(c+k)\,1!+k^2r(r+1)/(c+k)^2\,2!\ldots). \tag{3}$$

At first sight, this is merely trivial; if we write in (2) $y = kx$, we simply have a straight Poisson in y, and if we write ck for c in (3) we are back to (1). But it does make *some* comparison between different sets of data possible and admits of a process akin to the standardization of death-rates. Suppose we take an accident distribution, *any* accident distribution provided it is effectively a negative binomial distribution with a mean m, and adopt this as the standard form, viz. assume its k to be unity. We take now another accident distribution having a different mean m' which again is a negative binomial. Now put $k = m'/m$ and substitute for the c of the standard form $c/k = C$. We shall have again a negative binomial having the required mean. Does this effectively graduate the experience? Trials on some of Newbold's data (Newbold, 1926) showed that this very naïve plan did not give bad results, but they were not at all convincing because the range of means was narrow. She had, however, one pair of observations from the same factory, a manufactory of sweets, where the length of observations was the same for both sets but the means of accidents quite different. The sugar boilers had an accident mean of 2·50 and the other constants were $c = 0{\cdot}60902$, $r = 1{\cdot}52255$. The department of chocolate moulding had an accident mean of 3·94, with $c = 0{\cdot}36462$ and $r = 1{\cdot}4366$. Taking the sugar boilers as standard, k for moulders is 1·576, and the c should be 0·3864, which is not widely different from its empirical value 0·36462, while 1·4366 is within sight of 1·52255. Of course this is only what Darwin might have called a Tom Fool Experiment, but there may be something in the *idea*. At least, if the results of 'standardization' are widely discrepant from the observed results, it suggests heterogeneity.

An obvious cause of heterogeneity is selection. The second column of Table 1 gives the observed values of accident frequencies in 166 London Transport bus drivers (first year of observation) studied by Farmer & Chambers (1939). The second column gives the expected

Table 1. *Accidents of* 166 *bus drivers in first year of observation* (Farmer & Chambers)

No. of accidents	Observed	Expected 1	Expected 2	Expected 3
0	45	40·2	50·3	45·4
1	36	45·4	41·6	41·7
2	40	34·2	28·6	31·2
3	19	21·5	18·2	20·7
4	12	12·1	11·2	12·7
5	8	6·4	6·7	7·1
6 or more	6	5·5	9·4	7·2

Expected 1 is obtained by fitting from the observed mean and variance of the data. $P = 0{\cdot}38$ for 4 degrees of freedom. Expected 2 is fitted by 'standardization'; agreement in total and mean are forced. $P = 0{\cdot}19$ for 5 degrees of freedom. Expected 3 is fitted from the curtailed gamma function (the gamma function which is *assumed* to represent the proneness of the standard population) and gives the accident mean of the observed data. $P = 0{\cdot}64$ for 5 degrees of freedom.

numbers when c and r were deduced from the observed mean, 1·81, and variance, 2·91092. The c computed by 'standardization' on the sugar boilers was half the observed value, 0·8397 instead of 1·6519, and the observed r twice the 'invariant' 1·5225. The observed accident variance was 2·91092, but 'standardization' required it to be 3·9727. The 'expected' frequencies, although not so good as those directly calculated, were not bad (third column), but evidently the variance is hopelessly large. The results *ought* to be bad. Bus drivers are a highly select population (I shall refer to it when I return from this digression); the variability in proneness of the exposed to risk must be less than that of the confectionery workers, who were certainly not so stringently selected.

It is quite easy to determine the effects of selection, *if* proneness is obliging enough to have a gamma function distribution, thanks to the existence of Pearson's *Tables of the Incomplete Gamma Function.** Taking the sugar boilers as standard population, the mean of the bus drivers is reproduced by a curtailment of about 16% of the frequency. Using this, the process just indicated leads to the third set of 'expected' accidents which is an excellent fit. The arithmetic amused me and may, perhaps, amuse the reader, but I do not suggest that it is of much practical value; I have in fact tacitly dropped the notion

* The terms in (1) were obtained by integration of successive complete Γ functions; these must be replaced by the appropriate $I(u, p)$'s.

of a k constant for any one factory department, but changing from department to department. But if we take as standard population any population with a greater mean and variance than the population to be standardized, then—provided the accident distribution of the latter is fairly well fitted by a negative binomial—we should expect the process to give us a fairly good fit. Common sense, however, tells us both selection of workers and differences of environmental risk are involved. After all, we *know* nothing more than the accident distributions given by samples, and not very large samples, of two populations; too much room for guessing remains even for those whose algebra is better than mine.

Since the publication of Newbold's memoir, field workers have not bothered about a determinist approach on these lines and have used Newbold's stochastic results. To these, with certain exceptions to be noted (see p. 28), little has been added; it has, however, been pointed out that the algebra would have been simplified by working with factorial instead of power moments.

Newbold defined *statistical* proneness as meaning that in unit population $\overset{\infty}{\underset{0}{S}} f_{\lambda_r} = 1$, the accidents happening to the subfrequency f_{λ_r} will be distributed by a Poisson law with parameter λ_r, so that the 'accident universe' will be an aggregate of Poissons. What we have is a sample of persons from a universe so defined; from such a sample we can stochastically estimate the parameters of the 'universe'. By definition, the best available estimate of the mean $\bar{\lambda}$ is the mean of accidents, $\bar{A}$, and it is easy to see that if N (the number exposed to risk) is so large that $N-1$ does not differ appreciably from N, the best estimate of the variance of λ is the variance of accidents less the mean. In the particular case leading to equation (1) this is obvious. The variance of the gamma function is r/c^2, that of the accident distribution, $r/c^2 + r/c$. If $A_i = \lambda_i + \epsilon_i$, where ϵ_i is a sampling error and if λ_i and ϵ_i are not correlated, we have

$$r_{A\lambda} = \frac{\sigma_\lambda}{\sigma_A} = \sqrt{\left(1 - \frac{\bar{A}}{\sigma_A^2}\right)}.$$

If some other variable, x, say the scoring of a person in a test, is correlated with his accident score, then

$$r_{\lambda x} = \frac{r_{xA}}{\sqrt{\left(1 - \frac{\bar{A}}{\sigma_A^2}\right)}}.$$

These are the relations which have been most used by field workers, whose object has been to reach a test, or battery of tests, which correlates highly with λ.

The difficulty of non-comparability of data collected under different environments of course remains. For instance, suppose correlations of test scores and accidents vary significantly between different studies. Is that due to selection? Dr Irwin has pointed out to me how that question could be answered, if one had rather more adequate numerical data. But the variation of what I have called external risk remains. *Any* statistics of accidents must contain a number of events which have no connexion with the personal qualities of the exposed to risk. We may safely infer from what we already know that the proportion is not large, but it must vary with the occupation. I suppose a minute—and objective—record of *every* accident entered on the statistical statement *might* enable us to eliminate these, but it would be a statistician's paradise in which such information was available.

Between 1927 and 1941 no important additions were made to the field-worker's statistical tools, but at least two mathematical statisticians pointed out that a negative binomial distribution of accidents, although a necessary, was not a sufficient condition for the validity of the proneness hypothesis (Irwin, 1941, pp. 102–5). For reasons given by Irwin (pp. 106–7; see also Greenwood, 1941, pp. 107–8), the alternative hypotheses probably need not be entertained when the question is of the kind of industrial accidents which have furnished the data used, so far, by all field workers.

In 1941, Yule showed that if proneness, λ, is not increased by increased length of exposure to risk, then if we write $\bar{A}_k$ for the mean accident rate over a time of exposure k, then ${}_kv_\lambda^2 = \frac{{}_k\sigma_A^2 - \bar{A}_k}{\bar{A}_k^2}$ is invariant with respect to time of exposure; then ${}_kr_{A\lambda}$ is given by

$$ {}_kv_\lambda = {}_kr_{A\lambda}\frac{{}_k\sigma_A}{\bar{A}_k}. $$

Chambers, however, found from the data of bus drivers that ${}_kv_\lambda$ decreased with length of exposure, i.e. is changed by experience. For bus drivers in the first year ${}_kv_\lambda$ was 0·578, in the fifth year 0·444.

On the practical side much has been done in the last 20 years, especially by Farmer & Chambers. In their first report (1926) the subjects were girls employed in covering or packing sweets, dockyard and R.A.F. apprentices; in their last report (1939) bus drivers in the London area. These reports have certainly verified the existence of accident proneness and proved that this quality is correlated with the scoring of certain sensori-motor tests. One may conclude, as Irwin put it (1941, p. 107): '(1) That, apart from differences between drivers, accidents are certainly occurring at random. (2) The absence of a significant difference from year to year rules out the hypothesis of increasing liability to accident due to previous accidents.' The reader will remark that the first report held out more encouragement to job selectors than the last. The conclusion was reached then that 'the final weighted results show a difference of 48 % in accident rate between those above and those below the average in the tests' (Farmer & Chambers, 1926, p. 36). The study of bus drivers led to the following conclusions (Farmer & Chambers, 1939, p. 36). If the worst quartile of testees had been eliminated the subsequent accident rate on the retained would be 7 % less than it actually was for the total. If not only the worst quartile, but any other drivers who had three or more accidents in the first year were eliminated, a reduction of 44 % of the exposed to risk, the improvement would have been 13 %. There is, of course, no inconsistency between the results. Dockyard apprentices and employees of a wholesale confectioner do not have to pass the rigorous tests endured by would-be bus drivers. The eliminating tests are unlike those of the field psychologists, but they may well overlap them. Farmer & Chambers give the test scores of 128 of the 166 bus drivers. I find that the thirty-eight drivers who were not tested did not have accident scores significantly different from those of the 128, so we may regard the latter as fairly representative.

I make the correlation between accident score and test score for the first and fifth years 0·12 in each case; neither is 'significant' by itself, but, even assuming that there is really a positive correlation, its order of magnitude is too small to have much practical value if the regression is linear. Naturally what the field workers would like would be an unselected population of future motor-car drivers, all of whom are tested and all of whom are obliged to report every accident, however trivial. Outside a police state such a statisticians' paradise is a dream—or nightmare.

REFERENCES

CHAMBERS, E. G. (1941). *J.R. Statist. Soc. Suppl.* **7**, 89, 95.
FARMER, E. & CHAMBERS, E. G. (1926). *Rep. Industr. Fatig. Res. Bd*, no. 38. H.M.S.O.
FARMER, E. & CHAMBERS, E. G. (1939). *Rep. Industr. Hlth Res. Bd*, no. 84.
GREENWOOD, M. (1941). *J.R. Statist. Soc. Suppl.* **7**, 107.
GREENWOOD, M. & YULE, G. U. (1920). *J.R. Statist. Soc.* **83**, 255.
IRWIN, J. O. (1941). *J.R. Statist. Soc. Suppl.* **7**, 101.
NEWBOLD, E. M. (1926). *Rep. Industr. Fatig. Res. Bd*, no. 34. H.M.S.O.
NEWBOLD, E. M. (1927). *J.R. Statist. Soc.* **83**, 255.
PEARSON, K. (1917). *Biometrika*, **11**, 139.
SMITH, MAY (1943). *An Introduction to Industrial Psychology.* London.
YULE, G. U. (1941). *J.R. Statist. Soc. Suppl.* **7**, 91.