Inaccuracy of seven popular sphygmomanometers for home measurement of blood pressure

Eoin O'Brien, Fáinsiá Mee, Neil Atkins and Kevin O'Malley

Seven models, available commercially for the self-measurement of blood pressure, were subjected to a validation procedure in which three devices of each model were tested by observers who were trained to a high standard of accuracy. The models were the Omron HEM-400C, the Philips HP5308, the Healthcheck 'Cuffless' CX-5 060020, the Nissei Analogue Monitor, the Philips HP5306/B, the Systema Dr MI-150 and the Fortec Dr MI-100. The validation programme had a number of unique features which included assessment of interdevice variability before and after 1 month of home use, and a new form of analysis, which we term 'clinical', based on the likely influence of three grades of device inaccuracy on patient management. In the main validation phase, one device of each model was compared with simultaneous measurements made by two 'blinded' observers using a standard mercury sphygmomanometer (PyMaH Corporation, New Jersey, USA) in the same arm in 85 subjects with a wide range of blood pressures. Three models (the Healthcheck 'Cuffless' CX-5 060020, the Systema Dr MI-150 and the Fortec Dr MI-100) failed the interdevice variability tests and did not reach the main validation test. Two models (the Omron HEM-400C and the Philips HP5306/B) failed on the criteria set down by the American National Standard for Electronic or Automated Sphygmomanometers, as well as the 'clinical' criteria. The remaining two models (the Nissei Analogue Monitor and the Philips HP5308) were acceptable for the measurement of systolic blood pressure by both methods of analysis but failed in the 'clinical' analysis for diastolic blood pressure. The mercury sphygmomanometer was comfortably within the criteria for both methods of analysis.

Journal of Hypertension 1990, **8**:621–634

Keywords: Validation of blood pressure devices, measurement of blood pressure, self-measurement of blood pressure.

Introduction

Self-measurement of blood pressure has been recommended by a number of authorities as a useful technique in the management of hypertension [1–4]. Such recommendations are usually made on the assumption that the sphygmomanometers used for self-measurement are accurate.

Validation studies have been carried out by clinical researchers on a number of self-measuring home kits [5–12] and, although some of these devices have been passed as accurate [6,8,9,12], others have not [5,7,10–12]. The criteria used for determining accuracy in validation studies are often not clearly stated [8], and many of the validation tests used in the past would not meet contemporary requirements.

Devices for the self-measurement of blood pressure are usually marketed independently of any clinical considerations, and sales strategies are directed at the lay public. It is hardly surprising, therefore, that there have been a number of reports from consumer organizations [13–17].

In the present study, seven models for self-measurement of blood pressure, chosen by the Consumer's Association (UK), were subjected to a new validation procedure that included the criteria of the American National Standard for Electronic or Automated Sphygmomanometers (American Association for the Advancement of Medical

From the Blood Pressure Unit, Beaumont Hospital, Dublin 9, Republic of Ireland, in association with The Consumers' Association, London.

Sponsorship: A research grant from The Charitable Infirmary Trust is gratefully acknowledged, as is the support of the Research Committee of the Royal College of Surgeons in Ireland.

Requests for reprints to: Dr Eoin O'Brien, The Blood Pressure Unit, Beaumont Hospital, Dublin 9, Republic of Ireland.

Date of receipt: 23 August 1989; revised: 15 November 1989.

Instrumentation; AAMI) [18] as well as a number of additional features, such as a test of interdevice variability and validation after a period of use.

In analysing our results, we used the AAMI criteria [18] and the graphical method of displaying data proposed by Bland and Altman [19]. However, we made the final assessment of the devices using a method of analysis in which consideration is given to the likely effect that inaccurate measurements might have on clinical management. As most clinical decisions are based on diastolic blood pressure measurements, greater accuracy is required for diastolic than for systolic measurements.

Methods

The study consisted of six phases, observer training, observer assessment, interdevice variability before home use, home use, interdevice variability after home use and device validation. Each phase was completed in compliance with previously established criteria (Fig. 1). Two features of the programme merit some elaboration. First, considerable emphasis was placed on observer training. If the test standard, namely, the mercury sphygmomanometer and the observer, cannot be brought to the highest possible level of accuracy *before* the main validation procedure, further testing is pointless. Second, the main validation test was performed after the models had been subjected to a 1-month period of use, rather than, as is customary in validation studies, immediately after purchase of the devices, before exposure to any stresses of daily use that might alter their accuracy. In this paper we use the term 'model' to denote a particular brand of sphygmomanometer and the term 'device' to denote individual sphygmomanometers.

Three devices of each of seven models that were readily available in High Street stores were purchased by members of the Consumers' Association, with the vendor unaware that they were being purchased for testing. The characteristics of the seven models in the study are summarized in Table 1. The decision to select three devices of each model was governed by considerations of economics and feasibility. If all three devices were to yield comparable measurements at the time of purchase as well as after a period in use, it would suggest, at least, that the brand is manufactured so that it performs consistently. If, however, all three devices were to yield discordant measurements, further assessment would be pointless and the model could not be recommended. If one device was found to be discordant, with the remaining two consistent, further validation would be merited on the basis that one inaccurate device could have been included by chance. Such an occurrence may indicate that the overall production of that model is unsatisfactory. Standard mercury sphygmomanometers (PyMaH Corporation, New Jersey, USA), the components of which were carefully checked before the study, were used as a reference standard in the second interdevice variability test and in the main validation procedure.

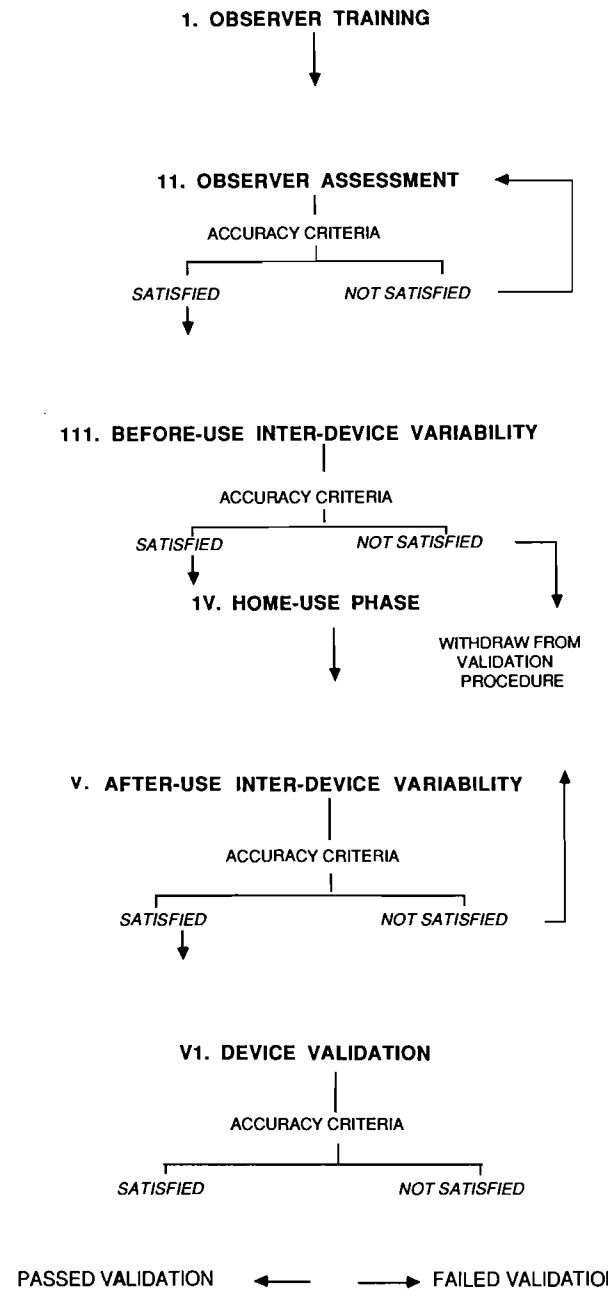


Fig. 1. The six text phases in the validation procedure.

Observer training and assessment

Eight nurses were trained in the technique of measuring blood pressure by the Korotkoff auscultatory technique, using Littmann stethoscopes (3M Company, Minnesota,

Table 1. Details of the sphygmomanometers tested.

Brand name	Model	Mode of operation	Retail price sterling (£)
Omron Digital Blood Pressure Monitor	HEM-400C	Digital and cuff oscillometry	70
Philips Blood Pressure Meter	HP5308	Gauge and microphone	40
Healthcheck 'Cuffless' Digital Blood Pressure Monitor	CX-5 060020	Digital and finger pulsation	80
Nissei Analogue Blood Pressure Monitor	Cuffed	Aneroid and stethoscope	30
Philips Blood Pressure Meter	HP5306/B	Digital and cuff oscillometry	70
Systema Digital Blood Pressure Meter	Dr MI-150	Digital and cuff oscillometry	40
Fortec Digital Blood Pressure Meter	Dr MI-100	Digital and cuff oscillometry	40

USA) and PyMaH mercury sphygmomanometers which had been serviced and checked for accuracy. The disappearance of the Korotkoff sounds was taken as indicating the diastolic blood pressure. The cuffs used throughout the study had bladders measuring 35 × 12 cm. As only one cuff was supplied with each device, and as no recommendations on cuff use in relation to arm size were given in the instruction leaflets, no attempt was made to select subjects according to their arm circumference. The nurses' training programme consisted of three parts: (1) training by an experienced observer (E.O'B.) using a multiheaded stethoscope; (2) training with test audiotape used for the International Study of Electrolyte Excretion and Blood Pressure (INTERSALT) project; and (3) training with the videotape 'Blood Pressure Measurement' produced in 1989 by the British Hypertension Society (available from the British Medical Journal).

Agreement between the observers was assessed using 'blinded' blood pressure measurements made simultaneously with a multiheaded stethoscope in 10 subjects with blood pressures ranging from 110/70 to 180/100 mmHg on three consecutive occasions. Trainees were not accepted for the study unless 85% or more of their systolic and diastolic blood pressure readings were within ± 5 mmHg of each other.

Before-use interdevice variability

Three devices of each model (21 sphygmomanometers) were available for interdevice variability testing. The interdevice variability test is based on the comparison of measurements made on opposite arms, and it was therefore necessary to select for inclusion in the test only those subjects in whom there was no substantial difference in blood pressure between arms. Subjects were excluded if their average interarm differences, measured simultaneously once by two observers with mercury sphygmomanometers, were > 5% for either systolic or diastolic blood pressure. As two of the three Systema Digital devices failed to function at the start of this test, this model was withdrawn.

Six observers were stationed in separate booths, each with three devices of each of the remaining six models. Two 'blinded' observers were used for each simultaneous measurement. Ten subjects with blood pressures ranging from 134/78 to 240/120 mmHg, in whom interarm differences in blood pressure were < 5%, were rotated amongst the tables and each device was tested for interdevice variability in the sequence: models 1, 3 and 2 on the right arm and models 2, 1 and 3 on the left arm, to ensure that each device was tested on each arm. The devices were inflated simultaneously when possible; when this was not practicable, they were activated simultaneously. One reading was taken with each device in 10 subjects and noted by the observer. One of the Fortec Digital devices failed to function during the test, and was withdrawn. Mean interdevice differences of > 4 mmHg for either systolic or diastolic blood pressure were considered unacceptable.

Home-use phase

Seventeen devices were available for this part of the procedure (three devices of each of five models and two of the Fortec Digital). These were given to 17 hypertensive subjects, each of whom was asked to measure blood pressure and record the results on a diary card twice a day for 1 month, according to the instructions provided with the model. These subjects were not given any special training or instruction, as they would not have received any guidance had they purchased the devices from a retail outlet. Any technical problems associated with the devices were noted during this phase.

After-use interdevice variability

The 17 devices were retested for interdevice variability after the home-use assessment to determine whether there had been any change with use. The same 10 subjects who had taken part in the first interdevice variability study underwent the repeat test. Three standard PyMaH mercury sphygmomanometers were included as a reference standard for the repeat interdevice assessment and for the subsequent validation test.

Device validation

Of the seven models originally submitted for testing, only four (the Omron, the Philips HP5308, the Nissei Analogue Monitor and the Philips HP5306/B) passed the interdevice variability tests before and after 1 month of home use, and these went forward to the validation test. One device for each model was arbitrarily selected from the devices used in the home test.

In order to permit simultaneous measurement, the test device was connected via a T-tube connector to a standard mercury manometer. In the devices with automatic

deflation (all except the Nissei Analogue Monitor), the deflation rate was controlled by the device. The Nissei was connected via a T-tube connector to a standard mercury manometer and an inflatable bulb which was deflated at 2 mmHg/s. We studied 85 subjects, ranging in age from 15 to 80 years, with blood pressures ranging from 88/58 to 260/138 mmHg. The cuff of the device was placed on the subject's arm, with the microphone or stethoscope head located over the brachial artery. The observers recorded pressure simultaneously using a double-headed stethoscope, each observer being blinded to the others' readings and to the test-device measurement, which was recorded by a third observer. Measurements were made simultaneously by each of the two 'blinded' observers using the same mercury sphygmomanometer and by a third observer with the test device in 85 subjects. The procedure was performed three times to give a total of 255 readings per device.

Analysis

Interdevice variability

The mean, standard deviation and mean difference were calculated for each device comparison (Tables 2 and 3).

Device validation

Three methods of analysis were used.

Clinical criteria. In this analysis we attempted to relate the acceptable degree of inaccuracy for devices for self-measurement of blood pressure to the circumstances of their use and the likely consequences of error during that use for the clinical management of an individual patient. Permitted error limits were graded in three categories (Table 4). The 255 measurements from each of the three observers were compared (Table 5) in the following sequence: first, the 255 measurements obtained by each of the mercury sphygmomanometer observers were compared to confirm observer agreement; then, the 255 measurements obtained by each of the mercury sphygmomanometer observers were compared with those made using the test sphygmomanometer. The permitted errors (Table 4) were then applied to the data and the model was given a 'pass' or 'fail' designation (Table 5). In order to display this method of analysis graphically, the pressure differences between the mercury standard and the test device were plotted against the mean pressures, and the error limits superimposed (Fig. 2, left panels). The number of points between the zero-line and the next line are grade I errors, those between the next two

Table 2. Interdevice variability before home use.

Model and device	n	Mean (mmHg)		s.d. (mmHg)		Mean diff. (mmHg)		Model and device	n	Mean (mmHg)		s.d. (mmHg)		Mean diff. (mmHg)	
		SBP	DBP	SBP	DBP	SBP	DBP			SBP	DBP	SBP	DBP	SBP	DBP
Omron								Fortec							
1	20	167.7	97.7	34.8	10.2	9.9	−1.6	1	20	159.6	91.5	36.6	13.7	3.0	−0.3
2	20	157.7	99.2	15.3	10.5			2	20	156.6	91.8	39.1	9.8		
1	20	167.7	97.7	34.8	10.2	14.2	0.8	1	20	159.6	91.5	36.6	13.7	−2.1	−9.7
3	20	153.5	96.8	16.0	12.1			3	20	161.7	101.2	41.4	37.9		
2	20	157.7	99.2	15.3	10.5	4.2	2.4	2	20	156.6	91.8	39.1	9.8	−5.1	−9.3
3	20	153.5	96.8	16.0	12.1			3	20	161.7	101.2	41.4	37.9		
HP5308								Nissei							
1	18	152.3	87.9	18.7	12.7	3.7	−3.5	1	20	173.6	101.2	29.3	13.6	−0.8	−0.5
2	18	148.7	91.4	20.9	10.8			2	20	174.4	101.7	27.9	12.5		
1	18	152.3	87.9	18.7	12.7	0.4	−2.4	1	20	173.6	101.2	29.3	13.6	1.6	−0.8
3	18	151.9	90.4	21.1	12.1			3	20	172.0	102.0	28.4	15.7		
2	18	148.7	91.4	20.9	10.8	−3.2	1.1	2	20	174.4	101.7	27.9	12.5	2.4	−0.3
3	18	151.9	90.4	21.1	12.1			3	20	172.0	102.0	28.4	15.7		
Cuffless								HP5306							
1	16	173.0	90.7	36.3	24.7	−3.0	−5.9	1	20	163.7	95.9	17.5	9.9	4.2	0.8
2	17	176.0	96.6	39.1	26.9			2	20	159.6	95.2	24.3	8.8		
1	16	173.0	90.7	36.3	24.7	2.4	−5.0	1	20	163.7	95.9	17.5	9.9	5.6	0.1
3	17	170.6	95.8	28.8	22.3			3	20	158.1	95.8	26.6	10.8		
2	17	176.0	96.6	39.1	26.9	5.4	0.9	2	20	159.6	95.2	24.3	8.8	1.4	−0.7
3	17	170.6	95.8	28.8	22.3			3	20	158.1	95.8	26.6	10.8		

n, Number of measurements obtained; diff., difference; SBP, systolic blood pressure; DBP, diastolic blood pressure.

Table 3. Interdevice variability after home use.

Model and device	n	Mean (mmHg)		s.d. (mmHg)		Mean diff. (mmHg)		Model and device	n	Mean (mmHg)		s.d. (mmHg)		Mean diff. (mmHg)	
		SBP	DBP	SBP	DBP	SBP	DBP			SBP	DBP	SBP	DBP	SBP	DBP
Omron								Nissei							
1	20	170.1	86.8	19.3	9.7	33.2	2.0	1	20	139.0	83.7	30.9	10.3	2.2	2.4
2	20	136.9	84.8	23.1	8.8			2	20	136.8	81.3	27.7	9.0		
1	20	170.1	86.8	19.3	9.7	29.0	1.9	1	20	139.0	83.7	30.9	10.3	1.3	1.9
3	20	141.1	84.9	21.3	12.9			3	20	137.7	81.8	29.8	9.1		
2	20	136.9	84.8	23.1	8.8	−4.2	−0.1	2	20	136.8	81.3	27.7	9.0	−0.9	−0.9
3	20	141.1	84.9	21.3	12.9			3	20	137.7	81.8	29.8	9.1		
HP 5308								HP5306							
1	20	136.0	81.8	23.8	10.9	0.1	−1.0	1	0						
2	20	135.9	82.8	24.4	11.3			2	10	135.1	81.3	25.8	15.6		
1	20	136.0	81.8	23.8	10.9	−1.9	1.5	1	0						
3	20	137.9	80.3	25.0	10.2			3	10	140.5	86.6	27.7	12.5		
2	20	135.9	82.8	24.4	11.3	−2.0	2.5	2	10	135.1	81.3	25.8	15.6	−5.4	−5.3
3	20	137.9	80.3	25.0	10.2			3	10	140.5	86.6	27.7	12.5		
Cuffless								Mercury							
1	14	144.4	92.9	27.2	12.6	−6.9	11.9	1	20	137.9	86.1	26.9	8.3	−0.3	0.2
2	15	151.3	81.0	32.6	19.2			2	20	138.2	85.9	28.4	6.9		
1	14	144.4	92.9	27.2	12.6	−10.5	−3.4	1	20	137.9	86.1	26.9	8.3	1.4	0.1
3	16	154.8	96.4	33.5	30.3			3	20	136.5	86.0	26.3	8.9		
2	15	151.3	81.0	32.6	19.2	−3.5	−15.4	2	20	138.2	85.9	28.4	6.9	1.7	−0.1
3	16	154.8	96.4	33.5	30.3			3	20	136.5	86.0	26.3	8.9		

n, Number of measurements obtained; diff., difference; SBP, systolic blood pressure; DBP, diastolic blood pressure.

lines are grade II errors and those outside the top and bottom lines are grade III errors.

AAMI standard. The AAMI standard [18] for device acceptability is that the mean difference of paired measurements made with the test device and the standard device should be ≤ 5 mmHg, with a standard deviation of ≤ 8 mmHg for either systolic or diastolic blood pressure.

Graphical analysis. In the presentation of validation data it is common practice to begin by producing a scatter plot of the two sets of blood pressure data. The scatter plot can be a useful first step, but is inefficient, as all the information is usually clustered near the line of equality. A better way of assessing the discrepancies is to plot the difference between the measurements against their average, as recommended by Bland and Altman [19] and shown in Fig. 2 (right panels). This plot shows the differences in blood pressure explicitly, and indicates whether the distribution of the differences varies according to the blood pressure level. The standard deviations and 95% limits of agreement can be superimposed on this plot, as in Fig. 2.

Results

All observers met the required limits of agreement. The four who were the most accurate were selected for the validation test.

Before-use interdevice variability

The data for this test are given in Table 2.

Two of the three Systema Digital devices failed to function at the start of the test, because of failure of the inflation bulb, and this model was therefore withdrawn from the study at this stage.

One of the inflation bulbs on one device of the Fortec Digital failed to function during the test: the two remaining devices were left in the test. The Philips HP5308 and the Nissei Analogue Monitor satisfied the agreement criteria for both systolic and diastolic blood pressures. One of the Philips HP5306/B devices only just failed for systolic blood pressure (mean difference 5.6 mmHg). The Om-

Table 4. Permitted error limits.

	Grade	Error (mmHg)	Readings permitted in error range (%)
Systolic blood pressure	I	≤10	100
	II	11–15	20
	III	>15	5
Diastolic blood pressure <60 mmHg	I	≤10	100
	II	11–15	20
	III	>15	5
Diastolic blood pressure >100 mmHg	I	≤5	100
	II	6–10	20
	III	>10	5
Diastolic blood pressure 60–100 mmHg	Graded linearly between diastolic blood pressure <60 and >100 mmHg		

The percentage of readings permitted is the maximum percentage allowed for each grade. For the first grade 100% of readings within ≤10 mmHg are acceptable, but for grades II and III >20 and >5%, respectively, fail the test. The least error is permitted in the diastolic range 80–120 mmHg as this is usually the critical area for decision-making in hypertension.

ron HEM-400C was unacceptable for measuring systolic blood pressure but was within acceptable limits for diastolic blood pressure recordings. There was an unacceptable level of interdevice variability with the Healthcheck 'Cuffless' and the Fortec Digital for both systolic and diastolic blood pressures. It was decided not to withdraw any models at this stage, even though the interdevice variability criteria were not met in all cases.

Home-use phase

In this phase of the study, the devices were subjected to the use for which they were designed before being submitted to the main validation test. The following is a summary of the 17 users' assessments of the devices and the comments of the nurse observers.

Omron

All three users were dissatisfied with this model, mainly because it frequently failed to give readings. The manufacturer's instructions were considered adequate by the users. The nurses found the cuff suitable for conical arms but not for the more usual cylindrically shaped arms.

Philips HP5308

Only one of the three users was satisfied with this model. One found it difficult to steady the device while measuring blood pressure. The manufacturer's instructions were considered adequate by the users. The nurses found it difficult to control deflation.

Healthcheck 'Cuffless'

All three users were satisfied with this model, which they found easy to operate, and the manufacturer's instructions were considered adequate. However, the nurses found it necessary to manipulate the battery connections on many occasions in order to activate the sphygmomanometer, and since many error codes were recorded, it was necessary to perform many repeat measurements in order to obtain a reading.

Nissei Analogue Monitor

Two of the users found it difficult to hear the Korotkoff sounds with the stethoscope, and the aneroid needle tended to stick in one device. The manufacturer's instructions were considered adequate by two of the three users. The nurses found the stethoscope unsatisfactory, especially for auscultation of the Korotkoff sounds (phase V).

Philips HP5306/B

The three users commented that they had to take a number of measurements before they could be satisfied that they had an accurate reading. Two of the three users found the manufacturer's instructions for use inadequate. The nurses considered the inflation bulb unsatisfactory.

Fortec Digital

Comments were available from only two users, as one device failed to function in the interdevice variability test. Both users found that this model functioned erratically, and one gave up trying to measure his blood pressure because of frustration in obtaining readings. Both the users and nurses found the inflation bulb difficult to control.

After-use interdevice variability

The data for this test are given in Table 3. The inflation bulb in the two Fortec Digital devices failed to function in the repeat interdevice variability test, and the device could not be assessed further.

Both the Philips HP5308 and the Nissei Analogue Monitor showed no change in interdevice variability after 1 month in use, and both models were passed forward to the main validation test without further consideration. The Healthcheck 'Cuffless' model showed considerable interdevice variability for both systolic and diastolic blood pressures which could not be attributed to a fault in one device, and it was not allowed forward for the main validation test. The inflation bulb in one of the Philips HP5306/B devices failed to function at the end of the home-use phase, and this device was identified as that contributing to the poor performance for systolic interdevice variability in the first test. As the interdevice variability of the two remaining devices was acceptable for both systolic and diastolic blood pressures this model was allowed forward to the main validation procedure. The Omron HEM-400C, which had shown poor interdevice agreement in the first test for systolic blood pressure, but acceptable diastolic agreement, again showed acceptable diastolic blood pressure values, whereas systolic agreement was poor. However, as the results of the test showed that device no. 1 was inaccurate, the Omron HEM-400C was allowed forward to the main test, with the observation that one out of the three devices purchased was discordant. The PyMaH control sphygmomanometer showed excellent interdevice agreement for both systolic and diastolic blood pressures.

Device validation

Analyses were performed on more than 250 measurements for each of the two mercury sphygmomanome-

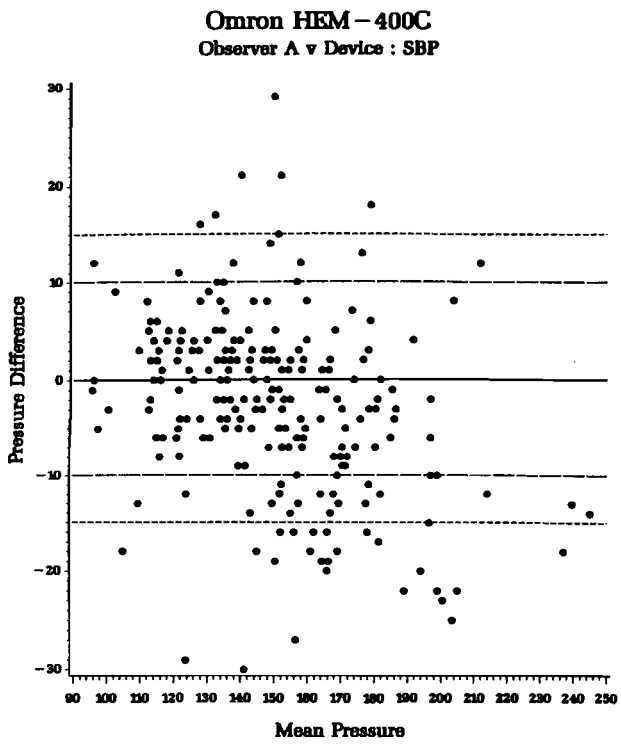
Table 5. Model validation.

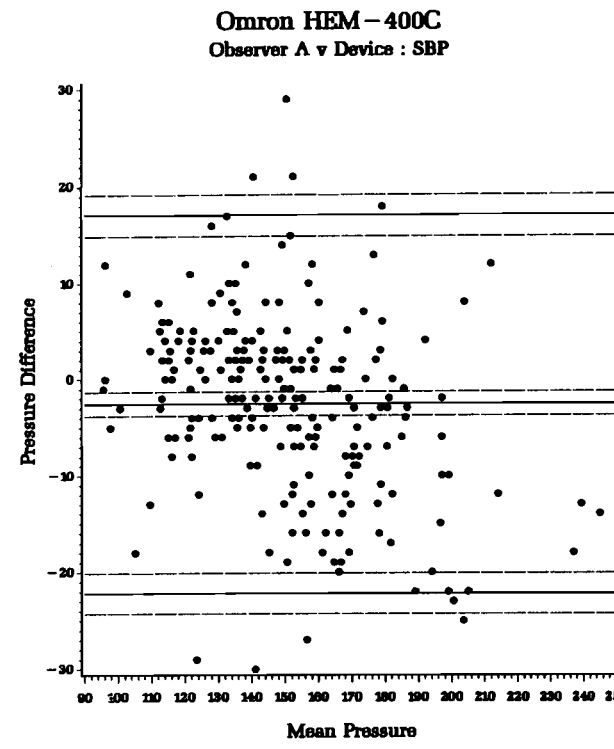
Obs./ device	n	Mean (mmHg)		s.d. (mmHg)		Differences (mmHg) Mean		s.d.		Error rate (%) Grade I		Grade II		Grade III		Pass–Fail	
		SBP	DBP	SBP	DBP	SBP	DBP	SBP	DBP	SBP	DBP	SBP	DBP	SBP	DBP	SBP	DBP
2	252	152.4	89.2	27.6	13.2	1.9	0.6	5.1	4.4	94.8	89.7	2.4	7.5	2.8	2.8	Pass	Pass
1	252	150.7	88.6	28.2	14.1												
Omron	252	148.1	88.6	25.1	14.1	−2.6	0.0	9.8	9.2	76.2	70.2	10.7	15.1	13.1	14.7	Fail	Fail
1	252	150.7	88.6	28.2	14.1												
Omron	252	148.1	88.6	25.1	14.1	−4.3	−0.5	8.8	8.2	79.0	72.6	9.9	13.1	11.1	14.3	Fail	Fail
2	252	152.4	89.2	27.6	13.2												
2	255	150.2	89.8	25.3	12.5	1.3	−0.0	4.7	3.8	97.6	94.1	1.6	5.1	0.8	0.8	Pass	Pass
1	255	148.9	89.8	25.6	13.5												
HP5308	255	146.1	85.7	26.2	13.3	−2.8	−4.2	5.4	4.8	92.9	76.1	4.3	17.6	2.7	6.3	Pass	Fail
1	255	148.9	89.8	25.6	13.5												
HP5308	255	146.1	85.7	26.2	13.3	−4.1	−4.1	5.3	4.5	91.0	74.1	5.9	20.8	3.1	5.1	Pass	Fail
2	255	150.2	89.8	25.3	12.5												
4	255	148.7	87.1	25.0	12.9	−0.1	1.4	3.7	2.8	99.2	94.1	0.0	5.5	0.8	0.4	Pass	Pass
3	255	148.9	85.7	24.8	12.8												
Nissei	255	148.8	81.4	25.9	13.1	−0.1	−4.3	4.4	4.6	96.9	71.4	1.6	22.4	1.6	6.3	Pass	Fail
3	255	148.9	85.7	24.8	12.8												
Nissei	255	148.8	81.4	25.9	13.1	0.1	−5.7	5.0	4.4	96.9	63.9	1.2	28.2	2.0	7.8	Pass	Fail
4	255	148.7	87.1	25.0	12.9												
4	255	152.9	87.8	26.0	13.9	−0.3	0.5	3.0	2.6	98.8	96.5	0.4	2.7	0.8	0.8	Pass	Pass
3	255	153.2	87.4	26.2	14.0												
HP5306	255	150.9	84.4	26.2	16.1	−2.2	−3.0	9.2	11.3	83.5	65.9	9.8	23.5	6.7	10.6	Fail	Fail
3	255	153.2	87.4	26.2	14.0												
HP5306	255	150.9	84.4	26.2	16.1	−1.9	−3.4	9.0	11.2	83.9	65.1	9.0	22.0	7.1	12.9	Fail	Fail
4	255	152.9	87.8	26.0	13.9												
4	255	150.0	87.6	25.7	12.2	0.3	1.3	3.1	2.7	99.2	94.9	0.4	4.7	0.4	0.4	Pass	Pass
3	255	149.7	86.3	25.4	12.0												
Mercury	255	150.9	87.4	25.9	12.0	1.1	1.2	3.0	2.7	99.2	96.5	0.4	3.1	0.4	0.4	Pass	Pass
3	255	149.7	86.3	25.4	12.0												
Mercury	255	150.9	87.4	25.9	12.0	0.9	−0.1	2.8	2.4	99.2	97.6	0.8	2.4	0.0	0.0	Pass	Pass
4	255	150.0	87.6	25.7	12.2												

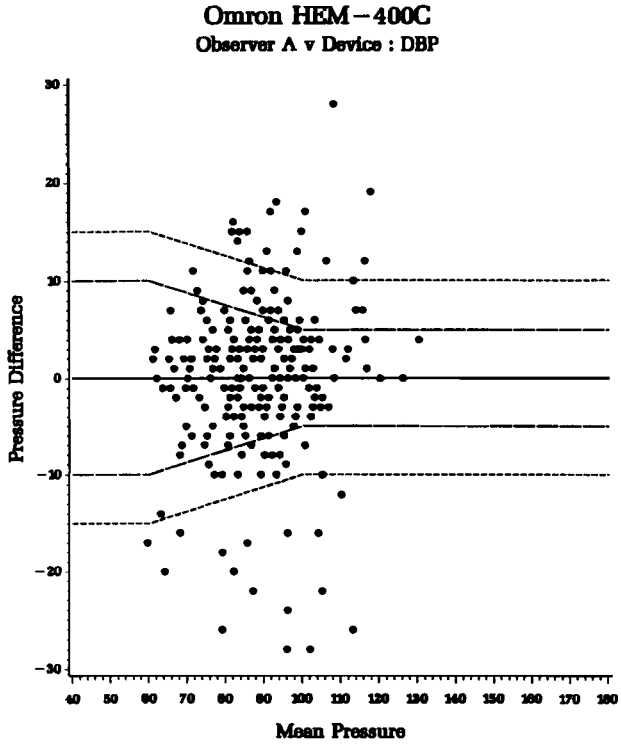
Obs., observer; n, number of readings; SBP, systolic blood pressure; DBP, diastolic blood pressure. Pass means ≤20% grade II errors and ≤5% grade III errors (Table 4).

ter observers and for the test device, first between each observer, and then between each observer and the test device. The results are presented in Table 5. The distribution of systolic blood pressure in the 85 subjects was <140 (n = 91 measurements), 140–179 (n = 125), 180–219 (n = 38) and >220 mmHg (n = 1), and of diastolic blood pressure <80 (n = 72), 80–99 (n = 137) and >100 (n = 46).

Our criteria for clinical accuracy, as outlined in Table 4, are presented graphically in Fig. 2 (left panels) as they apply to the data. The method of Bland and Altman [19] is shown graphically in Fig. 2 (right panels), and the confidence limits are also indicated. Our findings in the main validation test are summarized below.







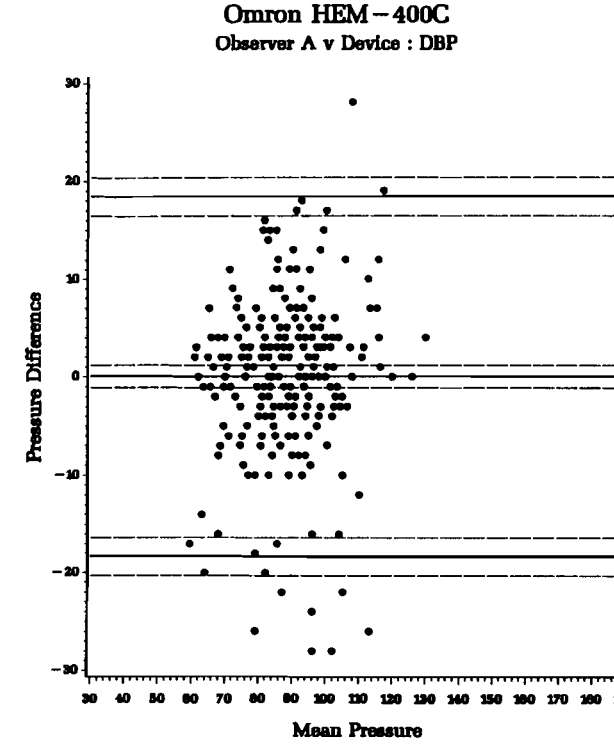
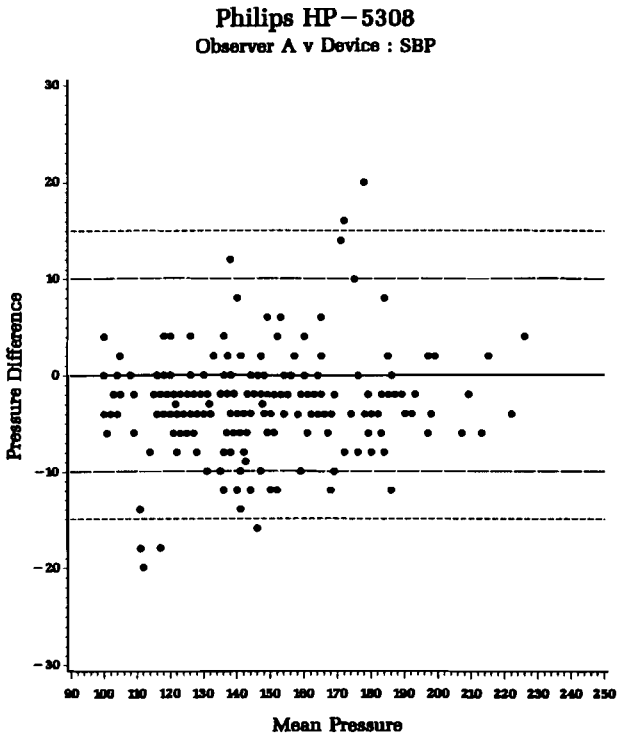
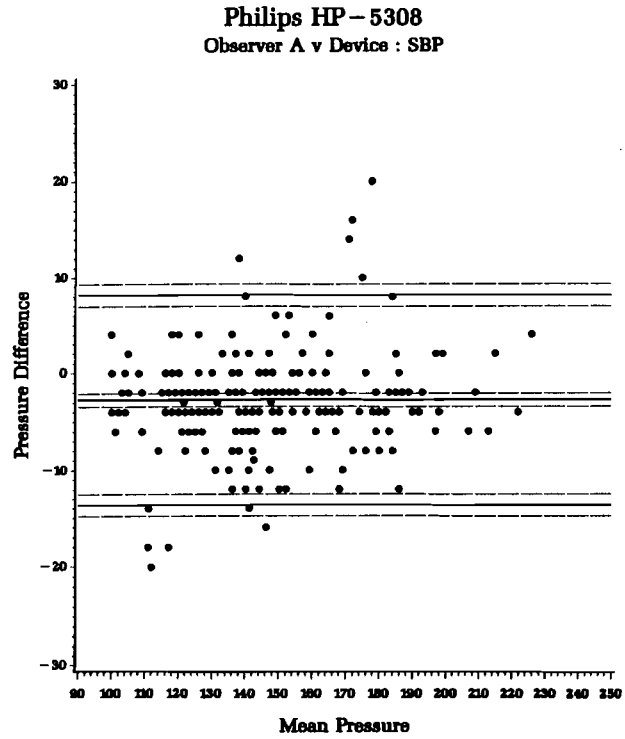


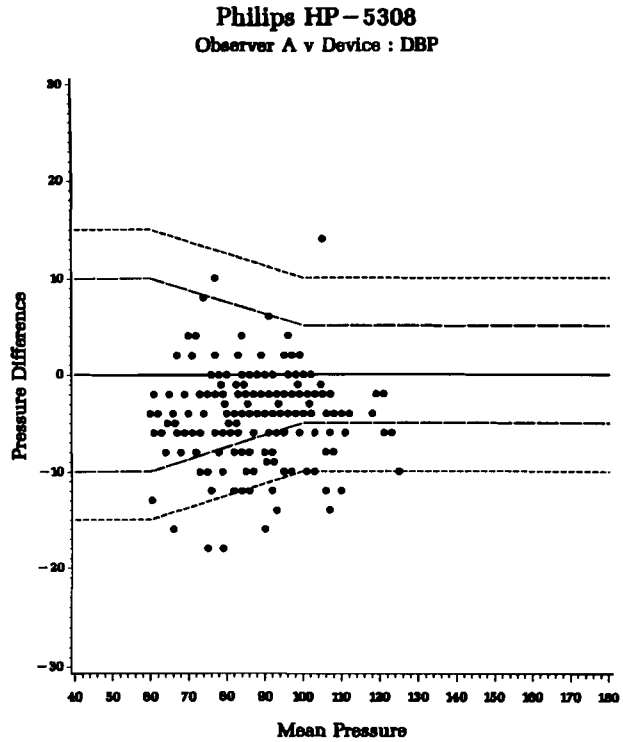
Fig. 2. Left panels: graphical presentation of 'clinical' criteria for accuracy for the Omron HEM-400C, the Philips HP5308, the Nissei Analogue Monitor, the Philips HP5306/B and the PyMaH mercury sphygmomanometers. Grade I errors are within the ± bands closest to zero, grade II errors in the next band and grade III errors in the outermost band. Right panels: graphical presentation of the same data by the method of Bland and Altman [19]. The data plotted in these graphs are for one observer only. Similar plots were obtained for the second observer. SBP, systolic blood pressure; DBP, diastolic blood pressure.



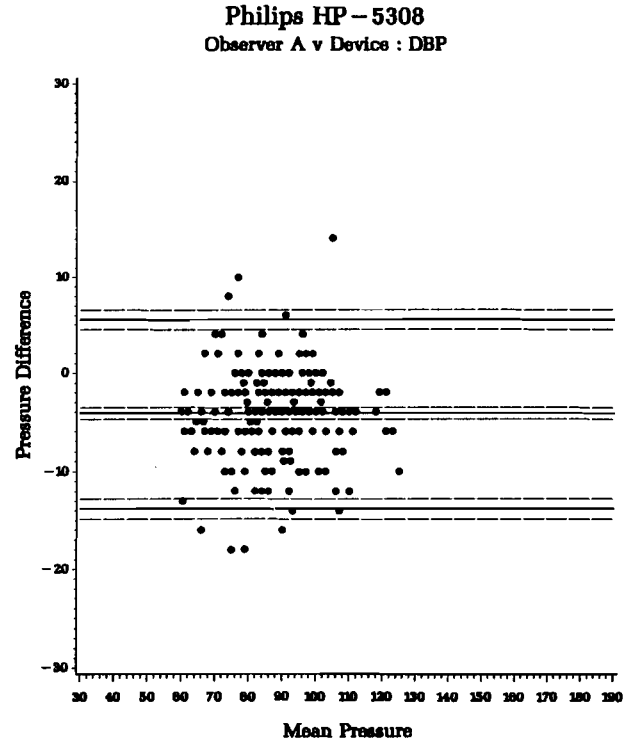
4.3% Grade II errors, 2.7% Grade III errors – PASS



Mean (−2.8) ±2 SD (5.4) with 95% Confidence Limits

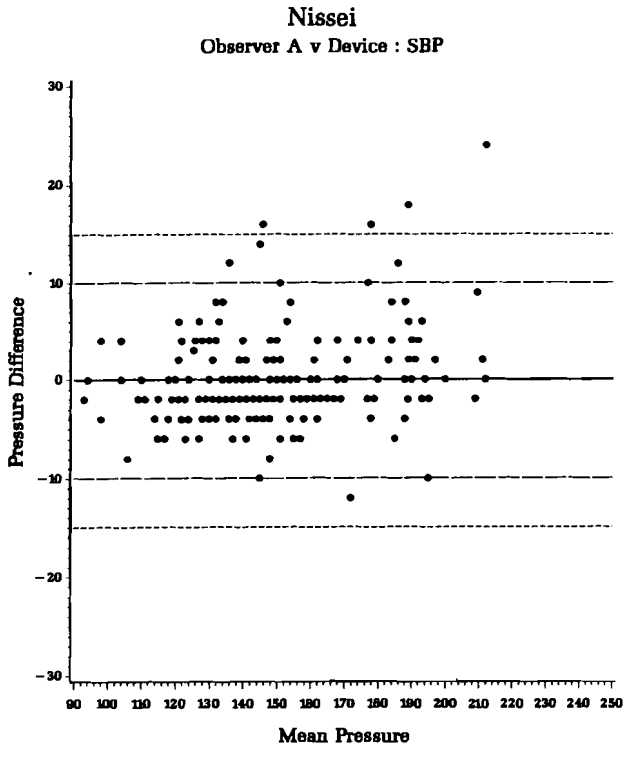


17.6% Grade II errors, 6.3% Grade III errors – FAIL

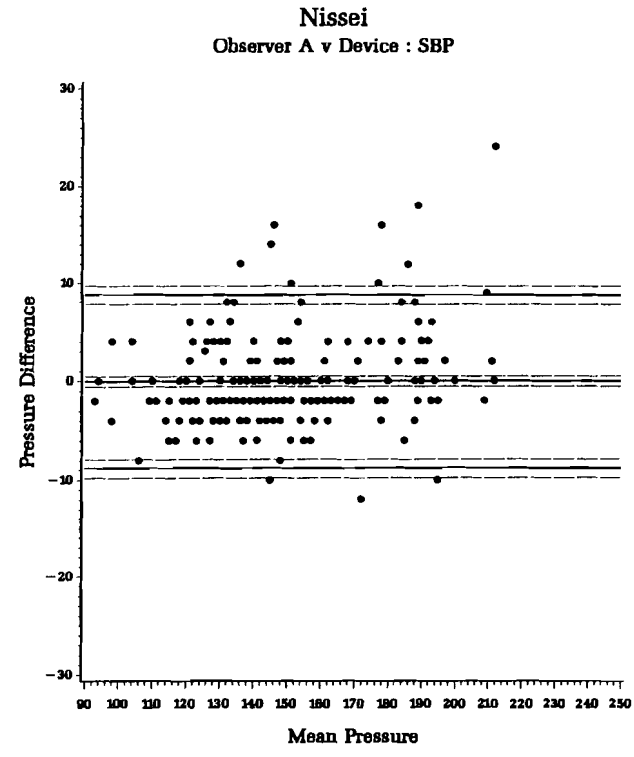


Mean (−4.2) ±2 SD (4.8) with 95% Confidence Limits

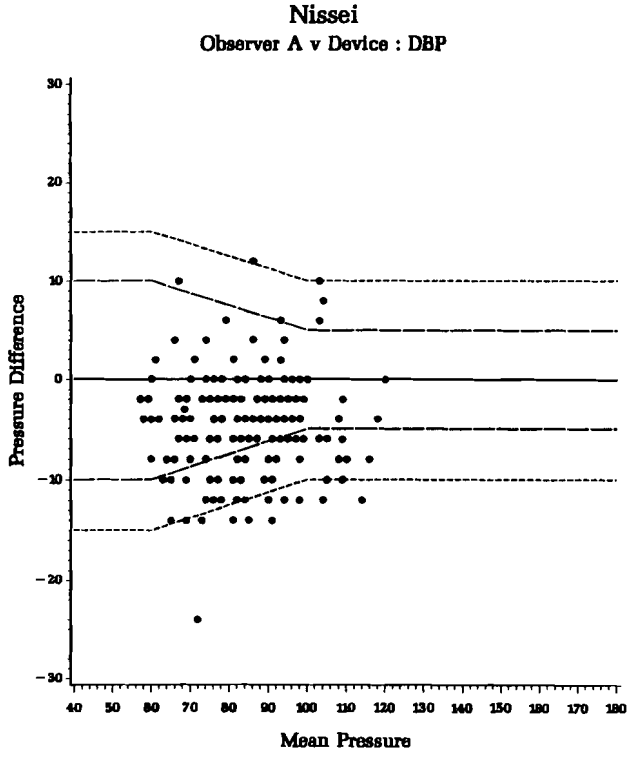
Fig. 2. (cont.: Philips HP5308)



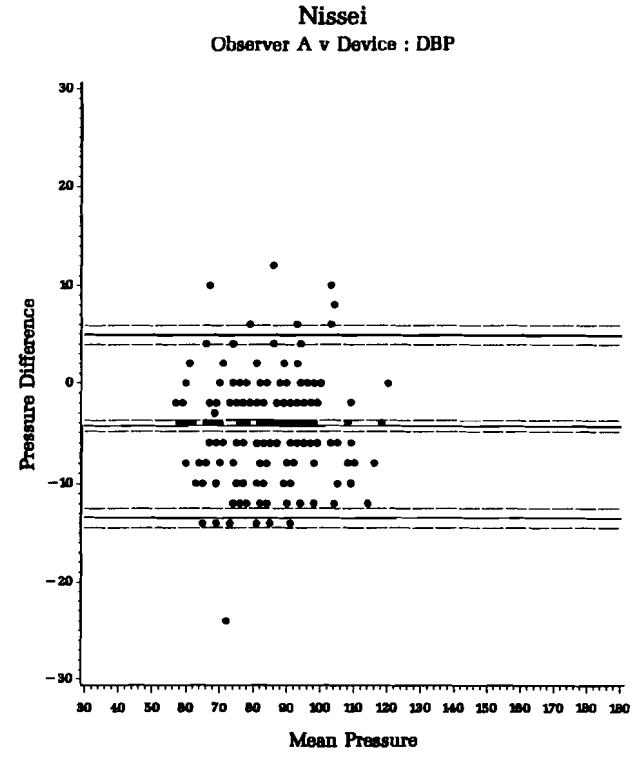
1.6% Grade II errors, 1.6% Grade III errors – PASS



Mean (−0.1) ±2 SD (4.4) with 95% Confidence Limits

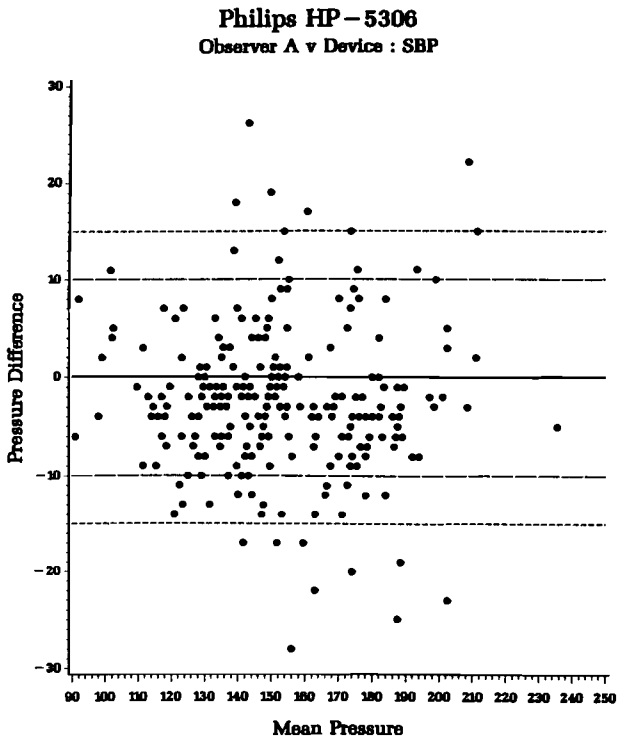


22.4% Grade II errors, 6.3% Grade III errors – FAIL

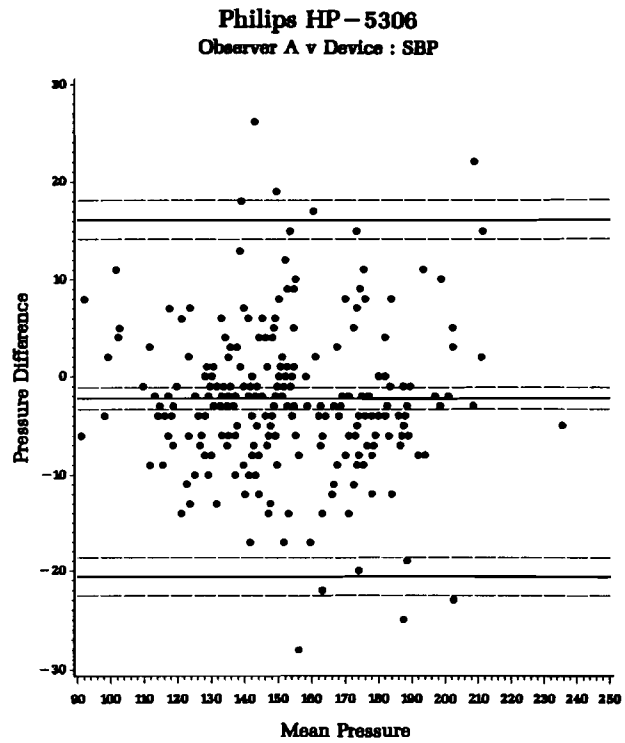


Mean (−4.3) ±2 SD (4.6) with 95% Confidence Limits

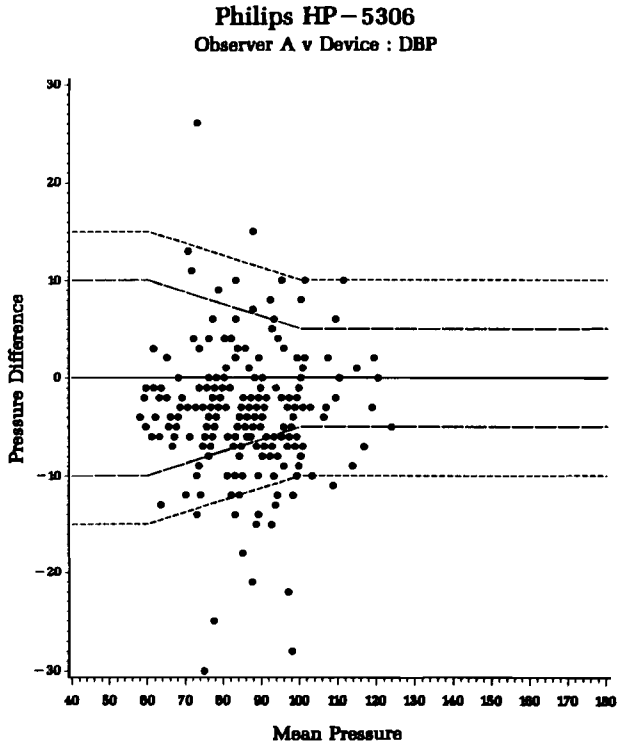
Fig. 2. (cont.: Nissei Analogue Monitor)



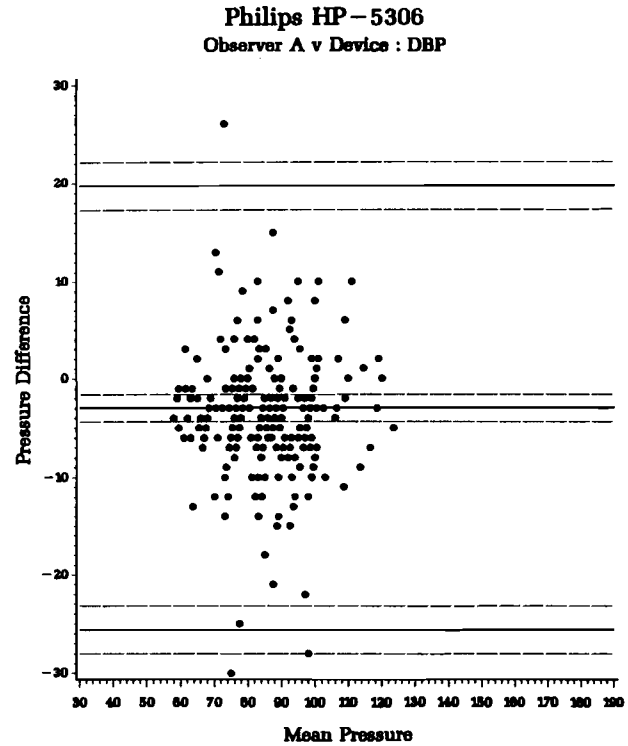
9.8% Grade II errors, 6.7% Grade III errors – FAIL



Mean (−2.2) ±2 SD (9.2) with 95% Confidence Limits

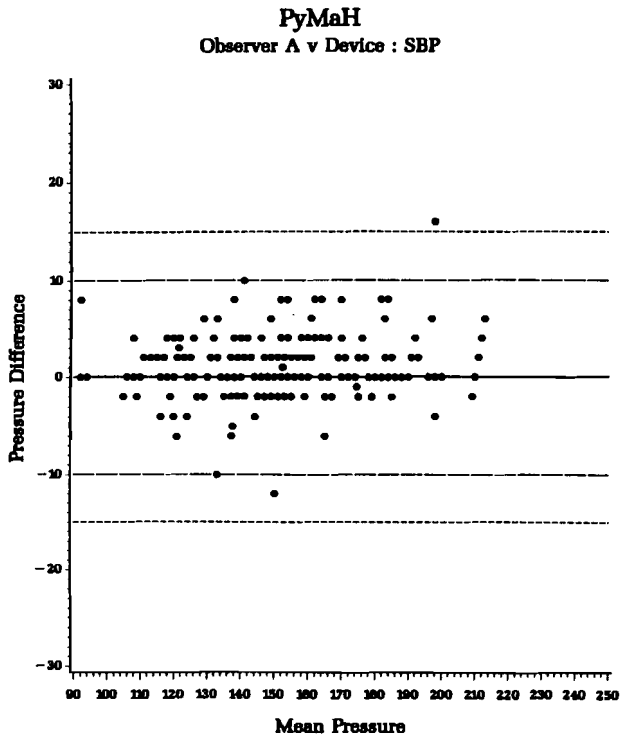


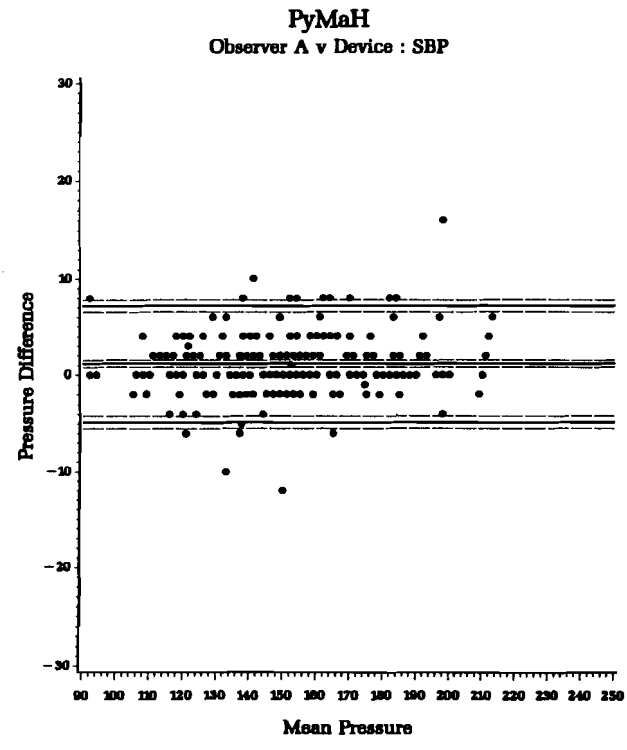
23.5% Grade II errors, 10.6% Grade III errors – FAIL

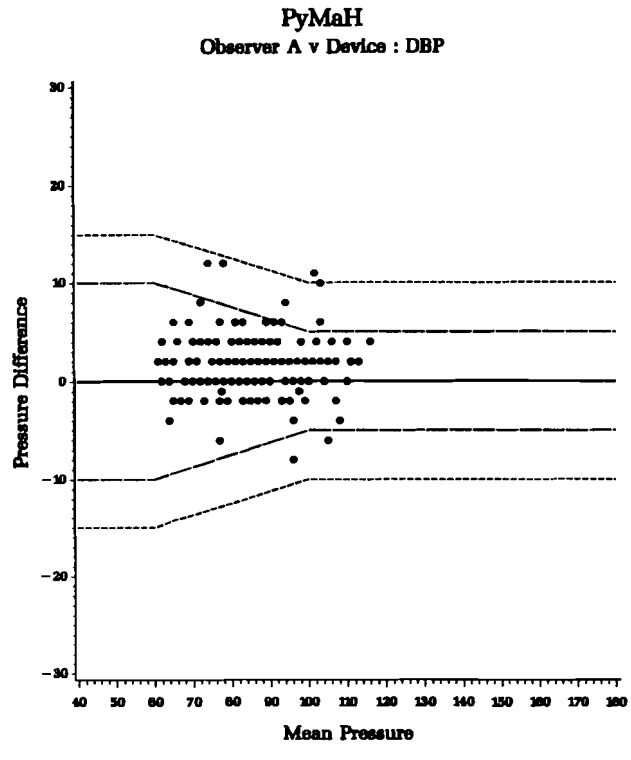


Mean (−3.0) ±2 SD (11.3) with 95% Confidence Limits

Fig. 2. (cont.: Philips HP5306/B)







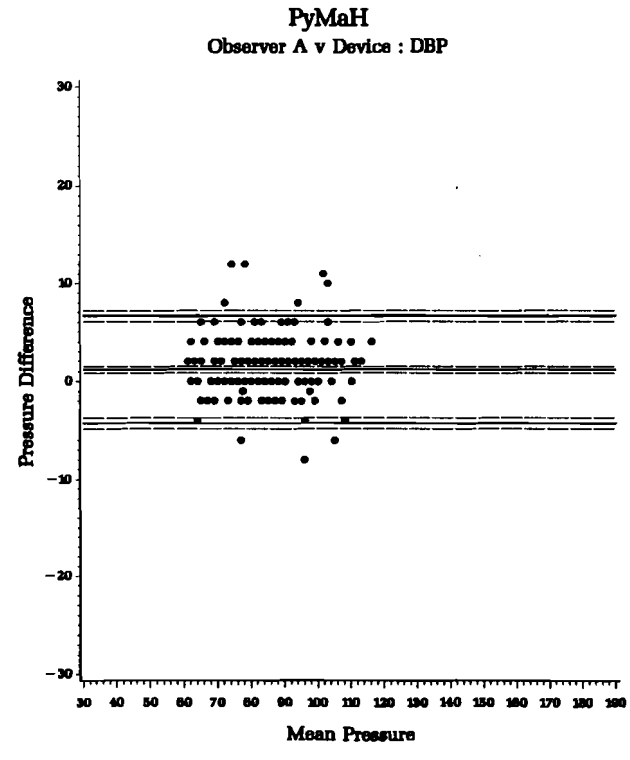


Fig. 2. (cont.: PyMaH mercury sphymomanometer)

Omron HEM-400C
This model failed the AAMI and the 'clinical' analysis.

Philips HP5308
This model satisfied the criteria of the AAMI standard. In our 'clinical' analysis it performed within the acceptable error limits for systolic blood pressure and was within the limit for grade II errors, but just outside the limit for grade III errors, for diastolic blood pressure.

Nissei Analogue Monitor
This model satisfied the criteria of the AAMI standard. It performed within the acceptable limits for systolic blood pressure for the 'clinical' criteria, but was well outside the limits for diastolic blood pressure for both grade II and grade III errors.

Philips HP5306/B
This model failed the AAMI standard and the 'clinical' analysis.

PyMaH
The standard sphygmomanometer was comfortably within the required limits for both systolic and diastolic blood pressures for both methods of assessment.

Discussion

The protocol design for this study had a number of unique features. Observer agreement was determined before the study began rather than afterwards, as recommended in the AAMI standard [18]. This seems sensible, since if observer agreement is shown to be unacceptable after the entire procedure has been completed, the results will be invalidated and the study must then be repeated. It is essential to train observers for validation studies in which small inaccuracies in the performance of the observers may have an important influence on the results. The achievement of an observer accuracy whereby 85% of measurements do not differ by more than 5 mmHg calls for a serious commitment to the training phase of the protocol. The accuracy of 'trained' observers cannot be assumed, and observers should be assessed before any validation study.

By assessing interdevice variability before and after a period of use, it is possible to identify those models that are either inaccurate at the time of distribution for sale or become inaccurate with use. There is nothing to be gained in putting discordant devices through the main validation procedure, which is costly and labour-intensive. In most previous studies it has been customary to perform validation tests on brand-new devices. However, because of past experience with aneroid manometers [25], it has been suggested that validation should also be conducted after a period of use. One model, the Systema Digital, failed to function in the first interdevice test, and the Healthcheck 'Cuffless' showed such interdevice variability after use that this model was not allowed forward to the main validation test. Indeed, if truly strict criteria had been applied, the Omron HEM-300C and the Philips HP5306/B would have been withdrawn after the second interdevice variability test; however, because only one of the three devices in each case appeared to be at fault, these models were included in the main validation test. Both subsequently failed the two methods of assessment, but even if they had proved accurate, their excessive interdevice variability (in all three devices of each model) would have adversely affected any positive recommendation. If manufacturers cannot produce a device so that it is accurate at the time of sale as well as after a reasonable period of use, then that product cannot be recommended to consumers.

In validation procedures it is customary to present inaccuracies in statistical terms which are often unrelated to the significance of those inaccuracies in clinical practice. Our accuracy requirements were based on the relevance of device accuracy to the clinical management of hypertension in the setting of self-measurement of blood pressure. The same reasoning would not apply for sphygmomanometers designed for use in hospital practice or research, where the requirement for accuracy would be greater. Our accuracy criteria were selected with the consideration that clinical decisions are based, for the most part, on diastolic blood pressure levels. In clinical practice, it is unlikely that measurement errors at the extremes of the systolic pressure range would call for a change in patient management, whereas relatively small errors in the measurement of diastolic blood pressure might have important consequences. For example, a 10 mmHg error at a systolic blood pressure level of 170 mmHg or a diastolic blood pressure level of 60 mmHg is unlikely to influence management much, whereas this level of inaccuracy in the diastolic pressure range 90–120 mmHg could have important implications. In self-measurement of blood pressure, sphygmomanometer inaccuracies of ≤ 10 mmHg, for all levels of systolic blood pressure, are unlikely to have any effect on management. However, errors of between 11 and 15 mmHg may have some influence on management, and errors >15 mmHg are likely to have relevance. For diastolic blood pressure, the same reasoning applies for pressures of ≤ 60 mmHg, but for pressures >100 mmHg errors of >5 mmHg could affect management, with errors of 10 mmHg becoming increasingly important. We have therefore graded the errors as acceptable for self-measurement on the basis of their likely influence on the management of a patient who might be using self-measurement. These criteria, which are slightly more strict than the AAMI standard, are more relevant to clinical practice.

In the statistical analysis we did not use the correlation coefficient, which can be very misleading in that it may be high when there are gross differences between two devices [19,26]. In this study, for example, r was never <0.9 for systolic and <0.7 for diastolic blood pressure, although there were substantial inaccuracies.

The most striking feature of this study was failure of any of these seven low- to medium-priced devices to meet the accuracy criteria for both systolic and diastolic blood pressures. It has been shown in this study that a mercury sphygmomanometer, with a good-quality stethoscope, is the most accurate device available for the measurement of blood pressure. We did not use a random-zero sphygmomanometer [20] as the standard in this study because it has been our experience, as well as that of other workers [21–24], that this device systematically underestimates diastolic blood pressure. It is possible that if the Nissei Analogue Monitor (which, like a mercury sphygmomanometer, is dependent on auscultation of the Korotkoff sounds) had been supplied with a good-quality stethoscope, diastolic blood pressure would have been recorded more accurately. The Philips HP5308 was the only other device that depended on detection of the Korotkoff sounds, in this case by the substitution of a microphone for the human ear. This model only narrowly failed to pass the validation test, but only one of the three users in the home-use phase was satisfied with it, and the nurse observers commented on the poor quality of its inflation bulb. All other models tested used the oscillometric method of blood pressure measurement.

The inferior quality of the components in some of the models also deserves comment. The quality of the inflation bulbs was unsatisfactory in the Philips HP5308, the Philips HP5306/B, the Systema Digital and the Fortec Digital, and even if these models had fulfilled the accuracy criteria, most would not have stood up to the rigours of a prolonged period of use.

The following conclusions can be drawn. First, the most accurate way to measure blood pressure is with a mercury sphygmomanometer and a good quality stethoscope. Mercury sphygmomanometers, such as the one used as the standard in this study, can be readily modified for self-measurement of blood pressure. We know of only one mercury device for home measurement (the PyMaH Home Care Blood Pressure Instrument with BP Scope), but unfortunately this is supplied with a poor quality stethoscope that makes it difficult to use the device accurately.

However, the use of a mercury or aneroid device with a stethoscope (such as the Nissei Analogue Monitor) requires that the subject has full use of both arms and has good hearing and sight. When these conditions cannot be met, a semi-automated device is required. Of those tested, the most accurate of these was the Philips HP5308. The manufacturers of blood pressure-measuring devices must be encouraged to produce a product that is as accurate as a mercury sphygmomanometer.

Acknowledgements

We acknowledge the co-operation of the Consumers' Association, London, which provided the devices for this study, and we are particularly grateful to Ms Sian Morrissy, Project Manager, who co-ordinated many aspects of the study.

References

1. HUNT JC, FROHLICH ED, MOSER M, ROCCELLA EJ, KEIGHLEY EA: **Devices used for self-measurement of blood pressure: revised statement of the National High Blood Pressure Education Program.** *Arch Intern Med* 1985, 145:2231–2234.
2. KAPLAN NM: **Misdiagnosis of systemic hypertension and recommendations for improvement.** *Am J Cardiol* 1987, 60:1383–1386.
3. CANADIAN COALITION FOR HIGH BLOOD PRESSURE PREVENTION AND CONTROL: **Recommendations on self-measurement of blood pressure.** *Can Med Assoc J* 1988, 138:1093–1096.
4. WORLD HYPERTENSION LEAGUE: **Self-measurement of blood pressure.** *Clin Exp Hypertens [A]* 1989, 11:167–171.
5. LABARTHE DR, HAWKINS CM, REMINGTON RD: **Evaluation of performance of selected devices for measuring blood pressure.** *Am J Cardiol* 1973, 32:546–553.
6. HUNYOR SN, FLYNN JM, COCHINEAS C: **Comparison of performance of various sphygmomanometers with intra-arterial blood pressure readings.** *Br Med J* 1978, 2:159–162.
7. BARKER WF, HEDIGER ML, KATZ SH, BOWERS EJ: **Concurrent validity studies of blood pressure instrumentation: the Philadelphia Blood Pressure Project.** *Hypertension* 1984, 6:85–91.
8. SLOAN PJM, ZEZULKA A, DAVIES P, SANGAL A, BEEVERS M, BEEVERS DG: **Standardized methods for comparison of sphygmomanometers.** *J Hypertens* 1984, 2:547–551.
9. STEINER R, LUSCHER T, BOERLIN H-J, SIEGENTHALER W, VETTER W: **Clinical evaluation of semiautomatic blood pressure devices for self-recording.** *J Hypertens* 1985, 3 (suppl 1):23–25.
10. PICKERING TG, CVETKOVSKI B, JAMES GD: **An evaluation of electronic recorders for self-monitoring of blood pressure.** *J Hypertens* 1986, 4 (suppl 5):S328–S330.
11. HAHN LP, FOLSOM AR, SPRAFKA JM, PRINEAS RJ: **Prevalence and accuracy of home sphygmomanometers in an urban population.** *Am J Public Health* 1987, 77:1459–1461.
12. EVANS CD, HAYNES RB, GOLDSMITH CH, HEWSON SA: **Home blood pressure-measuring devices: a comparative study of accuracy.** *J Hypertens* 1989, 7:133–142.
13. ANONYMOUS: **Home-blood pressure measuring kits.** *Which?* August 1989, pp 372–375.
14. ANONYMOUS: **Blood pressure kits.** *Consumer Rep* March 1979, pp 142–146.
15. ANONYMOUS: **Pressure point.** *Can Consumer* July 1982, pp 23–26.
16. ANONYMOUS: **Electronic blood pressure monitors.** *Can Consumer* February 1987, pp 21–24.
17. ANONYMOUS: **Blood pressure monitors.** *Consumer Rep* March 1987, pp 314–319.
18. ASSOCIATION FOR THE ADVANCEMENT OF MEDICAL INSTRUMENTATION: *American National Standard for Electronic or Automated Sphygmomanometers.* Arlington, Virginia: Association for the Advancement of Medical Instrumentation, 1987.
19. BLAND JM, ALTMAN DG: **Statistical methods for assessing agreement between two methods of clinical measurement.** *Lancet* 1986, i:307–310.
20. WRIGHT BM, DORE CF: **A random-zero sphygmomanometer.** *Lancet* 1970, i:337–338.
21. EVANS JG, PRIOR IAM: **Experience with the random-zero sphygmomanometer.** *Br J Prev Soc Med* 1970, 24:10–15.
22. LABARTHE DR, HAWKINS CM, REMINGTON RD: **Evaluation of performance of selected devices for measuring blood pressure.** *Am J Cardiol* 1973, 32:546–553.
23. DE GAUDEMARIS R, FOLSOM AR, PRINEAS RJ, LUEPKER RV: **The random-zero versus the standard mercury sphygmomanometer: a systematic blood pressure difference.** *Am J Epidemiol* 1985, 121:282–290.
24. PARKER D, LIU K, DYER AR, GIUMETTI D, LIAO Y, STAMLER J: **A comparison of the random-zero and standard sphygmomanometers.** *Hypertension* 1988, 11:269–272.
25. O'BRIEN ET, O'MALLEY K: **The sphygmomanometer.** In *ABC of Hypertension.* London: British Medical Association, 1981, p 4.
26. CONWAY C: **Home blood pressure recording.** *Clin Exp Hypertens [A]* 1986, 8:1247–1249.