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

\[ r = \frac{\text{cov}(H_{\text{mean}}, U_{\text{mean}})}{(S_{H_{\text{mean}}} S_{U_{\text{mean}}})} \]

If \( s^2_{\text{Dmean}} \) is the variance of differences between means over days by the two methods, we have

\[ s^2_{\text{Dmean}} = s^2_{H_{\text{mean}}} + s^2_{U_{\text{mean}}} - 2 \text{cov}(H_{\text{mean}}, U_{\text{mean}}) \]
\[ = s^2_{H_{\text{mean}}} + s^2_{U_{\text{mean}}} - 2 r s_{H_{\text{mean}}} s_{U_{\text{mean}}} \]

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

\[ s^2_D = s^2_{\text{Dmean}} + \frac{3(s^2_{EH} + s^2_{EU})}{4} \]

(paper in preparation).

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

<table>
<thead>
<tr>
<th>Pressure</th>
<th>Mean diff ((U-H)) (mmHg)</th>
<th>SD ((mmHg))</th>
<th>Limits of agreement ((mmHg))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systolic</td>
<td>+5.2</td>
<td>8.1</td>
<td>-11 to 21</td>
</tr>
<tr>
<td>Diastolic</td>
<td>+0.7</td>
<td>10.5</td>
<td>-20 to 22</td>
</tr>
</tbody>
</table>

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 bias 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


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 ORE

SIR—Altman and Bland have made a number of criticisms of our paper, 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
presentation of our results, and conclude that our comparison of the sphygmomanometers is invalid.

The sample used in our study is criticised for being too small and for covering a range of blood pressure which was too narrow. We acknowledge the small sample size but point out that the study was sufficiently sensitive to detect a systematic difference of \(5.2\text{mmHg}\) in systolic pressure as statistically significant at the 1% level and would have detected a systematic difference of \(2.6\text{mmHg}\) in diastolic pressure as statistically significant at the 5% level. We also acknowledge that the blood pressure levels of the subjects in our study were rather low. We were considering the use of the sphygmomanometers in an epidemiological setting. Given this context, we disagree with Altman and Bland’s assertion that “the principal use of such instruments is the detection and monitoring of hypertension”.

The major analysis that we performed was analysis of variance to test for systematic differences between the machines in both systolic and diastolic pressure. We did not present the full analysis of variance, but, as Altman and Bland infer, we obtained main effects for subjects, days, and machines and days \(\times\) machines interaction. The only statistically significant result was the machine’s main effect for systolic pressure \(F(1.91,\text{df}=8.3, \ p<0.01)\). Altman and Bland claim this analysis to be invalid because the measurement errors are unequal. From the four daily mean values for each machine they estimate the measurement error variances to be:

<table>
<thead>
<tr>
<th></th>
<th>Systolic</th>
<th>Diastolic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawksley</td>
<td>11.1</td>
<td>26.0</td>
</tr>
<tr>
<td>UA-231</td>
<td>29.4</td>
<td>100.7</td>
</tr>
</tbody>
</table>

They say it “seems unlikely, therefore, that the measurement errors are equal”. They have ignored the fact that each of their estimates has only three degrees of freedom, and the differences between these estimates do not begin to approach conventional levels of statistical significance. In fact, from the separate analyses of variance, the estimates of the measurement error variances, each with 39 degrees of freedom, are:

<table>
<thead>
<tr>
<th></th>
<th>Systolic</th>
<th>Diastolic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hawksley</td>
<td>91.8</td>
<td>41.1</td>
</tr>
<tr>
<td>UA-231</td>
<td>94.1</td>
<td>59.4</td>
</tr>
</tbody>
</table>

This confirms that there are no grounds for claiming unequal measurement error variances. Inserting these estimates into Altman and Bland’s method for estimating the standard deviation of the difference between individual measurements on the two machines, we get:

<table>
<thead>
<tr>
<th></th>
<th>Systolic</th>
<th>Diastolic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(SD=13.1)</td>
<td>(SD=9.7)</td>
</tr>
</tbody>
</table>

These values are more in keeping with our expectations than those provided on the basis of Altman and Bland’s estimates. Indeed, it would have been surprising if, as according to Altman and Bland, the variance of a pair of diastolic readings had been greater than that for a pair of systolic readings. We can then estimate that the ranges within which 95% of all differences between single measurements with the two machines would lie, are:

<table>
<thead>
<tr>
<th></th>
<th>Systolic</th>
<th>Diastolic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(-21.0) to (+31.4) mmHg</td>
<td>(-18.7) to (+20.1) mmHg</td>
</tr>
</tbody>
</table>

The range for systolic pressure is substantially larger than that calculated by Altman and Bland, while the range for diastolic pressure is slightly narrower. Altman and Bland call these ranges the limits of agreement and state that the “acceptability” of a new method can be based on these limits. Altman and Bland conclude these limits to be too wide to support our contention that the UA-231 produces readings comparable with the Hawksley. We do not understand how these ranges can be used, in isolation, to judge the acceptability of a new method, as they are meaningful only in the context of a comparison. These ranges should be compared with those found between pairs of measurements had both been made with the Hawksley. That there is no evidence of a difference in measurement error between the two machines implies that 95% of differences between pairs of measurements both made with the Hawksley would lie within a similar range, the only difference being that in the case of pairs of Hawksley measurements, the range would be symmetrical about zero. Therefore we cannot accept Altman and Bland’s assertion that these machines do not provide acceptably comparable blood pressure readings.

The objective of our Figure was to show that the average differences of 5.2 mmHg for systolic and 0.7 mmHg for diastolic were consistent across our range of blood pressures. We accept that Altman and Bland are correct in suggesting that our purpose would have been more clearly met by using a line of equality rather than the regression lines. They are also correct that the regression line for diastolic pressure was incorrectly drawn. In its correct position, it is indistinguishable from a continuation of the regression line for systolic pressure.

Finally, we cannot accept Altman and Bland’s assertion that we have arrived at the wrong conclusion concerning our evaluation of these instruments. However, we do not consider our study to have been a definitive test, and we accept that our comments about
observer bias were speculative. Our study was a
necessary and informative precursor to more extensive
comparisons which we are currently conducting.

J GALLACHER
P M SWEETNAM
J W G YARNELL
S ROGERS
MRC Epidemiology Unit (South Wales)
4 Richmond Road, Cardiff
CF2 3AS

Cigarette smoking and benign breast disease

SIR—Berkowitz et al report a reduced risk of benign breast disease (both fibrocystic disease and fibroadenoma) in current cigarette smokers on the basis of a case-control study involving hospital patients. We summarise here our findings in the Oxford-Family Planning Association (Oxford-FPA) contraceptive study.

The Oxford-FPA study involves the long-term follow-up of over 17 000 white, married women who were aged 25–39 years when they were recruited at 17 family planning clinics during the period 1968–74. During follow-up, all outpatient and inpatient hospital visits experienced by the study participants are identified and diagnoses are coded according to the 8th Revision of the International Classification of Diseases (ICD). Changes in contraceptive use are also recorded. Smoking habit, weight, height, social class, and many other variables were noted for each woman on entry to the study.

The table presents first hospital visit rates, both inpatient and total, for ICD codes 217 (benign neoplasm of the breast, principally fibroadenoma) and 610 (chronic cystic disease of the breast). The data are given combined because the results were closely similar for the two diagnostic groups. The rates are shown crude and adjusted (by indirect standardisation) for age (four groups), oral contraceptive use (current, past, never), social class (three groups), and Quetelet’s index (six groups). An earlier analysis had shown these to be the most important “risk factors” for benign breast disease in our study.

The data provide no indication of any reduction in the risk of benign breast disease in smokers; indeed, it is the non-smokers who have the lowest rates. It is unlikely that there has been sufficient change in the smoking habits of women within the cohort to have obscured a protective effect since we have shown smoking to be related quite clearly to a number of other diseases. It is, of course, true that the great majority of women in our study are still premenopausal, while Berkowitz et al found the negative association between current cigarette smoking and benign breast disease to be stronger in postmenopausal women; nonetheless, they did still find the association to hold in the premenopausal group. We suggest that the results obtained by Berkowitz et al reflect the use of hospital controls (see Baron), perhaps compounded by inexact age matching and the loss of subjects from socially deprived areas.

D MANT
M P VESSEY
M A SMITH
D YEATES
Department of Community Medicine and General Practice
Gibson Laboratories Building
Radcliffe Infirmary
Oxford OX2 6HE

References


First hospital visit rates (per 1000 woman-years) for benign breast disease (ICD codes 217, 610) according to smoking habit at entry to the Oxford-FPA study.

<table>
<thead>
<tr>
<th>Smoking habit</th>
<th>Inpatient visits</th>
<th>Total visits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of women</td>
<td>Crude rate</td>
</tr>
<tr>
<td>Never</td>
<td>317</td>
<td>2.7</td>
</tr>
<tr>
<td>Ex-</td>
<td>111</td>
<td>3.0</td>
</tr>
<tr>
<td>1–14 cigs/day</td>
<td>71</td>
<td>2.5</td>
</tr>
<tr>
<td>15+ cigs/day</td>
<td>163</td>
<td>6.0</td>
</tr>
</tbody>
</table>

*See text