European Society of Hypertension International Protocol for the validation of blood pressure monitors: a critical review of its application and rationale for revision
George S. Stergiou\textsuperscript{a}, Nikos Karpettas\textsuperscript{a}, Neil Atkins\textsuperscript{b} and Eoin O’Brien\textsuperscript{c}

**Objective** To perform a systematic review of validation studies of blood pressure measuring devices done using the European Society of Hypertension International Protocol (ESH-IP) since its publication in 2002.

**Methods** Major aspects of ESH-IP studies are described. A review of the ESH-IP performance, problems and violations in its application, and the effect of applying several more stringent validation criteria in an ESH-IP revision is carried out.

**Results** From January 2002 to June 2009, 104 validation studies had been conducted using the ESH-IP, 36 using the British Hypertension Society protocol and 28 using the US Association for the Advancement of Medical Instrumentation protocol. Among 78 studies reported up to June 2008, in 66 (85%) the tested device has passed the ESH-IP. In 19 validations a modification of the ESH-IP was performed to adapt for specific study needs (population or device). Protocol violations of the ESH-IP were identified in 23 studies (eight major violations). A test of several arbitrarily chosen changes in the ESH-IP validation criteria applied in the published studies showed the phase 2.1 criterion (BP differences ≤15 mmHg) and the phase 2.2 criteria to be the more stringent.

**Conclusion** The ESH-IP has succeeded in expanding the validation procedure worldwide by three to four-fold compared with the period before its publication. There is a need for protocol revision aiming to address issues that appeared in published studies, prevent protocol violations, and ensure complete data reporting. Standardization of the ESH-IP validation studies’ report and application of more stringent criteria should be considered.


Keywords: blood pressure, blood pressure measurement, device validation, European Society of Hypertension, international protocol, protocol

\textsuperscript{a}Hypertension Center, Third University Department of Medicine, Sotiria Hospital, Athens, Greece, \textsuperscript{b}Neil Atkins, dabl \textsuperscript{c}Educational Ltd. Blackrock, Co. Dublin and \textsuperscript{d}Eoin O’Brien, Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Ireland

Correspondence to Dr George S. Stergiou, MD, Hypertension Center, Third University Department of Medicine, Sotiria Hospital, 152 Mesogion Avenue, Athens 11527, Greece
Tel: +30 210 7763117; fax: +30 210 7719981; e-mail: gstergi@med.uoa.gr

Received 24 July 2009 Revised 23 August 2009 Accepted 24 August 2009

**Introduction**

The accuracy of the devices for blood pressure (BP) measurement is an important prerequisite for the reliable assessment of the level of BP and thereby for hypertension diagnosis, treatment initiation and titration, and long-term follow-up [1]. In 1987 the US Association for the Advancement of Medical Instrumentation (AAMI) published the first protocol for formal validation of all BP monitors [2]. This was followed in 1990 by the British Hypertension Society (BHS) protocol [3] and revised versions of these protocols were published in 1993 [4,5].

In February 2002, the European Society of Hypertension (ESH) Working Group on Blood Pressure Monitoring published the International Protocol (ESH-IP) for the validation of BP monitors [6]. The ESH-IP was developed on the basis of the analysis and experience of the data from 19 validation studies performed using the AAMI and BHS protocols at the Blood Pressure Unit in Dublin, Ireland [7]. The rationale for developing the ESH-IP was the simplification of the validation procedure, mainly by reducing the sample size required and by relaxing the BP range for inclusion [6]. In the rapidly expanding market of BP monitors for home, ambulatory and office measurements, it was anticipated that such a simplified protocol would facilitate greater use of the validation procedure by more centers throughout the world.

O’Brien and Atkins [7] performed the first review of 26 validation studies conducted using the ESH-IP until 2006 and made several comments and proposals to deal with issues raised when using the protocol in practice and areas that might need change in a future revision. It is 7 years since the publication of the ESH-IP for a systematic review of the use of the protocol by investigators in practice, the main results, the performance in complying with the protocol requirements and potential problems in its implementation. This analysis together with the earlier one [7] has provided evidence for improvement in the ESH-IP, which has undergone extensive revision.
Objective
To perform a systematic review of the use of the ESH-IP for validating BP measuring devices in terms of (i) number of validation studies performed, (ii) main results, (iii) performance in following the protocol’s requirements and criteria, (iv) problems in data reporting, (v) issues within the protocol that might need modification or clarification, and (vi) impact of applying more stringent criteria.

Methods
We carried out a systematic literature search for articles, in full paper or abstract form, reporting the results of validation studies of BP measuring devices using the ESH-IP since its publication in February 2002 until the end of June 2008. Sources were PubMed, abstract books of main meetings of the European, the International, and the American Society of Hypertension from 2003 to 2008 and the www.dableducational.org website. Keywords used were: blood pressure measurement, blood pressure, devices, validation, European Society of Hypertension, International Protocol, ESH-IP. When both a full paper and an abstract were found from the same validation study (same authors and device), the data from the full paper were used.

Results
General study features
A total of 78 [8–66] validation studies were found, of which 69 have been published as full papers and nine as abstracts only [44,45,51–53,58,59,64].

Publication
The vast majority of the full papers (43 papers reporting 58 validation studies) have been published in Blood Pressure Monitoring. The Journal of Human Hypertension has published four papers [46,49,65,66] and four other journals published one paper each [60–63], one of which reported the validation analysis of four devices [63].

Countries
The vast majority of studies (54 studies, 69%) were conducted in Europe. Studies have been performed in 18 countries – France (17), Italy (11), Turkey (11), Greece (eight), UK (seven), China (five), USA (three), Germany (three), Argentina, Bulgaria and Ireland (two each), and The Netherlands, Portugal, Russia, Singapore, Spain, Switzerland, and Venezuela (one each).

Investigators
A total of 26 different research groups performed the ESH-IP studies.

Manufacturers
Sixteen Omron devices (Omron Healthcare Europe, Hoofddorp, The Netherlands), were subjected to 24 validations, which were published in 17 papers [10,17,18,25–30, 38,39,44,45,47,53,54,63]; six Microlife devices (Microlife, Widnau, Switzerland) in eight validations published in seven papers [9,11,23,24,37,50,63]; three A&D devices (A&D Company, Ltd., Toshima Ku, Tokyo, Japan) in four validations published in three papers [8,19,20]; three Braun devices (Braun GmbH, Kronberg, Germany) in three validations in three papers [35,36,48] and 36 devices of 28 other manufacturers in 39 validations (one-to-two devices and studies per manufacturer) [12–16,21–23, 31–34,37,44,46,48,51–57,59–63,65].

Funding
Fourteen validation studies (18%) were funded by manufacturers [9,11,16,17,22,30,32,33,41,50,59,66], 15 by other resources [8,10,15,24,26,34,40,42,43,55,57] and in the remaining 49 (63%) funding was not declared. Among the full papers published, a declaration that the manufacturer had funded the study was made in 12 (24%) [9,11,15–17,22,32,33,41,49,50,66], a declaration of other sources of funding in 11 [8,10,12,15,24,34,40,42, 43,55,57] and the source of funding was not stated in the remaining 38 (75%) papers. In one study [66], industry employees were included as authors of the paper.

Arm versus wrist
Sixty-two studies (79%) validated devices that measure BP at the upper arm, 15 at the wrist [18,22,28,35–39, 45,48,54,58,66], and in one the type of device was not specified [51].

Home, office, and ambulatory devices
Fifty validations (64%) have been performed on devices designed for self-home monitoring, 16 for professional use in the office or clinic [8,12–18,50,62,64], 11 devices for ambulatory monitoring [40–43,51,59,61,65,66] and in one study the type of the device was not specified [53].

Oscillometric versus auscultatory
The vast majority of the studies (n=62, 79%) validated oscillometric devices. Auscultatory devices have been validated in nine studies [8,12–14,43,50,59], two of which estimated BP using a microphone [43,59] and in seven the device displayed the pressure of the deflating bladder and an observer measured BP using a stethoscope [8, 12–14,50]. One validation was performed on a device having auscultatory and oscillometric working simultaneously, with the latter methodology being the back-up for the former [55]. Another validation study evaluated a tonometric device [66]. Five validations did not specify the measurement method of the tested device [46,51,53, 61,64], of which only three were abstracts.

European Society of Hypertension International Protocol use compared with other protocols
The use of the ESH-IP was compared with that of the BHS and AAMI protocols by counting the validation studies conducted per year using each protocol starting from 2000 (2 years before the ESH-IP publication) until
June 2009 (the data analysis is based on studies to June 2008). The results are presented in Table 1 and Fig. 1.

Study conclusion

On the basis of the ESH-IP criteria, 66 of the 78 validations (84.6%) passed and 12 (15.4%) failed [8,15,17,32,37,42,54,59,62]. In four studies [17,59,62] (two reported only in abstract form [59]) it was concluded that the devices had passed the ESH-IP, the presented data suggested that the protocol criteria were not fulfilled. Twelve devices were validated using the ESH-IP in more than one study [8,10,11,17,18,23–25,27,28,38,39,42,43,47,50,59,63,65]. Nine devices were validated twice [8,10,11,17,18,25,42,43,50,59,63,65], one device three times [38,39], and one device four times [27,28,47]. Three devices that provide two different measurement modes (oscillometric and auscultatory) were assessed twice within the same study (once for each measurement mode) and were found to be accurate with both methods [43,50,59]. One nonmercury auscultatory device that has a button to mark the systolic and diastolic BP values was assessed twice (with and without using the mark button) in the same study and was found to fulfill the ESH-IP criteria only when the mark button was not used [8]. Three devices have been validated in adults twice by different investigators [11,27,28,42,63,65]. Two of these devices (Omron M6 or HEM-7001-E [27,28] and Microlife BP A100 Plus [11,63]) passed the ESH-IP in both validations, whereas the third device (TONOPORT V ambulatory monitor) failed the ESH-IP (in systolic BP measurement) in the first study [42], but passed in a subsequent study again in adults [65]. The authors of the second paper [65] speculated that characteristics of participants and external factors (differences in the experimental conditions that affect the stability of the sequential BP measurements) probably played a major role in the performance of the devices under investigation. Finally, six devices have been tested using the ESH-IP in different populations (e.g., children, elderly, obese, end-stage renal disease) and five of them passed all the validations [10,17,18,23–25,27,28,38,39,47]. The Omron 705IT was successfully validated in adults and in children [10,25], the Omron M5-I (HEM 757-E) in adults and in the elderly [18,25], the Omron M6 (HEM-7001-E) in adults, in obese and in the elderly [27,28,47], the Omron 637IT in adults, in obese and in the elderly [38,39], and the Microlife 3AC1-1 in adults and in those with end-stage renal disease [23,24]. However, the Omron HEM 907 passed in the elderly but was regarded as questionable in the adults [17,18].

Agreement of ESH-IP with AAMI protocol

The final result of the ESH-IP (pass/fail) was assessed by applying the mean ± SD BP difference criterion (5 ± 8 mmHg) of the AAMI protocol [2,4]. One device that passed the ESH-IP had mean systolic BP difference of more than 5 mmHg [31] and two other devices had a SD greater than 8 again for systolic BP [35,63]. None of the devices that passed the ESH-IP had a mean diastolic BP difference of more than 5 or SD greater than 8 mmHg. Four of the 10 devices that reported mean difference and SD and failed according to the ESH-IP criteria, had mean BP difference of less than 5 mmHg and SD of less than 8, for systolic or diastolic BP or both [8,15,59,62].

Recruitment and sample size

In 25 of the 69 studies published in a full paper (36.2%), only the number of subjects analyzed (in most cases 33) was given but the number of patients recruited and excluded was not reported, or this number was stated as being higher than 33 [66].
Use of Sphygmocorder
Only two validation studies (2.6%) [32,42], both conducted in a single center in Ireland, used the Sphygmocorder, whereas all the other investigators used live measurements by two observers.

Modifications of the ESH-IP by investigators
A protocol modification was defined as any intended deviation in the implementation of the protocol performed in a validation study aiming to adapt the protocol to specific study needs (specific study population or unusual device function), provided that the validity of the protocol requirements and the stringency of the criteria were not affected. The following modifications were noted:

(1) Recruitment criteria were modified in 13 validation studies aiming to include specific populations: four studies in the elderly [18,39,47], four in children and adolescents [10,15], two in obese participants [27,38], one in participants with arm circumference 32–42 cm [29], and one in patients with end-stage renal disease [24].

(2) The BP range and the study sample size were altered in one paper that reported the validation of three devices in patients aged 5–15 years (44 patients per device) [15] and, in another, in 197 participants aged 6–16 years [10], so as to allow for the high heterogeneity of this population in terms of body size and BP levels.

(3) To validate the A&D UM-101 professional auscultatory hybrid device [8], which has a vertical liquid-crystal display (LCD) resembling a mercury column and a button for the observer to mark the BP readings on the LCD display during cuff deflation, the tested device used by the supervisor (without using the mark button) was connected with a second device of the same type (Y tube), which was simultaneously used by one of the observers (in random order) always using the mark button so as to assess the effect of the mark button on BP measurement. The device comfortably passed the ESH-IP when the mark button was not used and failed when it was used.

(4) To validate the HealthSTATS BPro tonometric device, the investigators used an oscillometric device to calibrate the tested device on each study participant [66]. As this is a device for ambulatory BP measurement the device was tested in supine and standing position after validation in the sitting position.

Violations of the European Society of Hypertension International Protocol
An ESH-IP violation was defined as any deviation in the implementation of the protocol performed in a validation study, which was not performed to deal with a specific peculiarity of the study population or the tested device. Major violations are defined as those that affect the protocol integrity (requirements and stringency of criteria) and minor violations are those regarded as having only negligible impact.

A total of 21 types of violations of the ESH-IP were detected, appearing 33 times and involving 23 studies. Five types of violations (20%) were regarded as major and were found in eight studies [17,26,54,59,62,65,66] (one having two major violations [62]) of which six were published as full papers [17,26,54,62,65,66] and two only as abstracts [59]. In nine studies, more than one violation was found [10,14,16,31,33,54,62,65,66].

**Major violations**

(1) In four studies [17,59,62] (two reported only as an abstract [59]), the investigators stated that the devices had passed the ESH-IP, yet the presented data suggested that the ESH-IP criteria were not fulfilled. In all of these four studies, the devices failed at phase 2.1 of the ESH-IP and the results reported suggest that the investigators probably misunderstood the requirements of this criterion that need both ‘All of...’ and ‘Two of...’ to be fulfilled [6].

(2) Only one observer and an electronic device on the opposite arm used as a second observer (one study) [54].

(3) Consecutive instead of simultaneous observers’ BP measurements (two studies) [26,65].

(4) Results of phase 1 and 2.1 not presented (one study) [66].

(5) Results of phase 1 and 2.2 not presented (one study) [62].

**Minor violations**

(1) Recruitment
(a) Cuff size selected according to wrist circumference (one study) [31].

(b) During phase 2, 33 participants were randomly chosen from a population of 66 participants (one study) [19].

(c) One participant aged less than 30 years included (one study) [16]; one participant with an entry BP higher than 180 mmHg included (one study) [48]; four participants less than 30 years; and three with BP higher than 180 mmHg included (one study) [14].

(d) Fewer-than-required participants in the high BP range (one study) [24].

(2) Validation procedure
(a) In four studies, there was no supervisor and inter-observer BP differences were checked at the end of each participant’s study [10,16,33,65]. The tested device readings were retrieved from the device’s memory and participants were excluded if any inter-observer difference was more than 4 mmHg.

(b) An observer served as supervisor after completing BP measurements and additional measurements were taken if inter-observer difference was more...
than 4 mmHg after retrieving BP readings from the device’s memory [10,33]. In two studies, the observer acted as supervisor and additional measurements were taken if inter-observer BP differences (checked after completing all required measurements) were more than 4 mmHg (tested device readings retrieved from the device’s memory) [10,33].

(c) In three studies, the device failed in phase 1 and yet continued to be used in phase 2 [32,37,54].

(3) Analysis
(a) Different mean observer–device BP differences and SD reported in the paper and the abstract of the paper (abstract’s values better fit with the scatterplot) (one study) [31].

(4) Scatterplot
(a) Scatterplots not presented (two full papers) [46,66].
(b) Too many dots (readings) in the scatterplots (one study) [34].
(c) Same x-axis for systolic and diastolic BP, without BP range lines (one study) [30].
(d) Systolic and diastolic BP presented in the same plot (two studies) [33,55].

Comparison of pass/fail in different ESH-IP phases
Phase 1 versus phase 2
All three devices that failed in phase 1 also failed in phase 2 [32,37,54]. Six devices that passed the ESH-IP phase 1, failed in phase 2 [32,37,54,59,62]. Three studies, one in a full paper [62] and two reported in one abstract [59], did not report results for phase 1.

Phase 2.1 ‘All of’ versus ‘Two of’
Eight devices failed at both the ESH-IP phase 2.1 ‘Two of’ and ‘All of’ criteria. One device that passed the ‘All of’ criterion failed in the ‘Two of’ criterion and none did the reverse.

Phase 2.1 versus phase 2.2
Seven devices failed in the ESH-IP phase 2.1 and all of them also failed in phase 2.2. Of the devices that passed phase 2.1, only 1 failed in phase 2.2. Four validations, two in full papers [17,62] and two reported in one abstracts [59] had no results of phase 2.2.

Impact of applying several more stringent ESH-IP validation criteria
The high rate of successful ESH-IP validation studies indicates that technology has advanced and that it is now timely to introduce more stringent pass/fail criteria. To obtain some indication about which of the ESH-IP validation criteria were easily passed by the currently available accurate devices and which were only marginally passed, we tested several ‘arbitrarily chosen’ changes in all the validation criteria. The impact of applying these arbitrary criteria on the evaluation of devices that had passed the ESH-IP in published validation studies (n=66) was investigated. For the ESH-IP 2.1 phase the ‘All of’ thresholds of 60, 75, and 90 readings were arbitrarily set to 65, 80, and 95 readings. For the ESH-IP phase 2.2 the ‘2/3 within 5 mmHg’ threshold of 22 participants was arbitrarily set to 24 and the ‘0/3 within 5 mmHg’ threshold of three participants to 2. The number of devices that had passed the ESH-IP in published studies, but would fail if one or more tighter criteria at phase 2.1 and/or 2.2 were applied, are presented in Table 2.

It is clear that the most stringent of the tested criteria are those of phase 2.2 (Table 2, upper part: one tighter criterion applied). Although small changes in the phase 2.2 criteria have been tested (24 instead of 22 participants ‘with 2/3 BP differences ≤ 5 mmHg’ and two instead of three participants with ‘none of the differences of triplicate readings ≤ 5 mmHg’), these modifications appeared to have a big impact on the pass/fail ratio with 25–36% of the devices failing solely on the basis of one of these criteria. Regarding phase 2.1, the tightest criterion seemed to be that for readings with a difference of less than 15 mmHg (95 instead of 90). Interestingly, eight of the nine devices that failed this criterion also failed with the tighter criteria of phase 2. It should be noted that phase 2.2 deals with device accuracy in ‘individuals’ and the greater than 15 mmHg criterion of phase 2.1 with ‘very inaccurate BP readings’. It might be argued that these criteria (phase 2.1 ≤ 15 mmHg and phase 2.2) reflect the same phenomenon, which is the inability of the oscillometric technique to give an accurate measurement in some participants because of factors that remain unclear [69]. In contrast, the tighter phase 2.1 criterion for readings within 5 mmHg (65 readings instead of 60) had a small impact on the pass/fail ratio (five devices fail), and the less than 10 mmHg criterion (80 readings instead of 75) had absolutely no impact (no device fails) (Table 2). These criteria assess the performance of the measurement algorithm in participants in whom the device is relatively accurate. Thus, it seems that, in this respect, oscillometric technology has improved.

These data suggest that if a revision of the ESH-IP is to introduce more stringent criteria, then the phase 2.1 less than 5 mmHg and 10 mmHg readings thresholds appears to be the most appropriate modifications that will nearly double the proportion of devices that fail the ESH-IP yet about 75% will still pass. In contrast, any changes in the phase 2.1 less than 15 mmHg criterion or the phase 2.2 criteria should probably be modest because they can considerably increase the fail rate.

Discussion
Since 2002 when the ESH-IP was published, there has been an impressive increase in the publication of validation studies conducted using this protocol. At the
same time, the use of the BHS and the AAMI protocols has remained rather static (at the same levels as in the last 2 years before the ESH-IP publication, see Table 1), yet at a low rate with a maximum of six-to-seven validations reported using each protocol per year compared with at least 15 validations per year since 2006 using the ESH-IP (Table 1, Fig. 1). From January 2007 to June 2009, 29 studies have been published using the BHS or the AAMI protocols compared with 67 using the ESH-IP (Table 1, Fig. 1). It is clear that the ESH-IP now is the preferred validation protocol and that this area of research has greatly expanded since its publication.

The vast majority of ESH-IP validation studies (69%) have been conducted in Europe. It is reassuring that at least some studies have also been performed in the USA (n=3), China (n=5), and elsewhere (n=16). What is more important is that 26 different research groups have chosen to use this protocol and that devices of 32 different manufacturers have been subjected to ESH-IP validation, suggesting that the main objective of the development of this protocol, which was to expand the validation procedure to more devices and more researchers, has succeeded.

The large number of studies conducted using the ESH-IP allows a thorough and reliable evaluation of its performance in practice, as well as the detection of areas within the validation procedure and requirements that are in need of revision.

It is important to note that the proportion of the reported validation studies that fulfilled the ESH-IP criteria is impressively high (84.6%). Possible reasons to be considered are that advancement in technology has improved the accuracy of devices, that there might be a publication bias, whereby negative studies are not published, and finally that the ESH-IP criteria are too loose and need to be made more stringent.

<table>
<thead>
<tr>
<th>Number of tighter criteria applied</th>
<th>Change from 60 to 65</th>
<th>Change from 75 to 80</th>
<th>Change from 90 to 95</th>
<th>Change from 22 to 24</th>
<th>Change from 3 to 2</th>
<th>Devices that fail (out of 66 studies that passed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 2  Application of one or more tighter criteria for phases 2.1 and/or 2.2 of the ESH-IP and their impact on published successful validation studies

<table>
<thead>
<tr>
<th>Number of tighter criteria applied</th>
<th>Change from 60 to 65</th>
<th>Change from 75 to 80</th>
<th>Change from 90 to 95</th>
<th>Change from 22 to 24</th>
<th>Change from 3 to 2</th>
<th>Devices that fail (out of 66 studies that passed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>31</td>
</tr>
</tbody>
</table>

ESH-IP, European Society of Hypertension International Protocol.
Interestingly, one single journal, *Blood Pressure Monitoring*, has published the vast majority of papers reporting ESH-IP studies (84% of the validation studies reported in full papers). Despite this experience in a single journal, its peer review process did not prevent incomplete or inadequate reporting, and misreporting of important information was particularly common. Thus, standardization of the validation study report in a future revision of the ESH-IP seems to be necessary.

There are several important issues raised in this review that deserve special attention and should be considered in a revision of the ESH-IP.

**Reporting of recruited participants**

In many studies, the number of recruited, excluded and analyzed participants has not been reported but only that 33 participants were included. In practice, it is very unlikely that only 33 participants would be required. Even if indeed the first 33 are suitable to be distributed to 11 per BP range, it is not possible to have different participants for systolic and diastolic BP as stated in six studies [12,32,41,42,54,65] of which three [32,42,65] specifically state that they recruited only 33 participants.

**Recruitment on the basis of entry systolic or diastolic BP only**

It is unclear in the ESH-IP whether it is allowed to use in the validation analysis only systolic or diastolic pressures of some participants. Most of the studies analyzed 33 participants for both systolic and diastolic BP whereas at least six studies used in the analysis only systolic or only diastolic BP of some participants. A revision of the ESH-IP should better restrict analysis in only 33 participants who fulfill both the systolic and diastolic BP entry criteria. First, this approach seems to be feasible given that most published studies succeeded in finding 33 participants for both systolic and diastolic BP after recruiting about 40–45 participants. It is unclear in the ESH-IP whether it is allowed to use in the validation analysis only systolic or diastolic pressures of some participants. Most of the studies analyzed 33 participants for both systolic and diastolic BP whereas at least six studies used in the analysis only systolic or only diastolic BP of some participants. A revision of the ESH-IP should better restrict analysis in only 33 participants who fulfill both the systolic and diastolic BP entry criteria. First, this approach seems to be feasible given that most published studies succeeded in finding 33 participants for both systolic and diastolic BP after recruiting about 40–45 participants. Second, it is possible that other studies have been terminated after failure at phase 1, but never been published (publication bias). The result of the validation procedure is very important for physicians, participants and manufacturers and therefore deserves detailed evaluation irrespective of the final conclusion (pass or fail). In the revised ESH-IP, phase 1 could therefore be omitted.

**Phase 1 of the validation study**

The analysis of the currently reported validation studies provided limited information on the usefulness of phase 1. Among the reported validation studies, only three failed in phase 1 and they all continued to phase 2 in which they failed again [32,37,54]. Of course, it is possible that other studies have been terminated after failure at phase 1, but never been published (publication bias). The result of the validation procedure is very important for physicians, participants and manufacturers and therefore deserves detailed evaluation irrespective of the final conclusion (pass or fail). In the revised ESH-IP, phase 1 could therefore be omitted.

**Cuffs for reference and tested device measurements**

The cuff is an important part of BP monitors. Sixty of the validation studies (78.2%) did not report in detail the cuffs used, particularly for the observer measurements. A more standardized report of this information is needed.

**Extra features of the devices and ESH-IP modifications**

The familiarization phase seems to be important not only for the investigators to get used to the tested device but also to identify peculiarities of the device that might affect the validation procedure or the BP measurement. This information might guide investigators on how to apply the ESH-IP or design additional assessment in individual cases. It is important to realize that some novel features of the BP monitors can affect the BP measurement result. These features deserve special attention and validation, further to that of the straightforward BP measurement.

A typical example is the A&D UM-101 professional auscultatory hybrid device, which has a mark button for the observer to record the BP readings on the LCD...
display during cuff deflation [8]. This button aimed to prevent observer bias and terminal digit preference. However, this professional device was available on the market without the validity of the mark button being tested earlier. Unfortunately, a recent study has shown that the device comfortably passed the ESH-IP when the mark button was not used and failed when it was used [8]. This was attributed to the reaction time needed for the observer to press the button, which seemed to have a significant impact on device accuracy and resulted in systematic underestimation of BP. This problem was suspected in the familiarization phase and a protocol modification was made specifically to test the mark button.

Another example is the Microlife average mode (MAM) technology, which is implemented in several Microlife home BP monitors and gives a ‘weighted’ average of three consecutive measurements. However, one study in 152 hypertensives showed that in most cases usual average BP is comparable with that obtained by the MAM approach [68]. A third example is the Visomat Comfort 20/40 home monitor the manufacturer of which claims that a single cuff can be used in a wide range of arm circumference (23–43 cm) [49]. A typical ESH-IP validation showed the device to be accurate even in 20 of the 33 participants with large arms that required a larger cuff [49]. However, a 33-participant validation study cannot fully investigate the accuracy of this device that probably requires an adequate number of participants with small, medium and large arm size. The examples of the A&D UM-101 and the MAM technology clearly show that any novel feature of BP monitors, which has the potential to affect the BP measurement result, should be carefully assessed using appropriately modified protocols.

Another issue is that some novel electronic devices [43,50,55] have two methods for BP measurement (oscillometric and auscultatory). In these cases not only the oscillometric [55] but also the auscultatory BP measurement requires separate validation [43,50].

**Wrist and ambulatory devices**
A problem with the validation of ambulatory monitors is that these are validated only in static conditions whereas they are used in ambulatory conditions. A small 24-h monitoring study, e.g. in 20 participants, might be useful by reporting the proportion of erroneous readings in routine daily conditions. Regarding the wrist devices, some have been shown to be accurate in the strictly standardized conditions of the validation setting [18,22,28,35,36,38,39,45,48,58,66]. However, one study suggested that additional clinical validation at home might not confirm the accuracy of these devices [68].

**ESH-IP grading system for device accuracy**
In the earlier review of ESH-IP validation studies it has been suggested that a grading system might be considered in a revision of the protocol. As the sample size required for an ESH-IP validation has been considerably reduced it is questionable whether any grading system will have acceptable reproducibility. In particular, if phase 2.2, which requires 0–3 participants to pass, is taken into account in the grading system, then one or two individuals might change the device grade, which can be attributed to the chance of recruitment. Furthermore, although the manufacturers would probably favor a grading system, for the consumer this might not be very helpful and a pass/fail system would seem to be more meaningful.

**Publication bias**
It is quite possible that several validation studies with negative results were not published. The revised ESH-IP should encourage investigators to include in their initial agreement with the sponsor that the full report of the validation analysis will be published irrespective of the result (pass or fail). This is the currently accepted policy for all sponsored clinical trials and should be also applied in validation studies. This might be best achieved by registering validation studies on the dabl educational website (www.dableducational.org).

In conclusion, this systematic review has shown that, 7 years after its publication, the ESH-IP has proven to be successful in achieving its aims. The large number of studies published, devices tested, and investigators involved suggest that this protocol has succeeded in expanding the validation procedure worldwide by three- to four-fold compared with the period before its publication. The analysis of 78 reported studies suggests that there is a need for standardization of the ESH-IP validation studies’ report. A revision of the ESH-IP should include a form to fill in all the required information for a complete report of a validation study, which will guide investigators to strictly comply with the protocol and will prevent misreporting of the study results. In addition, several issues within the protocol need further clarification to prevent possible misunderstanding of the requirements and to avoid some of the violations. Aiming to address the issues discussed in this review, several modifications of the ESH-IP have been incorporated in its latest revision [70].

**References**


