Type and accuracy of sphygmomanometers in primary care: a cross-sectional observational study

Abstract

Background
Previous studies identified worrying levels of sphygmomanometer inaccuracy and have not been repeated in the era of digital measurement of blood pressure.

Aim
To establish the type and accuracy of sphygmomanometers in current use.

Design and setting
Cross-sectional, observational study in 38 Oxfordshire primary care practices.

Method
Sphygmomanometers were evaluated between 50 and 250 mmHg, using Omron PA350 or Scandmed 950831-2 pressure meters.

Results
Six hundred and four sphygmomanometers were identified: 323 digital (54%), 192 aneroid (32%), 79 mercury (13%), and 10 hybrid (2%) devices. Of these, 584 (97%) could be fully tested. Overall, 503/584 (86%) were within 3 mmHg of the reference, 77/584 (13%) had one or more error of 4–9 mmHg, and 4/584 (0.7%) had one or more error of more than 10 mmHg. Mercury (71/75, 95%) and digital (272/308, 88%) devices were more likely to be within 3 mmHg of the reference standard than aneroid models (152/191, 78%) (Fisher’s exact test P = 0.001).

Conclusion
Digital sphygmomanometers have largely replaced mercury models in primary care and have equivalent accuracy. Aneroid devices have higher failure rates than other device types; this appears to be largely accounted for by models from indiscernible manufacturers. Given the availability of inexpensive and accurate digital models, GPs should consider replacing aneroid devices with digital equivalents, especially for home visiting.

INTRODUCTION

Hypertension is the leading risk factor for cardiovascular disease and death worldwide. Its diagnosis and control require accurate blood pressure measurement, which depends upon both the instrument and the technique used. Inaccurate blood pressure measurement might lead to misdiagnosis, and either unnecessary or insufficient treatment, with both ethical and public health implications. The widely accepted ‘gold standard’ measuring device has been, until recently, the mercury sphygmomanometer; however, worldwide concern over increasing mercury levels led to a European Union directive that mercury sphygmomanometers be phased out.

Several protocols exist for the validation of blood pressure devices, but these do not guarantee sustained accuracy once in use: a hospital study employing a rigorous dynamic testing protocol across the whole blood pressure range reported error readings of >10 mmHg in 50% of aneroid devices and 10% of mercury devices. Primary care-based studies have shown similar results but have not been undertaken since the use of electronic devices became widespread. The revised general medical services contract for UK GPs (2004) includes a recommendation to ensure medical equipment is regularly maintained, calibrated, and replaced if faulty. Given that reported achievement for this standard is 97%, it might be expected that most sphygmomanometers in use are now accurate.

Therefore, this study aimed to assess the current state of equipment in primary care for monitoring blood pressure: the type of equipment being used, its accuracy, and, finally, the relationship between current performance and previous in-service testing.

METHOD

In January 2009, written invitations to participate in an anonymised cross-sectional survey of sphygmomanometer type and accuracy were sent to 83 practices in Oxfordshire Primary Care Trust. Participating practices were visited by a trained health technician, and practice staff identified all manometers, whether kept on site or in a doctor’s bag, or in use by staff attending the practice. Recruitment and testing were complete by July 2009.

Following pilot work confirming their equivalence, two types of pressure-testing device were deployed: the Omron PA 350 (for Omron digital devices, and any free-standing mercury, aneroid or digital device), and the more portable Scandmed 950831-2 Pressure Meter (for wall-mounted devices, and digital devices sharing the same manufacturer, A&D). As recommended by the British Hypertension Society (BHS), sphygmomanometers were inflated to 280 mmHg, then the pressure released to 250, 200, 150, 100, and 50 mmHg in turn.
and at each level the true pressure was read on the pressure tester’s digital display.

The dates of previous testing were noted: devices were taken to have been tested if there was a dated label on the device, or if the practice manager reported devices as having been tested and could specify a date. Correlation coefficients were used to examine whether errors at one pressure were correlated with errors at other pressures. To facilitate comparison with earlier studies, and communicate significance of errors, a ‘trafficlight’ classification of error readings was devised, in which devices were assigned a ‘green’ rating if all results fell within 0–3 mm Hg of the reference blood pressure (the BHS standard), an ‘amber’ if one or more readings deviated by 4–9 mm Hg, and ‘red’ if any of the results differed by 10 mm Hg or more. Fisher’s exact test was used to test whether this classification was associated with device type, manufacturer or branding, time since last testing, size of practice, or holder of the device. These exclusions left 584 (97%) devices that were fully evaluated.

Overall, 86% (503 of 584) of individual devices were within 3 mm Hg of the standard across the pressure range (‘green’ classification), 13% (77 of 584) had one or more error of 4–9 mm Hg (‘amber’), and <1% (4 of 584) had one or more error ≥10 mm Hg (‘red’). Significantly fewer mercury (4/75 [5%]) and digital devices (36/308 [12%]) failed the 3 mm Hg standard than aneroid (41/191 [21%]) (Figure 1). Compared to mercury, digital devices were no more likely to fail the BHS standard (relative risk [RR] = 2.19, 95% confidence interval [CI] = 0.80 to 5.97), but aneroid models were four times more likely to fail (RR = 4.02, 95%, CI = 1.49 to 10.8). The 10 hybrid devices were all within 3 mm Hg of the reference standard.

Device accuracy was similar between the most common manufacturers: all errors were within 3 mm Hg in digital devices manufactured by A&D/bosqo or Omron in 89% (92/103) and 88% (173/196) of cases respectively, and 83% (5/6) of others (P = 0.7 for difference between manufacturers). Similarly, in aneroid devices 86% (59/69) of Accoson devices, 85% (39/46) of Welch-Allyn devices, and 83% (38/46) of other recognised manufacturers (P = 0.96) had errors within 3 mm Hg. The 30 aneroid devices without a discernible manufacturer, of which 23 (77%) bore pharmaceutical advertising, were more likely to be classified as amber or red compared to other aneroid devices: 53% (16 of 30) versus 16% (25 of 161) (Fisher’s exact test P<0.001). The digital devices included three devices that bore the name of a high street retailer rather than a recognised manufacturer of manometers. A sensitivity analysis restricted to devices that were clearly from a recognised device manufacturer was therefore undertaken, in which 95% (71 of 75) of mercury devices were within 3 mm Hg.
mmHg, with 89% (270 of 305) of digital devices and 86% (136 of 161) of aneroid models now within this range (P < 0.07 for differences between device types).

A further sensitivity analysis evaluated absolute error across the pressure range (Figure 2). Mercury devices had a smaller error than aneroid and digital devices at every pressure (P < 0.001 at each pressure). The error given by a device at any preset pressure (50, 100, 150, 200, or 250 mmHg) was predictive of the error given by the same device at other pressures, with a significant correlation coefficient (P < 0.001) at every level. Thus, devices that substantially over-read at any one pressure were more likely to do so at other pressures, and similarly for under-reading.

There were 17 apparently brand new devices. Of the remaining 567, 147 (26%) had been tested within the previous 6 months, 173 (31%) within the past 6 to 12 months, 97 (17%) had been tested more than 12 months ago, and 150 (26%) had no record of previous testing. Overall, no significant difference was found between testing date and accuracy of device: a ‘green’ rating (within 3 mmHg) was achieved in 87% (178/203) of devices tested in the last 6 months, 89% (154/173) of those tested within 6 to 12 months, 87% (84/97) of those tested over 12 months before, and 80% (120/150) of those never tested (P = 0.1).

The accuracy of a device was not related to the size of practice (P = 0.7) or the holder of the device — whether this was a GP, practice nurse, or district nurse, or the device was on loan to a patient (P = 0.7).

**DISCUSSION**

**Summary**

This study shows that the most commonly used sphygmomanometers in UK primary care are now digital devices, which outnumber mercury models 4:1. Overall, one in seven devices failed the BHS standard requiring all devices to read within 0–3 mmHg of the true value. The performance of mercury and digital models was similar and significantly better than aneroid models. This difference appears to

### Table 1. Manufacturers of blood pressure devices in use in 38 general practices in Oxfordshire

<table>
<thead>
<tr>
<th>Type/brand</th>
<th>Number identified in survey (% of device type)</th>
<th>Number included in tests in this study</th>
<th>Reasons for exclusion from tests</th>
<th>Percentage (numbers) of tested device type classified amber/red</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury devices (13% of all devices identified)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accoson</td>
<td>79 (94)</td>
<td>70</td>
<td>Four: one with bubble in mercury, stored in drawer and unclaimed, last tested 1997, presumed not in use; three labelled as out of use since previous test, but wall-mounting and cost of decommissioning precluded removal from premises</td>
<td>5% (4/75 amber, 0/75 red)</td>
</tr>
<tr>
<td>Other manufacturers</td>
<td>5 (6)</td>
<td>5</td>
<td></td>
<td>0% (0/5 amber/red)</td>
</tr>
<tr>
<td>Aneroid devices (32% of all devices identified)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accoson</td>
<td>192 (24)</td>
<td>191</td>
<td></td>
<td>22% (42/191 amber, 4/191 red)</td>
</tr>
<tr>
<td>Welch-Allyn</td>
<td>13 (17)</td>
<td>13</td>
<td>15% (2/13 amber, 0/13 red)</td>
<td></td>
</tr>
<tr>
<td>Other device manufacturer</td>
<td>34 (42)</td>
<td>33</td>
<td>One with missing part (Stetho--- Big Ben)</td>
<td>18% (6/33 amber, 0/33 red)</td>
</tr>
<tr>
<td>Unknown manufacturer</td>
<td>7 (9)</td>
<td>7</td>
<td></td>
<td>43% (3/7 amber, 1/7 red)</td>
</tr>
<tr>
<td>Digital devices (53% of all devices identified)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omron</td>
<td>323 (41)</td>
<td>308</td>
<td>Two wrist monitors — incompatible with testers</td>
<td>12% (23/196 amber, 0/196 red)</td>
</tr>
<tr>
<td>A&amp;D/boso</td>
<td>104 (12)</td>
<td>103</td>
<td>One UBS11 wrist monitor — incompatible with testers</td>
<td>11% (11/103 amber, 0/103 red)</td>
</tr>
<tr>
<td>Other manufacturer</td>
<td>11 (3)</td>
<td>6</td>
<td>Five: either incompatible with testers (3) — H21186Wrist, KD 525, Rossmax, or mechanically failed (2) — Transtec, Prestige</td>
<td>17% (1/6 amber, 0/6 red)</td>
</tr>
<tr>
<td>High street pharmacist</td>
<td>10 (3)</td>
<td>3</td>
<td>Seven: either incompatible with testers (6), or mechanically failed (1), all Lloyd’s Pharmacy</td>
<td>33% (1/3 amber, 0/3 red)</td>
</tr>
<tr>
<td>Hybrid devices (2% of all devices identified)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A&amp;D</td>
<td>10 (10)</td>
<td>10</td>
<td></td>
<td>0% (0/10 amber/red)</td>
</tr>
</tbody>
</table>

---

*aSome A&D digital devices are branded as boso (personal communication: Mike Telford, A&D Europe). bThe brand name was ‘Lloyds Pharmacy’ on nine devices and ‘Boots’ one device. cThe A&D UM 101 meter combines aneroid and digital features (personal communication: Mike Telford, A&D Europe).*
be explained by the presence of a group of
donated aneroid sphygmomanometers with
no discernible manufacturer. No significant
differences were seen between devices from
the commonest manufacturers, nor
between those used by different professional
groups, which suggests a maturing of the
underlying technology. There was little
evidence that previous calibration made any
difference to current accuracy, although just
over half had been checked within the last
year.

Strengths and limitations
To the authors’ knowledge, this is the first
study to test sphygmomanometer accuracy
in a broad range of primary care practices in
the era of digital blood pressure
measurement. Testing was undertaken by a
single trained individual and so was unlikely
to have been affected by between-observer
variation, and included the vast majority of
devices in use in participating practices; only
3% of devices could not be fully tested,
because of either mechanical faults or
incompatibility with the testing equipment.
Some limitations should be noted: testing
was undertaken in one geographical area
that may not be representative of the country
as a whole. A minority of the practices
(seven) had participated in the earlier pilot
study and so were likely to perform better in

---

**Figure 1.** Errors in blood pressure measurements in mercury, aneroid, and
digital sphygmomanometers tested at 50, 100, 150, 200, and 250 mmHg. Eleven errors
11–50 mmHg are not shown for reasons of scale (five readings from one aneroid device
under-reading by 30 mmHg in all tests).

**Figure 2.** Absolute error (mean, ignoring direction of error; bars show 95% confidence intervals for
the mean) in blood pressure devices when tested at 50, 100, 150, 200, and 250 mmHg.
this study. In addition, practices that decided not to participate in the study may be less interested in the topic and therefore less likely to maintain their equipment. These potential biases could lead to an overestimation of true device accuracy in practice.

Comparison with existing literature
To date, mercury devices have been viewed as the ‘gold standard’ for sphygmomanometer accuracy.\(^6\) This study suggests that they have now been largely replaced by digital models in clinical practice, probably due to a combination of the drive to reduce mercury and the convenience of digital technology.\(^7\) Examination of the absolute errors observed demonstrates the superior performance of mercury models and suggests that a place for mercury devices remains, in terms of providing a reference standard.\(^8,9\) However, the clinically relevant measure of agreement (within or outside the 3 mmHg standard) highlighted by the ‘traffic light’ classification was similar for digital and mercury models.

The performance of digital devices from recognised manufacturers was consistently good, which is reassuring given their ubiquity. Many patients (and some professionals) now purchase low-cost digital devices from high street pharmacies, but it was not possible to test enough such devices in this study to assess their performance separately. In common with other investigators, a significantly higher failure rate was detected for aneroid sphygmomanometers, unless the analysis was restricted to aneroid devices from recognised manufacturers. A particularly high failure rate was found in those aneroid devices apparently received as gifts from pharmaceutical representatives. Such devices are likely to be low-cost, possibly inferior products, which have been given to clinicians under pharmaceutical industry regulations allowing donation and acceptance of inexpensive promotional items. The main advantage of aneroid sphygmomanometer devices is their portability, and one-fifth of the devices in the present study were definitely carried in a doctor’s bag. The mechanical construction of aneroid devices makes them vulnerable to physical damage, so that jolting, and the variation in temperature likely to occur in a doctor’s bag will affect their accuracy more than is the case for digital models.\(^1\)

Where monitors ‘failed’ on testing (over 3 mmHg deviation from the reference standard at one or more pressures), there was evidence that errors were systematic, leading to consistent over- or underdiagnosis or treatment in clinical practice.\(^4\) Conversely, instruments providing accurate readings at one pressure tended to be accurate across the range. Uncalibrated sphygmomanometer error potentially accounts for 20% of all undetected adult systolic hypertension. Where prevalence is low, overdetection is worse; for example, causing 63% of falsely detected systolic hypertension in 18–24 year-old females.\(^5\) Overdiagnosis could have an impact outside the clinical arena, due to the ‘sick patient’ effect, in which people assume a sick role and show more absenteeism after receiving a diagnosis of hypertension.\(^10\)

An unexpected finding was the lack of difference in performance between devices that had, or had not been tested previously. The negative findings might be due to the fact that the ‘never tested’ group included not only older devices (including the four most inaccurate devices) but also newer devices within the manufacturer’s warranty. On the other hand, within the ‘previously tested’ group, there were some manometers labelled as failing previous tests, which, worryingly, had been retained. Together, these findings emphasise that regular checks need to be acted upon but suggest that such checks could focus on just one or two pressures in a range of clinical interest, which might improve adherence to the recommendations.

Implications for practice and research
This study demonstrates that despite the increased use of digital devices, there remains an improved but still unsatisfactory prevalence of inaccurate, mostly aneroid manometers. The results suggest the need for replacement of inaccurate devices with high-quality, validated devices. Given the superior objectivity offered by digital devices, memory functions, improved portability, and falling costs, perhaps the time has now come to remove aneroid devices, at least from doctors’ bags. Further studies are needed to evaluate the in-service accuracy of the newer hybrid devices, although the small sample in the present study suggested possible high accuracy. The cost of replacing unreliable manometers is likely to be dwarfed by the cost of inadequately treated hypertension, or, in overtreated patients, the cost of inappropriate prescribing and adverse events.
REFERENCES


