STATISTICS IN MEDICINE, VOL. 2, 253–258 (1983)

BIMODALITY OF BLOOD PRESSURE LEVELS

I. C. McMANUS*

Department of Psychiatry, St. Mary's Hospital Medical School, London, W2, England and Department of Psychology, Bedford College, Regent's Park, London, NW1, England

SUMMARY

I examine distribution of systolic blood pressure by means of maximum likelihood fitting of various statistical models to the extensive data of Bøe *et al.*, and find evidence for a compound or mixed log-normal distribution, rather than the single log-normal distribution proposed by Pickering. I find a second distribution with a mean blood pressure above the ordinary range. I demonstrate further evidence for the existence of this subgroup in the abnormalities of regression to the mean found in the same data set.

KEY WORDS Blood pressure Bimodality Compound distributions Maximum likelihood fitting Regression to the mean

INTRODUCTION

During the 1950s and 1960s a major controversy raged concerning the distribution of blood pressure in populations.[1–4] Although expressed in the language of genetics (one gene *vs* polygenes), or causation (unifactorial *vs* multifactorial) or aetiology ('disease' *vs* 'continuum'), the main contention concerned statistical matters. Pickering *et al.*, in a series of influential papers,[5–8] suggested that there is not a disease entity called 'hypertension'; rather, blood pressure is a continuous variable, and the risk of complications arising from high blood pressure was simply proportional to the height of that blood pressure. Opponents of this theory, represented primarily by Platt,[2–4] argued that there are two or more subgroups of individuals within the population, with some subgroups having higher blood pressures than others (perhaps due to a single Mendelian gene).

In 1968, with the publication of the second edition of Pickering's *High Blood Pressure*,[1] the battle seemed fought and won by the 'single distribution school'. At that time, and later,[9] the main points of Pickering's statistical argument could be summarized thus: (i) blood pressure is distributed continuously and unimodally in the population, and (ii) because biological effects are more likely to act multiplicatively rather than additively, the distribution may be regarded as a log-normal rather than a simple normal (gaussian) distribution, a theoretical justification for this latter point having been recently presented.[10]

The most striking aspect of the entire controversy is that there have been almost no formal attempts at statistical testing of the goodness of fit of the various statistical distributions; the criterion of 'fit' has almost always been the appearance of the graphs. In the present paper I wish to present the results of formal maximum likelihood fitting[11] of various statistical models to the distribution of systolic blood pressure.

* Lecturer in Psychology as Applied to Medicine.

0277–6715/83/020253–06$01.00 *Received September 1982*

A normal distribution has two parameters, the mean (μ) and the standard deviation (σ). We may specify a particular normal distribution by $N(\mu, \sigma)$. In a mixed or compound distribution in which a proportion, p, of the population come from a normal distribution $N(\mu_1, \sigma)$, and the rest, q ($p+q=1$) come from a distribution with a different mean, $N(\mu_2, \sigma)$,[12] the total distribution $pN(\mu_1, \sigma)+qN(\mu_2, \sigma)$ will only be normal if $\mu_1 = \mu_2$. If μ_1 and μ_2 are widely separated then the distribution will have two clear modes. As μ_1 and μ_2 approach each other the distributions will merge,[13] so that only one mode will be apparent; and this merging of one distribution in the tail of the other will occur more readily if $p \ll q$, or vice versa. Nevertheless, despite there being only one visible mode, the distribution is still compound, since it contains individuals drawn from two separate populations, and it may be loosely referred to as bimodal and perhaps more correctly as biphasic, compound or mixed. The problem, given a particular data distribution, is to reverse the process, thereby recovering the original distributions, and providing adequate statistical evidence that the biphasic distribution is a better fit than a single distribution. In general the equations for such problems are difficult to solve;[14] however estimates of parameters may readily be found using the method of Newton–Raphson iteration on a computer.[15]

The problem in the present case is further complicated by the fact that there are good reasons for believing that a log-normal rather than a normal distribution is the better fit.[16-18] In the same way as we may have a biphasic normal distribution, so we may have a biphasic log-normal distribution, p of the population coming from a distribution $\Lambda(\mu_1, \sigma)$ and the remaining q from a distribution $\Lambda(\mu_2, \sigma)$, the population having a distribution $p\Lambda(\mu_1, \sigma)+q\Lambda(\mu_2, \sigma)$. It is an empirical question as to whether a single or a biphasic log-normal distribution, and whether normal or log-normal distributions (single or biphasic) are the better fits to the data. Most blood pressure distributions are skewed, and this gives a clear precedence to a simple log-normal distribution over a simple normal distribution. However biphasic normal distributions may readily be skewed, by the use of appropriate parameters.

METHOD

The data re-analysed in this paper are those of Bøe *et al.*,[19] who measured the B.P. of 67,976 persons aged over fourteen years in the city of Bergen in Norway. These individuals had presented themselves for mass X-raying (as required by law); they represented 93 per cent of those attending for X-ray, and 84·5 per cent of the 80,471 eligible persons in the population. The study was carried out in two areas, I (Northern and Central districts of the city) in 1950, and II (Southern districts of the city) in 1951. In the present paper I analyse only the systolic pressures for area I. The advantages of Bøe's results for the present investigation are that the sample is large, the study is relatively uncontaminated by selection biases, and was carried out before the advent of effective anti-hypertensive regimes[20, 21] which might be expected to modify the population distribution of blood pressure.

I analysed results separately in each of thirteen age groups for each sex. In each case I compared a biphasic log-normal model with a simple log-normal model. I obtained maximum-likelihood estimates of the parameters by means of a quasi-Newton–Raphson method, and assessed significance by the likelihood-ratio χ^2 test.

RESULTS

In 17 of the 26 subgroups there was statistically significant ($p < 0{\cdot}05$) evidence of a subgroup with raised B.P., and in 10 of these cases the improvement in fit was significant with $p < 0{\cdot}001$. I also fitted biphasic normal distributions, but these, in almost all cases, did not fit as well as biphasic log-

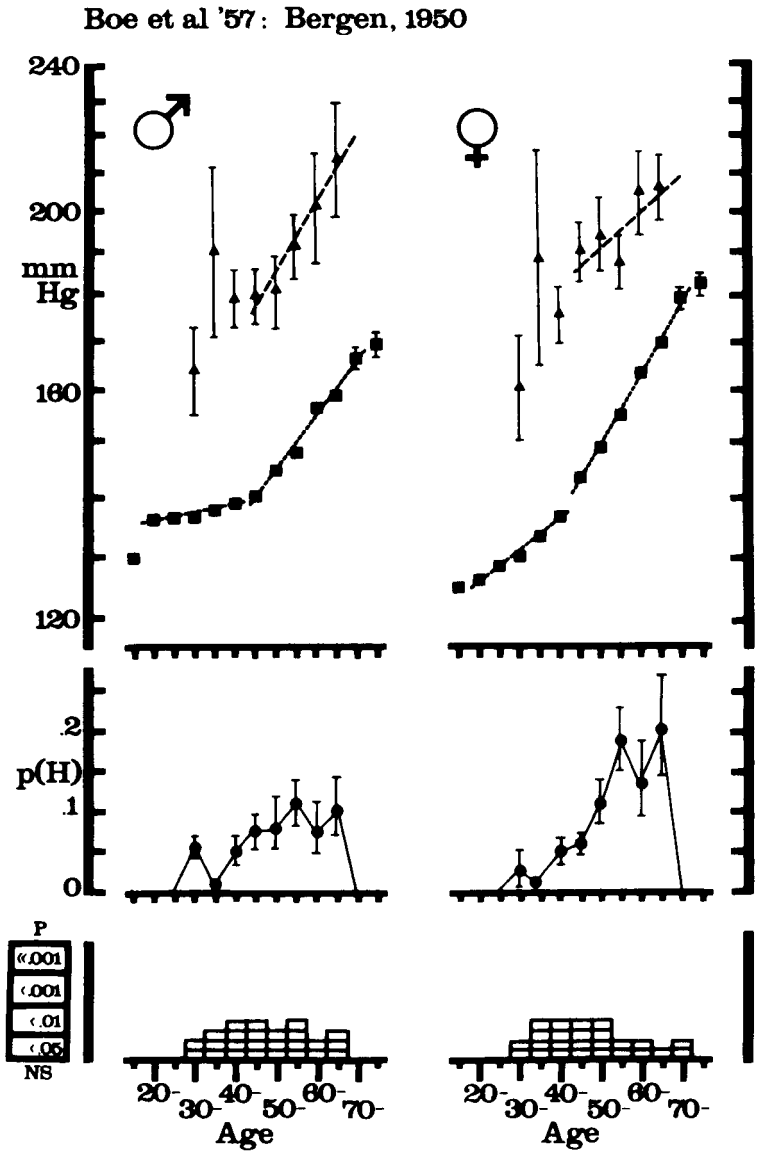

Figure 1. The top part of the figure shows the geometric mean systolic pressures of the high (▲) and low (■) blood pressure groups, separately for each sex, by age group (in five year intervals 15–19, 20–24, etc.). Approximate 95 per cent limits are plotted for the low B.P. groups, where these limits are larger than the size of the plotted symbol. The lines on the Figure represent weighted least squares regressions for the age groups 20 to 44 and 45 to 74. The ordinate is on a logarithmic scale. The middle part of the Figure (●) shows the estimated proportions (with approximate 95 per cent limits) of individuals in the higher B.P. group ($p(H)$), by age and sex. The bottom part of the figure shows for each of the 26 age × sex groups the statistical significance of the improvement in fit by using a compound log-normal distribution as opposed to a simple log-normal distribution. A single box indicates $p < 0{\cdot}05$, a double box $p < 0{\cdot}01$, a triple box $p < 0{\cdot}001$, and a quadruple box $p \ll 0{\cdot}001$ (i.e. $\chi^2_2 > 20, p < 4{\cdot}5 \times 10^{-5}$). The analysis of the data for females aged 70–74 indicated significant ($\chi^2 = 12{\cdot}2$, 2df, $p < 0{\cdot}01$) evidence for a compound distribution; however the parameters for the compound distribution (mean for low group = 163·7; mean for high group = 204·7; proportion in high group = 0·410) were so discrepant with other results, that this was presumed to be a type I error, and the simple distribution used

normal distributions; similarly models with unequal variances fit no better than models with equal variance. Figure 1 plots, separately for each of the groups, the fitted parameters of the distributions. A number of features of Figure 1 are apparent:

(i) The proportion of individuals in the 'High' group is minimal in the lower age-groups, but rises steadily to the age of 65–69. The proportion then falls away rapidly, presumably due to either increased mortality in the 'High' group, or increased morbidity (and hence failure to attend for examination).

(ii) There is evidence for a higher proportion of females than males in the 'High' group. There is no evidence of a sex difference in blood pressure of those in the 'High' group. However for those in the 'Low' group it is clear that female pressure is lower than male until the age of 45, after which the female pressure is higher.

(iii) If we consider just the 'Low' groups it appears that the change of B.P. with age can be considered as being in two components; firstly a relatively slow rate of increase up to the age of 45, and then a more rapid rate of increase after that. One may fit these two proportions of the curve by straight lines (representing exponential increases on a linear scale), and Figure 1 shows weighted regression estimates of these lines. The rate of increase in the 'High' group is much the same as that in the older part of the 'Low' group. Extrapolating backwards from the 'High' group suggests that

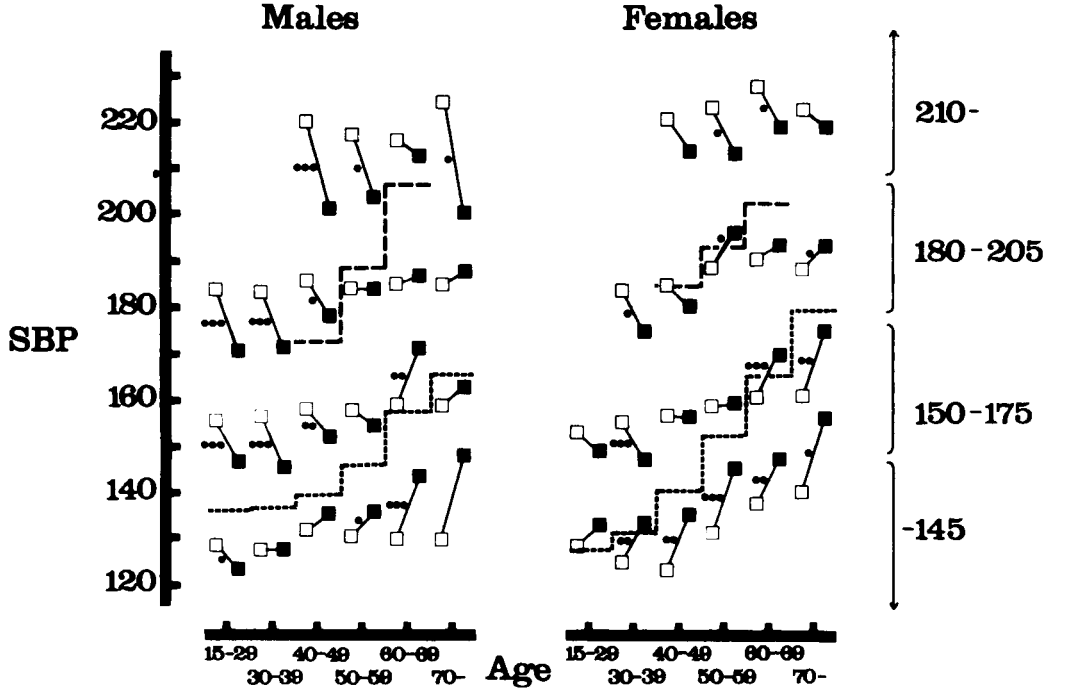

Figure 2. Shows the B.P. of the same group of subjects as Figure 1, remeasured after an interval of between one and two years. The open squares give the mean B.P. of the subjects on their first assessment, and on which basis they were stratified into four groups (< 145; 150–175; 180–205 and > 210 mm Hg). The solid squares indicate the mean B.P. of each age-stratum group at re-assessment, each being joined to its appropriate first assessment mean by a solid line. The dots alongside pairs indicate the statistical significance of the change in B.P.: ● , $p < 0{\cdot}05$; ●●, $p < 0{\cdot}01$; ● ●●, $p < 0{\cdot}001$. The dashed lines indicate the expected means from the values given in Figure 2 for the 'Low' group (– – – – – – – – – – –) and the 'High' group (————————). The overall population means are slightly above the position of the 'Low' group, but, to avoid confusion, have not been plotted separately. The ordinate is linear, since in the original work arithmetic means were calculated

the curves of the 'Low' and 'High' groups would meet at an absurdly low age, implying that the B.P.s of those in the 'High' group cannot be accounted for in terms of a higher rate of increase from an early age, but must rather be considered as showing evidence for a sudden and discrete rise in B.P. followed by a slower subsequent rise.

REGRESSION TO THE MEAN

If the hypothesis of a subgroup with higher blood pressure is correct then we should be able to find evidence of it on re-testing of the same individuals. If a variable has a relatively poor reliability of measurement (as does B.P.) then one finds 'regression to the mean'.[22] If a population is composed of two sub-groups then we would expect that individuals in each group would show regression *to their own particular sub-group mean*, not to the population mean.

In 1951–1952 Humerfelt[3] remeasured the B.P. of a stratified sample of 1550 of the individuals from area I of the survey of Bøe *et al.*[19] He classified these individuals into four groups on the basis of their initial systolic pressure (< 145; 150–175; 180–205 and > 210 mm Hg). Figure 2 shows, for each age, sex and blood pressure group the mean B.P. on the first occasion (1950) and on the second occasion (1951–1952). The lower dashed line on each graph approximately represents the population mean in the 1950 survey. In general the lowest groups (< 145 mm Hg) show regression upwards towards the population mean. Similarly the next group (150–175 mm Hg) usually shows regression towards the appropriate population mean. The third group (180–205 mm Hg) shows a very different pattern. In the lower age groups (< 49) there is regression towards the mean. However in the higher age groups (> 50) there is no regression, but instead there is evidence of regression *away* from the mean. One may readily explain this strange phenomenon if a proportion of these individuals is showing regression towards a mean other than the overall population mean; the logic is similar to that involved in the inference of the presence of an unknown planet by its pull of the orbits of the known and visible planets from their expected courses. The highest group (> 210 mm Hg) once more shows regression to the population mean; but the relatively small effect is probably best explained in terms of most of these individuals showing regression towards the mean of the 'High' B.P. group rather than to the overall mean.

DISCUSSION

This paper has presented evidence that systolic B.P. in the population is not distributed as a single continuous log-normal distribution, but instead shows evidence of at least two subgroups, one of which has a higher mean blood pressure than the other. Previous studies have probably failed to find such a result either due to their using young subjects,[1, 17] or their having too small a population sample.[4]

ACKNOWLEDGEMENTS

I am very grateful to Drs D. G. Beevers, J. K. Cruickshank, S. Jackson and D. N. J. Lockwood for their helpful discussion and for their critical readings of the paper.

REFERENCES

1. Pickering, G. *High Blood Pressure*, 2nd Edition, Churchill Limited, London, 1968.
2. Platt, R. 'The nature of essential hypertension', *Lancet*, **i**, 55–57 (1959).
3. Platt, R. 'Heredity in hypertension', *Lancet*, **i**, 899 (1963).
4. Morrison, S. L. and Morris, J. N. 'Epidemiological observations on high blood pressure without evident cause', *Lancet*, **ii**, 864–870 (1959).

5. Hamilton, M., Pickering, G. W., Roberts, J. A. F. and Sowry G. S. C. 'The aetiology of essential hypertension. 1. The arterial pressure in the general population', *Clinical Science*, **13**, 11–35 (1954).
6. Hamilton, M., Pickering, G. W., Roberts, J. A. F. and Sowry, G. S. C. 'The aetiology of essential hypertension. 2. Scores for arterial blood pressures adjusted for differences in age and sex', *Clinical Science*, **13**, 37–49 (1954).
7. Pickering, G. W., Roberts, J. A. F. and Sowry, G. S. C. 'The aetiology of essential hypertension. 3. The effect of correcting for arm circumference on the growth rate of arterial pressure with age', *Clinical Science*, **13**, 267–271 (1954).
8 Hamilton, M., Pickering, G. W., Fraser, J. A. F. and Sowry, G. S. C. 'The aetiology of essential hypertension. 4. The role of inheritance', *Clinical Science*, **13**, 273–304 (1954).
9. Pickering, G. 'Normotension and hypertension: the mysterious viability of the false', *American Journal of Medicine*, **65**, 561–563 (1979).
10. Makuch, R. W., Freeman, D. H. and Johnson, M. F. 'Justification for the log-normal distribution as a model for blood pressure', *Journal of Chronic Diseases*, **32**, 245–250 (1979).
11. Silvey, S. D. *Statistical Inference*, Chapman and Hall, London, 1975.
12. Murphy, E. A. 'One cause? Many causes? The argument from the bimodal distribution', *Journal of Chronic Diseases*, **17**, 301–324 (1964).
13. Schmitt, S. A. *Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics*, Addison-Wesley, Reading, Massachusetts, 1969.
14. Everitt, B. and Hand, D. J. *Finite Mixture Distributions*, Chapman and Hall, London, 1981.
15. Adby, P. R. and Dempster, M. A. H. *Introduction to Optimisation Methods*, Chapman and Hall, London, 1974.
16. Murphy, E. A. 'Evaluation of clinical data: improvement of efficiency by simple transformation', *Journal of Chronic Diseases*, **15**, 795–809 (1962).
17. Murphy, E. A., Thomas, C. B. and Bolling, D. R. 'The precursors of hypertension and coronary disease: statistical considerations of distributions in a population of medical students. II. Blood pressure', *John Hopkins Medical Journal*, **120**, 1–20 (1967).
18. Aitchison, J. and Brown, J. A. C. *The Lognormal Distribution*, Cambridge University Press, Cambridge, 1957.
19. Bøe, J., Humerfelt, S. and Wedervang, F. 'The blood pressure in a population', *Acta Medica Scandinavica*, Supplement 321 (1957).
20. Tidy, H. L. *A Synopsis of Medicine*, 10th Edition, John Wright and Sons, Bristol, 1954.
21. Bell, E. T. (Ed.) *Hypertension*, University of Minnesota, Minneapolis, 1951.
22. Nesselroade, J. R., Stigler, S. M. and Baltes, P. B. 'Regression towards the mean and the study of change', *Psychological Bulletin*, **88**, 622–637 (1980).
23. Humerfelt, S. B. 'An epidemiological study of high blood pressure', *Acta Medica Scandinavica*, Supplement 407 (1963).
24. Clark V. A., Chapman, J. M., Coulson, A. H. and Hasselblad, V. 'Dividing the blood pressures from the Los Angeles Heart Study into two normal distributions', *John Hopkins Medical Journal*, **122**, 77–83 (1968).