Accuracy of blood pressure monitors used for BP checks in retail pharmacies

Authors: James Hodgkinson¹, Constantinos Koshiaris², Una Martin³, Jonathan Mant⁴, Carl Heneghan², FD Richard Hobbs², Richard J McManus²

¹ Primary Care Clinical Sciences, University of Birmingham, Birmingham, UK
² Nuffield Department of Primary Care Health Sciences, NIHR School for Primary Care Research, University of Oxford, Oxford, UK
³ School of Clinical and Experimental Medicine, University of Birmingham, Birmingham, UK
⁴ Primary Care Unit, NIHR School for Primary Care Research, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Corresponding Author: Dr James Hodgkinson: j.a.hodgkinson@bham.ac.uk

Keywords: primary health care, hypertension, blood pressure monitors, calibration, community pharmacy services
Abstract

Background
Free blood pressure (BP) checks offered by community pharmacies offer a potentially useful opportunity to diagnose and/or manage hypertension, but the accuracy of the sphygmomanometers in use is currently unknown.

Aim
To assess the accuracy of validated automatic BP monitors used for BP checks in a UK retail pharmacy chain.

Design and Setting
52 pharmacies from one chain were visited in a range of locations (inner city, suburban, rural) in central England.

Method
Monitor accuracy was compared to a calibrated reference device (Omron PA-350), at 50 mmHg intervals across the range 0-300 mmHg (static pressure test), with a difference from the reference monitor of +/-3 mmHg at any interval considered a failure. The results were analysed by usage rates and length of time in service.

Results
Eight (13%) monitors failed (i.e. were more than 3mmHg from reference), all underestimating BP. Monitor failure rate from the reference monitor of +/-3 mmHg at any testing interval varied by length of time in use (2/38, 5% before 18 months vs. 4/14, 29% after 18 months; p=0.038) and to some extent but non-significantly by usage rates (4/22, 18% in monitors used more than once daily vs. 2/33, 6% in those used less frequently; p=0.204).

Conclusion
BP monitors within a pharmacy setting fail at similar rates to those in general practice. Annual calibration checks for blood pressure monitors are needed, even for new monitors, since these data indicate declining performance from 18 months onwards.
How this fits in

The accessibility of free blood pressure checks provided in community pharmacies provides an excellent opportunity to improve hypertension screening and management. However, even where clinically validated using a recognised protocol, there is very little evidence about how well the devices used maintain their accuracy over time in a potentially testing environment. We therefore assessed the accuracy of the monitors being used in blood pressure checks at pharmacies - the only study we are aware of that considers the accuracy of these monitors, and one of the very few to assess accuracy of monitors with respect to time in service. Based on our research findings, we recommend an annual calibration check is needed for this type of monitor, with evidence suggesting declining performance from 18 months onwards.
Background
High blood pressure (BP) is a key risk factor for the development of cardiovascular disease\(^1\) and a major cause of morbidity and mortality worldwide.\(^2\) An accurate BP monitoring device is fundamental to BP measurement in the diagnosis and control of hypertension.

Several protocols\(^3,4,5\) exist for the validation of BP measuring devices but these are generally undertaken and then published on brand new models and so do not guarantee sustained accuracy thereafter. The revised General Medical Services Contract for UK General Practitioners (2003)\(^6\) included a recommendation to ensure medical equipment is regularly maintained, calibrated and replaced if necessary. Typically new monitors are assumed to be accurate for two years and then annual checks are undertaken. However, it is not clear whether this is appropriate as the drift in accuracy over time of an automated sphygmomanometer is not known. The error rate is a function of random (variability) and systematic (bias) error and ultimately depends on the calibration interval of the device, and the conditions under which it is used.

Detection and control of hypertension are sub-optimal,\(^7\) and a recent study has identified that there are insufficient GPs in England to achieve high levels of detection while at the same time maintaining access to appointments with GPs.\(^8\) Community pharmacies are a good potential site for identifying cardiovascular risk factors and improving disease detection,\(^9\) including identifying people with hypertension, because of their accessibility and because many pharmacies provide free access to BP monitors. Pharmacists have been involved in successful community-based screening programmes developed to improve detection and treatment of hypertension, both in the UK and worldwide.\(^10,11,12,13,14,15,16\) Community pharmacy BP monitoring is readily available, widely demanded, and recommended by hypertension guidelines.\(^11,17,18\)

Evidence from the US\(^19\) suggests that many people with hypertension check their BP at a pharmacy. In the UK, the ‘Know Your Numbers’ campaign has ensured 1.5 million people have had their BP checked in the UK since 2001, mostly at community pharmacies.\(^20\)
However, the few studies evaluating publicly available monitors have shown them to be inaccurate.\textsuperscript{21,22,23,24,25,26,27} Whilst many of these publications are more than a decade old, there remains limited evidence of how well the monitors maintain their rigour in a potentially testing environment.

We therefore assessed the accuracy of validated\textsuperscript{28} automatic BP monitors in a UK retail pharmacy chain, where current policy was to retire monitors after two years in service for free BP checks without interim calibration checks.

\textbf{Methods}

The accuracy of all available digital sphygmomanometers (two models, both derived from the validated Kinetik BPM 1 Series) in 52 pharmacies was evaluated through comparison to a calibrated reference monitor (Omron PA-350), at 50mmHg intervals across the range 0-300mmHg as recommended by the British Hypertension Society. Following manufacturers’ protocol, a difference from the reference monitor of +/-3mmHg at any interval was considered a failure. In addition to this static pressure test, we also conducted a visual inspection (that the machine switched on and off, contained batteries, and had a readable display) and tested deflation, air leakage, and all cuffs in use. Pharmacies were visited in a range of locations (inner city, suburban and rural) in central England (Birmingham, Black Country, Herefordshire and Worcestershire) in order to achieve a sample of monitors of both high and low use. No pharmacies refused to participate. All testing was undertaken by JH.

The length of time in use and number of recorded uses was assessed in two ways:

(a) Asking pharmacy staff at the time of the calibration visit. Very occasionally this was calculated by reference to a log book, and some monitors had stickers on them that detailed the precise date they were first used for BP checks, but mostly was dependent on staff recall. Data on service duration, or length of time in use, was recorded only from the individual pharmacies.

(b) Collecting the number of BP checks for each pharmacy collated by the Head Office over a two year period – however, data were at the pharmacy level rather than concerning an individual monitor, so was of limited use where pharmacies had
multiple monitors. These data were also dependent on individual data input every time a BP check was done.

The association between monitor precision and both length of time in service and usage rate (using the data collected from fieldwork and that from the pharmacy Head Office, separately) was tested using linear regression. The linear models were fitted by using the mean of the differences as outcome, which was defined as the mean of the differences between the device being tested and the reference standard in all testing intervals ascending and descending. All model assumptions were checked. Failure rate by the different predictors was assessed using the Fisher exact test statistic.

No funding was received from either the pharmacy chain or the manufacturer of the monitor.

Results
Of a total of 61 monitors assessed, eight (13%) failed (i.e. were more than 3mmHg from reference), all underestimating BP. The largest disparity found was 8.3mmHg at a pressure of 300mmHg. The majority of failures were at higher pressures, and if a monitor failed at a given pressure it invariably failed at all higher pressures. Eight failed at 300mmHg, five failed from 250mmHg, four from 200mmHg, and one from 100mmHg. At the testing point nearest to the diagnostic threshold (150mmHg), two monitors failed (2/61, 3%).

Overall, the difference in BP between monitors and reference device increased with level of BP in a linear fashion to a mean 2mmHg difference at 300mmHg (Table 1). The relative difference did not appear to vary with pressure, as the percentage error was almost identical at each testing interval, albeit a little higher at the point closest to the diagnostic threshold in the testing process (150mmHg), where there was a mean difference of just over 1mmHg (Table 1).
71% of readings underestimated and 22% overestimated, and the ratio of readings underestimating compared to overestimating increased with pressure, as 80% of readings underestimated and 16% overestimated between 150-250mmHg, and 92% were underestimates and 5% overestimates at 300mmHg (Table 2). This pattern of underestimation is suggestive of a small systematic error across the monitors, but with random error (variability) in individual monitors.

Annual BP monitor usage estimates from pharmacy staff ranged from 104 to 1560 (compounded by variation from the ‘Know Your Numbers’ campaign) and from the pharmacy chain’s Head Office ranged from 14 to 276. Comparison of the ratio of uses between the two data sources was inconsistent, varying from a single pharmacy where usage data from Head Office was higher, to monitor usage from fieldwork estimates being more than 18 times greater. In view of this, both estimates were used separately and compared in sensitivity analyses.

Length of time in service varied from one week to almost three years, with a modal period of service duration of 6-12 months. However there was some uncertainty about the time in use of several monitors, with no clear estimate able to be provided for 9 (19%) of the devices.

Monitor failure rate (i.e. the presence of a difference from the reference monitor of +/- 3 mmHg at any testing interval) varied by length of time in use (2/38, 5% before 18 months compared to 8/61, 13% overall and vs. 4/14, 29% after 18 months; p=0.038). To some extent failure rate varied by usage rates (4/22, 18% in monitors used more than once daily compared to 2/33, 6% in those used less frequently) but this difference was not statistically significant (p=0.204). A greater failure rate (4/13, 31%) was observed for higher usage using the Head Office data, but as data were only available at the level of the pharmacy, and four failures happened in pharmacies where multiple devices were being used, this is difficult to interpret.

The association between monitor precision and length of time in service (Table 3) using linear regression showed that the mean difference for monitors in service for both 6-12 months and 13-18 months was around 0.35mmHg greater than that for those used for 0-5 months, increasing to over 0.70mmHg for monitors used for 19-24
months. However, no statistical significance could be observed for any of the categories.

A linear trend for predicted difference was found by amount of usage (for both the fieldwork data and that from the pharmacy’s Head Office), suggesting a drop in precision with high usage, but the results were not statistically significant. (Tables 4a and 4b)

Further sensitivity analyses were conducted to check the influence of outliers (monitors with a very high mean difference compared to the others), as some monitors could have been faulty from the outset, but results were consistent with the main analyses.

All monitors passed the fast deflation test (deflation from 260mmHg to 15mmHg should take less than 10 seconds) and only one failed the air leakage check (inflating to between 280-290mmHg and requiring the pressure to drop less than 6mmHg over one minute), but this device also failed the static pressure test.

A total of 52 normal (22-30cm), 45 large (30-42cm) and 17 extra-large (42-48cm) cuffs being used with the monitors were also assessed, all of which passed the fast deflation test and air leakage check.

Discussion

Summary of research findings
13% of monitors used for BP checks in community pharmacies that we tested, failed a standardised accuracy test, underestimating true BP, with the absolute pressure gap between the machines tested and the reference device increasing as BP rose. There was a trend for an increase in mean difference of monitors with longer service duration and higher usage but not to a statistically significant degree, perhaps due to lack of power. Failure rate of monitors after 18 months in use was significantly higher than for those with shorter service duration. Overall accuracy around the systolic
diagnostic threshold of 140mmHg was good with only small mean differences from reference.

**Strengths and limitations**
This is the only study we are aware of that considers the accuracy of monitors used in pharmacy BP checks, and one of the very few studies to assess accuracy of monitors with respect to time in service. As monitors in use were limited to two models, it provided an opportunity to consider calibration drift in the community for numbers of monitors not usually available on such a scale.

The weaknesses of the data surrounding length of usage and service duration have been previously noted. Both under- and over-estimation appeared dependent on who was asked. Length of time in service varied from one week to almost three years, which was surprising as the pharmacy chain in question has a policy of replacing the entire stock of monitors used for community testing every two years at the same time. The a priori intention in this study was to consider a cohort of monitors at the 18-24 month stage of the cycle which in theory presented a rare opportunity for a reasonable sample size for this type of research. In reality there was an even spread of service duration. Furthermore, the chain’s policy of jettisoning all monitors at a predetermined point, including replacement monitors (i.e. ones which may not have been in use for the full two year period), meant the sample size was relatively small, there being limited value in the study continuing further than it did, as all remaining monitors were about to be renewed.

As a consequence of the two year cycle, there was no formal calibration checking but several pharmacy staff anecdotally reported that they rejected any monitors they had concerns about (for example, where a monitor had been dropped or had liquids spilt on it, or alternatively if readings seemed spurious or likely to be wrong, albeit without formal testing against a reference device), sending them back to Head Office. This is reassuring and acts as a further level of quality control.

**Comparison with existing literature**
The failure rate of 13% we identified is very similar to that found by A’Court et al when testing monitors used in GP practices, namely 14% overall and 12%
specifically for digital devices. Interestingly, for technical reasons those investigators were not able to test the monitors included in this study, which suggests that the underlying technologies in oscillometric monitors may be similarly robust. Of note however, A’Court et al found no difference in accuracy by service duration over a two year period. Perhaps this discrepancy is due to differing usage rates: the estimates of usage from pharmacy staff correspond to the lower end of BP checks conducted in a typical GP surgery, where a BP machine would generally be used for many individuals per day.

26% (12/47) of automated BP monitors in use at a hospital were not within 3mmHg of a reference device, although only a minority of these machines were validated and it is unclear how long they had been in service. That study also found calibration errors in automated devices tend to underestimate pressure. Conversely, 5% of automated BP monitors in GP practices failed but these devices were all provided by one manufacturer. There are very few studies that assess the accuracy of BP monitoring equipment over time. One study of 14 identical ambulatory monitors over a six year period found that 90% of a standardised set of pressure readings were within 2mmHg on repeated measurement.

**Implications for clinical practice**

Provision of BP monitoring facilities in community pharmacies has the potential to improve the quality of care for patients but depends on accurate monitors as well as robust communication systems with the rest of the health service. This work supports current recommendations for testing monitor accuracy annually but suggests that leaving monitors for as long as two years from purchase to testing may lead to potential for error. The main issues detected in this study were of underestimation of BP. Whilst pharmacists would refer patients back to their GP rather than initiate treatment, reducing safety concerns associated with over-diagnosis, any under-diagnosis due to monitor underestimation of BP around clinical threshold levels represents a missed opportunity of pharmacy screening and management of hypertension, although most of the failures were at pressures higher than this. This effect could be exaggerated as previous research has highlighted community pharmacy measured BP is lower than clinic BP due to a reduced white coat
effect,\textsuperscript{33,34,35} despite a similar threshold for hypertension status being used by GPs and pharmacists.

This research has implications for patients who self-monitor their BP, as these type of monitors retail on the high street and, due to being validated and inexpensive, are very popular, selling up to millions of units.\textsuperscript{36} However, no advice is given (including in the manual provided) as to how long these monitors should be used for or how regularly they should be calibrated. Clearly, the amount of usage monitors will receive when purchased by one person to use in the home environment will be less than those used regularly in pharmacy BP checks, but the accuracy of such monitors cannot be assumed indefinitely, especially as at least 30\% of diagnosed hypertensives possess home BP measuring devices.\textsuperscript{37,38}

Successive governments' national policy have argued that expanding the range of services pharmacies provide will increase access and patient choice, reduce GP workload and lower NHS costs.\textsuperscript{39,40} This work shows that one building block – namely monitor accuracy – can be safely put into place. However, the large differences in the uptake of the BP checks in different pharmacies within the same chain, highlights the importance of implementation. We found a large difference in the reported uptake of the BP checks in different pharmacies, even in the same chain, highlighting that more could be done to encourage people to use this service or to understand why some facilities are apparently more attractive than others.

\textbf{Conclusions}
Community pharmacy BP checks present an excellent opportunity to improve hypertension diagnosis and management, but require accurate equipment. This study has shown BP monitors within this setting to be of similar accuracy to those in primary care, but that without similar calibration checks there is potential for error. These data indicate an annual calibration check is needed for this type of monitor, with evidence suggesting declining performance from 18 months onwards.
Acknowledgements

The authors would like to acknowledge the help of the pharmacies involved in this research, without whose input it would not have been possible. Also we acknowledge the representatives of Kinetik for providing us with the relevant valves and maintenance manual required to undertake the monitor testing.

Funding

This work represents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407–10347). The views expressed in this study are those of the authors and not necessarily of the NHS, the NIHR or the Department of Health. RJM is supported by an NIHR Professorship (NIHR-RP-02-12-015) and by the NIHR Collaboration for Leadership in Applied Health Research and Care (CLAHRC) Oxford at Oxford Health NHS Foundation Trust. FDRH is part funded as Director of the National Institute for Health Research (NIHR) School for Primary Care Research (SPCR), Theme Leader of the NIHR Oxford Biomedical Research Centre (BRC), and Director of the NIHR CLAHRC Oxford.

Competing interests

RJM has received research support in terms of blood pressure monitors from Omron and Lloyds Pharmacies and funding to attend and speak at the Japanese Society of Hypertension.

Ethical approval

The proposal required only approval from the University ethics board, as it involved only testing of machines and not any intervention with subjects – thus it was classified as service development not requiring wider ethical approval.
References

28 http://www.bhsoc.org/index.php?cID=246
Table 1: Mean difference against reference standard at all testing intervals

<table>
<thead>
<tr>
<th>Testing interval (pressure range, mmHg)</th>
<th>Mean difference (mmHg) (95% CI)</th>
<th>Percentage error (mean difference/ testing interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0 (-0.073 to 0.073)</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.15 (0.008 to 0.299)</td>
<td>0.30%</td>
</tr>
<tr>
<td>100</td>
<td>0.42 (0.21 to 0.62)</td>
<td>0.42%</td>
</tr>
<tr>
<td>150</td>
<td>1.15 (0.32 to 1.98)</td>
<td>0.76%</td>
</tr>
<tr>
<td>200</td>
<td>1.14 (0.84 to 1.44)</td>
<td>0.57%</td>
</tr>
<tr>
<td>250</td>
<td>1.41 (1.04 to 1.78)</td>
<td>0.56%</td>
</tr>
<tr>
<td>300</td>
<td>1.90 (1.48 to 2.33)</td>
<td>0.63%</td>
</tr>
</tbody>
</table>

Pressure is checked in 50mmHg increments up to 300mmHg and then down again. The data at each testing interval between 0-250mmHg going up has been combined with that going down.

Table 2: Overestimation and underestimation of pressure at all testing intervals

<table>
<thead>
<tr>
<th>Testing interval (pressure range, mmHg)</th>
<th>% of readings underestimating</th>
<th>% of readings overestimating</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>47</td>
<td>32</td>
</tr>
<tr>
<td>50</td>
<td>58</td>
<td>37</td>
</tr>
<tr>
<td>100</td>
<td>70</td>
<td>25</td>
</tr>
<tr>
<td>150</td>
<td>78</td>
<td>17</td>
</tr>
<tr>
<td>200</td>
<td>81</td>
<td>17</td>
</tr>
<tr>
<td>250</td>
<td>82</td>
<td>14</td>
</tr>
<tr>
<td>300</td>
<td>92</td>
<td>5</td>
</tr>
</tbody>
</table>

Pressure is checked in 50mmHg increments up to 300mmHg and then down again. The data at each testing interval between 0-250mmHg going up has been combined with that going down.

Overestimates and underestimates include any difference of 0.1mmHg or more from the reference device.
Table 3: Monitor precision and failure rate by length of time in service

<table>
<thead>
<tr>
<th>Service duration</th>
<th>No. of monitors tested</th>
<th>Failure rate (n)</th>
<th>Mean Bias (mmHg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5 months</td>
<td>10</td>
<td>10% (1)</td>
<td>0.38 (-0.19 to 0.95)</td>
</tr>
<tr>
<td>6-12 months</td>
<td>17</td>
<td>0% (0)</td>
<td>0.74 (0.30 to 1.17)</td>
</tr>
<tr>
<td>13-18 months</td>
<td>11</td>
<td>9% (1)</td>
<td>0.72 (0.17 to 1.26)</td>
</tr>
<tr>
<td>19-24 months</td>
<td>9</td>
<td>33% (3)</td>
<td>1.09 (0.49 to 1.69)</td>
</tr>
<tr>
<td>24+ months</td>
<td>5</td>
<td>20% (1)</td>
<td>1.16 (0.35 to 1.96)</td>
</tr>
<tr>
<td>Unsure</td>
<td>9</td>
<td>22% (2)</td>
<td></td>
</tr>
</tbody>
</table>

*‘Unsure’ equates to when staff were unable to provide evidence or estimate the service duration of a monitor

Table 4a: Monitor precision and failure rate by amount of usage (as reported by pharmacy staff)

<table>
<thead>
<tr>
<th>Category of usage</th>
<th>Amount of usage</th>
<th>No. of monitors tested</th>
<th>Failure rate (n)</th>
<th>Mean Bias (mmHg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>2 a day or more</td>
<td>12</td>
<td>17% (2)</td>
<td>1.05 (0.55 to 1.55)</td>
</tr>
<tr>
<td>Medium high</td>
<td>More than 1 a day up to 2 a day</td>
<td>10</td>
<td>20% (2)</td>
<td>0.68 (0.14 to 1.22)</td>
</tr>
<tr>
<td>Medium low</td>
<td>More than 4 times a week up to 1 a day</td>
<td>16</td>
<td>6% (1)</td>
<td>0.65 (0.23 to 1.07)</td>
</tr>
<tr>
<td>Low</td>
<td>'Very little' up to 4 times a week</td>
<td>17</td>
<td>6% (1)</td>
<td>0.67 (0.19 to 1.15)</td>
</tr>
<tr>
<td>Unsure</td>
<td>Undefined</td>
<td>6</td>
<td>33% (2)</td>
<td></td>
</tr>
</tbody>
</table>

*‘Unsure’ equates to when staff were unable to provide evidence or estimate the amount of usage of a monitor
Table 4b: Monitor precision and failure rate by amount of usage (Head Office data)

<table>
<thead>
<tr>
<th>Category of usage</th>
<th>Annual usage rate</th>
<th>No. of pharmacies where monitors tested</th>
<th>Failure rate (n)</th>
<th>Mean Bias (mmHg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>116 - 275.5</td>
<td>13</td>
<td>31% (4)</td>
<td>1.22 (0.72 to 1.71)</td>
</tr>
<tr>
<td>Medium high</td>
<td>86.5 – 108.5</td>
<td>10</td>
<td>10% (1)</td>
<td>0.87 (0.34 to 1.41)</td>
</tr>
<tr>
<td>Medium low</td>
<td>60.5 – 82.5</td>
<td>13</td>
<td>15% (2)</td>
<td>0.78 (0.32 to 1.24)</td>
</tr>
<tr>
<td>Low</td>
<td>13.5 – 59.5</td>
<td>16</td>
<td>6% (1)</td>
<td>0.47 (0.04 to 0.90)</td>
</tr>
</tbody>
</table>

N.B. Data available only at level of pharmacy (4 failures happened in pharmacies where multiple devices were being used, making it difficult to interpret the relationship with usage)