Validation up-date
Eoin O’Brien

With the increasing marketing of automated and semi-automated devices for the measurement of blood pressure, there is a need for potential purchasers to be able to satisfy themselves that such devices have been evaluated according to agreed criteria. Since their introduction a large number of blood pressure measuring devices have been evaluated according to one or both protocols. However, experience has demonstrated that the conditions demanded by the protocols are extremely difficult to fulfil. The European Society of Hypertension (ESH) protocol, named the International Protocol, which will be published shortly, is based on the data from 19 validation studies performed according to the AAMI and BHS protocols. Critical assessment of this data base of evidence has permitted rationalisation and simplification of validation procedures without loosing the merits of the much more complicated earlier protocols. This has been achieved by elimination of pre-validation phases, improving observer recruitment and training, minimising observer error during validation, reducing the number of subjects recruited, relaxing the range of blood pressures required and eliminating ‘hopeless’ devices early in the validation procedure. *Blood Press Monit* 6:275–280 © 2001 Lippincott Williams & Wilkins.

**Introduction**

With the increasing marketing of automated and semi-automated devices for the measurement of blood pressure, there is a need for potential purchasers 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 and aneroid sphygmomanometers in 1987 [2], which included a protocol for the evaluation of the accuracy of devices, this being followed in 1990 by the protocol of the British Hypertension Society (BHS) [3]; both protocols were revised in 1993 [4,5]. These protocols, which differed in detail, had a common objective, namely the standardisation of validation procedures to establish minimum standards of accuracy and performance, and to facilitate the comparison between one device and another [6].

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

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 [24–27]. The Working Group 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 has 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 were published [1–5], 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 analyse 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 allowed the 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, the proceedings of which have been published in order to invite comment and discussion [24–27].

The International Protocol has been drafted to be applicable to the majority of blood pressure measuring-devices on the market. The validation procedure is therefore confined to adults 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 help manufacturers in submitting blood pressure-measuring devices for validation in order to obtain the minimum approval necessary for them to be used in clinical medicine and that most devices on the market will in time be assessed according to the protocol for basic accuracy. This does not preclude the manufacturers of devices from subjecting their products to more rigorous assessment and validation.

Principal features of the ESH International Protocol
Our approach to simplifying previous validation procedures, so that the International Protocol has wide acceptability, has concentrated on the following areas.

Elimination of pre-validation phases
The main validation procedure of the existing BHS protocol has five phases: (1) before-use device calibration; (2) the in-use (field) phase; (3) after-use device calibration; (4) static device validation; and (5) a report of the evaluation [4]. Phases (1) to (3) were originally introduced to identify intra-device variability, but if a device has fulfilled the general requirements of the European Union directives [9–31] or the AAMI standard [5], it is not necessary to subject these devices to phases (1), (2) or (3) of the BHS protocol. These pre-validation phases are not, therefore, included in the present International Protocol, thereby resulting in a considerable reduction in time and labour.

Improving observer recruitment and training
The most fallible component of blood pressure measurement is the human observer, so consideration must be given to their education and certification. CD-ROMs are currently available to facilitate observer training and assessment [32,33].

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 blood pressures recorded during validation [33,34]. 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 a comparison between the blood pressure as measured by the device being tested and 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 that deflate at a rate of more than 5 mmHg, for example, do not permit accurate measurement by an auscultating observer, leading to an inaccurate comparison between the test and the reference device [4]. At a high deflation rate, an auscultating observer will tend to underestimate systolic and overestimate diastolic blood 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 reading. The device may have a facility for slowing the rate of deflation so that a simultaneous comparison can be performed, but this is not permissible as any modification of the usual operational mode may alter the accuracy.

Other factors that may preclude simultaneous same-arm testing are the confusion of noise from the device with Korotkoff sounds, a failure of the inflating mechanism to reach the required pressure, sudden deflation before the diastolic blood pressure 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 the 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 the International Protocol.

Minimizing observer error during validation
The supervisor’s role 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 systolic or the diastolic blood pressure 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 simply misreading the mercury column. Another change in the protocol has been to use the mean of the two observers’ results rather than analysing the results for each observer separately, these mean values being referred to simply as ‘observer measurements’.

Reduction in 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 an 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 lowering the number of subjects recruited from 85 to 33 is possible without affecting the accuracy of the validation [7–23].

### Table 1
Blood pressure ranges for entry blood pressure

<table>
<thead>
<tr>
<th></th>
<th>SBP</th>
<th>DBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>90–129</td>
<td>40–79</td>
</tr>
<tr>
<td>Medium</td>
<td>130–160</td>
<td>80–100</td>
</tr>
<tr>
<td>High</td>
<td>161–180</td>
<td>101–130</td>
</tr>
</tbody>
</table>

For the primary phase, 5 of the 15 subjects should have a systolic blood pressure (SBP) in each of the ranges. Similarly, 5 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 should have an SBP and DBP in each of the ranges.

### 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, an equal number of subjects being recruited to each range, facilitates the validation procedure without unduly affecting the results [7–23].

### Eliminating ‘hopeless’ devices
Our data support dividing the validation process into two phases: a primary phase in which three pairs of measurements are performed in 15 subjects in the stipulated pressure ranges, any device failing this phase being eliminated from further testing; and a secondary phase (Table 2a and 2b) for those devices passing the primary one, in which a further 18 subjects (giving a total of 33) are recruited, in whom comparisons must fulfil the criteria shown in Table 2b. 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 has been

#### Table 2a  Requirements to pass the primary phase

<table>
<thead>
<tr>
<th></th>
<th>Within 5 mmHg</th>
<th>Within 10 mmHg</th>
<th>Within 15 mmHg</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least one of</td>
<td>25</td>
<td>35</td>
<td>40</td>
</tr>
</tbody>
</table>

After 15 subjects have been tested, the data (45 comparisons) should be analysed to determine the number of comparisons falling within the 5, 10 and 15 mmHg error bands. At least 25 the number of comparisons falling within the 5, 10 and 15 mmHg error bands. At least 25 comparisons must be within 5 mmHg or at least 35 comparisons within 10 mmHg or at least 40 comparisons within 15 mmHg. If none of these counts reaches the criteria in the table, the device is deemed to have failed.

#### Table 2b  Requirements to pass the secondary phase. Part 1

<table>
<thead>
<tr>
<th></th>
<th>Within 5 mmHg</th>
<th>Within 10 mmHg</th>
<th>Within 15 mmHg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two of</td>
<td>65</td>
<td>80</td>
<td>95</td>
</tr>
<tr>
<td>All of</td>
<td>60</td>
<td>75</td>
<td>90</td>
</tr>
</tbody>
</table>

After completing all 33 subjects, the data (99 comparisons) should be analysed to determine the number of comparisons falling within the 5, 10 and 15 mmHg error bands. For the device to pass the criteria, there must be a minimum of 60, 75 and 90 comparisons within 5, 10 and 15 mmHg, respectively. Furthermore, there must be a minimum of 65 comparisons within 5 mmHg and 80 within 10 mmHg, or 85 comparisons within 5 mmHg and 95 within 15 mmHg. If none of these counts reaches the criteria in the table, the device is deemed to have failed.

#### Table 2c  Requirements to pass the secondary phase. Part 2

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>At least 2 out of 3 within 5 mmHg</td>
<td>0 out of 3 within 5 mmHg</td>
<td></td>
</tr>
<tr>
<td>At most 22</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

The data should now be analysed by subject to determine the number of comparisons per subject that fall within 5 mmHg. At least 22 of the 33 subjects must have at least two of their three comparisons within 5 mmHg (including those who have all three comparisons lying within 5 mmHg). At most, 3 of the 33 subjects can have all three of their comparisons over 5 mmHg apart.
shown to be 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 categories has been incorporated into the protocol. Ideally, 65, 80 and 95 of the 99 measurements should fall within 5, 10 and 15 mmHg respectively, but because a device might fail only marginally, 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 had to indicate clearly all modifications to the technological and software components of automated devices by changing the device number, but in addition that modified devices had to 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 [35].

Manufacturers have, however, from time to time expressed the view that the BHS stipulations are unreasonable in that they oblige the manufacturer to go to the unnecessary expense of re-evaluating a device that has undergone some design modifications without any alteration of the algorithm. Moreover, the stipulation may inhibit beneficial design modifications that need not involve adjusting an algorithm previously shown to have fulfilled the accuracy criteria of the protocol. This stipulation remains in the present protocol in principle 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, and (3) a system of model numbering that both acknowledges a common algorithm and (4) denotes the features of the modification [35].

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 especially happens in the higher ranges of pressure. Two simple measures to cope with this problem have been incorporated into this protocol:

1. The 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 of pressure, 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 systolic blood pressure measurements usually 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 that has been identified by an analysis of the Dublin database. Whereas the procedures in previous protocols were designed to determine whether a given device would, on average, give valid measurements for a population, there is in practice a need to know whether the device will provide accurate measurements for a particular individual. 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).

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 centres with proven expertise in the technique and meet 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 be fulfilled, that the relevant 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 be made to compensate subjects in the event of injury and
that full informed consent be obtained from all subjects [36].

A comparison between blood pressure-measuring systems that utilise indirect measurement and the direct intra-arterial measurement of blood pressure is, for several reasons, not recommended in this protocol. Systolic and diastolic blood pressure values obtained using the direct technique are different from those obtained by indirect methods [4,37]. In addition, clinical practice derives from data obtained by the indirect rather than the direct method. Importantly, ethical considerations preclude its use for device validation in healthy subjects [4,37]. There is also 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 systolic blood pressure as great as 24 mmHg for systolic and 16 mmHg for diastolic pressure have been observed when the blood pressure has been measured by both techniques on the same arm at the same time. Furthermore, these differences are random [4,37].

It is, however, recognized that valuable information on device performance may derive from intra-arterial comparisons in certain circumstances, and that as such circumstances will dictate the study design, it is not appropriate for this protocol to make stipulations other than to emphasize the importance of acknowledging any ethical issues that may arise.

Main differences between the International Protocol and other protocols

The International Protocol has been drafted to fulfil an urgent demand, namely to provide a validation procedure that is simple, is relatively easy to perform in a reasonable period of time and adheres to the accuracy criteria of previous protocols, so that the ever-increasing number of devices being marketed can be independently assessed for their basic accuracy. The protocol does not attempt to provide validation procedures for special devices with innovative features such as waveform analysis, nor does it provide specific procedures for validation in specific populations, for example infants, pregnant women, children and the elderly, or in special circumstances, such as during atrial fibrillation or exercise. Instead, the International Protocol provides basic procedures for validating blood pressure-measuring devices in adults, who constitute the majority of hypertensive patients; despite this, it could also form the basis of validation procedures in special groups and in special circumstances. There would, however, be no point performing such validations unless the device had first passed the International Protocol criteria. The main differences between the International Protocol and existing or proposed protocols are summarized in Table 3.

### References


