

The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems

E. O'BRIEN*, J. PETRIE†*, W. LITTLER‡*,
M. DE SWIET§*, P. L. PADFIELD¶*, K. O'MALLEY,
M. JAMIESON†, D. ALTMAN††, M. BLAND‡‡
and N. ATKINS

Blood Pressure Unit, Beaumont Hospital, Dublin, Ireland

†Clinical Pharmacology Unit, Department of Medicine and Therapeutics,
Aberdeen University, Polwarth Building, Forresterhill, Edinburgh, UK

‡Department of Cardiovascular Medicine, East Birmingham Hospital,
University of Birmingham, Birmingham, UK

§National Heart and Lung Institute, London, UK

¶Department of Medicine, Western General Hospital, Edinburgh, UK

††Imperial Cancer Research Fund Laboratories, London, UK

‡‡Department of Clinical Epidemiology and Social Medicine, St. George's
Hospital, London, UK

With the increasing manufacture of expensive systems for the measurement of ambulatory blood pressure there is a need for potential purchasers to be able to satisfy themselves that the systems have been evaluated according to agreed criteria. The British Hypertension Society has, therefore, drawn up a protocol of requirements for the evaluation of these devices. This protocol incorporates many features of the *American National Standard for Non-Automated Sphygmomanometers* but includes many additional features, such as strict criteria for observer training, interdevice variability testing before and after a month of ambulatory use, and a new system of analysis which permits the test system to be graded. It is recommended that manufacturers of ambulatory blood pressure measuring devices should obtain an unbiased evaluation according to a recognized standard before a device is marketed.

*Keywords: Protocol; device validation; ambulatory blood pressure;
British Hypertension Society.*

Introduction

Ambulatory blood pressure measurement is rapidly gaining acceptance as a useful procedure in the clinical management of hypertension [1, 2], in the assessment of antihypertensive drugs [3] and as a means of predicting outcome in hypertension

*Members of the British Hypertension Society Working Party on Blood Pressure Measurement.

Correspondence to: Dr E. O'Brien, The Blood Pressure Unit, Beaumont Hospital, Dublin 9, Ireland.

Reprinted with permission of the British Hypertension Society and Current Science Ltd. from the *Journal of Hypertension* 1990, 8, 607-619.

[4]. The procedure also provides data on the physiology of blood pressure behaviour [5]. Ambulatory blood pressure provides an assessment of blood pressure behaviour over time in the patient's normal environment and is likely to result in reappraisal of the clinical management of hypertension, which is presently based on conventional measurement techniques [6].

One consequence of the increased interest in ambulatory measurement has been the creation of a large market for ambulatory blood pressure measuring devices. In recent years the number of devices available commercially has risen rapidly, with more than ten now available on the international market and many others in the planning phase [7]. Ambulatory blood pressure measuring systems are expensive, often costing as much as £4000 (\$US 6800) for one recorder. Decoding facilities may cost as much again. Operational and maintenance costs may also be considerable. At present there is no obligation on manufacturers to comply with the few recommended standards that are available. There is no standard for automated blood pressure measuring devices in the UK. In the USA, the Association for the Advancement of Medical Instrumentation (AAMI) has produced a detailed standard for semi-automated and automated devices [8].

The present situation, therefore, is that manufacturers may market expensive blood pressure measuring devices without being obliged to provide evidence of their accuracy. Many validation studies of ambulatory blood pressure measuring devices have been performed with a variety of protocols and differing criteria for assessment, making comparison of the assessment difficult and comparison of one device with another almost impossible.

Because validation studies are time-consuming to perform, the time-lag between manufacture and publication of an independent evaluation in a reputable journal is often so long that manufacturers may be ready to market a modification of the original device and the outcome of any evaluation is thereby rendered obsolete and of little academic interest.

The British Hypertension Society (BHS) is of the opinion that evaluation of ambulatory blood pressure measuring systems must be standardized for the following reasons. (1) Continued uncontrolled marketing will inevitably result in the manufacture and sale of inaccurate devices; this has clear implications for clinical practice, the most important of which is inappropriate diagnostic and management decisions; and (2) without a standardized approach to evaluation, comparison of results between laboratories is not possible and work may have to be repeated with the consequent waste of scarce resources.

The BHS Working Party on Blood Pressure Measurement, having reviewed the possible approaches to the problem, concluded that while the AAMI standard [8] is the most comprehensive recommendation on validation available, it has a number of deficiencies; it does not cover all aspects of evaluation, e.g. interdevice variability, ambulatory assessment and patient acceptability are not included and there are deficiencies in the statistical methodology; it is obtainable only from the AAMI offices on payment of a fee, and is not, therefore, as accessible as a journal publication; it contains detailed recommendations for manufacturers of ambulatory devices which, though necessary in a standard, are not a requirement of an evaluation protocol. The Working Party decided, therefore, to prepare a protocol that would serve as a standardized procedure for the evaluation of ambulatory blood pressure measuring devices and to make recommendations for the adoption of this standard procedure. Though the Working Party's brief was to prepare a

protocol for ambulatory devices the principles of the procedure that follow can be applied to any automated or semi-automated blood pressure measuring device [9].

General considerations

The basis of this device evaluation is the comparison between blood pressure measured by the device being tested and simultaneous measurements made by an established technique, the 'gold standard'. The test methods must allow for variation in the design and technology of ambulatory measuring devices.

Two features of the programme described in this paper need elaboration. First, before embarking on what is a complex and labour-intensive protocol, the Working Party placed considerable emphasis both on observer training and on the capability of a number of devices of the model being tested to give consistent measurements. The AAMI standard recommends that two observers should measure blood pressure independently against the test device in the main validation phase [8]. The main advantage of using two trained observers is that the conclusion of the validation test is strengthened by having two independent standard measurements against which to judge the test device [9]. If the observers have already been shown to be in close agreement, it is only necessary to have one observer take the measurement. However, to minimize bias, it is advisable that separate observers each measure blood pressure in approximately half the subjects. This modification to the AAMI standard has the advantage of substantially reducing the cost of performing the main validation test.

In this protocol observer agreement is strictly assessed before the evaluation; if an observer is inaccurate re-training is easily accomplished at this stage. However, with the AAMI standard, observer agreement is assessed at the end of the validation, and in the event of the observers not being in agreement the entire procedure has to be repeated. It is preferable that observer agreement be assessed before the study begins so that this eventuality is avoided. If the test standard, namely the mercury sphygmomanometer and the observer, cannot be brought to the highest possible level of accuracy before the main validation procedure, further testing is pointless. Likewise, interdevice variability should be assessed before the validation test begins, since substantial differences between devices of the same model will render device validation impracticable.

Second, in drafting this protocol we attempted to determine the minimal criteria for a statistically valid assessment while also being alert to the demands that the validation tests impose on an assessment laboratory. Whereas it might be desirable to perform the main comparative validation when the device is new and repeat the test after a period of time in use, this would effectively nearly double the time and expense of the study. We compromised, therefore, by postponing the main validation test until the device has been in use for a period of time, and we arbitrarily chose a minimum period of 1 month. We believe this to be justified on the basis that the accuracy of a measuring device after use is of more relevance than immediately after purchase, before it has been subjected to the wear and tear stresses of daily use that might alter accuracy.

Unlike the AAMI standard [8], direct intra-arterial measurement is not included in the present protocol for the following reasons. First, systolic and diastolic blood pressure values obtained by the direct technique are different from measurements obtained by indirect methods [10, 11]. Second, clinical practice uses data obtained

by the indirect rather than the direct technique. Third; there are ethical considerations in the use of intra-arterial measurement which preclude its use for the evaluation of blood pressure measuring devices in most laboratories.

In an effort to minimize unnecessary testing, the programme has been designed so that the device passes through different phases of evaluation, entry to each test phase being dependent on the successful completion of the preceding phase.

A standard mercury sphygmomanometer, the components of which were carefully checked before the study, is used as a reference standard for all tests rather than a random zero sphygmomanometer [12] because there is evidence that the random zero sphygmomanometer systematically underestimates diastolic pressure [13–16].

The quality of the stethoscope is also crucial to the evaluation procedure. Stethoscopes with badly fitting ear-pieces and poor quality diaphragms preclude precise auscultation of Korotkoff sounds. The Littman stethoscope (3M Company, Minnesota, USA) or its equivalent is recommended.

The general principles of auscultatory measurement have been outlined in previous publications of the British Hypertension Society [17, 18].

In the protocol we use the term model to denote a particular brand of sphygmomanometer and the term device to denote individual sphygmomanometers.

Methods

The evaluation programme consists of six phases (figure 1): I, Observer training and assessment; II, Before-use interdevice variability assessment; III, In-use assessment; IV, After-use interdevice variability assessment; V, Device validation; and VI, Preparation of report.

Phase I: Observer training and assessment

Observer training

Two trained observers are required for the evaluation of a device. Each training day consists of two phases.

Film training. The observers, each of whom should understand the principles of blood pressure measurement, e.g. trained nurses, are retrained in blood pressure measurement using the British Hypertension Society video film *Blood Pressure Measurement* [19]. It is recommended that audiograms are obtained from the observers to detect any hearing deficit. The first part of the film training demonstrates the technique of blood pressure measurement and is followed by an assessment period in which the trainees can test themselves against a standard mercury sphygmomanometer as the mercury column falls against a background of recorded Korotkoff sounds. Observers should not move on to the next stage until they have satisfied this assessment. The video film lasts 30 min.

Expert training. In this phase, an expert in blood pressure measurement takes the trainee observers through the different stages of measurement as recommended by the British Hypertension Society [17]. Difficult aspects of interpretation, such as the auscultatory gap and bias, should be discussed and illustrated by example using a multi-aural stethoscope.

BHS protocol for blood pressure measuring devices

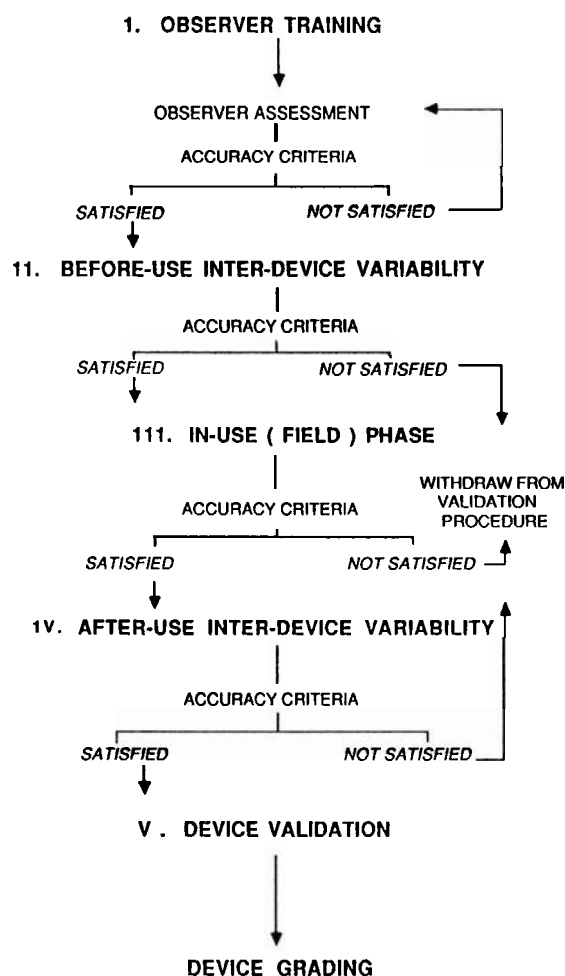


Figure 1. Validation procedure.

Observer assessment

Two (or more) observers are tested for accuracy against each other and against an expert observer. The expert observer should have had extensive experience in blood pressure measurement and he/she should have correctly interpreted 95% of a test sequence, e.g. the sequence in the British Hypertension Society video [19], before each training assessment. If the expert observer is inaccurate this will become apparent in the analysis.

The test procedure takes the following form (figure 2):

1. Two observers are seated at a bench fitted with temporary partitions so that each observer is isolated in a booth in which the only objects are a mercury column, a stethoscope, a pencil and a blank sheet of paper. When more than one observer is being trained and assessed it becomes difficult to prevent an observer who is unsure of a reading from gaining sight of a neighbouring observer's reading, and it is necessary, therefore, to separate the observers by a series of partitions.

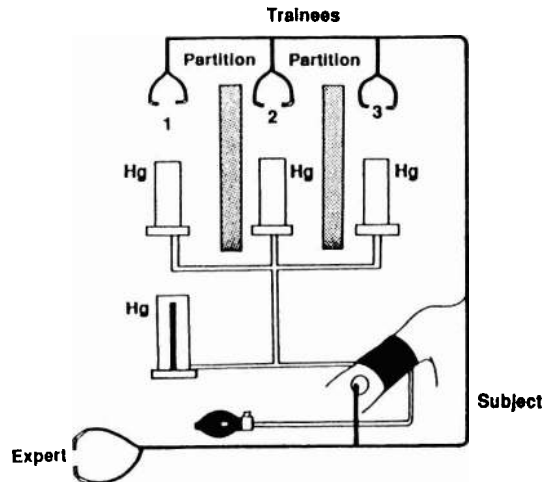


Figure 2. Testing observer agreement in three trainees.

2. In a similar adjoining booth an expert observer deflates a bladder attached to the arm of a subject.
3. The subjects' blood pressures should be in the range 110/60 to 240/120.
4. The bladder is connected to each of the columns of mercury in the observer booths so that all columns of mercury fall simultaneously for each of the blinded observers and the expert, all of whom write down their measurements.
5. Ten measurements are made by each observer on each of five subjects giving a total of 50 measurements for each observer.

Accuracy criteria for the procedure are:

1. 90% of systolic and diastolic differences between trainees and the expert may not differ by more than 5 mmHg and 98% by not more than 10 mmHg;
2. 85% of systolic and diastolic differences between each trainee may not differ by more than 5 mmHg and 95% by not more than 10 mmHg;
3. failure to achieve this degree of accuracy requires a repeat training and assessment session for the failed observer(s).

Familiarization session

As devices for ambulatory blood pressure measurement are complex, familiarization is important. The observers should be instructed in the use of the devices to be tested, preferably by a representative of the manufacturer. Practice measurements should be made on a number of subjects.

Phase II: Before-use interdevice variability assessment

If only one device is tested for validation, it is possible, in the event of the assessment proving unfavourable to the test device, that the device is unrepresentative of the product and the inaccuracy might have been due to poor calibration or to some other fault that might occur only occasionally [20]. It is also possible that

the first device to be tested might be accurate but unrepresentative. Because of these potential differences between machines we suggest that at least three devices should be tested for interdevice variability before proceeding to validation, and if differences emerge between devices further testing should not be conducted until the manufacturer has identified the source of error and provided three devices which are in agreement. The recommendation to select three devices is based on economic and feasibility considerations.

Semi-automated devices for blood pressure measurement should have a facility permitting connection with a mercury sphygmomanometer to check device calibration, and it is likely that future models of devices which presently do not readily lend themselves to calibration will provide this facility. Although calibration details vary from one system to another, the test is usually performed by connecting the device to a mercury sphygmomanometer with a Y-connector. The automatic pressure measuring system and the blood pressure detection mechanism (i.e. microphone, oscillometry, etc.) are disabled so that the device acts simply as a manometer. Pressures within the system are then compared throughout the pressure range on the mercury column.

Test requirements

Test requirements are:

1. 95% of measurements should be within 3 mmHg or 2%, whichever is the greater.
2. If the device fails to meet the manufacturer's calibration criteria it is not tested further.

Phase III: In-use (field) assessment

The three devices used for the interdevice assessment are next used to test the accuracy and performance of the device during and after the use for which it was designed, i.e. 24-h ambulatory monitoring. The purpose of this phase is to subject the ambulatory blood pressure measuring system to a period of fairly strenuous use before performing the main validation test. Each of the three devices are placed on six subjects with a wide range of pressure on 8 days over a 4-week period. At the end of this period the performance of the device is assessed.

Test requirements

Test requirements are:

1. Each of the three devices is to be worn for 24 h by eight subjects with a wide range of pressures using a total of 24 subjects.
2. 24-h ambulatory measurements are taken at 15 min intervals from 0900–2200 h (56 measurements), and at 30 min intervals from 2230–0830 h (19 measurements), giving a total of 75 readings for the 24 h;
3. 600 recordings per device are taken;
4. 1800 recordings per model are taken.

Table 1. Example of in-use assessment.

Subject	I	V	V/I%	R	A	V (D)	V (N)	D:N ratio
Ideal	75	75	100	0	0	60	15	3:1
1	94	65	69	3	26	52	13	4:1
2	97	68	70	0	29	55	13	4.2:1
3	89	75	84	1	13	61	14	4.4:1
4	84	75	89	0	8	61	15	4.1:1
5	94	75	79	0	10	61	15	4.1:1
6	98	75	77	1	12	70	15	4.7:1
7	88	75	85	1	13	60	15	4:1
8	93	75	82	1	15	62	15	4.1:1
9	79	70	89	0	9	55	15	3.7:1
10	84	70	83	1	12	55	15	3.7:1
11	100	69	69	2	29	54	15	3.6:1
12	88	75	85	0	13	60	15	4:1
13	84	71	85	0	13	59	12	4.9:1
14	82	75	91	0	6	61	15	4.1:1
15	85	73	86	0	12	58	15	3.9:1
16	89	75	84	0	13	63	13	4.1:1
17	82	64	78	0	18	51	13	3.9:1
18	84	75	89	1	8	60	15	4:1
19	93	69	74	0	24	55	14	3.9:1
20	90	69	77	1	20	54	15	3.6:1
21	81	75	93	0	5	61	15	4.1:1
22	81	73	95	0	4	60	15	4.1:1
23	79	75	95	0	4	60	15	4:1
24	95	73	77	3	19	59	14	4.2:1
Total	2104	1734	83	14	339	1405	346	4.1:1
Ideal	1800	1800	100	0	0	1440	360	3:1

I, number of inflations; V, number of valid readings; V/I%, percentage of valid readings per inflation; R, rejected; A, aborted; D, day; N, night.

Performance criteria

Performance criteria are:

1. Most ambulatory blood pressure measuring systems have programmed editing criteria and these are left in operation for this phase. If the instructions allow the operator to modify the editing programme, the programme recommended by the manufacturer is chosen.
2. The measurements obtained over each 24-h period are classified as follows (Table 1):
 - (a) *Inflations*. The total number of inflations made by the device.
 - (b) *Valid readings*. Those readings accepted by the system as genuine blood pressure measurements.
 - (c) *Rejected readings*. Those blood pressure readings that are rejected either by the recorder or the decoder as not being genuine blood pressure measurements.
 - (d) *Aborted readings*. Those occasions when an inflation fails to produce a reading of any kind.

(e) *Day/night readings.* Ratio of valid day/night readings.

3. If a device fails to record any pressure in the 24-h period and/or the subject may not have complied with instructions on a particular recording day, that 24-h recording is repeated.
4. At least 70% of readings should be valid for 22 of 24 recording days.
5. Failure to achieve this level of performance means that no further testing is carried out.

The purpose of this phase is to ensure that a period of time in use does not make the system inaccurate; it is not intended primarily as an assessment phase though the information on performance may be useful. However, there is little point in proceeding to the main validation test if the device performs extremely badly as an ambulatory recording system.

Patient/subject acceptability

In this assessment each subject is asked to make comments on the following aspects of device performance which are printed as headings, allowing five or six lines for comment on each topic. While this is not an elimination stage, the information may be helpful later in making an overall assessment of performance, and the comments may indicate areas of improvement for the manufacturer. The headings for comment are:

List any problems	Anxiety factors
General impression	Difficulty in using
Comfort/discomfort	Clarity of user instructions
Interference with sleep	Comparisons with other devices used by subject
Problems with noise	Suggestions for improving device

Subjects should be asked to keep a diary card in which they are asked to make a particular note of activity at the time of each ambulatory measurement.

IV: After-use interdevice variability assessment

At the end of the month of ambulatory assessment the three devices are retested for interdevice variability, as in the before-use interdevice variability test, to determine whether there has been any change in interdevice agreement after use.

If all three devices give measurements that are in agreement at the time of purchase as well as after a period in use, it suggests, at least, that the model is being manufactured to perform consistently. If, however, all three devices give discordant measurements, further assessment is pointless and the model cannot be recommended. If one device is discordant but the remaining two are consistent, further evaluation is reasonable on the basis that one inaccurate device might have been included by chance. The occurrence may indicate, however, that overall production of that model is not satisfactory and the finding should be included in the final report. If only one device is discordant it is removed while the other two are retained for the validation test. If all three devices are discordant no further testing is carried out.

Phase V: Device validation

If there has been no alteration in interdevice variability after the month of use, one device is arbitrarily selected from the three devices for the main validation test. Because blood pressure measuring devices are of varying designs it is necessary to allow some flexibility in the validation methodology. The following tests allow validation of devices with controllable and rapid deflation rates.

Subject selection

In the selection of subjects it is not sufficient to merely specify that subjects shall have blood pressures within a specified range (as required by the AAMI standard [8]) because there may be a tendency (arising out of convenience) to recruit more subjects in the lower pressure range than those with higher pressures.

Tests should be performed with the adult bladder supplied with the device; the dimensions should be noted and a similar bladder should be used for the comparative test. The circumference of the arms should be measured to ensure that the bladder being used is adequate for the subject, i.e. the bladder should be of sufficient length to encircle 80% of the arm circumference; only the cuff and bladder should be changed for obese arms, since it is important to ensure that the same microphone is used throughout the validation test.

Subject selection is also dependent on the circumstances under which the device will be used. If the device is intended for a special patient population, such as pregnant or paediatric patients, it must be validated in these groups; the recommendations in this paper are for adult patients. Likewise, patients with arrhythmias, such as atrial fibrillation, should not be included; if validation in these circumstances is required subject selection must be directed accordingly. Subjects in whom Korotkoff sounds persist to near zero should be excluded from the study.

Criteria for selection are as follows:

1. 85 subjects.
2. age range 15–80 years;
3. at least 15% of blood pressures in each of the following categories of systolic pressure (mmHg): 100–140, 140–180, 180–220, 220–240;
4. at least 20% of blood pressures in each of the following categories of diastolic pressure (mmHg): 60–80, 80–100, 100–120.

Devices with controllable deflation rates

This test is based on simultaneous same-arm measurement between the test device and a standard mercury sphygmomanometer. It is the ideal test, and blood pressure measuring devices should incorporate the facility for this form of validation. The test should be performed in a warm, quiet room. Procedure is as follows:

1. Connect the test device via a T-tube to a standard mercury manometer and pump bulb (figure 3).
2. Place the cuff of the test device on the subject's arm according to the manufacturer's instructions.
3. For auscultatory devices, place the microphone over the brachial artery.

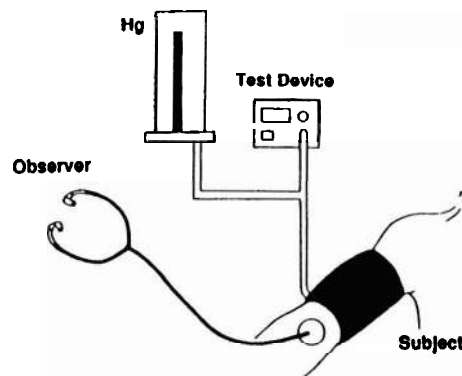


Figure 3. Design for simultaneous measurement between mercury standard and test devices.

4. For devices with ECG gating, place the electrodes according to instructions.
5. The inflation mechanism of the device is activated and one observer records pressure simultaneously with a stethoscope and mercury sphygmomanometer in 43 subjects and a second trained observer does likewise in the remaining 42 subjects (subjects should be distributed between the observers to ensure that each observer has a representative number of subjects with high and low pressures).
6. An independent observer charts the pressure reading of the device so that neither observer is aware of the other's reading.
7. Three measurements are made by each observer and tabulated for analysis of systolic and diastolic values; these measurements are not averaged as recommended in the AAMI standard because by so doing, the variability of error within a particular subject may be eliminated, thus incorrectly indicating greater accuracy.
8. A total of 255 measurements between observers and the test device are analysed.
9. Documentation must be provided for data omitted for legitimate technical reasons; once a subject is included, before the data-gathering phase, the data for that subject should not be excluded from the study if blood pressure values are obtainable; if blood pressure measurements from either the reference method or the ambulatory device are unavailable, data entry for that subject may be excluded with an accompanying explanation, and additional subjects must then be entered into the study to ensure a sample size of 85.

Accuracy criteria. The percentage of test device measurements differing from the mercury standard by 5, 10 and 15 mmHg or less are calculated (Table 2) and plotted (Appendix A; Figs 4 and 5); the device is then graded as A, B, C or D according to the criteria in Table 3. To reach a particular grade all three cumulative percentages should exceed the tabulated values. Though the mean, standard deviation of measurements and the mean and standard deviation of the differences are not used for grading purposes, they should be provided, as in Table 2, for information.

Table 2. Example of device validation for two paired measurements of systolic and diastolic pressures in 85 subjects.

Observers device	n	Mean ± s.d.	Difference ± s.d.	Readings(%)			Grade
				≤5	≤10	≤15	
Observer 1	127	152 ± 27	-1 ± 6	67	88	98	B
Device	127	151 ± 25					
Observer 1	127	91 ± 15	-2 ± 6	67	91	97	B
Device	127	88 ± 15					
Observer 2	128	151 ± 22	-2 ± 6	66	87	97	B
Device	128	148 ± 22					
Observer 2	128	90 ± 12	-1 ± 5	74	92	98	B
Device	128	89 ± 12					

Observers 1 and 2 measured blood pressure in 42 and 43 subjects, respectively.

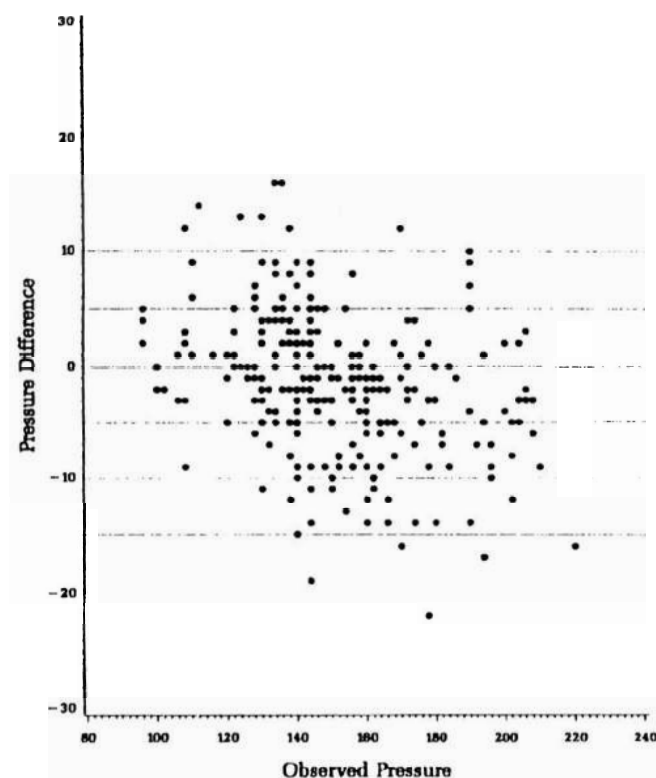


Figure 4. Plot of pressure difference and mean pressure for test device and observers in 85 subjects for systolic pressure ($n = 255$). Reference lines for 0, ± 5 , ± 10 and ± 15 mmHg differences are shown.

BHS protocol for blood pressure measuring devices

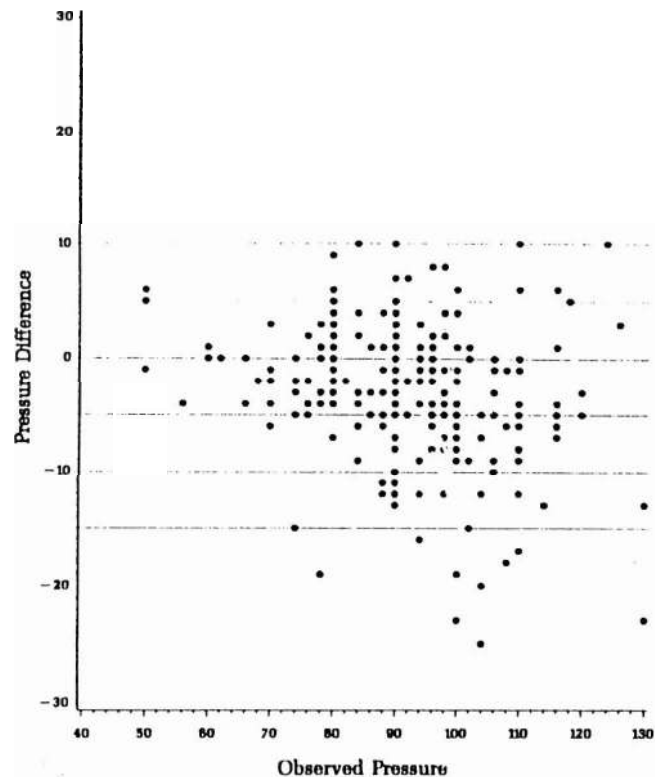


Figure 5. Plot of pressure difference and mean pressure for test device and observers in 85 subjects for diastolic pressure ($n = 255$). Reference lines for 0, ± 5 , ± 10 and ± 15 mmHg differences are shown.

Table 3. Grading criteria based on cumulative percentage of readings.

Grade	Difference between standard and test device (mmHg)		
	≤ 5	≤ 10	≤ 15
	80	90	95
	65	85	95
	45	75	90
	Worse than C	Worse than C	Worse than C

Devices with rapid deflation rates

The above test cannot be performed with devices that deflate at rates greater than 5 mmHg/s because faster rates do not allow a sufficiently accurate measurement by an auscultating observer, leading to an inaccurate comparison between the test and the reference device [21]. At fast deflation rates an auscultating observer will tend to underestimate systolic and overestimate diastolic pressure by recording the first definite pressure phase at which Korotkoff sounds are audible as the systolic value and the last definite phase of audible sounds as the diastolic. The device may have a facility for slowing the rate of deflation so that the above test can be

performed but it is important to determine that this modification of the usual operational mode does not alter the accuracy of measurement. Other factors that may preclude simultaneous same-arm testing are confusion between noise from the device and the Korotkoff sounds, failure of the inflating mechanism to reach the required pressure, and uneven deflation, making accurate auscultation impossible.

Alternatives to simultaneous measurements in the same arm are either simultaneous measurements in the opposite arms or sequential measurements in the same arm. We favour the latter approach because if simultaneous measurements are to be performed in opposite arms it is necessary first to determine that the interarm differences are small enough to prevent the introduction of error; this would require simultaneous measurements in both arms in all 85 subjects, a major undertaking in itself. Furthermore, sequential same-arm measurements are closer to simultaneous same-arm measurements than opposite arm measurements (Appendix A).

Sequential same-arm measurements between the test device and a standard mercury sphygmomanometer are carried out as follows:

1. A trained observer measures blood pressure with a stethoscope and a mercury sphygmomanometer deflating at 2 mmHg/s.
2. One minute later, measurement is made in the same arm with the test device, which is 'blinded' from the observer.
3. One minute later again, the observer repeats a measurement with the mercury sphygmomanometer.
4. Sequence 2–3 is repeated three times in 85 subjects to give 255 readings.
5. The difference is calculated. If the device pressure lies between the first and third pressure the difference is zero; otherwise, the nearer of the two readings is subtracted to give the difference.
6. Data are recorded and analysed as for devices with controllable deflation rates.

Phase VI: Report of evaluation

The final report should be prefaced with subject data that define the key characteristics of the subjects in the study; these data should include the number of subjects, the ranges of systolic and diastolic pressures and the numbers of subjects for each pressure level, age and arm circumference measurement.

Assessment of basic user information and service/maintenance facilities

All technical problems encountered during the validation tests are recorded, so that the information is available for reference, including a description of any problems encountered, the date of occurrence of any breakdown, date of repair, effect on validation procedure, comments on agency/manufacture efficiency, estimated costs of service and appropriate recommendations to the manufacturer for improving the equipment.

Basic information

The information provided in operational manuals is often deficient. Without appropriate specifications and operational instructions it is difficult to obtain an

optimal performance. The information listed in Appendix B should be provided and deficiencies in this regard should be listed in the report.

Acknowledgements

The report should state whether the equipment was purchased for the evaluation or if it was donated by the manufacturer. Data analyses should be carried out by the laboratory doing the evaluation. If the analyses were carried out by the manufacturers, this should be stated.

Discussion

The expanding role of ambulatory measurement is creating a large potential market and it is important to anticipate the consequences of uncontrolled proliferation of very expensive ambulatory systems which are not subjected to critical evaluation. The Working Party on Blood Pressure Measurement of the British Hypertension Society, which has previously made recommendations on the accurate measurement of blood pressure [17, 18, 22] was given a mandate by the society to prepare recommendations for the evaluation of ambulatory devices. When the Working Party began considering the problem it appeared that the AAMI standard [8] might be adopted with minor modifications as a standardized protocol for the general evaluation of ambulatory devices. However, careful consideration of the AAMI standard revealed a number of methodological and statistical problems and it was considered necessary to draft a protocol directed specifically at evaluation of ambulatory devices for clinical use, rather than dealing (as the AAMI has done very effectively) with manufacturing standards. In so doing, the Working Party acknowledges gratefully the strong influence that the AAMI has had in initiating thinking in this complex subject, and many of the AAMI recommendations are incorporated in this protocol.

One drawback of the AAMI standard is that it is not published in a medical journal. Of greater importance is the failure of the standard to provide a test for interdevice variability, a test for the device in the ambulatory setting and a test for accuracy after a period of use. This protocol addresses these areas as well as making recommendations on the information that should be supplied by the manufacturer, and permits an assessment of patient acceptability of the device. Though this protocol provides an assessment of performance during ambulatory use it is important to recognize that blood pressure measurements are usually made with the subject at rest; an ambulatory device that meets the criteria of the present protocol cannot be assumed to be accurate during physiological manoeuvres, such as exercise, isometric handgrip, Valsalva manoeuvre, etc. Moreover, the protocol does not test the device in the variety of positions in which ambulatory measurement may be made.

Since the AAMI standard was published, methods of statistical analyses in the evaluation of devices have also changed. Most notably, the correlation coefficient, once regarded as the basis of comparison for studies of one device against another, has been largely abandoned because it may suggest close accuracy when there are, in fact, gross differences between the devices being compared [23, 24]. Therefore, more suitable statistical methods are recommended here.

We regard the AAMI criteria of acceptable inaccuracy (mean difference of

± 5 mmHg with a standard deviation of 8 mmHg) [8] as too liberal. We have devised, therefore, a system of grading, grade A being an unachieved accuracy level to date, which it is hoped future ambulatory devices will attain.

The Working Party is conscious that following the recommendations in this protocol is an onerous task and has endeavoured to keep the procedures as simple as possible. Towards this end, the entire procedure has been designed to ensure that expensive and time-consuming tests are not performed on devices which do not meet certain basic accuracy criteria. For example, the most difficult test, the main validation test requiring the participation of 85 subjects with a wide range of pressures, is not performed until the device has been field tested and proven not to have developed interdevice variability during a period of ambulatory use. The procedure, nonetheless, is necessarily lengthy and requires considerable involvement of trained personnel and careful supervision, but if ambulatory measurement is to realise its full potential it is imperative that strict standards are applied without delay.

However, the adoption of these standards by the manufacturers of blood pressure measuring devices may not be easily effected. Manufacturers cannot be obliged to guarantee the accuracy of their product, though it is likely that the legislative harmonization being prepared by the Commission of the European Communities with regard to essential safety requirements of medical devices will be extended to other aspects of device performance, such as accuracy [25]. Also, we expect that reputable manufacturers will welcome the opportunity of having ambulatory blood pressure measuring devices evaluated independently according to a generally accepted protocol. Unfortunately, the presence of a national standard is not a guarantee of accuracy and it will be many years before there is any acceptable standard in Britain and Ireland. The British Standards Institution is presently preparing a standard for automated devices. The British Hypertension Society has made application to the Institution for a standard for semi-automated devices (personal communication to E.O'B). However, even if there was an acceptable standard for ambulatory devices, manufacturers would not be obliged to comply with it and the need for independent evaluation would still exist.

Manufacturers of ambulatory blood pressure measuring systems must be encouraged to have their product evaluated independently according to an approved evaluation procedure. This process, which will necessarily take time, could be influenced beneficially if editors of general medical, clinical pharmacology and hypertension journals critically evaluated the evidence supporting the accuracy of ambulatory blood pressure measuring systems used in research studies. Health authorities and sponsoring organisations should not continue to purchase equipment which has not been evaluated adequately.

References

1. PICKERING, T. G., HARSHFIELD, G. A., KLEINERT, H. D., BLANK, S., LARAGH, J. H. (1982) Blood pressure during normal daily activities, sleep and exercise. *JAMA*, **247**, 992-996.
2. MANCIA, G. (1989) Clinical use of ambulatory blood pressure monitoring. *Am J Hypertens*, **2**, 50S-54S.
3. O'BRIEN, E., COX, J., O'MALLEY, K. (1989) Editorial review: Ambulatory blood pressure measurement in the evaluation of blood pressure lowering drugs. *J Hypertens* **7**, 243-247.

4. PERLOFF, D., SOKOLOW, M., COWAN, R. M., JUSTER, R. P. (1989) Prognostic value of ambulatory blood pressure measurements: further analysis. *J Hypertens* 7 (Suppl. 3), S3-S10.
5. DRAYER, J. I. M., WEBER, M. A., DEYOUNG, J. L., WYLE, F. A. (1982) Circadian blood pressure patterns in ambulatory hypertensive patients: effects of age. *Am J Med* 73, 493-499.
6. O'BRIEN, E., O'MALLEY, K. (1988) Overdiagnosing hypertension. *Br Med J* 297, 1211-1212.
7. O'BRIEN, E., O'MALLEY, K., SHERIDAN, J. (1989) The need for a standardized protocol for validating non-invasive ambulatory blood pressure measuring devices. *J Hypertens* 7 (Suppl. 3), S19-S20.
8. ASSOCIATION FOR THE ADVANCEMENT OF MEDICAL INSTRUMENTATION (1987) *American National Standard for Electronic or Automated Sphygmomanometers* (Washington DC, AAMI, USA).
9. O'BRIEN, E., MEE, F., ATKINS, N., O'MALLEY, K. (1990) Inaccuracy of seven popular sphygmomanometers for home measurement of blood pressure. *J Hypertens* 8, 621-635.
10. MANCIA, G. (1983) Methods of assessing blood pressure values in humans. *Hypertension* 5 (Suppl. III), III5-III13.
11. VARDAN, S., MOOKHERJEE, S., WARNER, R., SMULYAN, H. (1983) Systolic hypertension, direct and indirect blood pressure measurements. *Arch Intern Med* 143, 938-938.
12. WRIGHT, D. M., DORE, C. F. (1970) A random-zero sphygmomanometer. *Lancet* 1, 337-338.
13. EVANS, J. G., PRIOR, I. A. M. (1970) Experience with the random-zero sphygmomanometer. *Br J Prev Soc Med* 24, 10-15.
14. LABARTHE, D. R., HAWKINS, C. M., REMINGTON, R. D. (1973) Evaluation of performance of selected devices for measuring blood pressure. *Am J Cardiol* 32, 546-553.
15. DE GUADAMARIS, R., FOLSOM, A. R., PRINEAS, R. J., LUEPKER, R. V. (1985) The random-zero versus the standard mercury sphygmomanometer: a systematic blood pressure difference. *Am J Epidemiol* 121, 282-290.
16. PARKER, D., LIU, K., DYER, A. R., GIUMETTI, D., LIAO, Y., STAMLER, J. (1988) A comparison of the random zero and standard sphygmomanometers. *Hypertension* 11, 269-272.
17. PETRIE, J. C., O'BRIEN, E., LITTLER, W. A., DE SWIET, M. (1986) Recommendations on blood pressure measurement. *Br Med J* 293, 611-615.
18. DE SWIET, M., DILLON, M. J., LITTLER, W., O'BRIEN, E., PADFIELD, P. L., PETRIE, J. C. (1989) Measurement of blood pressure in children: recommendations of a Working Party of the British Hypertension Society. *Br Med J* 299, 497.
19. JAMIESON, M., PETRIE, J., O'BRIEN, E., PADFIELD, P., LITTLER, W. A., DE SWIET, M. (1989) *Blood Pressure Measurement*. London: Video for the British Hypertension Society, British Medical Journal Publications.
20. FITZGERALD, D. J., O'CALLAGHAN, W. G., MCQUAID, R., O'MALLEY, K., O'BRIEN, E. (1982) Accuracy and reliability of two indirect ambulatory blood pressure recorders: Remler M 2000 and Cardiodyne Sphygmolog. *Br Heart J* 48, 572-579.
21. SANTUCCI, S., STEINER, D., ZIMLER, M., JAMES, G., PICKERING, T. (1988) Validation study of the Spacelabs models 90202 and 5200 ambulatory blood-pressure monitors. *J Ambulatory Monitoring* 1, 211-216.
22. O'BRIEN, E., PETRIE, J. C., LITTLER, W. A., DE SWIET, M. (1987) Standards for blood pressure measuring devices. *Br Med J* 244, 1245-1246.
23. CONWAY, J. (1986) Home blood pressure recording. *Clin Exp Hypertens [A]*, 8, 1247-1294.
24. BLAND, J. M., ALTMAN, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, i, 307-310.
25. JOINT EUROPEAN STANDARDS INSTITUTION (1989) *Proceedings of Workshop on the European Standardization of Medical Devices, Brussels 12-14 December 1988*. Brussels: CEN/CENELEC.

Appendix A: Statistical considerations

Introduction

Different observers and different devices never agree exactly, in the sense of giving the same blood pressure for all subjects. The comparison of two sets of blood pressure readings thus takes the form of assessing the amount of disagreement. Methods of comparison are described and illustrated in this appendix. However, statistical methods cannot indicate what is or is not acceptable agreement for an individual subject or a group of subjects; this decision must be based on clinical considerations.

Whether two observers or two devices are compared, the philosophy of the recommended approach is to consider the distribution of the differences between the blood pressure obtained for each individual subject. If more than two sets of measurements are available the same approach is used to compare each pair. Graphs are particularly useful. There is no place in this analysis for the calculation of correlation coefficients or hypothesis tests.

Initial analysis

In the presentation of evaluation data it is common practice to begin by producing a scatter plot of the two sets of blood pressure data (observer and test device). These plots usually show systolic and diastolic pressures separately, although they can both be shown in a single plot.

The scatter plot can be a useful first step, but it is inefficient as all the information is usually clustered near the line. We have, therefore, used a better way of assessing the discrepancies by plotting the differences between the measurements of the observer and the device, against their average, as in figures 4 and 5. This plot shows the differences in blood pressure explicitly, and also indicates whether the distribution of the differences varies according to the level of blood pressure. We use the average blood pressure here, as this is the best estimate of the true blood pressure for that patient at that time. This method of plotting, which can be extended to give more information (see below), is recommended in preference to the conventional scatter plot.

Quantification of agreement

The assessment of agreement is based on both the average differences between the methods of measurement and the variability in the differences. The average agreement between the two sets of blood pressure measurements is the mean of the differences from each subject (and is equal to the difference between the overall means). There are three approaches to the assessment of the variability component of agreement.

1. Proportion of differences that are greater than some reference value, say 10 mmHg, can be calculated. Reference values can be superimposed on the scatter diagram.
2. Values outside which a certain proportion, say 10%, of the observations fell can be calculated. This is done simply by ordering the data and taking the range of values left after a percentage of the sample is removed from each end. These values can also be superimposed on the scatter diagram.

Table 4. Effect of test methodology on grading analysis.

			Differences between standard and test device (mmHg)		
Grade			≤ 5	≤ 10	≤ 15
Simultaneous, same arm	SBP	A	89	97	99
	DBP	A	91	100	100
Simultaneous, opposite arm	SBP	B	74	95	98
	DBP	C	64	94	99
I. Sequential, same arm	SBP	B	69	91	98
	DBP	A	84	99	100
II. Sequential, opposite arm	SBP	A	85	98	99
	DBP	A	92	99	100

DBP, diastolic blood pressure; SBP, systolic blood pressure.

- Standard deviation of the intrasubject differences can be calculated. On the assumption that the differences will be normally distributed, which is usually reasonable for blood pressure data, the range of values expected to encompass most intrasubject differences can be calculated. For example, 90% of differences can be expected to lie between the mean \pm 1.645 s.d. These two values are called the 90% limits of agreement. They can also be superimposed on the scatter diagram.

Methods (1) and (2) do not require any assumptions about the distribution of the differences, but they are generally less reliable than those obtained using normal distribution theory, especially in small samples. However, if there are one or more outliers (extreme discrepancies between observers or methods), an empirical approach may be preferable. In this protocol, we have chosen to use the percentage of differences within certain limits (method 1), a simple approach that can be used for all phases of the evaluation. For the device validation phase (phase V) three of these assessments are made, relating to the percentage of differences within 5, 10 and 15 mmHg. A device is then graded according to these results using the criteria in Table 3.

Grading and method of validation

Table 4 illustrates the rationale for calculating the differences in the sequential test, using the grading criteria given in Table 3. In this analysis, 85 subjects with a wide range of blood pressure had simultaneous measurements taken with mercury sphygmomanometers by two trained observers in the same arm and simultaneously in the opposite arm by a third trained observer, the sequence being repeated three times so as to provide sequential measurements in both arms. The results in the table, therefore, are derived from 255 measurements.

In the first two lines, the grades for simultaneous same-arm measurements are shown and grade A status is reached for both systolic and diastolic pressures. This is the 'gold standard'. In the second pair of lines, data for simultaneous opposite-arm measurements are presented and a grade B rating is obtained for systolic and grade C for diastolic pressure; clearly this analysis is much inferior to the first. In

the third pair of lines (sequential I) the same-arm sequential measurements are analysed, the differences being calculated by comparing the mean of the first and third mercury measurements with the second measurement (which corresponds to a test device); a grade A rating is achieved for diastolic but only a grade B for systolic pressure. Clearly this is better than opposite-arm measurements but is not as good for systolic pressure as simultaneous same-arm measurements. However, this analysis is flawed mathematically in that the relationship between the first and third mercury measurement is assumed to be linear, which need not be so. In the fourth pair of lines (sequential II) the analysis is based on the assumption that the difference between the first and third blood pressure reading need not be linear, and the difference is calculated as follows. If the second (test) pressure lies between the first and third pressure the difference is zero; otherwise the nearer of the two readings is subtracted to give the difference. This correction technique restores the sequential analysis to parity with the simultaneous same-arm analysis by bringing the systolic rating to grade A.

Power

The calculation of an appropriate sample size for the device validation (phase V) is, to some extent, arbitrary. If the observed proportion of differences within 5 mmHg is 80%, then a 95% confidence interval for this proportion will be $\pm 5\%$ with a sample size of 85 subjects (225 observations), the size recommended in the AAMI standard [8]. We believe that a smaller sample may be acceptable, but we have decided to remain with the AAMI recommendations of 85 subjects until working data become available as the protocol is used, when it may be possible to make power calculations that would effect a reduction in this large sample size.

Appendix B: Basic information

Model identification: When manufacturers incorporate modifications into externally identical or indistinguishable versions of a model, this should be clearly indicated by model number and full details as to how the model differs from earlier versions should be provided. In particular, the likely effect of all such modifications on the performance and accuracy of the model should be stated.

Costs: Cost of the recorder, decoder, computer analysis facilities and all components should be listed. The consumables needed for device operation and their cost should be provided.

Compliance with standard(s): Standard adopted by the manufacturer should be stated.

Validation studies and results: Results of validation assessments by the manufacturer, and/or by independent laboratories should be summarized to provide the following details: the method of validation, the number of subjects, any special features in subject selection, e.g. pregnancy, childhood, the range of blood pressures, the heart rate change, the accuracy requirements and the statistical analysis employed. Full references for all published validation studies should be listed together with the addresses of the laboratories.

Instructions for use: These should be clearly stated in a step-by-step layout. Illustrations are helpful in this context.

Patient instruction card: A card should be provided for distribution to patients

using the ambulatory recorder, which gives simple operational instructions together with instructions as to what precautions to take in the event of the device malfunctioning.

Power supply: Mains voltage and frequency must be shown and whether or not a transformer is needed to adapt the decoder. If the latter applies, the frequency must also be converted as the movement of certain parts may be affected, with resultant inaccuracies. The most suitable batteries for the device should be listed and those capable of being recharged should be indicated. The number of recordings obtainable for a set of batteries, or per charge, and the warning system for battery failure should be indicated.

Instructions for care and maintenance: The operator should be given clear instructions on the day-to-day care of the equipment and the need for regular maintenance. Product warranty information should be provided. Ambulatory devices should have full warranty cover for at least 1 year after the date of purchase.

Service facilities: Location of national and international service facilities should be listed. It is regrettable that some manufacturers appoint agents who, though competent with certain ranges of medical devices, have little or no knowledge of specialized blood pressure measuring equipment. Potential purchasers should be aware of this problem and check that the agent is competent to provide the necessary facilities. An estimate of the cost of routine servicing out of warranty together with an estimate of the costs of transporting the equipment for servicing should be given. Maintenance contracts are available for some ambulatory systems and details of these should be provided.

Dimensions: Dimensions of the recorder and its total weight with batteries, pump, etc., should be indicated. Means of attachment, waist-belt, shoulder-strap, or bag, etc., should also be stated.

List of components: All major components of the system should be listed. Dimensions of the bladders supplied and the dimensions of the range of bladders available should be indicated. A 35×13 cm bladder is strongly recommended for routine use in most adults by the British Hypertension Society [17].

Method(s) of blood pressure measurement: Basic method of pressure detection, for example, auscultatory or oscillometric, should be stated and if more than one method is used the indications for changing methods and the means of denoting this on the recording should be stated. With Korotkoff sound detecting devices, the use of either phase IV or V as the diastolic end-point must be disclosed. If data are derived from recorded measurements, such as mean pressure, the method of calculation must be stated.

Artefact editing: Some ambulatory devices have inbuilt systems for editing artefactual measurements. The method of doing this and the rationale should be stated. Reliable and accurate devices should require only minimal editing and this should be performed automatically by the device. It should not be necessary for the operator to have to screen the device measurements for bizarre recordings that are likely to be artefactual. We have refrained, therefore, from making recommendations on artefact editing.

Facility for checking device accuracy: Blood pressure measuring devices should be provided with a facility for accuracy assessment against a reference system whereby simultaneous measurement can be performed on the same arm with the device which is being tested and the reference system. Some ambulatory systems function with rapid deflation rates but in some models it is possible to switch to

a slower deflation rate. These devices should be tested using the rapid deflation rate, as switching to a slower mode may give results which do not reflect the accuracy of the device in use. Special consideration has to be given to the method of testing and to the interpretation of data with these devices.

Facility for device recalibration: The manufacturer should state the intervals at which recalibration becomes necessary and a simple method for checking accuracy should be provided. If recalibration is required, the manufacturer should state whether this can be done by the owner, and if so, how.

Factors affecting accuracy: Many factors may affect the accuracy of ambulatory recordings, such as arm movement, exercise, arm position, cuff or cloth friction. All these factors should be listed by the manufacturer.

In patients with cardiac arrhythmias, it is difficult and sometimes impossible to obtain an accurate measurement of blood pressure with a standard mercury sphygmomanometer. In these subjects the likelihood of obtaining an accurate ambulatory record is remote, and unless sound validation of accuracy is available for arrhythmias it should be assumed that ambulatory devices are probably inaccurate in these patients. The manufacturer's literature should carry a statement along the following lines: 'This instrument has not been validated in patients with arrhythmias'.

Operator training requirement: Some ambulatory systems require considerable expertise on the part of the operator if accurate measurements are to be obtained, whereas other systems require relatively little instruction. These requirements should be stated.

Computer analysis: Some ambulatory systems are compatible with personal computer systems. The exact requirements for linking with computer systems and their cost should be stated. If the ambulatory system is dependent on its own computer for plotting and analysis this should be made clear and the cost of the computer facility, if it is an optional extra, should be stated.

Clear instructions should be provided for setting recording conditions (e.g. frequency of recordings during defined periods, on/off condition of digital display); retrieving recordings and saving data to disk; retrieving data from disk; displaying numerical data and graphics; importing data into statistical/graphic/spreadsheet software programs; printing results (excerpts or total).

The manufacturer should list compatible computers (PC or other) and printers together with memory requirements, compatible graphic adaptors, additional software or hardware requirements (including interfaces and cables if these are not supplied).

Problem list and solutions: Finally, a list of common operational problems should be listed with the means of detection and remedy.