Limitations of current validation protocols for home blood pressure monitors for individual patients

William Gerin[a], Amy R. Schwartz[b], Joseph E. Schwartz[c], Thomas G. Pickering[a], Karina W. Davidson[a], Jonathan Bress[d], Eoin O'Brien[e] and Neil Atkins[e]

Background Automatic blood pressure monitoring conducted at home is increasingly used in the diagnosis and management of hypertension. We assessed the adequacy of existing British Hypertension Society (BHS) and Association for the Advancement of Medical Instrumentation (AAMI) validation standards for automatic blood pressure monitoring devices.

Subject and methods A theoretical study and an empirical test are presented to estimate the proportion of persons for whom a blood pressure monitor validated according to existing BHS and AAMI standards would be inaccurate.

Results The results suggest that a major limitation of both protocols is the lack of attention given to the number of individual patients for whom a monitor may be inaccurate. A blood pressure monitor that meets the AAMI and BHS validation criteria may report blood pressures in error by more than 5 mmHg for more than half of the people.

Conclusions A validation standard that does not take account of the person-effects on error will lead to a substantial proportion of persons using self-monitors that are systematically inaccurate for that person. *Blood Press Monit* **7:** 313–318 © 2002 Lippincott Williams & Wilkins.

Blood Pressure Monitoring 2002, **7**:313–318

Keywords: hypertension, validation, measurement, blood pressure, blood pressure monitoring

[a]Mount Sinai School of Medicine, New York, NY; [b]Yale University School of Medicine, New Haven, CT; [c]State University of New York, Stony Brook, Stony Brook, NY; [d]Joan and Sanford I. Weill Medical College of Cornell University, New York, NY, USA and [e]Blood Pressure Unit, Beaumont Hospital, Dublin, Ireland.

Sponsorship: Preparation of this manuscript was partially supported by National Institutes of Health, Bethesda, MD, USA, Grants HL47540, HL67677, and HL04458, and partially supported by American Heart Association, Greenfield, TX, USA, Grant 9750544N. Note that portions of the data analyzed for this manuscript have been reported elsewhere.

Potential conflicts of interest: None

Correspondence and requests for reprints to William Gerin, PhD., The Zena and Michael A. Wiener Cardiovascular Institute, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1030, New York, NY, USA, 10029.
Tel: +1 212 659 5610; fax: +1 212 659 5615;
e-mail: william.gerin@msnyuhealth.org

Received 09 July 2002 **Revised** 23 September 2002
Accepted 14 October 2002

Introduction

The effective management of hypertension is critically dependent on the accuracy of blood pressure measurement, since deviations of only a few mmHg may determine whether or not antihypertensive treatment is recommended. Readings taken in the patient's usual environment may be more useful than those taken in the physician's office for diagnosing hypertension and for adjusting therapy [1–5]. One reason for this is that office measurements may not accurately represent the patient's blood pressure during daily life (e.g., 'white-coat hypertension' [2]). Automated self-monitoring thus provides a promising technology for improving blood pressure control [6–8]. Self-monitors allow measurements to be taken in a variety of situations, including at home and at work, and with much greater frequency than is possible at the physician's office. A sizeable literature has demonstrated the superiority of ambulatory monitoring over office blood pressures for predicting target organ damage and prognosis [9,10]; self monitoring may have similar predictive power [11–13] and indeed, is increasingly being used in the diagnosis and treatment of hypertension. Its use has been endorsed by both the sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure [1] and the World Health Organization International Society of Hypertension guidelines [14].

Several consensus reports have recommended that a monitor be considered valid for an individual patient only if it is accurate to within 5 mmHg, and that this be determined by comparing the averages of several readings taken by the automatic monitor and a trained human listener using a stethoscope and mercury column; the difference between the monitor's and observer's readings is referred to as the 'error' [15,16]. In a recent editorial, The Council for High Blood Pressure Research of the American Heart Association noted that few of these instruments have

had adequate validation, and emphasized the importance of accurate blood pressure measurement [17]. Although there is a large literature on the validation of monitors, to our knowledge no one has addressed the clinically relevant issue of their accuracy in individual patients. The purpose of this paper is to demonstrate that the existing monitor validation standards allow for unacceptably high rates of potentially clinically detrimental errors in a substantial number of patients.

Methods

AAMI and BHS validation protocols

The Association for the Advancement of Medical Instrumentation (AAMI) published the first standard for the evaluation of automated blood pressure measurement devices in 1987 [18]. This was followed by a similar standard published in 1990 by the British Hypertension Society (BHS) [19]. The standards were updated and revised in 1992 and 1993 [20–22]. Both protocols require an assessment based on 85 subjects, with three blood pressure measurements recorded for each person, for a total of 255 measurements [23]. The measurement recorded by a trained human observer using a mercury column sphygmomanometer is regarded as the 'gold standard' blood pressure. Readings are taken by the monitor and by a human observer either sequentially or simultaneously. Each monitor reading is compared with the observer's corresponding reading and the difference, or 'error' is computed. The BHS validation protocol also involves assessments to confirm the accuracy of the observers [23].

For the AAMI validation standard, monitor accuracy is determined by calculating the mean and standard deviation (SD) of the errors (i.e., the spread of errors around the mean). It requires that the mean error across the 255 readings (three from each of 85 subjects) be $\pm$ 5 mmHg or less for both systolic and diastolic pressures and that the SD of the errors be 8 mmHg.

The original BHS protocol, while also basing its results on three comparisons from each of 85 subjects, required in addition that 65, 85, and 95% of the readings (when simultaneous comparisons are made) have errors less than or equal to 5, 10, and 15 mmHg respectively. The later version (using sequential readings) requires only 50, 75, and 90% of readings have errors less than 5, 10, and 15 mmHg respectively.

It is important to emphasize that the analyses assess only the distribution of the errors across the *255 individual readings*, not the distribution of average errors observed on *each of the 85 participants*. Implicitly, this approach assumes that there is no between-person variance; that is, no tendency for the monitor to be more accurate for some persons and less accurate for others. If true, then as more readings are taken on any single person, the average error for that person will converge toward the *population* mean error, which would be no more than 5 mmHg for either systolic or diastolic pressure. However, if this assumption is false, and the errors tend to cluster *within persons* (i.e., the spread of the errors within a single person is less than that for the entire population), then as more readings are gathered on a particular person, the average error will converge toward *that person's* true error; if that person's error is more than 5 mmHg, the monitor will do a poor job of providing an accurate average for that person no matter how many measurements are taken. We will demonstrate that errors do tend to cluster within persons, and that the existing protocols, by concentrating solely on the *population* mean error, allow the approval of monitors which are inaccurate for a substantial proportion of people.

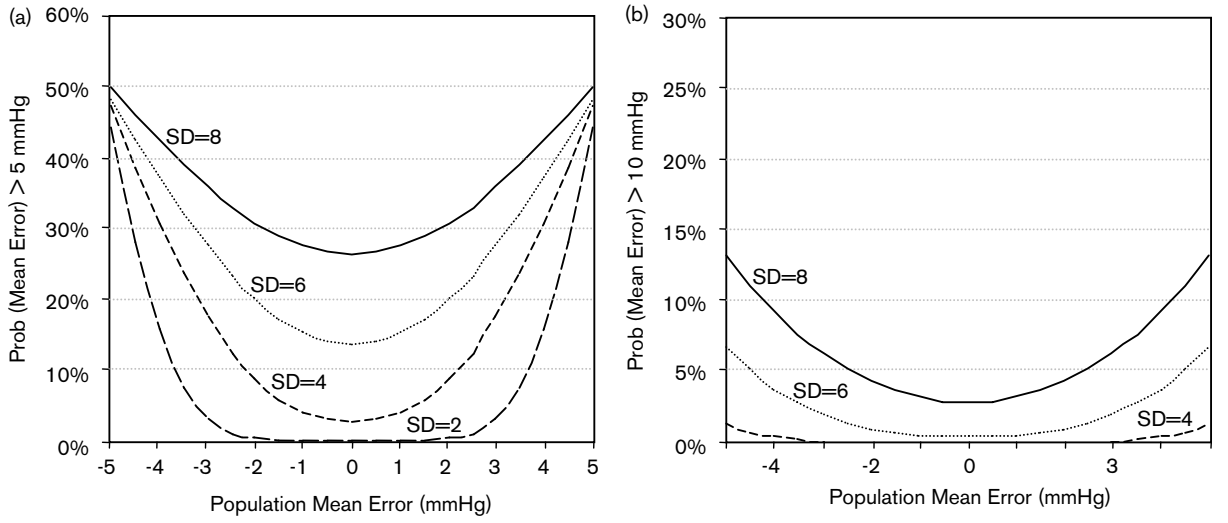
Results

Analytic model

To develop a model describing the proportion of persons receiving accurate readings from a 'validated' monitor, we begin with the unlikely assumption that the blood pressure monitor is equally biased (or unbiased) for all persons; that is, the (true) mean error is identical for everyone and thus there is zero between-person variance. We further assume that the errors are normally distributed, an assumption that is consistent with the available data. Figure 1a shows how the probability that the mean of the three readings will have an error exceeding 5 mmHg varies as a function of the overall, or population, mean error and SD (i.e., of all 255 readings) of the monitor. Figure 1b shows the probability that the mean error exceeds 10 mmHg. One sees that as the population mean error approaches 5 mmHg, the outer limit of acceptable values under both the AAMI and BHS criteria, more than 40% of persons will have a mean error greater than 5 mmHg, even when the population SD is as low as 2 mmHg. Similarly, if the SD of the errors is 8 mmHg, then more than 25% of persons will have average measurements in error by at least 5 mmHg, even when the population mean error is zero.

Unfortunately, the estimates shown in Figure 1 are almost surely too optimistic. To the extent that errors do cluster within persons, the probability of taking three readings with an average error of more than 5 (or 10) mmHg *exceeds* the estimates shown in Figure 1. The increased risk depends on the ratio of the between-person variance (the variability due to differences among people) to the total variance of the errors. This ratio is known as the intra-class correlation (ICC). As the ICC approaches 1.0, errors are more likely to cluster within persons, and, for any given population mean error and SD, the proportion of persons with average errors outside the acceptable range will increase. Figure 2 illustrates this, showing that as the ICC approaches 1.0, well over 50% of persons may have average

Fig. 1



Expected proportion of mean errors (based on three within-person measurements) that will exceed 5 mmHg (a) or 10 mmHg (b) in magnitude as a function of the population mean error and SD, assuming no between-person variance (i.e., intra-class correlation equals zero).

measurements that are in error by more than 5 mmHg. Below, we present empirical evidence suggesting that ICC values between 0.25 and 0.75 are probably realistic.

Empirical study

To empirically estimate the proportion of persons for whom a 'validated' monitor produces inaccurate readings, we examined the raw data from three published validation studies [24–26]. Although both simultaneous and sequential monitor-observer readings were taken in each study, we report only the data measured using the simultaneous readings so as to provide the most optimistic results (only negligible differences were observed between the simultaneous and sequential methods). Both auscultatory (Schiller BR-102 [Schiller AG, Baar, Switzerland] in auscultatory mode and CH-Druck [Disetronic Medical Systems, Burgdorf, Switzerland]) and oscillometric (Schiller BR-102 in oscillometric mode and Profilomat II [Disetronic Medical Systems]) automated monitors were assessed.

While the specific monitors analyzed are used as ambulatory, rather than self, monitors, the relevant aspects of the validation protocols are identical. The summary statistics of two auscultatory monitors, the CH-Druck and Schiller BR-102, for both systolic and diastolic pressures and also the Schiller BR-102 using the oscillometric mode for diastolic blood pressure only were within the validation criteria. For those monitors that passed the validation criteria, 19 to 36% of the *individual readings* had errors outside the range –5 to +5 mmHg, while 4 to 10% had errors exceeding 10 mmHg in magnitude (see Table 1).

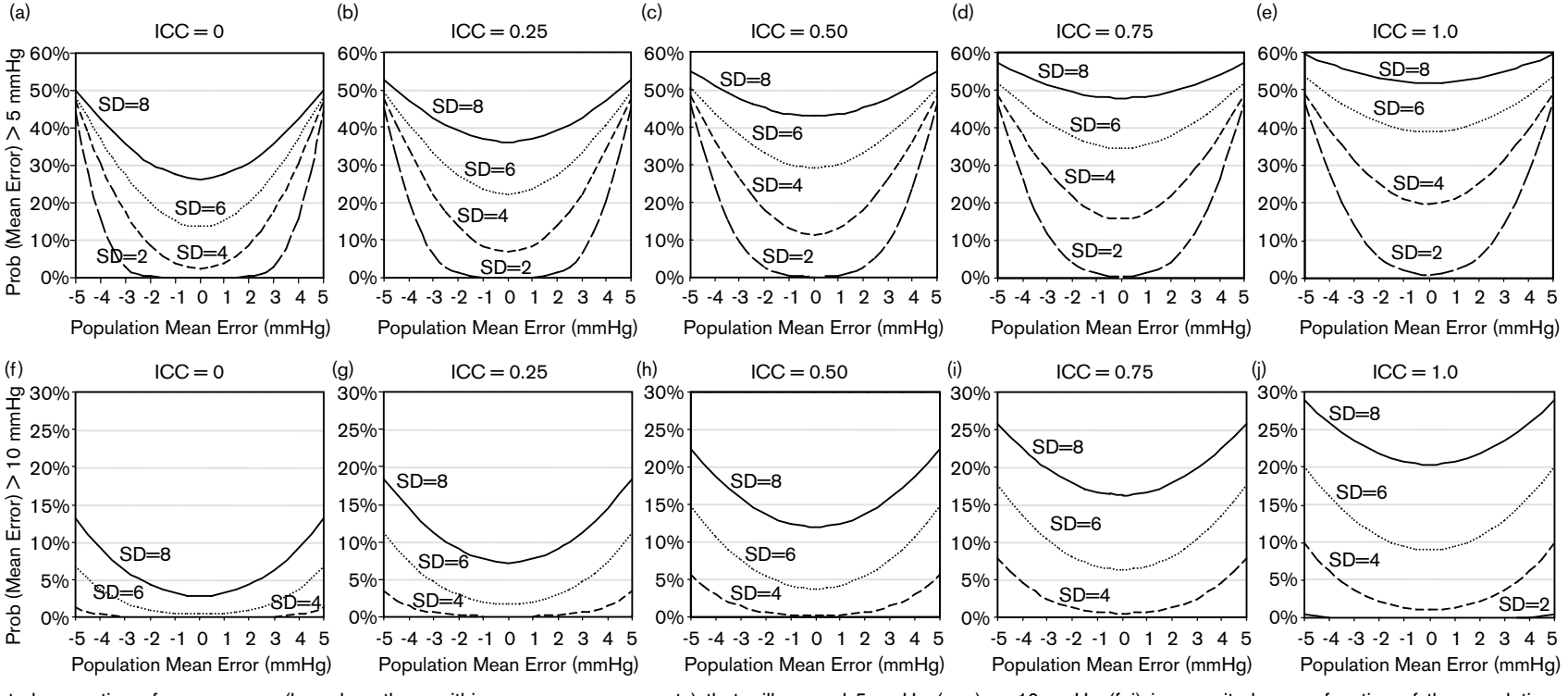
We next determined the proportion of *persons* for whom each monitor failed to validate, that is, for whom the average error exceeded 5 mmHg. For the monitors that passed the validation criteria, 20–38% of individuals had average measurements that were inaccurate (by more than 5 mmHg), and 2–5% of individuals had average measurements that were inaccurate by more than 10 mmHg.

Analysis of the individual errors, using repeated measures analysis of variance, yielded estimates of between-person variance that accounted for 26–63% of the total variance (ICC = 0.26–0.63). This indicated that the errors of the measurements do tend to cluster within persons.

The existence of substantial between-person variance also implies that even if the overall mean error of a monitor is acceptable, each person has his or her own true mean error which in many cases may be outside the range of –5 to +5 mmHg. For such persons, regardless of how many measurements are taken their (unacceptable) mean error will never converge to an acceptable level.

In the validation studies analyzed here, all readings for each individual were taken at a single session. We note that it is therefore not possible to determine if errors for an individual during one session were more similar to each other than would be errors for that individual for readings taken at different sessions; this could occur because of factors such as cuff placement. If there were session-

Fig. 2



Expected proportion of mean errors (based on three within-person measurements) that will exceed 5 mmHg (a–e) or 10 mmHg (f–j) in magnitude as a function of the population mean error, SD, and intra-class correlation (ICC).

Table 1 **Distribution of measurement errors for several automatic blood pressure monitors**

	Auscultatory				Oscillometric			
	CH-Druck		Schiller BR-102		Schiller BR-102		Profilomat II	
	SBP[1]	DBP[2]	SBP	DBP	SBP	DBP	SBP	DBP
Mean error ($n = 255$ readings)	−2.77	−2.41	−4.05	−4.27	−5.08	−1.97	+1.31	+1.57
SD of error ($n = 255$ readings)	5.03	4.59	3.97	4.68	9.69	6.15	9.21	9.64
Readings within 5 mmHg ($n = 255$)	81%	78%	67%	67%	41%	64%	41%	52%
Persons within 5 mmHg ($n = 85$)	80%	79%	68%	62%	49%	74%	44%	49%
Readings within 10 mmHg ($n = 255$)	93%	96%	93%	96%	71%	90%	75%	79%
Persons within 10 mmHg ($n = 85$)	95%	98%	95%	95%	76%	95%	79%	78%
Variance accounted for by 'person' (ICC)	43%	63%	37%	55%	31%	26%	66%	71%
BHS/AAMI Validation	Pass	Pass	Pass	Pass	Fail	Pass	Fail	Fail

[1]Systolic blood pressure; [2]diastolic blood pressure.

effects, these effects would not be distinguishable from person-effects in these data, leading to over-estimation of the person-effect (ICC). We therefore recommend that future studies assess the between-session variability of error.

Discussion

Recommendations

It is reasonable to expect that when a patient purchases a blood pressure monitor declared valid by the AAMI and BHS standards, it will provide accurate blood pressure measurements for that individual. Under the current validation criteria, however, such reliance may be unwarranted, and it is possible that more than half of patients will have an average error greater than 5 mmHg, and more than one in four will have an average error greater than 10 mmHg. The current validation standards for ensuring accuracy are thus inadequate to ensure that individuals receive accurate blood pressure measurements.

We propose that there be two stages of validation: first, the *model* of the monitor in question should be validated at the population level, and second, the *particular* monitor unit should be validated in the physician's office for the intended user.

Stage 1: population validation

The population validation protocol should include an assessment of the proportion of persons that receive accurate readings from the tested blood pressure monitor. Analyses of the data generated by such a protocol must take into account the variation between persons in monitor accuracy (the ICC). Failure to do so provides insufficient information to allow clinical decision-making regarding the use of the monitor.

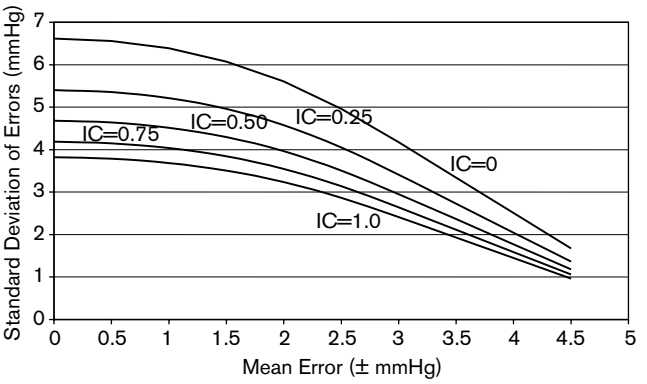
Ideally, 100% of users would receive accurate measurements; however, technical and cost limitations make this unlikely. Selecting a cut-off point at which a monitor is declared acceptable is arbitrary; however, we suggest that it is reasonable for devices to be approved that provide 85% of users with average measurements accurate to within 5 mmHg, and 95% of users with measurements accurate to within 10 mmHg. Accuracy for the remaining 5–15% could then be ensured through the use of the individual validation protocol (see below). In any case, reporting of the percentage of persons who receive average measurements accurate to within 5 and 10 mmHg should allow available monitors to be compared, and the most consistently accurate to be selected for use. Figure 3 illustrates varying combinations of mean error, SD, and ICC that ensure that 85% of persons receive average readings accurate to within 5 mmHg. It is worth noting that of the monitors for which validation data are summarized in Table 1, none meet these criteria.

Stage II: validation for individual users

We recommend that each patient has his or her monitor validated by a person trained in blood pressure measurement. Validation should be performed by connecting the monitor to a sphygmomanometer using a T-connector and then comparing several blood pressure readings assessed by both the monitor and the observer. As individual needs vary, physicians should use clinical judgment in determining what level of accuracy is necessary for a particular patient and circumstance. Monitors that fail to reach that standard should be exchanged for another monitor, and the Stage II process re-initiated.

One of the advantages of a population validation protocol is that observer accuracy can be assessed and confirmed. However, the accuracy of persons who might perform monitor validations for the individual users is difficult to ensure. For this reason, monitors that show little between-person variability (i.e., have small ICCs) may be preferred. Obviously, as the percentage of persons for whom the monitor gives accurate readings in the population validation study increases, the need for individual validation decreases.

Fig. 3



Combinations of mean error, standard deviation, and intra-class correlation that ensure 85% of persons have mean errors (across three readings) that are no greater than $\pm$ 5 mmHg.

Conclusions

Current validation protocols fail to describe the extent to which approved self-monitors will generate accurate readings in individual patients. The combination of a validation protocol that takes account of between-person variability, and individual validation of the 'fit' between each patient and his or her monitor, should ensure that self-monitors are able to contribute to accurate hypertension management, diagnosis, and treatment.

References

1 The Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure. The sixth report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Arch Intern Med* 1997; **157**:2413–2446.

2 Pickering TG, James GD, Boddie C, Harshfield GA, Blank S, Laragh J. How common is white coat hypertension?. *JAMA* 1988; **259**:225–228.

3 Coats AJ, Radaelli A, Clark SJ, Conway J, Sleight P. The influence of ambulatory blood pressure monitoring on the design and interpretation of trials in hypertension. *J Hypertens* 1992; **10**:385–391.

4 Myers MG. Ambulatory blood pressure monitoring: guidelines for use in clinical practice. *Can J Cardiol* 1996; **12**:1271–1275.

5 Mancia G, Zanchetti A, Agabiti-Rosei E, Benemio G, De Cesaris R, Fogari R, *et al.* Ambulatory blood pressure is superior to clinic blood pressure in predicting treatment-induced regression of left ventricular hypertrophy: Study on Ambulatory Monitoring of Blood Pressure and Lisinopril Evaluation (SAMPLE). *Circulation* 1997; **95**:1464–1470.

6 Gerin W, Holland J, Pickering TG. Home blood pressure telemonitoring and telephone transmission for better diagnosis and treatment. *Curr Hypertens Rep* 1999; **1**:1–10.

7 Rogers MAM, Small D, Buchan DA, Butch CA, Stewart CM, Krenzer BE, *et al.* Home monitoring service improves mean arterial pressure in patients with essential hypertension. A randomized, controlled trial. *Ann Intern Med* 2001; **134**:1024–1032.

8 Vetter W, Hess L, Brignoli R. Influence of self-measurement of blood pressure on the responder rate in hypertensive patients treated with losartan: results of the SVATCH Study. Standard versus Automatic Treatment Control of COSAAR in Hypertension. *J Hum Hypertens* 2000; **14**:235–241.

9 Perloff D, Sokolow M, Cowan RM. The prognostic value of ambulatory blood pressures. *JAMA* 1983; **249**:2792–2798.

10 Verdecchia P, Porcellati C, Schillaci G, Borgioni C, Ciucci A, Battistelli M, *et al.* Ambulatory blood pressure. An independent predictor of prognosis in essential hypertension. *Hypertens* 1994; **24**:793–801.

11 Jula A, Puukka P, Karanko H. Multiple clinic and home blood pressure measurements versus ambulatory blood pressure monitoring. *Hypertension* 1999; **34**:261–266.

12 Ohkubo T, Imai Y, Tsuji I, Nagai K, Kato J, Kikuchi N, *et al.* Home blood pressure measurement has a stronger predictive power for mortality than does screening blood pressure measurement: a population-based observation in Ohasama, Japan. *J Hypertens* 1998; **16**:971–975.

13 Sega R, Trocino G, Lanzarotti A, Carugo S, Cesana G, Schiavina R, *et al.* Alterations of cardiac structure in patients with isolated office, ambulatory, or home hypertension: data from the general population (Pressione Arteriose Monitorate E Loro Associazioni [PAMELA].Study). *Circulation* 2001; **104**:1385–1392.

14 Guidelines Subcommittee. 1999 World Health Organization International Society of Hypertension Guidelines for the Management of Hypertension. *J Hypertens* 1999; **17**:151–183.

15 The National High Blood Pressure Education Program Coordinating Committee. National High Blood Pressure Education Program Working Group report on ambulatory blood pressure monitoring. *Arch Intern Med* 1990; **150**:2270–2280.

16 The Scientific Committee. Consensus document on non-invasive ambulatory blood pressure monitoring. *J Hypertens* 1990; **8(Suppl 6)**:S135–S140.

17 Jones DW, Frohlich ED, Grim CM, Grim CE, Taubert KA. Mercury sphygmomanometers should not be abandoned: an advisory statement from the Council for High Blood Pressure Research, American Heart Association. *Hypertension* 2001; **37**:185–186.

18 Association for the Advancement of Medical Instrumentation. American standard for electronic and automated sphygmomanometers. Washington, DC: AAMI; 1987.

19 O'Brien E, Petrie J, Littler W, de Swiet M, Padfield PL, O'Malley K, *et al.* The British Hypertension protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems. *J Hypertens* 1990; **8**:607–619.

20 White WB, Berson AS, Robbins C, Jamieson MJ, Prisant LM, Roccella E, *et al.* National standard for measurement of resting and ambulatory blood pressure with automated sphygmomanometers. *Hypertension* 1993; **21**:504–509.

21 Association for the Advancement of Medical Instrumentation. American national standard: electronic or automated sphygmomanometers. Arlington, VA: AAMI; 1993.

22 O'Brien E, Petrie J, Littler W, de Swiet M, 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(Suppl 2)**:S43–S62.

23 O'Brien E, Atkins N. A comparison of the British Hypertension Society and Association for the Advancement of Medical Instrumentation protocols for validating blood pressure measuring devices: can the two be reconciled?. *J Hypertens* 1994; **12**:1089–1094.

24 O'Brien E, Mee F, Atkins N, O'Malley K. Accuracy of the CH-Druck/Pressure Scan ERKA ambulatory blood pressure measuring system determined by the British Hypertension Society protocol. *J Hypertens* 1993; **11(Suppl 2)**:1–7.

25 Mee F, Atkins N, O'Brien E. Evaluation of the Profilomat II ambulatory blood pressure system according to the protocols of the British Hypertension Society and the Association for the Advancement of Medical Instrumentation. *Blood Press Monit* 1998; **3**:353–361.

26 O'Brien E, Mee F, Atkins N. Evaluation of the Schiller BR-102 ambulatory blood pressure system according to the protocols of the British Hypertension Society and Association for the Advancement of Medical Instrumentation. *Blood Press Monit* 1999; **4**:35–43.