

Working Group on Blood Pressure Monitoring of the European Society of Hypertension International Protocol for validation of blood pressure measuring devices in adults

Eoin O'Brien^a, Thomas Pickering^b, Roland Asmar^c, Martin Myers^d, Gianfranco Parati^e, Jan Staessen^f, Thomas Mengden^g, Yutaka Imai^h, Bernard Waeberⁱ and Paolo Palatini^j and with the statistical assistance of Neil Atkins^a and William Gerin^k, on behalf of the Working Group on Blood Pressure Monitoring of the European Society of Hypertension

Blood Pressure Monitoring 2002, 7:3–17

Keywords: protocol, device validation, blood pressure, blood pressure measurement, British Hypertension Society (BHS), Association for the Advancement of Medical Instrumentation (AAMI)

^aBlood Pressure Unit, Beaumont Hospital, Dublin, Ireland; ^bHypertension Center, New York Hospital, Cornell Medical Center, New York, USA; ^cSociété Française D'Hypertension Artérielle, Fillale de la Société Française de Cardiologie, Paris, France; ^dDivision of Cardiology, Sunnybrook and Women's College Health Sciences Centre, Toronto, Canada; ^eInstituto Scientifico Ospedale San Luca, IRCCS, Istituto Auxologico Italiano Milan, Italy; ^fKatholieke Universiteit Leuven, Hypertensie en Cardiovasculaire Revalidatie Eenheid, Inwendige Geneeskunde-Cardiologie, Leuven, Belgium; ^gUniversity Clinic Bonn, Department of Internal Medicine, Bonn, Germany; ^hSecond Department of Internal Medicine, Tohoku University School of Medicine, Sendai, Japan; ⁱCentre Hospitalier Universitaire Vaudois, Division D'Hypertension, Department de Medecine Interne, Lausanne, Switzerland; ^jDipartimento di Medicina Clinica e Sperimentale, Università di Padova, Padua, Italy; ^kMount Sinai School of Medicine, New York, USA

Correspondence and requests for reprints to Professor E. O'Brien, Blood Pressure Unit, Beaumont Hospital, Dublin 9, Ireland.

Received 13 November 2001 Accepted 15 November 2001

Introduction

With the increasing marketing of automated and semi-automated devices for the measurement of blood pressure, potential purchasers need to be able to satisfy themselves that such devices have been evaluated according to agreed criteria [1]. With this in mind, the Association for the Advancement of Medical Instrumentation (AAMI) published a standard for electronic or aneroid sphygmomanometers in 1987 [2] that included a protocol for evaluating the accuracy of devices, this being followed in 1990 by the protocol of the British Hypertension Society (BHS) [3]. Both of these were revised in 1993 [4,5]. These protocols, which differed in detail, had a common objective, namely the standardization of validation procedures to establish minimum standards of accuracy and performance, and to facilitate the comparison of one device with another [6].

Since their introduction, a large number of blood pressure measuring devices have been evaluated according to one or both protocols. Experience has, however, demonstrated that the conditions demanded by the protocols are extremely difficult to fulfil. This is especially so because of the large number of subjects who have to be recruited and the ranges of blood pressure required. The time required to complete a validation study is such that it is difficult to recruit trained staff for the duration of an investigation. These factors have made validation studies difficult to perform and very costly, with the result that fewer centers are prepared to undertake them. This is particularly unfortunate as more devices than ever before are in need of independent validation.

When the BHS dissolved its Working Party on Blood Pressure Measurement, the Working Group on Blood Pressure Monitoring of the European Society of Hypertension (ESH) undertook to produce an updated protocol, which it has named the International Protocol. The ESH Working Group on Blood Pressure Monitoring is composed of experts in blood pressure measurement, many of whom

have considerable experience in validating blood pressure measuring devices.

In setting about its objective, the ESH Working Group recognized the urgent imperative to provide a simplified protocol that does not sacrifice the integrity of the earlier protocols. When the AAMI and BHS protocols [2–5] were published, the relevant committees did not have evidence from previous studies on which to base their recommendations. The ESH Working Group has had the advantage of being able to examine and analyze the data from 19 validation studies performed according to the AAMI and BHS protocols at the Blood Pressure Unit in Dublin [7–23]. A critical assessment of this database of evidence has permitted a rationalization and simplification of validation procedures without losing the merits of the much more complicated earlier protocols. The basic recommendations of the simplified International Protocol have been presented at meetings of the ESH Working Group, and the proceedings of these meetings have been published in order to invite comment and discussion [24–27].

The International Protocol has been drafted in such a way as to be applicable to the majority of blood pressure measuring devices on the market. The validation procedure is therefore confined to adults over the age of 30 years (as these will constitute the majority of subjects with hypertension), and does not make recommendations for special groups, such as children, pregnant women and the elderly, or for special circumstances, for example exercise. It is anticipated that the relative ease of performance of the International Protocol will encourage manufacturers to submit blood pressure measuring devices for validation in order to obtain the minimum approval necessary for a device to be used in clinical medicine, and that, in time, most devices on the market will be assessed according to the protocol for basic accuracy. It does not preclude manufacturers of devices from subjecting their products to more rigorous assessment and validation.

Validation procedure

Summary

The validation team should consist of four persons experienced in blood pressure measurement: two observers and a supervisor (generally nurses), and an ‘expert’ (a doctor overseeing the entire procedure). If the doctor can be present throughout the entire validation procedure, he/she can take over the role of supervisor, thereby reducing the number of personnel to three. The validation procedure consists of the following steps

1. *Observer training and assessment.* Two observers are trained in accurate blood pressure measurement.
2. *Familiarization session.* The validation team becomes familiar with the workings of the device and the accompanying software.

3. *Validation measurements.* Observer and device measurements are recorded on subjects in two phases. In the first phase, 15 subjects are recruited; devices passing this primary phase proceed to the secondary phase, for which a further 18 subjects are recruited.
4. *Analysis.* An analysis of the recorded measurements is carried out after each phase.
5. *Reporting.* The results are presented in tabular and graphical forms.

Observer training and assessment

Consideration must first be given to the technique of blood pressure measurement, which should be as follows throughout the validation procedure.

Blood pressure measurement technique

A standard mercury sphygmomanometer, the components of which have been carefully checked before the study, is used as a reference standard. It is appreciated that terminal digit preference is a problem with conventional mercury sphygmomanometry, and care should be taken to reduce this in the observer training session. The Hawksley random-zero sphygmomanometer only disguises digit preference, and its accuracy has been questioned [7,28]; therefore, its use is not recommended in validation studies. All blood pressures should be recorded to the nearest 2 mmHg.

Blood pressure should be measured with the arm supported at heart level [29]; the level of the manometer does not affect the accuracy of measurement, but it should be at eye level and within 1 m of the observer. The quality of the stethoscope is also crucial to performing the evaluation procedure. Stethoscopes with badly fitting earpieces and poor-quality diaphragms preclude precise auscultation of the Korotkoff sounds. A well-maintained quality stethoscope is recommended.

Observer training

The first prerequisite for this validation test is to ensure that the observers have adequate auditory and visual acuity, and that they have achieved the required accuracy as laid out below. It is, however, possible that observers who fulfil these criteria at the outset of the study will not do so at the end, and if this happens the observers must be re-assessed for accuracy. To avoid this, analysis should be performed as the study proceeds to detect any drift in agreement between the observers.

Observers may be trained in the following ways:

1. By fulfilling the test requirements of the CD-ROMs produced by the BHS or the Société Française d’Hypertension Artérielle as described in Appendix B [30,31].

2. By formal training and assessment is described in Appendix B [32, 33].
3. By using an audio-visual method for validation, such as the Sphygmocorder [34,35] as discussed in Appendix B.

Familiarization session

As automated devices for blood pressure measurement may be complex, it is important that the personnel performing a validation study are fully conversant with the equipment. The observers, having satisfied the training criteria, should next be instructed in the use of the device to be validated and any accompanying computer software. For uncomplicated devices designed to provide a straightforward blood pressure measurement, the familiarization session should consist of performing a series of practice measurements on volunteers. A more formal session should, however, be applied to complex devices such as systems for measuring 24-h blood pressure. This session has two functions: first, it serves as a familiarization period for the personnel performing the validation study, and second, any technical peculiarities of the device being tested, which might influence the validation procedure, may be identified.

Validation measurements

General considerations

Device validation should be performed at room temperature without disturbing influences such as telephones and beeps in the area.

Some automated devices have more than one method of measuring blood pressure. It may be claimed for a particular device, for example, that electrocardiogram gating may be used when more accurate measurement is required. In these circumstances, validation must be performed with and without electrocardiogram gating. Similarly, some Korotkoff sound-detecting devices provide an oscillometric back-up when sound detection fails. When both systems generate simultaneous readings, only one comparative validation is required, but when the oscillometric method is a back-up to the auscultatory method and provides a separate measurement, both systems of measurement must undergo individual validation.

Arm circumference and bladder dimensions

The circumference of the arms should be measured to ensure that the bladder being used is adequate for the subject. Measurements made with the test device should use the appropriate bladder according to the manufacturers' instructions. Standard mercury manometer measurements must be taken with a bladder of sufficient length to encircle 80% of the arm circumference [29]. If a

test device recommends different cuff sizes, the appropriate cuff/bladder should be used, but no other part of the apparatus should be changed. It is important to ensure, when assessing auscultatory devices, that the same microphone(s) are used throughout the validation test.

Devices for measuring blood pressure at the wrist

The International Protocol may be used to validate devices that measure blood pressure at the wrist. There is little literature regarding the accuracy of devices for wrist measurement, and most studies have shown these devices to be inaccurate [1]. Measurements of blood pressure at the wrist using oscillometric devices generally overestimate blood pressure compared with conventional sphygmomanometry on the upper arm, and the differences can be substantial [36–38].

It must, however, be emphasized that although a device designed for measuring blood pressure at the wrist may be accurate when tested according to the International Protocol, it may be inaccurate for the self-measurement of blood pressure if the instructions to have the wrist at heart level during measurement are not strictly followed. Devices for self-measurement that measure blood pressure at the finger are not recommended because vasoconstriction of the digital arteries can introduce substantial errors.

Subject selection

In selecting 33 subjects (15 for the phase 1, and a further 18 for phase 2) with a wide range of blood pressure it is probable that there will be a representative range of arm circumference, and subjects should not be selected on the basis of their arm circumference. Subjects may be taking antihypertensive medication but must not present in atrial fibrillation or any sustained arrhythmia.

<i>Number</i>	Phase 1 Fifteen subjects Phase 2 Thirty-three subjects
<i>Sex</i>	Phase 1 At least five male and five female Phase 2 At least 10 male and 10 female
<i>Age range</i>	All subjects should be at least 30 years of age
<i>Arm circumference</i>	Distribution by chance
<i>Blood pressure range</i>	As in Table 1

There are three ranges for systolic (SBP) and three for diastolic (DBP) blood pressure, with 11 subjects in each range to provide 99 pairs of measurements. To optimize recruitment, it is recommended that subjects for the high-diastolic and low-systolic groups should be recruited first. The emphasis should then be placed on filling the remaining high-systolic and low-diastolic groups. Finally, the remaining gaps in the middle groups should be filled. The blood pressure used in this categorization is the entry

Table 1 Blood pressure ranges for entry blood pressure (BPA)

	SBP	DBP
Low	90–129	40–79
Medium	130–160	80–100
High	161–180	101–130

For the primary phase, five of the 15 subjects should have a systolic blood pressure (SBP) in each of the ranges. Similarly, five of the 15 subjects should have a diastolic blood pressure (DBP) in each of the ranges. For the secondary phase, 11 of the 33 subjects (including the first 15 subjects) should have SBP and DBP in each of the ranges. It is recommended that recruitment should commence by targeting subjects likely to have pressures in the low-systolic and high-diastolic ranges, then progressing to complete the high-systolic and low-diastolic ranges so that it will be easy to complete the recruitment with the remaining medium ranges.

blood pressure at the time of the validation procedure (BPA), rather than that at the time of recruitment for validation.

Observer measurement

Measurements can be either assessed live using two observers or recorded and later re-assessed using the Sphygmocorder [34,35]. Measurements made simultaneously by two observers must be checked by the validation supervisor. If the systolic and diastolic measurements are no more than 4 mmHg apart, the mean value of the two observer measurements for both systolic and diastolic blood pressures is used. Otherwise, the measurement must be taken again.

When the Sphygmocorder is used, two observers should assess the recording separately. If their opinion differs, they should re-assess the values together until agreement is reached. Further references to 'observer measurement' indicate either the mean of two observer measurements or the agreed measurement using the Sphygmocorder. At least 30 s should be allowed between each measurement to avoid venous congestion, but not more than 60 s or variability may be increased.

Procedure

1. The subject is introduced to the observers, and the procedure is explained. Arm circumference, sex, date of birth and current date and time are noted. The subject is then asked to relax for 10–15 min (in order to minimize anxiety and any white-coat effect, which will increase variability).
2. Nine sequential same-arm measurements using the test instrument and a standard mercury sphygmomanometer are recorded as follows:

BPA Entry blood pressure, observers 1 and 2 each with the mercury standard. The mean values are used to categorize the subject into a low, medium or high range separately for SBP and DBP (Table 1).

BPB Device detection blood pressure, observer 3. This blood pressure is measured to allow the test instrument to determine the blood pressure characteristics of the subject; more than one attempt may be needed with some devices; this measurement is not included in the analysis. If the device fails to record a measurement after three attempts, the subject is excused.

BP1 Observers 1 and 2 with the mercury standard.

BP2 Supervisor with the test instrument.

BP3 Observers 1 and 2 with the mercury standard.

BP4 Supervisor with the test instrument.

BP5 Observers 1 and 2 with the mercury standard.

BP6 Supervisor with the test instrument.

BP7 Observers 1 and 2 with the mercury standard.

3. Documentation must be provided for data omitted for legitimate technical reasons. Once a subject has been included, the data for that subject should not be excluded from the study if blood pressure values are obtainable; if blood pressure measurements using either the reference method or the test instrument are unavailable, data entry for that individual may be excluded, with an accompanying explanation. Additional individuals must then enter into the study to ensure a sample size of 33.

Analysis

For a detailed discussion on the statistical methods used in the protocol, see Appendix D. A software program has been designed specifically to analyze the data (Société Française d'Hypertension Artérielle, Paris).

Accuracy criteria

The BHS protocol introduced the concept of classifying the differences between test and control measurements according to whether these lay within 5, 10 or 15 mmHg, or were over 15 mmHg apart. The final grading was based on the number of differences falling into these categories. This protocol seeks to keep this concept but expand its flexibility.

Differences are always calculated by subtracting the observer measurement from the device measurement. When comparing and categorizing differences, their absolute values are used. A difference is categorized into one of four bands according to its rounded absolute value for SBP and DBP:

- 0–5 mmHg These represent measurements considered to be very accurate (no error of clinical relevance).
- 6–10 mmHg These represent measurements considered to be slightly inaccurate.

- 11–15 mmHg These represent measurements considered to be moderately inaccurate.
- > 15 mmHg These represent measurements considered to be very inaccurate.

The analysis is based on how values in these bands fall cumulatively into three zones:

- Within 5 mmHg This zone represents all values falling in the 0–5 mmHg band.
- Within 10 mmHg This zone represents all values falling in the 0–5 and 6–10 mmHg bands.
- Within 15 mmHg This zone represents all values falling in the 0–5, 6–10 and 11–15 mmHg bands.

Subject measurements

For assessment of accuracy, only measurements BP1 to BP7 are used. The mean of each pair of observer measurements is calculated; this is denoted as observer measurement BP1, BP3, BP5 or BP7. Each device measurement is flanked by two of these observer measurements, and one of these must be selected as the comparative measurement.

From these, further measurements are derived as follows.

1. The differences BP2–BP1, BP2–BP3, BP4–BP3, BP4–BP5, BP6–BP5 and BP6–BP7 are calculated.
2. The absolute values of the differences are calculated (i.e. the signs are ignored).
3. These are paired according to the device reading.
4. If the values in a pair are unequal, the observer measurement corresponding to the smaller difference is used.
5. If the values in a pair are equal, the first of the two observer measurements is used.

When this has been completed, there are three device readings for SBP and three for DBP for each subject. Each of these six readings has a single corresponding observer measurement, a difference between the two and a band for that difference as described above.

Experience with existing protocols has demonstrated that the overall outcome of a device can be apparent from a very early stage. This is particularly so with poor devices and is in accordance with statistical expectancy – the larger the error, the smaller the sample size required to prove it. To persist with the validation of a device that is clearly going to fail is an unnecessary waste of time and money, and an inconvenience to participating subjects. A mechanism for eliminating poor devices at an appropriate stage is therefore introduced by dividing the validation process into two phases. In the primary phase, three pairs of measurements are performed on 15 subjects in the pressure ranges given in Table 1, any device failing this

Table 2a Requirements to pass phase 1

Measurements	Within 5 mmHg	Within 10 mmHg	Within 15 mmHg
At least one of	25	35	40

After completing the assessment on 15 subjects, the data (45 comparisons) should be analyzed to determine the number of comparisons falling within the 5 10 and 15 mmHg error bands. At least 25 comparisons must lie within 5 mmHg, at least 35 within 10 mmHg or at least 40 within 15 mmHg. If none of these counts are reached the device is deemed to have failed.

Table 2b Requirements to pass phase 2.1

Measurements	Within 5 mmHg	Within 10 mmHg	Within 15 mmHg
Two of	65	80	95
All of	60	75	90

After completing all 33 subjects, the data (99 comparisons) should be analyzed to determine the number of comparisons falling within the 5, 10 and 15 mmHg error bands. For the device to pass, there must be a minimum of 60, 75 and 90 comparisons falling within 5 10 and 15 mmHg, respectively. Furthermore, there must be a minimum of either 65 comparisons within 5 mmHg and 80 comparisons within 10 mmHg, or 65 comparisons within 5 mmHg and 95 comparisons within 15 mmHg, or 80 comparisons within 10 mmHg and 95 comparisons within 15 mmHg.

Table 2c Requirements to pass phase 2.2

Subjects	2/3 within 5 mmHg	0/3 within 5 mmHg
At least	22	
At most		3

The data should now be analyzed per subject to determine the number of comparisons per subject falling within 5 mmHg. At least 22 of the 33 subjects must have at least two of their three comparisons lying within 5 mmHg. (These include those who have all three comparisons within 5 mmHg.) At most, three of the 33 subjects can have all three of their comparisons over 5 mmHg apart.

phase (Table 2a) being eliminated from further testing. Devices passing this proceed to a secondary phase in which a further 18 subjects (giving a total of 33) are recruited (Table 2b).

Assessment of phase 1

Once there are five subjects in each of the six blood pressure ranges (Table 1), recruitment should be stopped and an assessment performed. Data from only the first five subjects in each range are used. (In filling these ranges, some ranges may be over-subscribed because of subjects having different SBP and DBP ranges.) This will yield 45 sets of measurements for both SBP and DBP.

1. The number of differences in each zone is calculated as described above.
2. A continue/fail grade is determined according to Table 2a (see also Table 3 for an example).
3. If the device fails, the validation is complete; if the device passes, it proceeds to phase 2.

Assessment of phase 2

This phase determines how accurate the device will be for individual measurements (Phase 2.1) and for individual

Table 3 Example of device validation table in report

Phase 1		within 5 mmHg	within 10 mmHg	within 15 mmHg	Recommendation		
Required	One of	25	35	40			
Achieved	SBP	22	35	43	Continue		
	DBP	35	42	44	Continue		
Phase 2.1		within 5 mmHg	within 10 mmHg	within 15 mmHg	Recommendation	Mean difference	Standard deviation
Required	Two of	65	80	95			
	All of	60	75	90			
Achieved	SBP	52	79	90	Fail	3.4 mmHg	8.4 mmHg
	DBP	77	90	94	Pass	-0.6 mmHg	6.9 mmHg
Phase 2.2		2/3 within 5 mmHg	0/3 within 5 mmHg	Recommendation			
Required		≥ 22	≤ 3				
Achieved	SBP	17	4	Fail			
	DBP	28	2	Pass			

The device passes for diastolic blood pressure (DBP, but fails for systolic blood pressure (SBP), thereby failing overall.

subjects (Phase 2.2) by determining the number of differences within 5, 10, and 15 mmHg, and then determining the accuracy.

After all ranges have been filled, there will be 99 sets of measurements for both SBP and DBP.

1. The number of differences in each zone as described above is calculated.
2. A pass/fail grade for phase 2.1 is determined according to Table 2b (see Table 3 for example).
3. For each of the 33 subjects, the number of measurements falling within 5 mmHg is determined.
4. A pass/fail recommendation for phase 2.2 is determined according to Table 2c (see Table 3 for example).
5. If the device passes both phase 2.1 and phase 2.2, it passes the validation and can be recommended for clinical use. If it does not, it fails and is not recommended for clinical use.

Reporting

Statistical report

The report should be prefaced with subject data in order to describe the key characteristics of the subjects in the study. An example of a device validation is shown in Table 3.

1. *Sex distribution.* The number of males and females.
2. *Age distribution.* The mean, standard deviation and range of the subjects' ages.
3. *Arm circumference distribution.* The mean, standard deviation and range of the subjects' arm circumferences and, when different cuff sizes are used, the number of subjects on which each size was used.
4. *Blood pressure.* The mean, standard deviation and range of the subjects' entry SBP and DBP (BPA).

The report should then give the results of the validation.

Phase 1

The number of differences falling in the Within 5 mmHg, Within 10 mmHg and Within 15 mmHg zones (Table 2a), together with the requirements, should be reported in text and tabular form as in Table 3. The basis on which the decision to continue or stop at this stage should be stated.

Phase 2

The number of differences falling in the Within 5 mmHg, Within 10 mmHg and Within 15 mmHg zones (Table 2b), together with the requirements, should be reported in text and tabular form as in Table 3. The number of subjects with all three differences, at least two differences and no differences falling within 5 mmHg (Table 2c) should be reported in text and tabular form as in Table 3. The mean and standard deviation of the observer and device measurements and the differences should be stated. The basis on which the decision to pass or fail the device should be stated.

Graphical representation

Difference-against-mean plots should be presented for the data at the phase at which the study ceased. Phase 1 data should be plotted for devices failing at that stage, and phase 2 data for those passing. The *x*-axis of these plots represents blood pressures in the systolic range 80–190 mmHg and the diastolic range 30–140 mmHg, and the *y*-axis values from -30 to +30 mmHg. Horizontal reference lines are drawn at 5 mmHg intervals from +15 to -15 mmHg. The mean of each device pressure and its corresponding observer pressure is plotted against their difference using a point. Differences greater than 30 mmHg are plotted at 30 mmHg. Differences less than -30 mmHg are plotted at -30 mmHg. The same scales should be used for both SBP and DBP plots. An example is shown in Fig. 1 [39].

Problems

Any problems encountered during the validation procedure, the date of their occurrence, the date of any repairs to the device and the effect of these on the validation procedure should be recorded.

Operational report

The following information should be provided with blood pressure measuring devices, and the final report should acknowledge that such information is available, and although this need not be presented in detail, any deficiencies should be listed in the report.

Basic information

The information provided in operational manuals is often deficient. Without appropriate specifications and opera-

tional instructions, it is difficult to obtain optimal performance.

List of components

All major components of the system should be listed. The dimensions of the bladders supplied and those of the range of bladders available should be indicated.

Method(s) of blood pressure measurement

The basic method of pressure detection (e.g. 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, whether phase IV or phase V is being used for the diastolic endpoint must be disclosed. If data are derived from recorded measurements, such as mean pressure, the method of calculation must be stated.

Factors affecting accuracy

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

Operator training requirements

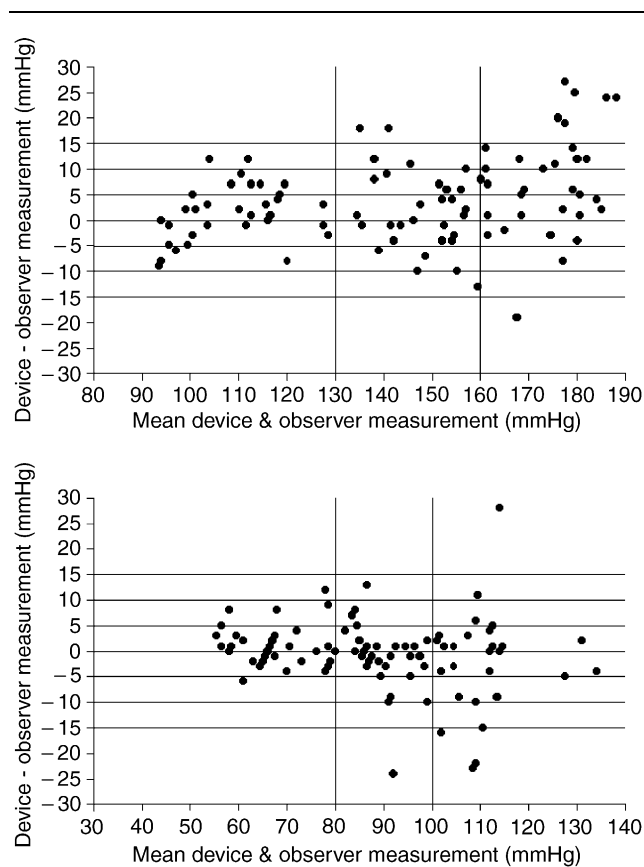
Some automated systems require considerable expertise on the part of the operator if accurate measurements are to be obtained, whereas other systems require relatively little instruction. These requirements should be stated.

Computer analysis

Some automated systems are compatible with personal computer systems. The exact requirements for linking with computer systems and their approximate cost should be stated. If the automated 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 and the on/off condition of the digital display); retrieving recordings and saving data to disk; retrieving data from disk; displaying numerical data and graphics; exporting data to statistical, graphic and spreadsheet software programs; and printing the results (partial or complete). If data cannot be exported, information on how they are stored should be available to facilitate the external analysis of several monitoring events. The manufacturer should list compatible computers (PC or other) and printers together with memory requirements, operating systems, compatible graphic adaptors and additional software or hardware requirements (including interfaces and cables if these are not supplied).

Fig. 1



Devices passing and failing phase 2.1 The x -axis represents the mean of the device and observer measurements. Both systolic blood pressure (upper plot) and diastolic blood pressure (lower plot) ranges should be plotted on the same scale. Recruitment limits are indicated by the vertical lines. The y -axis represents the difference between the device and observer measurements. The 5 mmHg bands from +15 to -15 mmHg are indicated by the horizontal hatched lines. The 99 comparisons are presented in a difference-against-mean scatterplot. In this example, the systolic blood pressure plot depicts a poor device whereas the diastolic blood pressure plot depicts an accurate one.

Acknowledgements

The report should state whether the equipment was purchased for the evaluation or donated or loaned by the manufacturer. The data analysis should ideally be carried out by the laboratory doing the evaluation. If it has been done by the manufacturers, this should be stated. Any consultancies or conflict of interest should be acknowledged by the investigator.

References

- O'Brien E, Waeber B, Parati G, Staessen G, Myers MG, on behalf of the European Society of Hypertension Working Group on Blood Pressure Monitoring. Blood pressure measuring devices: validated instruments. *BMJ* 2001; **322**:531-536.
- Association for the Advancement of Medical Instrumentation. *The national standard of electronic or automated sphygmomanometers*. Arlington, VA: AAMI; 1987.
- O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, O'Malley K, et al. The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems. *J Hypertens* 1990; **8**:607-619.
- O'Brien E, Petrie J, Littler WA, de Swiet M, Padfield PL, Altman D, et al. The British Hypertension Society protocol for the evaluation of blood pressure measuring devices. *J Hypertens* 1993; **11**(suppl 2): S43-S63.
- Association for the Advancement of Medical Instrumentation. *American national standard. Electronic or automated sphygmomanometers*. Arlington, VA: AAMI; 1993.
- O'Brien E, Atkins N. A comparison of the BHS and AAMI protocols for validating blood pressure measuring devices: can the two be reconciled? *J Hypertens* 1994; **12**:1089-1094.
- O'Brien E, Mee F, Atkins N, O'Malley K. Inaccuracy of the Hawksley random-zero sphygmomanometer. *Lancet* 1990; **336**:1465-1468.
- O'Brien E, Mee F, Atkins N, O'Malley K. Inaccuracy of seven popular sphygmomanometers for home-measurement of blood pressure. *J Hypertens* 1990; **8**:621-634.
- O'Brien E, More O'Ferrall J, Galvin J, Costello R, Lydon S, Sheridan J, et al. An evaluation of the Accutracker II non-invasive ambulatory blood pressure recorder according to the AAMI standard. *J Ambul Monit* 1991; **4**:27-33.
- O'Brien E, Atkins N, Mee F, O'Malley K. Evaluation of the SpaceLabs 90202 according to the AAMI standard and BHS criteria. *J Hum Hypertens* 1991; **5**:223-226.
- O'Brien E, Mee F, Atkins N, O'Malley K. Accuracy of the Del Mar Avionics Pressurometer IV determined by the British Hypertension Society Protocol. *J Hypertens* 1991; **9**(suppl 5):S1S-7.
- O'Brien E, Mee F, Atkins N, O'Malley K. Accuracy of the Novacor DIASYS 200 determined by the British Hypertension Society Protocol. *J Hypertens* 1991; **9**(suppl 5):S9S-15.
- O'Brien E, Mee F, Atkins N, O'Malley K. Accuracy of the Takeda TM-2420/TM-2020 determined by the British Hypertension Society Protocol. *J Hypertens* 1991; **9**(suppl 5):S17S-23.
- O'Brien E, Mee F, Atkins N, O'Malley K. Accuracy of the SpaceLabs 90207 determined by the British Hypertension Society Protocol. *J Hypertens* 1991; **9**(suppl 5):S25S-31.
- O'Brien E, Mee F, Atkins N, O'Malley K. Accuracy of the SpaceLabs 90207, Novacor DIASYS 200, Del Mar Avionics Presurometer IV and Takeda TM-2420 ambulatory systems according to the AAMI and BHS criteria. *J Hypertens* 1991; **9**(suppl 6):S332S-333.
- O'Brien E, Mee F, Atkins N, Halligan A, O'Malley K. Accuracy of the SpaceLabs 90207 ambulatory blood pressure measuring system in normotensive pregnant women determined by the British Hypertension Society Protocol. *J Hypertens* 1993; **11**(suppl 5):869-873.
- O'Brien E, Mee F, Atkins N, O'Malley K. Short report: Accuracy of the Dinamap portable monitor, Model 8100, determined by the British Hypertension Society protocol. *J Hypertens* 1993; **11**:761-763.
- O'Brien E, Mee F, Atkins N, O'Malley K. Accuracy of the CH-Druck/Pressure ERKA ambulatory blood pressure measuring system determined by the British Hypertension Society protocol. *J Hypertens* 1993; **11**(suppl 2):1-7.
- O'Brien E, Mee F, Atkins N, O'Malley K. Accuracy of the Profilat ambulatory blood pressure measuring system determined by the British Hypertension Society protocol. *J Hypertens* 1993; **11**(suppl 2):S9S-15.
- Mee F, O'Brien E, Atkins N, O'Malley K. Comparative accuracy of the CH-Druck, Profilat, SpaceLabs 90207, DIASYS 200, Pressurometer IV and Takeda TM-2420 ambulatory blood pressure measuring (ABPM) devices determined by the British Hypertension Society (BHS) protocol. *J Hum Hypertens* 1993; **7**:98.
- O'Brien E, Mee F, Atkins N, Thomas M. Evaluation of three devices for self-measurement of blood pressure: according to the revised British Hypertension Society Protocol: the Omron HEM-705CP, Phillips HP5332, and Nissei DS-175. *Blood Press Monit* 1996; **1**:55-61.
- Mee F, Atkins N, O'Brien E. Evaluation of the Profilat II® ambulatory blood pressure system according to the protocols of the British Hypertension Society and the Association for the Advancement of Medical Instrumentation. *Blood Press Monit* 1999; **3**:353-361.
- O'Brien E, Mee F, Atkins N. Evaluation of the Schiller BR-102 ambulatory blood pressure system according to the protocols of the British Hypertension Society and the Association for the Advancement of Medical Instrumentation. *Blood Press Monit* 1999; **4**:35-43.
- O'Brien E. Criteria for validation of devices. *Blood Press Monit* 1999; **4**:279-293.
- Asmar R, Zanchetti A. On behalf of the Organizing Committee and participants. Guidelines for the use of self-blood pressure monitoring: a summary report of the first international consensus conference. *J Hypertens* 2000; **18**:493-508.
- O'Brien E. Proposals for simplifying the validation protocols of the British Hypertension Society and the Association for the Advancement of Medical Instrumentation. *Blood Press Monit* 2000; **5**:43-45.
- O'Brien E, De Gaudemaris R, Bobrie E, Agabiti Rosei E, Vaisse B. Proceedings from a conference on self blood pressure measurement: devices and validation. *Blood Press Monit* 2000; **5**:93-100.
- Conroy R, O'Brien E, O'Malley K, Atkins N. Measurement error in the Hawksley random zero sphygmomanometer: What damage has been done and what can we learn? *BMJ* 1993; **306**:1319-1322.
- O'Brien E, Petrie J, Littler WA, de Swiet M, Padfield PD, Dillon MJ, et al. *Blood pressure measurement: recommendations of the British Hypertension Society*. 2nd edn. London: BMJ Publishing Group; 1997.
- Blood Pressure Measurement*. CD-ROM. London: British Hypertension Society/BMJ Books; 1998.
- Société Française d'Hypertension Artérielle. *La prise de la pression artérielle au cabinet médical*. Société Française d'Hypertension Artérielle; 1998.
- O'Brien E, Mee F, Atkins N, O'Malley K, Tan S. Training and assessment of observers for blood pressure measurement in hypertension research. *J Hum Hypertens* 1991; **5**:7-10.
- Mee F, Atkins N, O'Brien E. Problems associated with observer training in blood pressure measurement. *J Hum Hypertens* 1994; **8**:293.
- O'Brien E, Atkins N, Mee F, Coyle D, Syed S. A new audio-visual technique for recording blood pressure in research: the Sphygmocorder. *J Hypertens* 1995; **13**:1734-1737.
- Atkins N, O'Brien E, Wesseling KH, Guelen I. Increasing observer objectivity with audio-visual technology: the Sphygmocorder. *Blood Press Monit* 1997; **2**:269-272.
- Forsberg SA, de Guzman M, Berlind S. Validity of blood pressure measurement with cuff in the arm and forearm. *Acta Med Scand* 1979; **188**:389-396.
- Blackburn H, Kihlberg J, Brozek J. Arm versus forearm blood pressure in obesity. *Am Heart J* 1965; **69**:423-427.
- Rytand DA, Boyer SH. Auscultatory determination of arterial pressure at wrist and ankle. *Am J Med* 1954; **17**:112-115.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**:307-310.
- European Commission for Standardisation. European Standard EN 1060-1 (British Standard BSSEN 1060-1:1996). Specification for Non-invasive sphygmomanometers. Part 1. General requirements. European Commission for Standardisation; 1995.
- European Commission for Standardisation. European Standard EN 1060-2 (British Standard BSSEN 1060-2:1996). Specification for Non-invasive sphygmomanometers. Part 2. Supplementary requirements for mechanical sphygmomanometers. European Commission for Standardisation; 1995.
- European Commission for Standardisation. European Standard EN 1060-3 (British Standard BS EN 1060-3: 1997). Non-invasive sphygmomanometers. Part 3. Supplementary requirements for electro-

- mechanical blood pressure measuring systems. European Committee for Standardization; 1997.
- 43 O'Brien E. Modification of blood-pressure measuring devices and the protocol of the British Hypertension Society. *Blood Press Monit* 1999; **4**: 53–54.
 - 44 European Commission for Standardisation. *European Standard EN 540. Clinical investigation of medical devices for human subjects*. European Commission for Standardisation; 1993.
 - 45 West NW, Sheridan JJ, Littler WA. Direct measurement of blood pressure. In: O'Brien E, O'Malley K, editors: *Handbook of hypertension*. Vol. 14. *Blood pressure measurement*. Amsterdam: Elsevier; 1991. pp. 315–333.
 - 46 Parati G, Casadei R, Gropelli A, Di Rienzo M, Mancia G. Comparison of finger and intra-arterial blood pressure monitoring at rest and during laboratory testing. *Hypertension* 1989; **13**: 647–655.
 - 47 O'Brien E, Atkins N, Mee F, O'Malley K. Comparative accuracy of six ambulatory devices according to blood pressure levels. *J Hypertens* 1993; **11**: 673–675.

Appendix A. Comparison with previous protocols

Our approach to simplifying previous validation procedures has concentrated on the following areas:

Elimination of pre-validation phases

The main validation procedure of the existing BHS protocol has five phases: (i) before-use device calibration; (ii) the in-use (field) phase; (iii) after-use device calibration; (iv) static device validation; and (v) report of the evaluation (4). Phases (i)–(iii) were originally introduced to identify intra-device variability, but if a device has fulfilled the general requirements of the European Union directives [40–42] or the AAMI standard [5], it is not necessary to subject these devices to phases (i), (ii) or (iii) of the BHS protocol. These pre-validation phases are thus not included in the present protocol, thereby resulting in considerable reduction in time and labor.

Improving observer recruitment and training

The most fallible component of blood pressure measurement is the human observer, and consideration must be given to the role of the education and certification of observers. CD-ROMs are available to facilitate the training and assessment of observers [30,31].

The Sphygmocorder, a device that provides an audio recording of Korotkoff sounds with a video recording of a mercury column, has been designed to provide objective evidence of validation blood pressures [34,35]. The Sphygmocorder removes the expensive need to employ two observers and a supervisor throughout the validation procedure and has greatly facilitated device validation.

Use of simultaneous or sequential comparisons

The basis of device evaluation is the comparison between blood pressure measured by the device being tested and measurements made by trained observers using a mercury sphygmomanometer and stethoscope to auscultate the Korotkoff sounds. With most automated devices, a number

of factors may make it difficult or impossible to perform simultaneous comparison on the same arm.

Devices, for example, that deflate at a rate of more than 5 mmHg per second do not permit accurate measurement by an auscultating observer, leading to inaccurate comparison between the test and reference devices [4]. At fast deflation rates, an auscultating observer will tend to underestimate SBP and overestimate DBP 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 simultaneous comparison can be performed, but this is not permissible as any modification of the usual operational mode may alter its accuracy.

Other factors that may preclude simultaneous same-arm testing are confusion of noise from the device with Korotkoff sounds, failure of the inflating mechanism to reach the required pressure, sudden deflation before DBP can be confirmed and uneven deflation, making accurate auscultation impossible. The most important objection to simultaneous comparisons is that true simultaneous measurement cannot be achieved with oscillometric devices, which now constitute virtually all automated devices available for blood pressure measurement.

Simultaneous opposite-arm comparisons are not permitted because the blood pressure difference between the arms is a variable rather than a constant factor, and the measurements are not truly simultaneous. To overcome the problems associated with simultaneous measurements in either the same or opposite arms, sequential testing is advocated in this protocol.

Minimizing observer error during validation

The role of the supervisor has been modified from that in the BHS protocol [4] so that he or she observes the result of each paired measurement made by observers 1 and 2, and if either the SBP or DBP values are more than 4 mmHg apart, the supervisor will simply state that the measurement must be taken again, without giving a reason, so that neither observer will be biased when re-taking the blood pressure. In this way, errors will be minimized. Experience has shown, for example, that errors of 10 mmHg can be made by observers simply misreading the mercury column. Another change in the protocol has been to use the mean of the two observers' results rather than analyzing the results for each observer separately, these mean values being referred to simply as 'observer measurements'.

Reduction in the number of subjects recruited

Reducing the number of subjects required for validation would greatly simplify the procedure, and there are now

sufficient data from the many validation studies performed to review the number of subjects required [7–23].

The first AAMI protocol required a sample of 85 subjects, the paired measurements being averaged to give a total of 85 paired comparisons [2]. The BHS protocols [3,4] and the revised AAMI protocol [5] did not average the values, leaving 255 sets of measurements for analysis. In the current protocol, reducing the number of paired measurements to 99 (which allows for easy conversion to equivalent percentage values) brings the sample size back to the original AAMI recommendation but reduces the number of subjects to 33. Reducing the number of subjects results, of course, in some loss in measurement independence, but an analysis of 19 validation studies has shown that reducing the number of subjects recruited from 85 to 33 is possible without affecting the accuracy of the validation (Appendix D) [7–23].

The subjects should be at least 30 years of age in order to ensure those recruited are representative of the age range in which most hypertensive patients lie.

Relaxing the range of blood pressures

Experience has shown that recruiting subjects at the extremes of high and low pressure is impractical. Furthermore, as blood pressure variability is greater at these extremes, sequential comparisons may become unreliable. The relaxation of these requirements to those shown in Table 1 above, with an equal number of subjects being recruited to each range, facilitates the validation procedure without unduly affecting the results.

Eliminating ‘hopeless’ devices

Our data support dividing the validation process into two phases. In the primary phase, three pairs of measurements are performed in 15 subjects in the stipulated pressure ranges, any device failing this phase being eliminated from further testing. Devices passing this phase then proceed to a secondary phase, a further 18 subjects (to provide a total of 33) being recruited, in whom comparisons must fulfil the criteria shown in Table 2. These alterations do not substantially alter the results of the validation studies examined, but by eliminating ‘hopeless’ devices at an early stage, the validation process is simplified and unnecessary testing avoided [7–23].

Expression of validation results

In this protocol, the BHS grading system and AAMI assessment according to the mean and standard deviation of the differences have been abandoned in favour of a straightforward pass/fail system. Moreover, a degree of tolerance in deciding the pass/fail category has been incorporated into the protocol. Ideally 65, 80 and 95 of the 99 measurements should lie within 5, 10 and 15 mmHg respectively, but because a device might only marginally

fail, a tolerance factor whereby one of the above targets is not achieved for five measurements is allowed.

Algorithm integrity and design modification

The first BHS protocol emphasized the importance of manufacturers indicating by a change in model number any modifications made to blood pressure measuring devices [3]. The revised BHS protocol, published in 1993, went further by stipulating not only that manufacturers must indicate clearly all modifications to the technological and software components of automated devices by changing the device number, but also that modified devices must be subjected to renewed validation [4]. These stipulations were influenced by consequences that had resulted from changes made by manufacturers to the algorithms of devices for measuring ambulatory blood pressure [43]. Manufacturers have, however, from time to time expressed the view that the BHS stipulations were unreasonable, in that they obliged the manufacturer to go to the unnecessary expense of re-evaluating a device that had undergone some design modifications without any alteration of the algorithm. Moreover, the stipulation might inhibit beneficial modifications to device design, which need not involve adjusting an algorithm previously shown to have fulfilled the accuracy criteria of the protocol. This stipulation remains in principle in the present protocol but can be waived if a manufacturer of a device that has previously fulfilled the accuracy criteria of the protocol can provide the following: (1) independent evidence that the algorithm in the modified device is identical to that in the originally validated model; (2) evidence that the proposed modifications cannot alter the performance of the algorithm; a system of model numbering that (3) acknowledges a common algorithm and (4) denotes the features of the modification [43].

Intra-subject variability

The influence of intra-subject variability is substantial and can disadvantage devices, particularly when sequential measurements differ by over 10 mmHg, as happens especially in the higher pressure ranges. Two simple measures to cope with this problem have been incorporated into this protocol.

1. *Exclusion of subjects with extremely high and extremely low pressures.* Not only do measurements in these ranges tend to vary considerably, but also large differences, which would be substantial in the mid-range pressures, are in practice unlikely to affect treatment at these extremes.
2. *Tolerance for comparative differences over 15 mmHg.* It must be accepted that sequential measurements may vary quite considerably in some subjects, especially at high pressures, and that these are not errors. An analysis of previous studies has shown that sequential SBP measurements typically lie within 5, 10 and 15 mmHg

of each other 75, 93 and 97% of the time, respectively. The mean difference is typically 1 mmHg, with a standard deviation of around 5 mmHg.

Suitability of the device for individuals

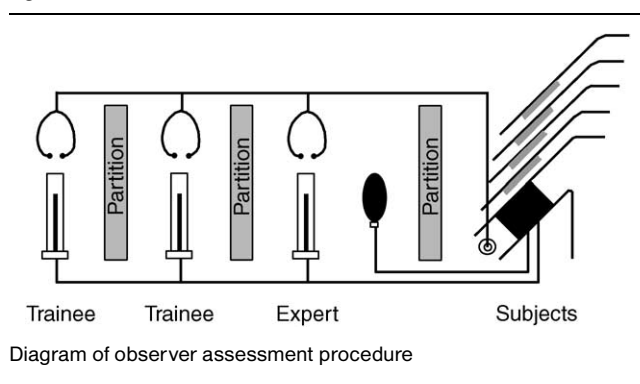
There is a fundamental paradox in the design of previous protocols, which has been identified by an analysis of the Dublin database [personal communication from Gerin W and Pickering T, 2001]. Whereas the procedures in previous protocols were designed to determine whether a given device would, on average, provide valid measurements for a population, there is in practice a need to know whether the device will give accurate measurements for a particular subject. The protocol therefore introduces a tertiary phase whereby the device is assessed according to the number of subjects in whom it gives accurate measurements in addition to its overall accuracy. (Table 2c).

Appendix B. Observer training

Self-assessment

The observers, usually nurses who understand blood pressure measurement, are retrained in blood pressure measurement using a CD-ROM such as that produced by the BHS or the Société Française d'Hypertension Artérielle [30,31]. These demonstrate the technique of blood pressure measurement and permit an assessment period during which trainees can test themselves against a standard mercury sphygmomanometer in which the mercury column falls against a background of recorded Korotkoff sounds. Observers should not move on to the next stage until they have satisfied the test requirements of the CD-ROM. It is helpful for an expert in blood pressure measurement to take trainee observers through the different stages of blood pressure measurement [29]. Difficult aspects of interpretation, such as the auscultatory gap and observer bias, should be discussed and illustrated by example. It is recommended that observers have audiograms to detect any hearing deficit.

Fig. 2



Observer assessment

As an alternative to self-assessment, observers can be tested formally as in the BHS protocol [4].

Trainee 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 50 numbered cards on which to write down assessments (Fig. 2). The rationale for this procedure is that when more than one observer is being trained and assessed, it becomes difficult to prevent an observer who is unsure of a reading gaining sight of a neighbouring observer's reading. It is therefore necessary to separate observers by a series of partitions.

1. The expert observer occupies a similar adjoining booth, the only difference being the presence of a hand bulb to inflate and deflate the cuff on the arm of the subject.
2. Five subjects with a range of blood pressure from about 110/60 to 170/100 mmHg are seated behind a partition. The 'supervisor' places the cuffs in random order on the arms without the expert or trainee observers being aware of the order. When the stethoscope head and cuff are in place, the 'supervisor' gives a verbal cue to the observers and the expert observer operates the cuff and deflates it at a rate of 2 mmHg/s.
3. As the inflatable bladder is connected to each of the columns of mercury in the observer booths, all the columns of mercury fall simultaneously for each of the blinded observers and for the expert, all of whom write down their measurements. Using a series of manometers, time must be allowed for each manometer to deflate fully and the mercury meniscus to return to zero.
4. Ten measurements are made by each observer on each of five subjects, giving a total of 50 measurements for each observer.

The accuracy criteria for the test procedure are the following.

1. Forty-five systolic and diastolic differences between each trainee and between trainees and expert should differ by not more than 5 mmHg, and 48 by not more than 10 mmHg.
2. Failure to achieve this degree of accuracy necessitates a repeat training and assessment session for the failed observer(s).

Audio-visual techniques

Training observers as described above is a labour-intensive procedure, and even when observers are instructed to a high degree of accuracy, there is the problem of maintaining that accuracy throughout the study [32,33].

A need has been recognized, therefore, for an electronic audio visual system to measure blood pressure in validation studies that is not dependent on observers but will nevertheless retain the traditional auscultatory methodology using the mercury sphygmomanometer. An example of such is the Sphygmocorder which was developed for this purpose and, since it was first described in 1995 [34], a number of improvements have been made to the system [35]. This system is being developed for commercial distribution.

Appendix C. Intra-arterial comparison

The ESH Working Group agrees with the stipulations of the previous BHS protocol that intra-arterial comparisons should not be recommended for general validation, while acknowledging that intra-arterial comparisons may in some instances give information that cannot be obtained non-invasively [4]. If, however, intra-arterial comparisons are to be performed, they should be confined to centers with proven expertise in the technique, and the requirements of EN 540, *Clinical investigation of medical devices for human subjects*, which requires among other stipulations that the World Medical Declaration of Helsinki is fulfilled, that the Ethics Committee be provided with information to assess whether the risks to subjects, who cannot be expected to derive any direct therapeutic benefit, can be justified by the collective benefit, that provisions have been made to compensate subjects in the event of injury, and that full informed consent is obtained from all subjects [44].

A comparison between blood pressure measuring systems that utilize indirect measurement and those using the direct intra-arterial measurement of blood pressure is not recommended in this protocol. Apart from ethical considerations, there are several reasons for this. Systolic and diastolic blood pressure values obtained by the direct technique differ from measurements obtained by indirect methods [4,46]. Clinical practice derives from data obtained by the indirect rather than the direct technique. There is considerable beat-to-beat variation in blood pressure, which is not reflected in indirect readings. Blood pressures measured directly and indirectly from the same artery are rarely (if ever) identical. Discrepancies in SBP as great as 24 mmHg and in DBP as much as 16 mmHg have been observed when blood pressure has been measured by both techniques on the same arm at the same time. In addition, these differences are random, displaying no schematic pattern [4,45].

It is, however, recognized that valuable information on device performance may derive from intra-arterial comparisons in certain circumstances, such as validating devices that analyse beat-by-beat blood pressure non-invasively, but the International Protocol would need to be modified procedurally to allow intra-arterial comparisons and to test

device performance in tracking fast beat-by-beat blood pressure changes (46).

Appendix D. Statistical considerations

Sample size

The AAMI published its first protocol for the validation of blood pressure measuring devices in 1987 [2]. The accuracy component of the protocol basically consisted of a comparison of the mean of three test device measurements with simultaneous observer measurements, measuring blood pressure with a mercury sphygmomanometer, on each of eighty-five subjects. The selection of 85 subjects was made on the ability to detect a somewhat arbitrary error of 5 ± 8 mmHg at a significance level of 0.05 and a power of 0.98. The calculation was based on independent, rather than paired, samples for comparison, thereby allowing for the fact that devices and observers may not measure blood pressure on exactly the same heart beat even when using simultaneous readings.

A blood pressure measuring device could pass the AAMI protocol, but still be inaccurate. The BHS protocol identified two difficulties [3,4]. The first was that only average measurements were used in the analysis whereas individual measurements would be identified in practice. The second was that, in using means and standard deviations, the percentage of measurements required to be reasonably accurate, that is lying within 5 mmHg, was insufficient. Paradoxically, few outlying measurements are permitted in the normal model whereas a more relaxed approach may be necessary in practice as variability can be considerable in some subjects and may make a truly accurate reading appear otherwise.

When the first BHS protocol was published in 1990 [3], the requirement to take three simultaneous measurements on each of 85 subjects was therefore retained, but the measurements were no longer averaged, thus giving 255 pairs of measurements for comparison. The accuracy criteria were based on the percentage of measurements lying within 5, 10 and 15 mmHg. Furthermore, the possibility of device-induced bias was highlighted with a recommendation that bracketing sequential measurements be used as an alternative to simultaneous measurement. A grading system was introduced to describe accuracy [3].

In its revised protocol in 1993, the AAMI also recommended that measurements no longer be averaged; it also permitted the sequential technique when simultaneous measurements were not feasible [5]. The 5 ± 8 mmHg accuracy criteria were retained.

It has proved extremely difficult to recruit 85 subjects within the pressure range requirement of the previous protocols; in practice, more than 100 subjects have been

needed to fulfil the pressure range stipulations. A number of factors were considered in reducing sample size.

1. The original statistical criteria were based on 85 measurements [2] whereas later protocols used 255 [3–5].
2. For grading results, percentage values are conceptually the most appropriate.
3. The more practical the study is, the more easily and more often it will be performed.

Having performed 19 studies [7–23], we were able to re-analyze the data from these studies to check the validity of new proposals. Taking all factors into consideration, the most appropriate sample size was 99 measurements, which provides more than the 85 measurement pairs required in the previous BHS and AAMI protocols [4,5] but with a sample size of only 33 rather than 85 subjects. Although there is some loss of measurement independence, the results compare well with independent measurements [7–23].

The selection of 33 subjects is based on two factors. First, each subject has three measurements, each of which is used individually. This gives 99 sets of measurements, which is larger than the 85 sets of measurements accepted as the minimum necessary in the AAMI and BHS protocols [4,5].

In comparing the variance of all 99 differences (total variance) with the 33 differences obtained from the mean differences for each subject (the between-subject variance), the F-test consistently yields a significantly lower variance for the 33 subject mean differences than that for the 99 measurement differences. If between-subject variance were the main cause of total variance, these would not differ significantly. If, on the other hand, a device gave practically the same average error with each subject, the between-subject variation would be close to zero, and the F-test would show a very significant result.

Tests on data from previous experiments yield results of probabilities of the between-subject variance and the total variance being the same as lying between 0.1 and 0.01 for SBP and between 0.2 and 0.02 for DBP. There should therefore be little difference between using single comparisons on 99 subjects and three comparisons on each of 33 subjects.

The use of 99 subjects allows for an even distribution of blood pressures as these can easily be broken into three ranges. It is also close to 100, which allows targets to be considered as being approximate to percentages.

Table 4(a–c) demonstrates how the choice-specific values for 5, 10 and 15 mmHg bands are more flexible and

preferable to a mean and standard deviation method of validation.

Pressure ranges

Prior to the introduction of the 1993 BHS protocol [4], there was no specific recommendation on the range of blood pressure required for validation. As a consequence, these varied greatly from one validation to the next. As most devices fared worse in the high pressure ranges, this reduced the reliability and comparability of results [47].

To redress this problem, specific ranges were introduced. In particular, at least eight subjects in both the hypotensive and severe hypertensive ranges had to be recruited. The reasoning behind the inclusion of the hypotensive range was not only to assess accuracy in subjects with hypotension, but also to give some indication of accuracy for devices measuring ambulatory blood pressure during sleep when the values can fall to low levels. Subjects with severe hypertension were included because such levels are quite common in hypertension clinics.

In practice, however, these two groups have proved extremely difficult to find. The prevalence of persistent hypotension is very low, and whereas severely hypertensive patients were to be found in specialist clinics, the validation study had to be performed before blood pressure-lowering drugs were prescribed, which was often ethically impractical. Furthermore, during the resting laboratory phase of the validation procedure, blood pressures in such subjects tended to fall below the required level required. Next and importantly, blood pressure in these subjects tended to be highly variable, making comparisons unreliable.

Finally, the division of subjects according to blood pressure level did not lead to independent analysis. Indeed, tertile analysis, included in the 1993 BHS protocol [4], was used only as a guide to accuracy, and the final recommendation was based on the overall analysis. The reason for this was that most devices fared poorly in the upper tertile, with greater variability in this range being at least partly responsible [47].

Given these difficulties and the fact that the comparisons in these extremes are diluted in the overall analysis, specific requirements to include them are omitted in the recommendations in this protocol. Although the range of pressures has been reduced, all subjects must fit into a specific category, whereas in the earlier protocols 10% of subjects could lie in any range (Table 1) [4].

Recommendations

The non-parametric recommendation system, shown in Table 2a–c, considers both the subject/measurement and subject accuracy. White-coat hypertension and the morning

Table 4a Percentage of comparisons of devices satisfying Association for the Advancement of Medical Instrumentation criteria that fall within a 5 mmHg error band

Within 5 mmHg		Standard deviation (mmHg)							
		1	2	3	4	5	6	7	8
Mean (mmHg)	0	100.0%	98.6%	90.1%	78.5%	68.0%	59.3%	52.3%	46.6%
	1	100.0%	97.4%	88.3%	77.1%	67.0%	58.7%	51.8%	46.3%
	2	99.8%	93.1%	82.9%	73.0%	64.2%	56.7%	50.5%	45.4%
	3	97.6%	84.0%	74.2%	66.6%	59.8%	53.7%	48.4%	43.8%
	4	84.0%	69.1%	62.8%	58.5%	54.1%	49.7%	45.6%	41.8%
	5	50.0%	50.0%	49.9%	49.3%	47.6%	45.0%	42.2%	39.3%

This table shows the expected percentage of errors of at most 5 mmHg for devices passing with mean absolute differences of 0–5 mmHg and standard deviations of 1–8 mmHg.

Table 4b Percentage of comparisons of devices satisfying Association for the Advancement of Medical Instrumentation criteria that fall within a 10 mmHg error band

Within 10 mmHg		Standard deviation (mmHg)							
		1	2	3	4	5	6	7	8
Mean (mmHg)	0	100.0%	100.0%	99.9%	98.6%	95.1%	90.1%	84.3%	78.5%
	1	100.0%	100.0%	99.8%	98.3%	94.7%	89.6%	83.9%	78.2%
	2	100.0%	100.0%	99.5%	97.4%	93.4%	88.3%	82.7%	77.1%
	3	100.0%	100.0%	98.9%	95.7%	91.2%	86.0%	80.7%	75.4%
	4	100.0%	99.8%	97.6%	93.1%	88.0%	82.9%	77.9%	73.0%
	5	100.0%	99.3%	95.0%	89.2%	83.8%	78.9%	74.4%	70.1%

This table shows the expected percentage of errors of at most 10 mmHg for devices passing with mean absolute differences of 0–5 mmHg and standard deviations of 1–8 mmHg. The gray area indicates the improvement that might be obtained if the standard deviation is related to the mean, which in this instance is set so that the expected number of differences within 10 mmHg will be at least 85%.

Table 4c Percentage of comparisons of devices satisfying Association for the Advancement of Medical Instrumentation criteria that fall within a 15 mmHg error band

Within 15 mmHg		Standard deviation (mmHg)							
		1	2	3	4	5	6	7	8
Mean (mmHg)	0	100%	100%	100%	100%	99.6%	98.6%	96.5%	93.6%
	1	100%	100%	100%	100%	99.6%	98.4%	96.3%	93.4%
	2	100%	100%	100%	99.9%	99.4%	98.1%	95.8%	92.8%
	3	100%	100%	100%	99.8%	99.0%	97.4%	94.9%	91.8%
	4	100%	100%	100%	99.6%	98.5%	96.4%	93.6%	90.4%
	5	100%	100%	99.9%	99.3%	97.6%	95.0%	91.9%	88.5%

This table shows the expected percentage of errors of at most 15 mmHg for devices passing with mean absolute differences of 0 mmHg to 5 mmHg and standard deviations of 1–8 mmHg. It shows that devices with 88.5% of measurements within this range could pass. One of the problems with the AAMI protocol is that, by setting the error independently for mean and standard deviation, it permitted a very liberal level of accuracy. From Table 4a, it can be seen that where a device barely passes with a mean error of 5 mmHg and a standard deviation of 8 mmHg, one could not expect even 40% of measurements to be accurate. Even devices passing more comfortably would have more than half of their expected measurements classified as inaccurate. The acceptable standard deviation must be inversely related to the mean error. In practice, however, this tends not to be the case as standard deviation tends to increase with error. This makes practical parametric passing criteria problematic.

alarm response are just two examples in which single measurements are crucial. It is much easier for devices to pass when only average subject measurements are used, and it would be wrong to assume that devices being recommended for use under such protocols are also accurate for individual measurements.

It must also be recognized that measurements near the extremes of the pressure range are more variable. Decisions should not therefore be based on small differences at these limits, and zones are used to allow for this. The target requirements are based on existing protocols and the evidence from previous validation data.

When comparing a device measurement with its preceding and succeeding observer measurements, the nearer observer measurement is used. This poses a dilemma only if the two observer measurements are equally close except for sign, for example a device measurement of 150 mmHg and observer measurements of 146 and 154 mmHg. One choice would indicate that the device overestimates pressure whereas the other would indicate that it underestimates pressure. The protocol recommends that whichever of the two observer measurements was taken first is selected. This eliminates bias, and it is likely that the overestimating and underestimating selections will balance out over the 99 measurements.

Membership of European Society of Hypertension Working Group on Blood Pressure Monitoring

Roland Asmar, Société Française D'Hypertension Artérielle, Fillale de la Société Française de Cardiologie, 15, rue de Cels-75014, Paris, France.

Lawrie Beilin, Department of Medicine, University of Western Australia, Australia, GPO Box x2213, 35 Victoria Square, Perth WA 600, Australia.

Denis L. Clement, Afdeling Hart-en Vaatziekten, Universitair Ziekenhuis, De Pintelaan 185, B-9000 Gent, Belgium.

Peter De Leeuw, Interne Geneeskunde, Academisch Ziekenhuis, P. Debyelaan 25, postbus 5800, 6202 AZ Maastricht, The Netherlands.

Robert Fagard, Katholieke Universiteit Leuven, Hypertensie en Cardiovasculaire Inevalidatie Eenheid, Inwendige Geneeskunde-Cardiologie, U.Z. Gasthuisberg, Herestraat 49, 3000 Leuven, Belgium.

Yutaka Imai, The Department of Clinical Pathology and Therapeutics, Tohoku University Graduate School of Pharmaceutical Science and Medicine, 1-1 Seiryō-Cho, Aoba-Ku, Sendai 980-8574, Japan.

Jean-Michel Mallion, Médecine Interne et Cardiologie, Chef de Service, Centre Hospitalier Universitaire de Grenoble, B.P. 217 - 38043 Grenoble Cedex, France.

Giuseppe Mancia, Università Degli Studi di Milano-Bicocca, Cattedra di Medicina Interna, Ospedale San Gerardo Dei Tintori, Via Donizetti, 106, 20052 Monza, Italy.

Thomas Mengden, University Clinic Bonn, Department of Internal Medicine, Wilhelmstrasse 35 5311 Bonn, Germany.

Martin G. Myers, Division of Cardiology, Sunnybrook and Women's College Health Sciences Centre, 2075 Bayview Avenue, Toronto, Ontario M4N 3M5, Canada.

Eoin O'Brien (Chairman), Blood Pressure Unit, Beaumont Hospital, Dublin 9, Ireland.

Paul Padfield, Department of Medicine, Western General Hospital, Edinburgh EH4 2XU, UK.

Gianfranco Parati, Istituto Scientifico Ospedale San Luca, IRCCS, Istituto Auxologico Italiano, 20149 Milan, Via Spagnoletto 3, Italy.

Paolo Palatini, Dipartimento di Medicina Clinica e Sperimentale, Università di Padova, Via Giustiniani 2, I-35128 Padua, Italy.

Thomas G. Pickering, Director, Integrative and Behavioral Cardiovascular Health Program, Mount Sinai Medical Centre, New York, NY 10029-6574, USA.

Josep Redon, Hypertension Clinic, Internal Medicine, Hospital Clinico, University of Valencia, Avda Blasco Ibañez, 17. 46010. Valencia, Spain.

Jan Staessen, Katholieke Universiteit Leuven, Hypertensie en Cardiovasculaire Revalidatie Eenheid, Inwendige Geneeskunde-Cardiologie, U.Z. Gasthuisberg, Herestraat 49, 3000 Leuven, Belgium.

George Stergiou, Hypertension Center, Third University Department of Medicine, Sotiria Hospital, Athens, Greece.

Gert van Montfrans, Academisch Medisch Centrum, Interne Ziekten, Meibergdreef 9, AZ 1005 Amsterdam, The Netherlands.

Paolo Verdecchia, Dipartimento di discipline Cardiovascolari, Ospedale R. Silvestrini, Perugia, Italy.

Bernard Waeber (Secretary), Centre Hospitalier Universitaire Vaudois, Division D'Hypertension, Departement de Medecine Interne, 1011 Lausanne, Switzerland.

William White, Section of Hypertension and Vascular Diseases, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06030-3940, USA.

Statistical Assistance:

Neil Alkins, Blood Pressure Unit, Beaumont Hospital, Dublin 9, Ireland.

William Gerin, Mount Sinai School of Medicine, New York, NY, USA.