Bias and Variability in Blood Pressure Measurement With Ambulatory Recorders

Giuseppe Pannarale, Gwynn Bebb, Sue Clark, Ann Sullivan, Clare Foster, Andrew J.S. Coats

This study sought to determine whether patient characteristics such as age, sex, blood pressure, and pulse pressure differently affect the accuracy of an oscillometric (SpaceLabs 90207) and a microphonic (TM2420 version 7) blood pressure monitor. Blood pressure recorded by two oscillometric and two microphonic ambulatory monitors was compared with simultaneous readings by two pairs of trained, blinded observers using random-zero sphygmomanometry. One hundred and eighteen subjects (53 men and 65 women, aged 17 to 94 years; systolic pressure, 89 to 211 mm Hg; diastolic, 44 to 116 mm Hg) were studied. There were no significant differences within each observer pair or between the two observer pairs as well as no correlation between interobserver differences and patient characteristics. The differences between the monitor and trained observers' readings were 2.8±9.9 mm Hg systolic and 3.9±6.8 mm Hg diastolic for the SpaceLabs and 5.0±5.2 mm Hg systolic and 3.4±6.1 mm Hg diastolic for the TM2420. Patient characteristics that predicted measurement error were defined by multiple regression. For oscillometry, systolic measurement error was highly correlated with systolic pressure, pulse pressure, and subject age. The diastolic error was significantly correlated with pulse pressure, diastolic pressure, and subject sex. For the oscillometric monitor, patient characteristics accounted for 36.6% of the variation of the systolic error and 34.7% of the variation of the diastolic error. For the microphonic monitor, only age correlated with diastolic error, and no significant correlations were seen with systolic error. Patient characteristics accounted for only 1.2% of the systolic and 8.9% of the diastolic error. We conclude that the systematic variation of accuracy with blood pressure level and pulse pressure of the oscillometric monitor would bias and reduce the spread of blood pressure values in population studies. It may also affect the accuracy in patient groups with high or low pressures and wide or narrow pulse pressures. For the microphonic method, accuracy is affected by age. These differences in accuracy necessitate specific validations in patient groups in which problems may be anticipated for both types of monitor. (Hypertension. 1993;22:591-598.)

KEY WORDS • blood pressure determination • blood pressure monitors • oscillometry

The use of automated methods of blood pressure measurement is increasing rapidly in intensive care, during surgery, for home measurements of blood pressure, and for 24-hour ambulatory monitoring. There are many difficulties with accurate manual reading of blood pressure,1,2 including digit preference, threshold avoidance, and knowledge of the previous reading, so this trend is probably one to be welcomed. However, validations of these monitors have mainly looked at average performance and with a few exceptions3 have not examined the possibility of patient factors being associated with increased measurement error.

There are two main systems for determining blood pressure automatically. The first is based on microphonic detection of the Korotkoff sounds. This in theory should produce readings very similar to those of an observer using a stethoscope and standard mercury sphygmomanometer. For this to be the case, however, the sensitivity of the microphone must be similar to the human ear, and the monitor must be able to screen out irrelevant and background noise in a way similar to the human brain. In certain conditions, the automatic monitor may have more difficulty than an observer. One example is when altered arterial compliance changes the nature of the Korotkoff sounds.4,5 Reflected pressure waves may produce arterial noises that may interfere with Korotkoff sound interpretation. These conditions are often seen in elderly patients, especially those suffering from isolated systolic hypertension.4,5

The second major method in use is oscillometry. Here, blood pressure is calculated from pressure oscillations detected in the arm cuff as the pressure is decreased. The method determines mean arterial pressure directly from the point of maximum oscillation. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) are not measured directly but are calculated using an algorithm based on a putative relation between the oscillations, mean arterial pressure, SBP, and DBP derived from measurement of blood pressure in many individuals. From direct comparisons of different oscillometric monitors,6 different manufacturers would ap-
peer to use different algorithms. In conditions in which the relation between SBP, DBP, and mean arterial pressure is altered, however, the algorithm may no longer adequately describe this relation. These include isolated systolic hypertension, in which pulse pressure is wider than normal, and heart failure, in which pulse pressure is narrower.

Therefore, we undertook the present study to determine whether the accuracy of oscillometric and microphonic blood pressure measurement as used in two widely evaluated and commonly used ambulatory blood pressure recording systems was affected by patient characteristics such as age, sex, blood pressure level, and pulse pressure.

The two monitors chosen have been among the most extensively validated in the literature. Published validations of the oscillometric monitor, the SpaceLabs 90207 and the similar but larger 90202, have so far been confined to general populations including hypertensive individuals. The same is true of the microphonic monitor, the TM2420 version 7, except for one validation in pregnancy published by our group. The validation were carried out largely according to the recommendations of the Association for the Advancement of Medical Instrumentation (AAMI), with adaptations necessitated by the inclusion of a larger patient number and a greater number of subjects in an older age range so that we could test the monitors’ performances over a wider range of patient characteristics. The British Hypertension Society validation protocol was not used because of doubts about the validity of the statistical methods it recommends to compare consecutive blood pressure readings.

**Methods**

The procedures in this study were approved by the local ethics review committee, and the study was conducted in accordance with institutional guidelines. All subjects gave informed consent for this study.

**Subjects**

One hundred and eighteen subjects were recruited from normal volunteers and the wards and outpatients in two centers: 30 from the National Heart and Lung Institute and 88 from the John Radcliffe Hospital. Details are given in Table 1. The range of subject blood pressures and the distribution of the subjects’ ages conform to those recommended by the AAMI for validations of blood pressure monitors, with extra subjects in the elderly group and the low blood pressure group as well as those demonstrating the isolated systolic blood hypertension pattern.

**Monitors**

Two machines of each of the latest available versions of the SpaceLabs 90207 (SpaceLabs Co, Redmond, NJ) and TM2420 version 7 (A&D Co, Tokyo, Japan) monitors were used. All four monitors had been used for ambulatory blood pressure monitoring for at least 1 month before their use in this validation and had functioned well with no problems. The design of the SpaceLabs 90207, which measures blood pressure oscillometrically, has been described in detail elsewhere. It functions in a similar way to its predecessor, the 90202, but is much quieter and considerably lighter (349 compared with 710 g). The

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean Measurement</th>
<th>Range</th>
<th>Percentage of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>55.8±21.3</td>
<td>17-94</td>
<td>100.0</td>
</tr>
<tr>
<td>Age subgroups</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17-24</td>
<td></td>
<td></td>
<td>8.5</td>
</tr>
<tr>
<td>25-44</td>
<td></td>
<td></td>
<td>22.0</td>
</tr>
<tr>
<td>45-64</td>
<td></td>
<td></td>
<td>29.7</td>
</tr>
<tr>
<td>65+</td>
<td></td>
<td></td>
<td>39.8</td>
</tr>
<tr>
<td>SBP, mm Hg</td>
<td>135.9±26.2</td>
<td>89-211</td>
<td></td>
</tr>
<tr>
<td>DBP, mm Hg</td>
<td>77.2±11.7</td>
<td>44-116</td>
<td></td>
</tr>
</tbody>
</table>

SBP indicates systolic blood pressure; and DBP, diastolic blood pressure. The population comprised 53 men and 65 women; 79 subjects were normotensive and 20 had isolated systolic hypertension, defined by the mean of three clinic readings giving an SBP>160 mm Hg and DBP<90 mm Hg. Seven subjects were hypertensive, defined by the mean of three clinic readings giving an SBP>150 mm Hg and DBP<95 mm Hg. Twelve subjects had heart failure.

**Validation**

Patients sat down for 5 minutes while the procedure was explained to them, their arm circumference was measured, and blood pressure was checked twice in both arms with a standard mercury sphygmomanometer connected by a T-tube to cuffs on both arms. A trained observer read the blood pressure simultaneously on each arm blinded to the other observer. If blood pressure in the two arms differed by more than 5 mm Hg systolic or diastolic on either reading, the subject was excluded at this stage because of a greater than 5 mm Hg interarm difference.

All subjects recruited had arm circumferences in the range of 25 to 32 cm, so a standard-sized cuff and bladder from each monitor were used throughout. These cuffs have bladders measuring 12 x 22 cm, the size currently recommended by the American Heart Association for these arm circumferences. This cuff size results in at least two thirds of the arm being encircled. Subjects sat with the arm supported at heart level.

A random-zero sphygmomanometer (Hawksley, London, England) was used to avoid the problems of bias and last digit preference. The cuff of the SpaceLabs 90207 was fitted on one arm of the subject, and the other arm was fitted with the cuff of the TM2420 connected by a T-tube to the random-zero sphygmomanometer. The arm used for each monitor was chosen using random numbers. Two trained observers in each center (pairs A and B) made three sets of simultaneous blood pressure readings on each subject using a double-headed stethoscope. The observers were blinded to both the readings of the monitors, the other observer's readings, and the true value of the random-zero sphygmomanometer reading until after the set of readings on each patient was complete.
The above arrangement was chosen so that each monitor could be used in its normal running mode. The SpaceLabs 90207 deflates in 8 mm Hg bleed steps, making accurate same-arm, simultaneous comparisons difficult and probably less reliable.11 The TM2420 deflates at 3 mm Hg/s. Early models of the TM2420 (up to and including version 5) worked by automatically releasing pressure in the cuff after DBP was registered by the monitor so that the cuff was inflated for the minimum necessary time. This feature makes it impossible for the observers to measure DBP independently without modifying the mode of action. The latest model of the monitor (version 7) deflates in 3 mm Hg steps down to 35 mm Hg for initial readings, allowing the monitor readings to be checked by an observer without modification.

Statistical Methods

Data were analyzed for differences between each pair of observers and between the two pairs of observers. Comparisons between the monitors and observers were based on the mean of three readings in each subject. The observers’ blood pressure value was the average of the pairs of observers’ readings. Paired and unpaired Student’s t-tests were used to test for group differences as appropriate. Bland-Altman plots25 were used to show the distribution of the differences between the methods at all pressures and the mean and standard deviation of the differences. Because we were comparing the accuracy of two monitors, we used the mean of observers’ readings as the “gold standard” for the x axis in the Bland-Altman plot. Multiple regression was used to look at patient characteristics predicting measurement error. Significances were corrected to account for the multiple comparison effect. Analysis of variance was used to look at the level of error at different levels of blood pressure.

Results

Interobserver Differences

The standard deviation of the difference (SDD) between the observers was 3.2 mm Hg for SBP and 2.9 mm Hg for DBP. The mean differences between observers were 0.57 mm Hg for SBP and 0.13 mm Hg for DBP (Fig 1). Differences were not significant. There was good agreement between the two pairs of observers: mean intercenter differences in observer differences were −1.03 mm Hg for SBP and 0.26 mm Hg for DBP (NS).

Monitor-Observer Differences

The average of three readings for each subject recorded by each monitor was compared with the average of the simultaneous readings by the observers. The SDD for the TM2420 was 5.2 mm Hg for SBP and 6.1 mm Hg for DBP. The mean differences between methods were 5.0 mm Hg for SBP and 3.4 mm Hg for DBP (within the AAMI recommendations for accuracy21) (Fig 2a and 2b). The SDD for the SpaceLabs 90207 was 9.9 mm Hg for SBP (not within the AAMI recommendations for accuracy) and 6.8 mm Hg for DBP (within the AAMI recommendations for accuracy). The mean differences between observers and the SpaceLabs were −9.9±10.6 mm Hg for the highest tertile of SBP (>160 mm Hg), −5.2±10.6 mm Hg for the highest tertile of DBP (>90 mm Hg), +7.5±6.7 mm Hg for the lowest tertile (<120 mm Hg), +3.1±9.5 mm Hg for the middle tertile, and +4.6±6.3 mm Hg for the lowest tertile (<45 mm Hg). The errors were also correlated with level of SBP, pulse pressure, and the age of the subject. For DBP, 34.7% of the error was accounted for. As a comparison, the lack of correlation of error with level of pulse pressure or pulse pressure is shown for the microphonic monitor in Fig 3b and 3d.

Relation Between Measurement Error and Patient Characteristics

Patient characteristics that predicted measurement error were defined by multiple regression (only significant if P<.01). All individual readings were used for this analysis. The results are shown in Table 2. For oscillometry with the SpaceLabs, the measurement error of SBP was highly correlated with SBP level (Fig 3a). With the use of repeated-measures analysis of variance, the errors were +7.5±6.7 mm Hg for the lowest tertile (<120 mm Hg), +3.1±9.5 mm Hg for the middle tertile, and −5.2±10.6 mm Hg for the highest tertile of SBP (>160 mm Hg, P<.0001). The error was also correlated with pulse pressure and the age of the subject. For DBP, the error was significantly correlated with pulse pressure (Fig 3c). The errors were +0.8±6.2 mm Hg for the lowest tertile (<45 mm Hg), +4.6±6.3 mm Hg for the middle tertile (45 to 70 mm Hg), and +7.2±6.9 mm Hg for the highest tertile of pulse pressure (>70 mm Hg, P<.0001). The error was also correlated with level of DBP and with the sex of the subject. The patient characteristics tested accounted for 36.6% of the error in SBP. For DBP, 34.7% of the error was accounted for. As a comparison, the lack of correlation of error with level of blood pressure or pulse pressure is shown for the microphonic monitor in Fig 3b and 3d.
For the microphonic method, only age significantly correlated with DBP error: old (>65 years), 5.2±4.9 mm Hg; young (<35 years), -1.75±8.4 mm Hg; P<.0001 (Fig 4a). Only 1.2% of the error for SBP and 8.9% of the error for DBP were accounted for by patient characteristics. As a comparison, the lack of

<table>
<thead>
<tr>
<th>Difference in SBP</th>
<th>TM2420 Version 7</th>
<th>Significance</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of SBP</td>
<td>0.488</td>
<td>NS</td>
<td>0.056</td>
<td>NS</td>
</tr>
<tr>
<td>Level of DBP</td>
<td>0.098</td>
<td>NS</td>
<td>1.274</td>
<td>NS</td>
</tr>
<tr>
<td>Age</td>
<td>4.098</td>
<td>NS</td>
<td>2.199</td>
<td>NS</td>
</tr>
<tr>
<td>Pulse pressure</td>
<td>0.874</td>
<td>NS</td>
<td>2.327</td>
<td>NS</td>
</tr>
<tr>
<td>R²</td>
<td>0.488</td>
<td>NS</td>
<td>0.056</td>
<td>NS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Difference in DBP</th>
<th>TM2420 Version 7</th>
<th>Significance</th>
<th>F</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of SBP</td>
<td>0.200</td>
<td>NS</td>
<td>2.199</td>
<td>NS</td>
</tr>
<tr>
<td>Level of DBP</td>
<td>32.83</td>
<td>&lt;.0001</td>
<td>1.703</td>
<td>NS</td>
</tr>
<tr>
<td>Pulse pressure</td>
<td>0.157</td>
<td>NS</td>
<td>0.687</td>
<td>NS</td>
</tr>
<tr>
<td>R²</td>
<td>0.200</td>
<td>NS</td>
<td>0.000</td>
<td>NS</td>
</tr>
</tbody>
</table>

SBP indicates systolic blood pressure; and DBP, diastolic blood pressure. R² is percentage of variance of measurement error accounted for by patient characteristics. Levels of significance were adjusted for effects of multiple comparison.
significant correlation of age and DBP error for the oscillometric monitor is shown in Fig 4b.

**Discussion**

The overall results of the validation showed that the TM2420 version 7 passed the AAMI criteria and the SpaceLabs 90207 just failed on the SDD of SBP. One possible factor in this difference is that the microphonic monitor was compared on the same arm, whereas the oscillometric monitor was on the opposite arm. The reason for this is that the mode of deflation makes strictly simultaneous measurement of blood pressure impossible with the SpaceLabs monitor. A number of strategies have been adopted by different authors to cope with this. Clinically, the monitor is usually set to deflate in 8 mm Hg steps. Several validations have been done with the rate slowed down to bleed steps of 4 mm Hg, allowing same-arm comparisons. Others have made the assumption that the systolic reading was the median of the 8 mm Hg step in which the Korotkoff sound first appeared, ie, 4 mm Hg higher than the pressure at which the first sound was heard. The same reasoning was used for diastolic, ie, 4 mm Hg lower than the pressure at which the last sound was heard. One author has used sequential same-arm measurements.

All strategies have problems but are useful as long as one is aware of any likely effect on the results. Our method would be unlikely to substantially increase the variability of the monitor-observer differences because we had prescreened our subjects to exclude any significant interarm differences. In any case, the oscillometric method does not use the same heartbeats to estimate SBP and DBP as do the microphonic or auscultatory techniques; hence, strictly same-beat simultaneous comparisons are impossible for any oscillometric monitor. As the minute-to-minute differences in SBP and DBP are frequently more than 5 mm Hg, whereas the simultaneous opposite-arm blood pressure differences in our subjects were all less than 5 mm Hg, the use of a simultaneous opposite-arm comparison is seen to be preferable to a sequential same-arm comparison in which minute-to-minute changes in blood pressure are also incorporated, as these latter differences are larger than the interarm differences in our study. Some authors advocate a secondary correction of sequential readings to make them more similar to simultaneous readings, but this correction can artificially manufacture 33% of comparisons as having zero differences between methods, because it states that whenever the middle of three readings lies between the other two, the difference will be arbitrarily considered zero. This has been criticized, and we would not recommend its use.

The impossibility of exactly contemporaneous same-beat comparisons between oscillometric and auscultatory methods may explain a small increase in the variability of the error with the SpaceLabs monitor that resulted in its failure to pass the AAMI standards for the SDD of SBP. It would not explain, however, the systematic differences in errors in SBP and DBP seen with the SpaceLabs monitor that were related to absolute level of SBP, pulse pressure, or both.

The systematic difference seen is probably related to the way in which the oscillometric method works. Mean arterial pressure can be determined with some accuracy from the point of maximum oscillations detected in the cuff. SBP and DBP are then calculated from this using.
The errors for the microphonic monitor were more random and less clearly related to patient factors. It may be that there are other factors untested in this study that predict error with this monitor, but this remains speculative. No reports we are aware of before this study have detected such factors, in contrast to the reports with the oscillometric SpaceLabs monitor. We did find a significant correlation (albeit weaker) between age and DBP in both the TM2420 monitor and the SpaceLabs 90207, which is used to detect non-Korotkoff sound noise in the oscillometric SpaceLabs monitor. We did find a significant correlation (albeit weaker) between age and DBP in both the TM2420 monitor and the SpaceLabs 90207.

Cates et al. found that only 12% of the SBP error variance ($R^2$) and 10% of the DBP error variance were explained by the factors they studied (which did not include the actual level of SBP, DBP, or pulse pressure), whereas we found that between 34% and 36% of the variance could be explained when these factors were included. Only a relatively small amount of the variance of the microphonic method could be explained by patient characteristics. The greater variation in accuracy we found with the SpaceLabs monitor and the greater proportion of the error related to patient characteristics we have found compared with the previous report by Cates and colleagues probably reflect the fact that we set out to test the monitors in extreme cases of high and low blood pressure, high and low pulse pressure, and more elderly subjects. We also had a higher proportion of subjects with isolated systolic hypertension. The worse performance of the oscillometric monitor is likely to be even more marked in nonresting conditions, during which this type of monitor has been shown to be even less reliable.11

The systematic and patient characteristic–dependent errors seen with the SpaceLabs monitor are especially important for epidemiologic studies and for studies involving particular patient groups with high or low blood pressures and wide pulse pressures. In the lowest and highest tertiles of SBP (89 to 120 mm Hg and 160 to 211 mm Hg, respectively), the differences between the observers and the SpaceLabs were +7.5±6.7 and −5.2±10.6 mm Hg, respectively. For DBP, the error was significantly correlated with pulse pressure. The errors were +0.8±6.2 mm Hg for the lowest tertile (pulse pressures of <45 mm Hg) and +7.2±6.8 mm Hg for the highest tertile of pulse pressure (>70 mm Hg). These are not small differences, and although perhaps less important for general clinical practice with many hypertensive individuals falling in the middle tertiles, they would make substantial differences to some studies. Studies of the distribution of blood pressure in the general population, for example, would show a reduced spread in the normal distribution of SBP, as lower pressures would be underestimated and high pressures overestimated. The effects on DBP would be more complex. Other important effects are in studies of special groups, such as isolated systolic hypertensive patients, in which high SBPs are combined with wide pulse pressures, both exaggerating the errors in SBP estimation with the SpaceLabs monitor. Pregnant women, in whom blood pressures tend to be low, present another potentially problematic group. There is clearly a need for specific validations in these special groups.28 The problems of population studies cannot be remedied in this way, as there will inevitably be subjects with low and high pressures and varying pulse pressures. In population studies, mean arterial pressure, which the oscillometric method may record more accurately, may be a more useful measure to report.

It should be emphasized that as different oscillometric monitors have different algorithms, these results cannot be generalized except in as much as validations are required in any special patient group before a particular monitor is used. There may need to be different algorithms built into monitors for different blood pressure levels and pulse pressures. It is probably also undesirable to mix both methods in one recording, particularly when the nature of the method used to make a particular reading is not made clear.
did meet the AAMI standards in the elderly, although we found it to be slightly less accurate than in a younger population. The errors in the TM2420 monitors are in part related to static accuracy. The two monitors used passed the AAMI standards for static accuracy, but both read approximately 3 mm Hg high across the pressure range. This was probably the reason why mean differences for both SBP and DBP measured with the monitors deviated from the observers by 3 to 5 mm Hg. Although this level of difference is acceptable clinically, it would be preferable for it to be zero. Other TM2420 monitors we have tested had no or very little static error, and we found no drift with time in the static accuracy with this monitor type.

We used the random-zero sphygmomanometer in this validation instead of the standard mercury column recommended by the AAMI. Readings with the random-zero sphygmomanometer are on average 1 to 2 mm Hg lower for both SBP and DBP than readings with the conventional sphygmomanometer. This difference is consistent and is apparently due to the greater height of the mercury column in the random-zero sphygmomanometer. The potential advantages of using the random-zero sphygmomanometer are less threshold avoidance, reduced digit preference, and absence of knowledge of the previous reading. Although this underreading is undesirable, the difference is small, and we felt that the potentially much more serious and unpredictable problems of observer bias and digit preference (seen even in the article describing the British Hypertension Society Validation Protocol itself) outweighed this known and predictable problem. The only publication reporting greater differences between a standard mercury column and a random-zero sphygmomanometer is unusual in that the first part of the publication reports a well-controlled trial with two observers using a standard column and one using a random-zero sphygmomanometer and shows very good agreement with a very small, systematic difference totally in agreement with previous work; the second part with only one observer using the standard mercury column showed much greater and very erratic differences (up to 30 mm Hg), which could not be explained. Without two observers acting as a control, this could well be due to an inaccurate observer.

We have shown that oscillometry as represented by the SpaceLabs 90207 differs systematically from the trained observers’ readings. It might be argued that the readings from the oscillometric monitor are actually more representative of “true” arterial pressure. Studies comparing both oscillometry and microphonic methods with intra-arterial readings do not bear this out, as both indirect methods show similar levels of divergence from the direct. Our knowledge of the risk of differing levels of blood pressure is also based on the Korotkoff method. The limitations of this method of blood pressure measurement, to which the microphonnic method is closely related, are well known. Those of the oscillometric method will also become known as the technique is used more widely, but for the moment, until studies of mortality and morbidity show that one method is superior to the other in predicting risk, we feel that the method we know most about, the Korotkoff method, should remain the clinical gold standard against which other methods must be measured.

References
