## REVIEW

# Determining which automatic digital blood pressure device performs adequately: a systematic review

Y Wan<sup>1</sup>, C Heneghan<sup>2</sup>, R Stevens<sup>2</sup>, RJ McManus<sup>3</sup>, A Ward<sup>2</sup>, R Perera<sup>2</sup>, M Thompson<sup>2</sup>, L Tarassenko<sup>4</sup> and D Mant<sup>2</sup>

<sup>1</sup>Department of Health Statistics, Fourth Military Medical University, Xi'an, China; <sup>2</sup>Department of Primary Health Care, University of Oxford, Oxford, UK; <sup>3</sup>Department of Primary Care Clinical Sciences, University of Birmingham, Birmingham, UK and <sup>4</sup>Institute of Biomedical Engineering, University of Oxford, Oxford, UK

The aim of this study is to systematically examine the proportion of accurate readings attained by automatic digital blood pressure (BP) devices in published validation studies. We included studies of automatic digital BP devices using recognized protocols. We summarized the data as mean and s.d. of differences between measured and observed BP, and proportion of measurements within 5 mm Hg. We included 79 articles (10783 participants) reporting 113 studies from 22 different countries. Overall, 25/31 (81%), 37/41 (90%) and 34/35 (97%) devices passed the relevant protocols [BHS, AAMI and ESH international protocol (ESH-IP), respectively]. For devices that passed the BHS protocol, the proportion of measured values within 5 mm Hg of the observed value ranged from 60 to 86% (AAMI protocol 47-94% and ESH-IP 54–89%). The results for the same device varied

Keywords: blood pressure; validation; sphygmomanometer

## Introduction

Hypertension is an important worldwide public health problem and is a major risk factor for cardiovascular disease, the leading cause of death and disability worldwide.<sup>1,2</sup> Testing for hypertension is the most commonly performed screening test in medical practice and is important not least because individuals with a normal blood pressure (BP) at the age of 50 still run a 90% life-time risk of developing hypertension.<sup>3</sup> Furthermore, trial evidence indicates that even small differences in systolic BP of 2–4 mm Hg are clinically important, thus accurate measurement is vital.<sup>4–6</sup>

Errors in measurement occur through three main sources: device errors, user errors and patient errors.<sup>7</sup> Device errors are potentially the most

E-mail: Carl.heneghan@dphpc.ox.ac.uk

Received 1 November 2009; revised 3 March 2010; accepted

4 March 2010; published online 8 April 2010

significantly when a different protocol was used (Omron HEM-907 80% of readings were within 5 mm Hg using the AAMI protocol compared with 62% with the ESH-IP). Even devices with a mean difference of zero show high variation: a device with 74% of BP measurements within 5 mm Hg would require six further BP measurements to reduce variation to 95% of readings within 5 mm Hg. Current protocols for validating BP monitors give no guarantee of accuracy in clinical practice. Devices may pass even the most rigorous protocol with as few as 60% of readings within 5 mm Hg of the observed value. Multiple readings are essential to provide clinicians and patients with accurate information on which to base diagnostic and treatment decisions.

*Journal of Human Hypertension* (2010) **24,** 431–438; doi:10.1038/jhh.2010.37; published online 8 April 2010

fundamental of these three sources, and, in particular, accurate measurement is underpinned by the requirement for an instrument that measures pressure accurately in patients across the range of pressures.<sup>8</sup> For many years, the standard instrument for BP measurement was a mercury sphygmomanometer, but in recent years this has been superseded in many clinical settings by automated electronic devices.

The need to validate the accuracy of such devices has led to the development of various protocols. (Box 1).<sup>9,10</sup> These protocols stipulate the minimum testing population with an acceptable range of mean and s.d. of errors or an s.d. below a threshold, which is dependent on the level of mean difference, but do not specify the proportion of individuals within those populations that should have accurate readings.<sup>11</sup> However, it is possible for a BP monitor to meet the above validation criteria, but still record BP measurements in error by >5 mm Hg in half of the individuals assessed.<sup>6</sup> It has, therefore, been argued that validation standards should be based, in part, on the percentage of individual readings within an

Correspondence: Dr C Heneghan, Department of Primary Health Care, University of Oxford, Rosemary Rue Building, Headington, Oxford OX3 7LF, UK.

432

Box 1 Summaries of protocols for validating automated sphygmomanometers

British Hypertension Society Protocol (1993)

- Sample size: 85 subjects with 3 measurements on each Selection: to represent a range of blood pressures from  $< 90/60 \,\mathrm{mm \, Hg}$  to  $> 180/110 \,\mathrm{mm \, Hg}$ , arm
- circumference and other characteristics by unbiased sampling
- Comparison: sequential measurement in the same arm Criteria: the acceptable differences<sup>a</sup> vary according to
- the true blood pressure—see full protocol for details Statistical methods: treat as 255 independent measurements

#### Association for the Advancement of Medical

Instrumentation (American National Standard) (2002)<sup>b</sup>

- Sample size: at least 85 subjects with 3 measurements each
- Selection: a range of blood pressures, at least 10% . < 100/60 mm Hg and 10% > 160/100 mm Hg, and 10%to be on subjects on arm sizes  $<25 \,\mathrm{cm}$  and 10%>35 cm, other characteristics by unbiased sampling
- Comparison: simultaneous measurement in the same . arm
- Criteria: difference has mean < 5 mm Hg and s.d. <8 mm Hg
- Statistical methods: treat as 255 independent measurements

European Society of Hypertension International Protocol (2002)

- Sample size: 33 subjects with 3 measurements each
- Selection: in three blood pressure bands from <130/80 to  $> 160/100 \,\mathrm{mm}\,\mathrm{Hg}$ , with additional requirements on age and sex
- Comparison: sequential measurement in the same arm
- Criteria: at least 2/3 of subjects have at least 2/3 of differences  $<5 \,\mathrm{mm}\,\mathrm{Hg}$ ; no >3 out of 33 subjects have differences > 5 mm Hg
- Statistics: account for both subject level and observation level variation (see 'criteria')

<sup>a</sup>These summaries refer to the protocols for validating automated sphygmomanometers against auscultatory measurements in the general adult population. <sup>b</sup>Also provides alternative, approximately equivalent, criteria to

be used with an alternative statistical method.

c'Differences' refer to differences between the test value and comparison values, usually taken by an auscultatory method by two independent observers.

accurate range.<sup>12,13</sup> To understand the potential effect of such a change in policy, we aimed to systematically examine the proportion of accurate readings attained by automatic digital BP devices in published clinical validation studies.

### Methods

#### Eligibility and search strategy

We included all published studies of automatic digital BP devices that (1) validated studies of BP devices using a recognized protocol in a clinic setting or in the community; (2) used a sphygmomanometer with upper arm BP measurement; (3) included adult patients ( $\geq 18$  years of age) with no significant severe intercurrent health problems and

(4) included data on the mean and s.d. of the BP difference between calibrated standard sphygmomanometer (mercury or random-zero sphygmomanometers) and (semi-)automatic (oscillometric or aneroid) sphygmomanometer. There were no language restrictions.

Reasons for study exclusion were devices used in special clinical situations (for example pregnancy, at altitude, during exercise, in trauma patients, neonates, children and serious diseases), measurements not from upper arm BP monitoring devices (for example wrist and ABPM), comparison with other sphygmomanometers (for example direct central pressure and ambulatory intra-arterial pressure) and published papers reporting no mean difference or s.d.. We searched Ovid versions of MEDLINE and EMBASE, and Cochrane Library through Wiley Interscience from January 1989 to June 2008. Medical Subject Headings and synonym terms used were sphygmomanometer, BP, hypertension, calibration and monitor device. We also searched the reference lists of identified papers and reviews.

#### Data abstraction

Two authors reviewed the titles and abstracts of all identified articles and selected full-text papers. Articles clearly not meeting the criteria were excluded at this stage. The remaining articles were then reviewed in detail by three authors for inclusion. We extracted data on the percentage of BP measurements within 5 mm Hg, mean and s.d. of difference between calibrated standard sphygmomanometer and (semi-) automatic sphygmomanometer. We also obtained information on detail of process, including study design, sample size, device characteristics and the relevant guidelines used to assess the device accuracy.

One author (YW) extracted data and this was rechecked by a second author (CH) independently, disagreements were resolved by discussion with a third author (RP). The reviewers were not masked to any aspect of the studies (for example journal type, author names or institution).

#### Data analysis

We used Review Manager (RevMan) Version 5.0. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2008 for the statistical analysis and forest plots (sorted by mean effect and device type for systolic BP and diastolic BP).

We summarized data as mean and s.d. of differences between measured and observed BP (the observer control measurement used in the protocol), and proportion of measurements within 5 mm Hg of observed BP and ordered the rows of the figures by the percentage of measurements with errors <5 mm Hg. When this information was not available, we replaced it by a bias-corrected modelled value, defined as  $0.27 + 0.76 \times P$ , where P is the

percentage of errors < 5 mm Hg assuming a normal distribution, and the bias correction coefficients (0.27 and 0.76) were derived by linear regression on the trials that report all outcomes. We analysed the effect of year of protocol by separating studies into four groups (before 1995, 1996–2000, 2001–2005 and after 2006) and used analysis of variance and the  $\eta^2$  to estimate the amount of variation explained by year of protocol These calculations were performed using STATA (Intercooled STATA 10 for Windows).

We also calculated the additional number of measurements required assuming a normal distribution for a selection of devices to achieve 95% of readings within 5 mm Hg based on the assumption of a mean difference equal to zero and a dispersion of 2.5 mm Hg (similar to the top-rated device for each protocol).

## Results

From 5912 potentially relevant records, 79 articles (10783 participants) reporting 113 studies (including subgroups for different devices and protocols) met the eligibility criteria (Figure 1). Of these, 96 studies were recommended devices by the various protocols, 13 were not recommended and 14 devices were validated in the community (including six not recommended by the protocol).

Studies were from 22 different countries: US (23); UK (20); France (15); Italy (10); Ireland and Poland (6 each); Japan (5); Canada (4); Russia, Germany and Turkey (3 each); Denmark, Greece, Spain and

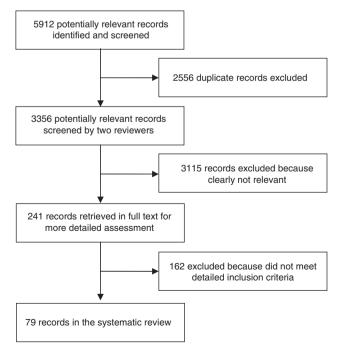


Figure 1 Flowchart of search results.

Switzerland (2 each); and Argentina, Brazil, China, Israel, Netherlands, Sweden and Tanzania (1 each).

Studies used five different protocols for validating devices: 35 used the ESH international protocol (ESH-IP), 9 the BHS protocol, 19 ANSI/AAMI protocol and 22 used both the BHS and ANSI/AAMI protocols. The results from these are presented in Figures 2–4, respectively. A further 16 studies used different protocols: 3 the American Heart Association BP measurement guidelines, 2 the Health Insurance Portability and Accountability Act protocol and 11 studies did not state which protocol they used. All of these 16 were excluded from the present analysis, and on re-analysis, four aneroid devices were removed from the analysis.

For each study, the citation, device tested and mean and s.d. of differences are shown, together with the reported proportion of measurements within 5 mm Hg of the observed value. Some studies and/or devices appear more than once in which results were reported using different sampling schemes and/or protocols or where more than one group validated a device.

Overall, 25/31 (81%, 95% CI: 67–95%), 37/41 (90, 81–99%) and 34/35 (97, 92–99.5%) of devices passed the relevant protocol (BHS, AAMI and ESH-IP, respectively). No particular pattern was seen either in terms of the date of assessment or of manufacturer in terms of likelihood of a machine in terms of accuracy (data not shown).

Devices passing the BHS protocol had between 60 and 86% of measured values within 5 mm Hg of the observed value (Figure 2). Devices failing the BHS protocol (marked with a †) had between 35 and 59% of test device values within 5 mm Hg. Devices passing the AAMI protocol had between 47 and 94% of measured values within 5 mm Hg of the correct value (Figure 3) and for the four devices failing 36–58% within 5 mm Hg. For the ESH-IP protocol devices passing the protocol, it had 54– 89% within 5 mm Hg, and for the one device reported to fail the International protocol, 52% were within 5 mm Hg (Figure 4).

Some studies appear twice as they undertook the protocol, but using different techniques. For example, in the Reinders' study,<sup>14</sup> when automated measurements using the BHS protocol were taken simultaneously with the reference standard, 96% were within 5 mm Hg of the reference value; when taken sequentially, 80% were within 5 mm Hg.

In devices tested in community/clinic-based studies (studies performed from a population-based perspective), between 35 and 46% of measured values were within 5 mm Hg of the observed value. We found only two devices tested using both a protocol and in the community. For the Omron HEM-705CP, 86% of measurements were within 5 mm Hg of the observed value under the BHS protocol<sup>15</sup> compared with 46% within 5 mm Hg in the community-based study.<sup>16</sup> For the IVAC model 4200, these figures were 59% under the BHS

Mean SBP Difference±SD Mean DBP Difference±SD (% <5mmHq) (% <5mmHq) Mean SBP Difference±SD Study or Subgroup **Device Name** Imai 1995 (Korotkoff sound)\* Terumo ES-H51 -0.6+3.4(86)-0.2+4.0(86)Iglesias 2002 Omron HEM-705 CP -1.1±4.7(86)  $-0.4 \pm 4.0(91)$ Iglesias 2002\* Omron M1 -1.3±5.3(83) -0.5±3.9(91) Zaetta 2007 SAA-102 0.4±5.0(81) -2.0±6.0(71) Kobalava 2005 A&D UA-704 -1.9±4.3(79) -1.4±4.0(87) Bortolotto 1999 Omron HEM-722C 0.8±5.0(76) 0.4±8.0(71) A&D TM-2564G Kobalava 2005  $-1.9\pm4.9(76)$  $-1.2\pm4.2(83)$ De Greeff 2007 Dinamap ProCare monitor -0.3±6.9(75) -4.0±5.9(60) Jones 2001\* Welch Allyn Vital Signs -1.0±5.0(74)  $-3.0\pm4.0(72)$ Imai 1995 (cuff-oscillometric)\* Terumo ES-H51  $-0.6\pm6.1(72)$  $0.0\pm 4.9(72)$ De Greeff 2007 Health and Life device (HL888HA) 0.5±7.3(72) -2.1±7.8(60) Coleman 2006\* Omron 705IT (HEM-759-E) 0.6±6.0(70) -3.2±6.6(61) Mano 2002\* DIXTAL DX-2710  $0.0\pm6.3(69)$ -3.8±6.2(63) Coleman 2008\* Omron M7 (HEM-780-E) -0.8±6.5(69) 1.3±5.0(70) De Greeff 2008\* Tensoval duo control -2.4±6.0(68) 0.9±6.3(73) Verdecchia 2004 A&D UA-774 (UA-767Plus) -10+71(68)-1.4±7.6(69) Westhoff 2008 Stabil-O-Graph -1.9±6.6(66) 0.0±5.8(65) O'Brien 1993 † Dinamap 8100 -1.0±7.0(66) -6.0±7.0(47) Bortolotto 1999 Omron HEM-735C 0.2±8.0(66)  $0.9 \pm 8.0(74)$ Alpert 2004 Pharma-Smart PS-2000 0.2±7.0(65) -0.1±6.7(66) Cuckson 2002\* Microlife BP 3BTO-A -1.7±7.4(64) -2.1±6.3(68) A&D 11A-705 0.4±7.7(63) 0.3±8.2(65) Verdecchia 2006 1.5±7.6(62) De Greeff 2008\* Rossmax (ME 701 series) -1.2±6.1(71) Alpert 2007\* Welch Allyn Spot Vital Signs Lxi -0.9±7.2(60) -2.2±6.7(60) Golara 2002 Omron-MIT -2.1±7.1(60) -2.4±5.8(70) Shuler 1998\*\* IVAC model 4200 1.6±7.4(59) 2.0±7.5(54) Beaubien 2002\* <sup>+</sup> Dinamap1846XT 0.0±8.0(59) -3.0±7.7(56) Dieterle 1998\*\* Visomat OZ2 -3.7±7.5(47)  $-4.8\pm5.6(54)$ A&D UA 777 -5.5±8.3(39) Shahriari 2003\* † -6.8±6.8(36) Shahriari 2003\*\* Omron M4  $-5.5 \pm 8.5(36)$ -3.0±5.9(60) Braam 2002 † Welch Allvn Signs Monitor 52000 5.3±6.7(35) 7.5±7.1(43) -5 -10 ń 10

Mean difference (mmHg)

Figure 2 Means and s.d. of errors in systolic BP in devices passing, or failing (marked dagger), the British Hypertension Society protocol.

protocol and 39% in the community-based study.<sup>17</sup> When we studied the performance of the monitors for diastolic BP, results were broadly similar.

Figure 5 shows the mean and single s.d. of errors in BP monitors in repeated studies of the same monitor using either the same or different protocols. The results for the same device could vary substantially when a different protocol was used. For example, the Omron HEM-907 device could vary by 18% of readings within 5 mm Hg depending on which protocol was used: 80% (using the AAMI protocol)<sup>18</sup> compared with 62% (ESH-IP protocol).<sup>19</sup> The same device could also vary by 22% of readings within 5 mm Hg when the same protocol was used: 62% (El Assaad 2002)<sup>19</sup> compared with 84% (De Greef 2007).20

There was no significant difference between the three different protocols (P = 0.20). However, the year of protocol was significantly correlated with the proportion of measurements (r = 0.257, P < 0.01). Overall, 19% of the variation could be explained by changes over time  $(\eta^2 = 2581.11/13600.77 = 0.19;$ F = 6.513, P < 0.001). Subgroup analysis by protocol

showed the ESH-IP to be the contributor to this correlation (the majority of its studies were later with 64% after 2005), whereas the BSH and AAMI (earlier studies) were not.

Even if a given device could be improved to have a mean zero, it would still be necessary to reduce the variation. For example, for a device with 74% of BP measurements <5 mm Hg [the Welch Allen Vital Signs (s.d.: 5)], a further six BP measurements would be required to reduce the variation, so that 95% of readings fall within 5 mm Hg. To obtain the same precision for the Pharma Smart PS-2000 (s.d.: 7), a device with 66% < 5 mm Hg, a further 11 measurements are required (based on the assumption that in each individual the BP differences between the tested device and the reference method are random). For the ESH-IP protocol when using a device with 81% <5 mm Hg [Artasan CS410 (s.d.: 5.9)], a further eight BP measurements are required to reduce variation, so that 95% of readings are within 5 mm Hg; whereas for the Omron M6 Comfort (s.d.: 7.4) with 68% < 5 mm Hg, a further 12 measurements are required. The mean difference

Y Wan et al	Determining which automatic digital BP device performs adequately
	Y Wan et al

Study or Subgroup	Device Name	Mean SBP Difference±SD (% <5mmHg)	Mean DBP Difference±SD (% <5mmHg)	Mean SBP Difference±SD
aetta 2007 (AAMI method 2)	SAA-102	0.1±3.2(94#)	-2.7±3.5(83#)	
arows 2007 (AAMI method 2)	HoMedics BPA-300	-0.9±4.0(93)	-0.5±3.9(92)	
arows 2007 (AAMI method 1)	HoMedics BPA-300	-0.6±4.3(93)	-0.5±4.3(92)	
ai 1995 (Korotkoff sound)*	Terumo ES-H51	-0.6±3.4(86)	-0.2±4.0(86)	
Brien 1990	Nissei Analogue Monitor	-0.1±4.4(84#)	-4.3±4.6(68#)	
esias 2002*	Omron M1	-1.3±5.3(83)	-0.5±3.9(91)	
etta 2007 (AAMI method 1)	SAA-102	0.1±4.6(82#)	-2.7±5.2(73#)	
goza 2000*	A&D UA-767	0.0±5.4(82)	-0.4±4.8(80)	
nite 2001	Omron HEM-907	1.6±4.4(80)	3.5±4.6(64)	
ster 1994*	Omron HEM-706	-1.2±5.1(77#)	-4.4±5.4(65#)	
rtolotto 1999*	Omron HEM-722C	0.8±5.0(76)	0.4±8.0(71)	
Greeff 2007*	Dinamap ProCare monitor	-0.3±6.9(75)	-4.0±5.9(60)	
nes 2001*	Welch Allyn 'Vital Signs'	-1.0±5.0(74)	-3.0±4.0(72)	
right 2001*	BPM-100	-0.2±5.1(74)	-1.4±4.7(78)	
pert 2004 (AAMI method 2)	Pharma-Smart PS-2000	0.1±5.9(73#)	-0.3±6.1(72#)	
ai 1995 (cuff-oscillometric)*	Terumo ES-H51	-0.6±6.1(72)	0.0±4.9(72)	
e Greeff 2007*	Health and Life (HL888HA)	0.5±7.3(72)	-2.1±7.8(60)	
Brien 1990	Philips HP5308	-2.8±5.4(72#)	-4.2±4.8(68#)	
war 1998	Omron HEM-737 IntelliSense	0.8±6.6(71)	1.0±5.5(77)	
leman 2006*	Omron 705IT (HEM-759-E)	0.6±6.0(70)	-3.2±6.6(61)	
ano 2002*	DIXTAL DX-2710	0.0±6.3(69)	-3.8±6.2(63)	
leman 2008*	Omron M7 (HEM-780-E)	-0.8±6.5(69)	1.3±5.0(70)	
Greeff 2008*	Tensoval duo control	-2.4±6.0(68)	0.9±6.3(73)	
pert 2004 (AAMI method 1)	Pharma-Smart PS-2000	0.1±7.0(67#)	-0.3±6.6(69#)	
rtolotto 1999*	Omron HEM-735C	0.2±8.0(66)	0.9±8.0(74)	
r 2005	MABIS Smart Read (04-310-001)	-1.2±7.6(64#)	-1.4±7.5(64#)	
nite 2003	Accutorr PlusTM (Datascope Corp.)	-0.1±7.5(64)	2.5±5.2(71)	
ickson 2002*	Microlife BP 3BTO-A	-1.7±7.4(64)	-2.1±6.3(68)	
r 2005	MABIS Smart Speed Plus	-0.4±7.7(64#)	-1.5±6.7(68#)	
Greeff 2008*	Rossmax (ME 701 series)	1.5±7.6(62)	-1.2±6.1(71)	
r 2005	MABIS Semi-Automatic	2.7±7.8(62#)	0.4±7.4(65#)	
attu 2001*	BPM-100Beta	-0.6±7.0(61)	-1.5±4.8(82)	
pert 2007*	Welch Allyn Spot Vital Signs Lxi	-0.9±7.2(60)	-2.2±6.7(60)	
olara 2002*	Omron-MIT	-2.1±7.1(60)	-2.4±5.8(70)	
uler 1998*	IVAC model 4200	1.6±7.4(59)	2.0±7.5(54)	
aubien 2002*	Dinamap 1846XT	0.0±8.0(59)	-3.0±7.7(56)	
Brien 1990†	Philips HP5306/B	-2.2±9.2(58#)	-3.0±11.3(52#)	
Brien 1990†	Omron HEM-400C	-2.6±9.8(56#)	0.0±9.2(59#) ←	
eterle 1998*	Visomat OZ2	-3.7±7.5(47)	-4.8±5.6(54)	
ahriari 2003*†	A&D UA 777	-5.5±8.3(39)	-6.8±6.8(54#)	
nahriari 2003* †	Omron M4	-5.5±8.5(36)	-3.0±5.9(69#)	

Figure 3 Means and s.d. of errors in systolic BP in devices passing, or failing (marked dagger), the AAMI protocol.

between devices that passed the protocols was relatively stable: BHS range: -2.4-1.5 mm Hg, AAMI: 2.8-2.7 mm Hg and ESH-IP: -3.2-2.0 mm Hg.

## Discussion

Devices that passed international validation protocols for BP accuracy vary greatly in quality. Taking 'percentage of readings within 5 mm Hg of the observed value' as a measure of quality, we found wide variations in quality among devices that passed protocol testing, whether the protocol was BHS, AAMI or ESH-IP, with slight improvements in performance using the ESH-IP protocol. This is a reflection in part of the numbers of individuals required to fulfil the protocols, which affects their ability to distinguish between machines of differing accuracies: the BHS protocol requires testing on 85 patients, and, therefore, has the power to distinguish differences between devices of > 15%, whereas the international protocol requires only 33 patients (less than the other protocols), and can only detect differences between devices of > 24%.

Mean difference (mmHg)

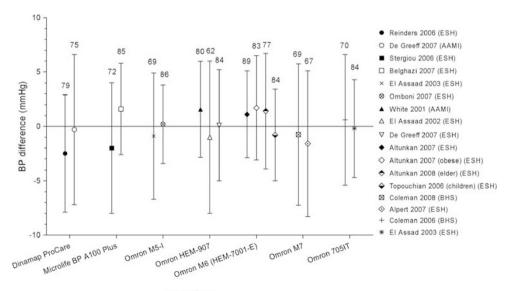
Moreover, devices passing a protocol are not necessarily more accurate than devices that failed, and devices that pass may perform substantially worse when tested in a community setting. In addition, a device that passes validation does not mean that it provides accurate readings in all patients; inaccuracy is more likely in older and diabetic patients.<sup>21,22</sup> 435

Determining which automatic digital BP device performs adequately Y Wan et al

Study or Subgroup	Device Name	Mean SBP Difference±SD (% <5mmHg)	Mean DBP Difference±SD (% <5mmHg)	Mean SBP Difference±SE
ltunkan 2007 (adults)	Omron M6 (HEM-7001-E)	1.1± 4.0(89)	-0.5±3.5(89)	<u> </u>
elghazi 2007	Spengler KP7500 D	-1.4±8.6(87)	-0.1±3.5(94) 🔶	
mboni 2007	Omron M5-I (HEM-757-E)	0.2±3.6(86)	0.2±3.9(86)	
e Greeff 2007	Artsana CS 610	-1.4±4.8(86)	-0.9±3.5(91)	
Cotte 2008	Citizen CH-432B	0.7±4.4(85)	-3.6±4.0(74)	
elghazi 2007	Microlife BP A100 Plus	1.6±4.2(85)	0.5±2.8(92)	
opouchian 2006	Omron M6 (HEM-7001-E)	-0.8±4.2(84)	-1.9±3.8(85)	
l Assaad 2003	Omron 705IT	-0.2±4.5(84)	-2.0±4.8(75)	
aensel 2005	TONOPORT V	-0.7±4.6(84)	-0.8±4.4(81)	
e Greeff 2007	Omron HEM-907	0.1±5.1(84)	-1.9±4.2(81)	
elghazi 2007	Omron M1 Plus (HEM-4011C-E)	-1.4±5.5(84)	-0.4±4.8(81)	
ltunkan 2007 (obese)	Omron M6 (HEM-7001-E)	1.7±4.8(83)	-0.8±4.3(81)	
ini 2008	Artsana CS410	-0.9±5.9(81)	-1.0±3.1(93)	·
leinders 2006	Dinamap ProCare monitor	-2.5±5.4(79)	0.5±4.5(77)	
i 2006	Oregon Scientific BPW810	-2.2±4.8(78)	-0.5±3.7(82)	
e Greeff 2008	Tensoval duo control	-1.5±5.3(78)	1.5±6.1(78)	
opouchian 2005	SEINEX SE-9400	-0.9±5.2(77)	-1.7±4.7(80)	
/ilton 2006	PMS mandaus	-3.2±3.8(77)	-1.8±2.9(88)	
ongo 2003	A&D UA-787	1.0±5.3(77)	0.7±5.3(79)	· .
ltunkan 2008	Omron M6 (HEM-7001-E)	1.4±5.3(77)	-1.4±4.5(78)	
lolly 2004	Braun BP 2550 (UG)	-1.6±4.8(76)	2.2±3.8(79)	
tergiou 2007	Microlife WatchBP Home device	-0.3±5.6(76)	-2.4±4.8(75)	
opouchian 2005	Microlife BP 3AC1-1	-1.3±5.6(75)	-0.4±4.8(82)	
ongo 2002	A&D UA-631 (UA-779 Life Source)	2.0±5.0(73)	1.0±3.0(94)	
lesthoff 2008	Stabil-O-Graph	0.0±5.6(73)	1.0±5.7(76)	
tergiou 2006	Microlife BPA100 Plus	-2.0±6.0(72)	-3.1±4.1(72)	
e Greeff 2008	Rossmax (ME 701 series)	-0.1±6.2(71)	-2.8±5.3(73)	· · · · · ·
era 2007	HEM-780REL	0.5±7.7(71)	0.4±4.7(77)	l
oleman 2005	Omron MX3 Plus (HEM-742-E)	-1.2±5.7(69)	-1.6±4.7(76)	
I Assaad 2003	Omron M5-I	-0.9±5.8(69)	-0.8±4.8(84)	
elghazi 2007	Omron M6 Comfort (HEM 7000-E)	-2.1±7.4(68)	0.1±4.9(74)	· · · · · · · · · · · · · · · · · · ·
lpert 2007	Omron M7 (HEM-780-E)	-1.6±6.7(67)	-0.1±4.0(83)	
I Assaad 2002	Omron HEM-907	-1.0±7.0(62)	-5.0±6.0(53)	
loin 2002	Calor® TensioSense BrasTM	0.0±9.0(54)	0.0±7.0(63)	
)'Brien 2003†	ROSSMAX blood pressure monitor	( )	-1.8±5.0(72)	

Note: †: failed to pass validation protocol (judgment for both SBP and DBP).

Figure 4 Means and s.d. of errors in systolic BP in devices passing, or failing (marked dagger), the ESH-IP protocol.



Mean difference (mmHg)



Figure 5 Mean and single s.d. of errors in BP monitors in repeated studies of the same monitor.

436

In general, we found devices performing badly do so because of wide variation between differences (large s.d.) rather than systematic over- or underestimation of BP (Figures 2-4). This is reflected in the relatively stable mean difference, which leads to a large number of additional measurements for devices relative to only a small drop off in precision (that is lower proportion of measurements within 5 mm Hg). This effect is negated by multiple readings over multiple clinic visits. Yet, as a BP approaches the threshold for diagnosis, this inaccuracy can substantially under- or overestimate the burden of hypertension.<sup>13</sup> In addition, community clinic-based validations were fewer and the results much worse. This is particularly of concern, as this setting is where devices are most often used for the assessment and ongoing management of hypertension.

However, it is worth noting that the change from mercury to electronic BP measurement in practice has led to no consistent change in mean BP after their introduction, but there was a large and significant fall in terminal digit preference suggesting improved precision of recording with electronic devices.<sup>23</sup>

This study had a number of limitations: the reproducibility of protocol testing is currently unknown. We found no studies, which repeated earlier validations. Some studies may have been missed. Although the search of 5912 potentially relevant records was comprehensive, with analysis of 3356 potentially relevant records, studies in which machines failed the validation are less likely to be published, particularly if the manufacturer is involved in the funding. However, missing such studies would have the effect of worsening the results and hence would not alter the conclusions. All validation studies found used the standard cuff size. A significant proportion of adults will need to use the larger cuff and yet there is no evidence as to whether this will be accurate. The analysis was limited by the lack of reporting of proportional data in some cases. Development of an international consensus on minimal standards of reporting, including representations of proportional data, would go some way to allowing comparison of devices under different validation protocols. Our estimates of the proportion within 5 mm Hg of the observed value may not be accurate, because they assume a study-specific normal distribution that leads to some degree of inaccuracy, but were sufficient to inform the ordering of the figures.

The existing protocols have several limitations. First, the procedure specifications are sometimes ambiguous. For example, the AAMI protocol allows a choice of data collection methods (simultaneous vs sequential) and a choice of data-analysis methods, either of which may affect the final results. Problems with validation protocol methods not being detailed enough has earlier been cited as a problem for interpretation.<sup>24</sup> Second, not all protocols directly address the margin of error between measurement and observed value, and those that do allow margins of error as wide as 10 or 15 mm Hg, result in measurements that are up to half the time >5 mm Hg from the observed value in some monitors that pass protocols. Third, we found few examples of the protocols being tested in community settings (that is, real clinics rather than artificial settings), and the results from these studies were generally poor. This is of concern because it may suggest that most protocol studies overestimate device quality, although we found too few community studies to draw firm conclusions. Finally, the British and American protocols are difficult or costly to obtain, which could discourage further evaluation of BP monitors. Further information on protocol publications is available from the BHS website (http://www.bhsoc.org/Blood\_pressure\_Publications. stm).

Protocols that include measurements for a certain proportion of readings to be within a certain value and have unambiguous procedures could encourage the manufacture of higher quality BP monitors. Monitors need to be re-evaluated in community settings ('real clinics') to verify their accuracy in practice. The advent of a central, open access repository of evaluation studies would be beneficial to clinics and healthcare providers when choosing digital monitors.

Current protocols for validating BP devices give no guarantee of accuracy in clinical practice. Devices may pass even rigorous protocols with as few as 60% of readings within 5 mm Hg of the observed value. The limited evidence available suggests that performance in clinical practice will be even worse than under protocol testing. BP standards are essential to provide clinicians and patients with accurate information on which to base diagnostic and treatment decisions.

## **Conflict of interest**

The authors declare no conflict of interest.

## Acknowledgements

This work has received financial support from a National Institute of Health Research (NIHR) programme grant: Development and implementation of new diagnostic processes in primary care. The Department of Primary Health Care in Birmingham and Oxford are part of the NIHR school of primary care research, which provide financial support for senior staff who contributed to this paper. Professor Lionel Tarassenko has received funding from the Oxford Biomedical Research Centre.

## References

- 1 O'Brien E, Asmar R, Beilin L, Imai Y, Mallion JM, Mancia G et al. European Society of Hypertension recommendations for conventional, ambulatory and home blood pressure measurement. J Hypertens 2003; 21(5): 821–848.
- 2 Wolf-Maier K, Cooper RS, Banegas JR, Giampaoli S, Hense HW, Joffres M *et al.* Hypertension prevalence and blood pressure levels in 6 European countries, Canada, and the United States. *JAMA* 2003; **289**(18): 2363–2369.
- 3 Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo Jr JL *et al.* The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* 2003; **289**(19): 2560–2572.
- 4 ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). JAMA 2002; **288**(23): 2981–2997.
- 5 Julius S, Jamerson K, Gudbrandsson T, Schork N. White coat hypertension: a follow-up. *Clin Exp Hypertens A* 1992; **14**(1–2): 45–53.
- 6 Pater C. Beyond the Evidence of the New Hypertension Guidelines. Blood pressure measurement—is it good enough for accurate diagnosis of hypertension? Time might be in, for a paradigm shift (I). *Curr Control Trials Cardiovasc Med* 2005; **6**(1): 6.
- 7 Padwal R, Straus SE, McAlister FA. Evidence based management of hypertension. Cardiovascular risk factors and their effects on the decision to treat hypertension: evidence based review. *BMJ* 2001; **322**(7292): 977–980.
- 8 NICE. Clinical guideline 34: hypertension: management of hypertension in adults in primary care: partial update. 2008. Available from: URL: http://www. nice.org.uk/CG034guidance.
- 9 O'Brien E, Petrie J, Littler W, de SM, Padfield PL, Altman DG *et al.* An outline of the revised British Hypertension Society protocol for the evaluation of blood pressure measuring devices. *J Hypertens* 1993; 11(6): 677–679.
- 10 O'Brien E, Pickering T, Asmar R, Myers M, Parati G, Staessen J et al. Working Group on Blood Pressure Monitoring of the European Society of Hypertension International Protocol for validation of blood pressure measuring devices in adults. *Blood Press Monit* 2002; 7(1): 3–17.
- 11 Schwartz AR, Haas DC, Gerin W, Pickering TG. Accurate measurement of blood pressure. *JAMA* 2003; **289**(21): 2792–2794.
- 12 Fournier A, Safar M. Accurate measurement of blood pressure. *JAMA* 2003; **289**(21): 2793–2794.

- 13 Jones DW, Appel LJ, Sheps SG, Roccella EJ, Lenfant C. Measuring blood pressure accurately: new and persistent challenges. *JAMA* 2003; **289**(8): 1027–1030.
- 14 Reinders A, Jones CR, Cuckson AC, Shennan AH. The Maxi Stabil 3: validation of an aneroid device according to a modified British Hypertension Society protocol. *Blood Press Monit* 2003; **8**(2): 83–89.
- 15 Iglesias BP, Mayoral SE, Lapetra PJ, Iborra OM, Villalba AF, Cayuela DA. Validation of two systems of selfmeasurement of blood pressure, the OMRON HEM-705 CP and OMRON M1 (HEM 422C2-E) models. Aten Primaria 2002; 30(1): 22–28.
- 16 Lewis JE, Boyle E, Magharious L, Myers MG. Evaluation of a community-based automated blood pressure measuring device. *CMAJ* 2002; **166**(9): 1145–1148.
- 17 Shuler CL, Allison N, Holcomb S, Harlan M, McNeill J, Robinett G *et al.* Accuracy of an automated blood pressure device in stable inpatients: optimum vs routine use. *Arch Intern Med* 1998; **158**(7): 714–721.
- 18 White WB, Anwar YA. Evaluation of the overall efficacy of the Omron office digital blood pressure HEM-907 monitor in adults. *Blood Press Monit* 2001; **6**(2): 107–110.
- 19 El Assaad MA, Topouchian JA, Darne BM, Asmar RG. Validation of the Omron HEM-907 device for blood pressure measurement. *Blood Press Monit* 2002; 7(4): 237-241.
- 20 De GA, Reggiori F, Shennan AH. Clinical assessment of the DINAMAP ProCare monitor in an adult population according to the British Hypertension Society Protocol. *Blood Press Monit* 2007; **12**(1): 51–55.
- 21 van Popele NM, Bos WJ, de Beer NA, van Der Kuip DA, Hofman A, Grobbee DE *et al.* Arterial stiffness as underlying mechanism of disagreement between an oscillometric blood pressure monitor and a sphygmomanometer. *Hypertension* 2000; **36**(4): 484–488.
- 22 van Ittersum FJ, Wijering RM, Lambert J, Donker AJ, Stehouwer CD. Determinants of the limits of agreement between the sphygmomanometer and the SpaceLabs 90207 device for blood pressure measurement in health volunteers and insulin-dependent diabetic patients. J Hypertens 1998; **16**(8): 1125–1130.
- 23 McManus RJ, Mant J, Hull MR, Hobbs FD. Does changing from mercury to electronic blood pressure measurement influence recorded blood pressure? An observational study. Br J Gen Pract 2003; **53**(497): 953–956.
- 24 Pickering T. Hypertension: a companion to Braunwald's Heart Disease. Black HR and Elliott W (eds). *Home Monitoring of Blood Pressure*, Chapter 6. Saunders Elsevier: Philadelphia, PA, USA, 2006, p 70.

This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 License. To view a copy of this license, visit http://creativecommons.org/licenses/ by-nc-nd/3.0/

## 438