

Letters to the Editor

Comparison of methods of measuring blood pressure

SIR—Gallagher *et al*¹ recently reported the results of an evaluation of an automatic sphygmomanometer, the Copal UA-231, including a comparison with the Hawksley “random zero” manual sphygmomanometer. Aspects of the design, and especially the analysis of their study, severely weaken the validity of their findings.

A single operator made pairs of measurements on each of 14 normal subjects on four different days. The restriction to normal volunteers means that the sample excludes hypertensive subjects. No serious evaluation of a blood pressure measuring device should exclude hypertensives, because the principal use of such instruments is the detection and monitoring of hypertension. Further, the sample size of 14 is very small for such an evaluation. Although the repetition of the comparison on four days mitigates the small sample size slightly, it is no substitute for an adequate sample.

Two types of statistical analysis were used: analysis of variance to test the overall difference in mean blood pressure between the instruments, and regression and correlation of the within-patient means by each instrument. A scatter diagram of the within-patient means was plotted, showing for diastolic and systolic pressures separately the regression lines of the UA-231 readings on the Hawksley.

The analysis of variance was presumably a three-way analysis with the main effects of subjects, days, and methods and the days \times methods interaction—we infer this from 91 degrees of freedom for the residual. Only test statistics and P values were given in the paper. No estimate was given of the variability of between-method differences, or of the measurement error. In fact the analysis of variance requires the assumption that the measurement errors by the two methods are the same, which is something that one would usually try to examine in such a study. We can do this approximately as the authors have given means for the same group of 14 patients on each of four days. Provided we assume no systematic day effects we can estimate the within-subject variances, which will be 14 times the variances of the four daily means. The estimated variances are: for systolic, 11.1 for Hawksley and 29.4 for UA231; for diastolic 26.0 for Hawksley and 100.7 for UA231. It seems unlikely, therefore, that the measurement errors are equal, so that the analysis of variance is invalid.

The authors also calculated correlation coefficients and performed linear regression. There are several reasons why these analyses are highly misleading for

such data.^{2,3} Firstly, the correlation coefficient is a measure of linear association, not of agreement: one cannot infer good agreement from a high correlation. It is not how close the points lie to the best-fitting straight line that is important, but how close they lie to the line of equality. Secondly, even if the two instruments were identical the expected value of the slope of the regression line would not be 1.0 unless there was no measurement error.² Lastly, these analyses were performed on the average of four measurements per subject. Because a lot of the measurement error will be removed, this will lead to an inflated correlation (perhaps considerably so) and will also affect the slope of the fitted line. It is reasonable to analyse the average of several readings only if that is the standard clinical measurement, which is not the case here. (Incidentally, the plotted regression line for diastolic blood pressure looks wrong, as 10 of the 14 points lie above the line.)

The comparability of methods of measurement is an estimation problem: statistical significance is irrelevant. A more sensible approach to the analysis of such data is possible using very simple methods. For unreplicated data, the mean and standard deviation of the between-method differences convey the most useful information, along with a plot of the differences against the average values by the two methods.^{2,3} For equally replicated data, the within-subject means can be used to estimate the mean difference between machines, but the standard deviation must be adjusted to allow for the averaging because we want to know the comparability of *single* measurements.³ The variability of the differences between the methods for single measurements will be considerably greater than that shown for averages in their figure. The mean differences of 5 and 1 mm Hg for systolic and diastolic blood pressures cannot properly be interpreted without some measure of variability about these mean differences. A useful way to express this variability is by the 95% limits of agreement, given by the mean difference \pm twice the standard deviation of the between-method differences.³

Unfortunately, we cannot carry out the appropriate calculations exactly from the information given in the paper, but we can make a reasonable approximation. As we have already observed, we have estimates of the within-subject variances, s_{EH}^2 and s_{EU}^2 . The authors have also given standard deviations over 14 subjects for each method for each day. The averages of the four daily variances give reasonable estimates of the between-subject variances of single measurements, s_H^2 and s_U^2 . The variances of means of four measurements, which we can denote by $s_{H\text{mean}}^2$ and $s_{U\text{mean}}^2$, are given by

$$s_{H\text{mean}}^2 = s_H^2 - 3s_{EH}^2/4 \quad \text{and} \quad s_{U\text{mean}}^2 = s_U^2 - 3s_{EU}^2/4.$$

The observed correlation between means of four measurements enables us to estimate the covariance, because

$$r = \text{cov}(H_{\text{mean}}, U_{\text{mean}}) / (S_{H_{\text{mean}}} S_{U_{\text{mean}}})$$

If $s_{D_{\text{mean}}}^2$ is the variance of differences between means over days by the two methods, we have

$$\begin{aligned} s_{D_{\text{mean}}}^2 &= s_{H_{\text{mean}}}^2 + s_{U_{\text{mean}}}^2 - 2 \text{cov}(H_{\text{mean}}, U_{\text{mean}}) \\ &= s_{H_{\text{mean}}}^2 + s_{U_{\text{mean}}}^2 - 2 r s_{H_{\text{mean}}} s_{U_{\text{mean}}} \end{aligned}$$

We estimate the variance of the differences between individual readings by the two methods by

$$s_D^2 = s_{D_{\text{mean}}}^2 + 3(s_{EH}^2 + s_{EU}^2)/4$$

(paper in preparation).

We can calculate approximate 95% limits of agreement³ as the mean difference $(U-H) \pm 2s_D$ (see table).

Pressure	Mean diff (U-H) (mmHg)	s _D (mmHg)	Limits of agreement (mmHg)
Systolic	+5.2	8.1	-11 to 21
Diastolic	+0.7	10.5	-20 to 22

The limits of agreement show the estimated range within which the differences between single readings by the two sphygmomanometers would fall on about 95% of occasions. Considerations of acceptability of a new method can be based on these limits of agreement. In this case they are quite wide, and we do not think that they support the contention of Gallagher *et al* that the UA-231 produces readings comparable with the Hawksley machine. Further, the UA-231 readings appear to be much less reproducible than those of the standard Hawksley machine. Our calculations, derived from the authors' summary statistics, are only approximate, but we think that they tell us far more about likely differences between the instruments than do regression lines and correlation coefficients.

The use of a single observer for both methods is in general a sensible approach in method comparison studies. However, in this case the observer was blind to the results by the automatic method, so it might have been better to compare the two methods simultaneously using both arms. This would not have introduced any bias⁴ and would have allowed a truer comparison of the methods. Gallagher *et al* say that the mean difference is "no greater than might be expected between observers using the Hawksley instrument", but this is not relevant. They also suggest

that "A different systematic difference between instruments might have been found with another observer". While this may be true for small samples just because of random variation, if the authors really believe in systematic differences between observers their evaluation of this new method by one observer is meaningless. In this context, their final comment that there is a "lack of observer bias" is most odd, given that only one observer was involved in this study.

The authors have not established that the UA-231 automatic sphygmomanometer is comparable with the Hawksley machine. Method comparison studies are frequently analysed wrongly, particularly using correlation and regression. This is not merely an academic point. Incorrect analyses can often, as here, lead to incorrect conclusions. We hope that this Journal and its referees will be alerted to the possibility of such errors in future papers which compare different methods of measurement.

References

- Gallagher JEJ, Yarnell JWG, Rogers S, Sweetnam P. Automatic measurement of blood pressure: evaluation of the Copal UA-231 automatic sphygmomanometer. *J Epidem Comm Hlth* 1985; **39**: 220-3.
- Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983; **32**: 307-17.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; *i*: 307-10.
- Gould BA, Hornung RS, Kieso HA, Altman DG, Raftery EB. Is the blood pressure the same in both arms? *Clin Cardiol* 1985; **8**: 423-6.

DOUGLAS G ALTMAN
Division of Medical Statistics,
Clinical Research Centre,
Watford Road,
Harrow
Middlesex HA1 3UJ

J MARTIN BLAND
Department of Clinical Epidemiology and Social
Medicine,
St George's Hospital Medical School,
Cranmer Terrace,
London SW17 0RE

SIR—Altman and Bland have made a number of criticisms of our paper,¹ which reported the results of a comparison between an automatic sphygmomanometer, the Copal UA-231, and the Hawksley "random zero" manual sphygmomanometer. Altman and Bland principally criticise our sample and the statistical analysis and