Sources of inaccuracy in the measurement of adult patients’ resting blood pressure in clinical settings: a systematic review

Noa Kallioinen, Andrew Hill, Mark S. Horswill, Helen E. Ward, and Marcus O. Watson

Background: To interpret blood pressure (BP) data appropriately, healthcare providers need to be knowledgeable of the factors that can potentially impact the accuracy of BP measurement and contribute to variability between measurements.

Methods: A systematic review of studies quantifying BP measurement inaccuracy. Medline and CINAHL databases were searched for empirical articles and systematic reviews published up to June 2015. Empirical articles were included if they reported a study that was relevant to the measurement of adult patients’ resting BP at the upper arm in a clinical setting (e.g. ward or office); identified a specific source of inaccuracy; and quantified its effect. Reference lists and reviews were searched for additional articles.

Results: A total of 328 empirical studies were included. They investigated 29 potential sources of inaccuracy, categorized as relating to the patient, device, procedure or observer. Significant directional effects were found for 27; however, for some, the effects were inconsistent in direction. Compared with true resting BP, significant effects of individual sources ranged from –23.6 to +33 mmHg SBP and –14 to +23 mmHg DBP.

Conclusion: A single BP value outside the expected range should be interpreted with caution and not taken as a definitive indicator of clinical deterioration. Where a measurement is abnormally high or low, further measurements should be taken and averaged. Wherever possible, BP values should be recorded graphically within ranges. This may reduce the impact of sources of inaccuracy and reduce the scope for misinterpretations based on small, likely erroneous or misleading, changes.

Keywords: blood pressure determination, clinical deterioration, hypertension, measurement, vital signs

Abbreviations: BP, blood pressure; DBP, diastolic blood pressure; SBP, systolic blood pressure

INTRODUCTION

The measurement of blood pressure (BP) is a common procedure that is relied upon in a variety of healthcare settings. In hospital care, its uses include identifying clinical deterioration [1], informing vasoactive drug titration [2] and guiding goal-directed treatment [3]. In general practice, high BP values are used as a basis for the diagnosis of hypertension [4]. Inaccurate or misleading BP values, therefore, can be detrimental to the quality of healthcare received by patients.

Several guidelines have been published with the aim of improving the accuracy of BP measurements by standardizing the associated procedures [4–6]. These have primarily addressed measurements taken at the upper arm and have commonly included recommendations in relation to patient posture, cuff size, arm height, cuff deflation rate and the number of repeated measurements. Studies comparing BP measurements taken with strict adherence to guidelines vs. ‘usual technique’ have reported marked variation and differential treatment decisions between the two methods [7,8]. However, even after training on standardized procedures, BP measurement may be limited in its accuracy [9].

To interpret BP data appropriately, it would be beneficial for healthcare providers to be knowledgeable of the factors that can impact the accuracy of a BP measurement and contribute to variability between measurements. The current systematic review supplements existing guidelines by identifying empirically evaluated potential sources of inaccuracy and summarizing the evidence for each. In so doing, it also addresses the limitations of previous reviews of the BP measurement literature, which have included sources of inaccuracy that lack empirical evidence, or reported estimated effects based on small numbers of empirical studies [10–14]. The resulting list of empirically evaluated potential sources of inaccuracy could be incorporated into...

Journal of Hypertension 2017, 35:421–441

*School of Psychology, The University of Queensland, St. Lucia, 2Clinical Skills Development Service, Metro North Hospital and Health Service, Herston, 3The Prince Charles Hospital, Metro North Hospital and Health Service, Chermside and 4School of Medicine, The University of Queensland Mayne Medical School, Herston, Queensland, Australia

Correspondence to Andrew Hill, Clinical Skills Development Service, Metro North Hospital and Health Service, Herston 4006, QLD, Australia. Tel: +61 7 3636 4843; fax: +61 7 3646 6501; e-mail: Andrew.Hill@health.qld.gov.au

Received 23 February 2016 Revised 13 September 2016 Accepted 5 November 2016

J Hypertens 35:421–441. Copyright © 2017 Wolters Kluwer Health, Inc. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. This work cannot be changed in any way or used commercially without permission from the journal.

DOI:10.1097/HJH.0000000000001197
evidence-based training packages and used as a reference for clinicians when interpreting BP readings.

**METHODS**

A systematic literature review was performed to determine and categorize empirically evaluated potential sources of inaccuracy in the measurement of BP in adult patients, and determine the range of estimated effects of these sources of inaccuracy on measured BP values. The scope of this review was limited to studies relevant to upper arm BP measurement in clinical settings, such as wards and offices. Consequently, it did not directly address home BP monitoring, 24-h ambulatory BP monitoring or measurement at locations other than the upper arm.

**Initial exploratory searches**

To determine a list of potential sources of inaccuracy to investigate, the online databases CINAHL, The Cochrane Library, Medline and PsychINFO were searched using the terms: ‘vital sign measurement’, ‘vital sign AND (error OR accuracy)’, ‘blood pressure determination’, ‘blood pressure measurement’, ‘blood pressure error’, ‘blood pressure accuracy’ and ‘sphygmomanometer error’. Articles obtained from these searches aided the identification of specific sources of inaccuracy and the creation of individual search terms for each (Table 1). Two additional potential sources of inaccuracy (also included in Table 1 and the final searches) were added in response to reviewer comments (i.e. short interval between repeated measurements and acute meal ingestion).

**Final searches**

Individual searches were undertaken for each identified source of inaccuracy, using the Medline and CINAHL databases (which had been found to be the most relevant databases in the initial exploratory searches) via EBSCOHost. The searches covered journal articles from the inception of each database through to June 2015 and were limited to English language publications relating to humans. The search terms used for each potential source of inaccuracy are presented in Table 1. Reference lists from acquired empirical articles and systematic reviews were also searched for articles that may have not been identified directly from the database searches.

<table>
<thead>
<tr>
<th>Potential source of inaccuracy</th>
<th>Search terms</th>
<th>Number of results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute ingestion of food</td>
<td>blood pressure AND (food OR meal OR ingest+ OR eat) AND acute effect+ NOT (alcohol OR caffeine+ OR nicotine)</td>
<td>143</td>
</tr>
<tr>
<td>Acute alcohol use</td>
<td>blood pressure AND alcohol AND acute NOT withdraw+</td>
<td>274</td>
</tr>
<tr>
<td>Acute caffeine use</td>
<td>blood pressure AND (caffeine OR coffee OR ‘energy drink’) AND acute</td>
<td>139</td>
</tr>
<tr>
<td>Acute nicotine use</td>
<td>blood pressure AND (nicotine OR smoke+ OR cigarette+) AND acute effect+</td>
<td>215</td>
</tr>
<tr>
<td>Bladder distension</td>
<td>blood pressure AND ‘bladder distension’</td>
<td>23</td>
</tr>
<tr>
<td>Cold exposure</td>
<td>blood pressure AND cold exposure+</td>
<td>238</td>
</tr>
<tr>
<td>Paretic arm</td>
<td>blood pressure AND (paretic OR paralysed arm OR paralyzed arm)</td>
<td>9</td>
</tr>
<tr>
<td>White-coat effect</td>
<td>blood pressure AND (white coat effect OR white-coat effect)</td>
<td>294</td>
</tr>
<tr>
<td>Indirect measurement</td>
<td>aneroid AND mercury AND blood pressure</td>
<td>48</td>
</tr>
<tr>
<td>General device inaccuracy</td>
<td>sphygmomanometer+ AND agreement</td>
<td>101</td>
</tr>
<tr>
<td>Aneroid device inaccuracy</td>
<td>aneroid AND mercury AND blood pressure</td>
<td>48</td>
</tr>
<tr>
<td>Automated device inaccuracy</td>
<td>auto+ AND manual AND blood pressure</td>
<td>231</td>
</tr>
<tr>
<td>Pseudohypertension</td>
<td>pseudohypertension OR pseudo-hypertension</td>
<td>76</td>
</tr>
<tr>
<td>Rest period duration</td>
<td>(blood pressure measurement OR measuring blood pressure) AND before AND (wait+ OR rest+)</td>
<td>88</td>
</tr>
<tr>
<td>Body position</td>
<td>blood pressure AND (body position+)</td>
<td>224</td>
</tr>
<tr>
<td>Leg position</td>
<td>blood pressure AND (leg cross+ OR leg position+)</td>
<td>54</td>
</tr>
<tr>
<td>Unsupported back</td>
<td>blood pressure AND (back support+ OR supported back OR unsupported back OR back unsupported)</td>
<td>13</td>
</tr>
<tr>
<td>Unsupported arm</td>
<td>blood pressure AND (arm support+ OR supported arm OR unsupported arm OR arm unsupported)</td>
<td>37</td>
</tr>
<tr>
<td>Arm position</td>
<td>blood pressure AND arm position+</td>
<td>117</td>
</tr>
<tr>
<td>Cuff size</td>
<td>blood pressure AND (cuff+ size+ OR bladder+ size+)</td>
<td>221</td>
</tr>
<tr>
<td>Cuff tightness</td>
<td>blood pressure AND (loose+ OR tight+) AND cuff</td>
<td>13</td>
</tr>
<tr>
<td>Clothing effect</td>
<td>blood pressure AND (clothes OR clothing)</td>
<td>167</td>
</tr>
<tr>
<td>Stethoscope placement</td>
<td>blood pressure AND stethoscope AND (placement OR location)</td>
<td>4</td>
</tr>
<tr>
<td>Talking during measurement</td>
<td>blood pressure measure+ AND (talk OR talks OR talking)</td>
<td>28</td>
</tr>
<tr>
<td>Stethoscope head</td>
<td>blood pressure AND diaphragm AND bell</td>
<td>7</td>
</tr>
<tr>
<td>Deflation rate</td>
<td>blood pressure AND deflat+ rate</td>
<td>24</td>
</tr>
<tr>
<td>Interval between repeated measurements</td>
<td>(‘blood pressure measurement’ OR ‘blood pressure measurements’ OR ‘blood pressure readings’) AND interval AND time</td>
<td>142</td>
</tr>
<tr>
<td>Number of measurements</td>
<td>blood pressure AND number of measurements</td>
<td>65</td>
</tr>
<tr>
<td>Interarm difference</td>
<td>blood pressure AND (arm difference OR inter-arm difference OR interarm difference)</td>
<td>134</td>
</tr>
<tr>
<td>Auscultatory gap</td>
<td>auscultatory gap</td>
<td>8</td>
</tr>
<tr>
<td>General observer inaccuracy</td>
<td>blood pressure AND (observer error+ OR observer factors)</td>
<td>40</td>
</tr>
<tr>
<td>Korotkoff sound interpretation</td>
<td>korotkoff OR korotkov AND (sound+ OR phase+)</td>
<td>255</td>
</tr>
<tr>
<td>Terminal digit bias</td>
<td>blood pressure AND (terminal digit OR end digit OR final digit OR last digit OR terminal-digit OR end-digit OR final-digit OR last-digit)</td>
<td>50</td>
</tr>
</tbody>
</table>
Inclusion criteria
Publications that contained all of the following were eligible for inclusion in the final analysis:

1. Results from an empirical study relevant to the measurement of adult patients’ resting BP at the upper arm in a clinical setting (e.g. ward or office);
2. Identification of one or more specific potential source(s) of inaccuracy in the measurement of BP; and
3. Quantification of the independent effects of one or more source(s) of inaccuracy on the measured value of SBP and/or DBP, or the prevalence of such an inaccuracy.

Search outcome
The Medline and CINAHL database searches on each individual potential source (Table 1) yielded a total of 3147 articles, excluding duplicates. The abstracts and/or titles of the pooled articles were then evaluated for relevance to the inclusion criteria, after which 783 full texts were obtained. Of these, a total of 328 publications were summarized in this review (Fig. 1) after articles that did not meet the inclusion criteria were excluded (e.g. because they included child participants, confounded multiple potential sources of inaccuracy or did not use a quantifiable measure).

Quality appraisal
Due to the diverse nature of research questions and study methodologies reported, the Standard Quality Assessment Criteria for Evaluating Primary Research Papers from a Variety of Fields were considered appropriate for quality appraisal [15]. These criteria can be used to assess the quality of various types of empirical studies by rating aspects of the design, participants, analysis and results. Summation of the individual scores for each area was used to represent the overall quality of a study, expressed as a percentage of the maximum possible score applicable to the study. These quality indicators are not directly comparable due to the many differences between studies. However, they can be used as a guide when interpreting the quality of evidence from individual studies (Supplementary Tables 1–29, http://links.lww.com/HJH/A702).

Data extraction
After selection of articles from title and abstract review, the full-texts of potentially relevant articles were analysed. Information relating to country of origin, observers, participants, devices and procedures was extracted from articles meeting the inclusion criteria, along with the effects of the specific source of inaccuracy on patients’ SBP and/or DBP, or its prevalence. Extracted information varied depending on the source of inaccuracy, with some sources requiring more categories of data to be obtained than others. These procedures were conducted by one reviewer (N.K.) and checked by another (M.O.W.), with disagreements resolved through discussion.

Synthesis
Meta-analysis was not considered appropriate due to the diversity of research questions addressed by the articles (a total of 29 potential sources of inaccuracy were identified), and marked variation in methodologies, reporting detail and study populations. Instead, study results were aggregated in a table outlining the empirically evaluated potential sources of inaccuracy, and the ranges of their reported mean effects across studies (Table 2). Note that ranges

![Flow chart showing study selection.](http://links.lww.com/HJH/A702)
expressed in mmHg (i.e. results for all potential sources of inaccuracy except device calibration error and terminal digit preference for zero) only include statistically significant effects. However, detailed descriptions of all studies that met the inclusion criteria (including those that lacked significance testing) are included in the data supplement, in which they are organized into tables according to the relevant potential source of inaccuracy (Supplementary Tables 1–29, http://links.lww.com/HJH/A702). For ease and consistency of communication, we adopted the following conventions in our narrative descriptions of the size of potential inaccuracies in BP measurement: small = less than 5 mmHg, moderate = 5–15 mmHg and large = greater than 15 mmHg.

---

**TABLE 2. Empirically-evaluated potential sources of inaccuracy in the measurement of adults’ resting blood pressure (note that ranges expressed in mmHg include only significant effects and hence exclude results from studies reported in Supplementary Tables that lacked significance testing, [http://links.lww.com/HJH/A702](http://links.lww.com/HJH/A702))**

<table>
<thead>
<tr>
<th>Potential source of inaccuracy</th>
<th>Range of reported significant mean effects (in mmHg) unless specified</th>
<th>Suppl. Table number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient-related</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Acute meal ingestion [16,17]</td>
<td>−6 to −5 −1.9</td>
<td>1</td>
</tr>
<tr>
<td>2. Acute alcohol use [18–35]</td>
<td>−23.6 to +24</td>
<td>2</td>
</tr>
<tr>
<td>3. Acute caffeine use [40–74]</td>
<td>+3 to +14</td>
<td>3</td>
</tr>
<tr>
<td>4. Acute nicotine use or exposure [70,75–103]</td>
<td>+2.8 to +25</td>
<td>4</td>
</tr>
<tr>
<td>5. Bladder distension [104–106]</td>
<td>+4.2 to +33</td>
<td>5</td>
</tr>
<tr>
<td>6. Cold exposure [107–115]</td>
<td>+5 to +32</td>
<td>6</td>
</tr>
<tr>
<td>7. Arm lower than heart level [255,269,271–278]</td>
<td>+2 to +5</td>
<td>7</td>
</tr>
<tr>
<td>8. White-coat effect [120–160]</td>
<td>−12.7 to +26.7</td>
<td>8</td>
</tr>
<tr>
<td>Device-related</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Device model bias</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mercury models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vs. invasive criterion [167–170]</td>
<td>−10.6 to −4</td>
<td>9A</td>
</tr>
<tr>
<td>Aneroid models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vs. invasive criterion [162,171,172]</td>
<td>−9.7 to −4.0</td>
<td>9B</td>
</tr>
<tr>
<td>vs. noninvasive criterion [173–179]</td>
<td>−0.8 to −1.7</td>
<td>9C</td>
</tr>
<tr>
<td>Automated models</td>
<td></td>
<td></td>
</tr>
<tr>
<td>vs. invasive criterion [162,167,168,171,182–189]</td>
<td>−23 to +6</td>
<td>9D</td>
</tr>
<tr>
<td>vs. noninvasive criterion [167,176,190–227]</td>
<td>−3.7 to +16.53</td>
<td>9E</td>
</tr>
<tr>
<td>Procedure-related</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Insufficient rest period [249,250]</td>
<td>+4.2 to +11.6</td>
<td>11</td>
</tr>
<tr>
<td>12. Body position</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing [251,252]</td>
<td>−2.9 to +5</td>
<td>12</td>
</tr>
<tr>
<td>Supine [251–259]</td>
<td>−10.7 to +9.5</td>
<td>12</td>
</tr>
<tr>
<td>13. Legs crossed at knees [260–266]</td>
<td>+2.5 to +14.89</td>
<td>13</td>
</tr>
<tr>
<td>14. Unsupported back [267]</td>
<td>No significant effects reported</td>
<td>14</td>
</tr>
<tr>
<td>15. Unsupported arm [268–270]</td>
<td>+4.87 to +6.9</td>
<td>15</td>
</tr>
<tr>
<td>16. Arm lower than heart level [255,269,271–278]</td>
<td>+3.7 to +23</td>
<td>16</td>
</tr>
<tr>
<td>17. Incorrect choice of cuff size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smaller cuff [279–283]</td>
<td>+2.08 to +11.2</td>
<td>17</td>
</tr>
<tr>
<td>Larger cuff [279–283]</td>
<td>−3.7 to −1.45</td>
<td>17</td>
</tr>
<tr>
<td>18. Cuff placed over clothing [284–287]</td>
<td>No significant effects reported</td>
<td>18</td>
</tr>
<tr>
<td>19. Stethoscope under cuff [288,289]</td>
<td>+1.0 to +3.1</td>
<td>19</td>
</tr>
<tr>
<td>20. Talking during measurement [290–295]</td>
<td>+4.0 to +9.4</td>
<td>20</td>
</tr>
<tr>
<td>21. Use of stethoscope bell vs. diaphragm [267,296–299]</td>
<td>−3.8 to +1.54</td>
<td>21</td>
</tr>
<tr>
<td>22. Excessive pressure on stethoscope head [300]</td>
<td>No significant effects reported</td>
<td>22</td>
</tr>
<tr>
<td>23. Fast cuff deflation rate [301–303]</td>
<td>−9 to −2.6</td>
<td>23</td>
</tr>
<tr>
<td>24. Short interval between measurements [304–306]</td>
<td>No significant effects reporteda</td>
<td>24</td>
</tr>
<tr>
<td>25. Reliance on a single measurement [147,258,259,275,307,308]</td>
<td>+3.3 to +10.4</td>
<td>25</td>
</tr>
<tr>
<td>Observer-related</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. Observer hearing deficit [328]</td>
<td>−1.55 to −0.11</td>
<td>27</td>
</tr>
<tr>
<td>28. Korotkoff Phase IV (vs. V) for DBP [334,335]</td>
<td>N/A</td>
<td>28</td>
</tr>
</tbody>
</table>

Note: Only one study found a significant effect.

a No study found a significant directional effect (although one found reduced variation in SBP).
RESULTS
A total of 328 full-text articles were summarized in this review. Collectively, these studies investigated a total of 29 distinct empirically evaluated potential sources of inaccuracy in the measurement of adult BP in clinical settings. These sources were organized into four categories, similar to those used in previous reviews: patient-related, device-related, procedure-related and observer-related [11,13,14]. Results for the potential sources of inaccuracy within each category are summarized in Table 2 and outlined below. Each potential source of inaccuracy has been numbered (from 1 to 29) to facilitate cross-referencing between text and tables.

Patient-related sources of inaccuracy
BP is a dynamic variable: its value naturally changes over time and within different regions of the body. This review identified eight distinct empirically evaluated patient-related sources of inaccuracy that may cause additional variation, contributing to an underestimation or overestimation of the patient’s ‘true’ resting BP.

1. Acute meal ingestion
Two studies were included that reported the effect of acute meal ingestion on BP (Supplementary Table 1, http://links.lww.com/HJH/A702) [16,17]. One study reported the effects of a mixed meal at two specific intervals after ingestion, finding moderate significant decreases in both SBP and DBP at 180 min, but no significant effect at 60 min. The other study assessed the mean effect of a light breakfast meal over a 2-h period after ingestion (averaging over measurements taken every 15 min) finding no significant effect on SBP and a small decrease in DBP.

2. Acute alcohol use
Eighteen studies were included that reported the effect of acute alcohol ingestion on BP (Supplementary Table 2, http://links.lww.com/HJH/A702) [18–35]. Alcohol dosages ranged from 0.4 to 1 g/kg (relative to the participant’s body weight), and one study reported an absolute dosage of 41 g [18]. Significant effects of acute alcohol ranged from large decreases to large increases in SBP and moderate decreases to large increases in DBP (Table 2). Only a small number of studies found no significant effects. The studies employed inconsistent time intervals between alcohol consumption and measurement (ranging from 5 min to 4 h). However, with the exception of only two studies [21,28], all significant increases were measured within 60 min of consumption, and all significant decreases from 60 min to 4 h postconsumption. Of the exceptions, one study found significant decreases in both SBP and DBP 30 min after ingestion [21]. The other study was not directly comparable because it reported the maximum BP response up to 3-h postconsumption, rather than using a fixed time interval [28].

3. Acute caffeine use
Three systematic reviews and one narrative review were identified that assessed the short-term effect of caffeine consumption on BP [36–39]. A total of 35 empirical studies that were eligible for inclusion in this review were identified, some having been included previously in other reviews (Supplementary Table 3, http://links.lww.com/HJH/A702) [40–74]. Common caffeine dosages evaluated were 3.3 mg/kg (relative to the patient’s weight) and 250 mg (absolute). Others ranged from relative dosages of 1.5–6.0 mg/kg and absolute dosages of 67–400 mg. All reported significant mean effects on SBP or DBP were increases, ranging in magnitude from small to moderate (Table 2). Some small caffeine dosages, such as 100 and 200 mg, were not consistently reported to have a significant effect on BP [59,61]. Studies also employed different time intervals between consumption and measurement (with those intervals that were fixed ranging from 30 to 270 min). Significant effects were reported for intervals as long as 180 min after consumption.

4. Acute nicotine use or exposure
A total of 30 articles measuring the short-term effect of cigarette smoking, nicotine ingestion or passive cigarette smoke exposure were evaluated (Supplementary Table 4, http://links.lww.com/HJH/A702) [70,75–103]. The reported significant effects ranged from small to large increases for both SBP and DBP (Table 2). The majority of studies reported dosages in mass of nicotine, ranging from 0.4 to 13 mg. Other dosages included one and four cigarettes over 1 h, both of unspecified nicotine content [80,94]. Most studies examined the effect of conventional cigarette smoking; however, the effects of electronic cigarettes, bidi cigarettes, cigars, Arabian pipes, nicotine tablets, nicotine patches, nicotine gum and snuff tobacco were also reported. Studies also used different time intervals between nicotine use and measurement (including during, immediately after and up to 60 min later). Significant effects of acute nicotine use were measured as far as 20–30 min after smoking, 30 min after snuff chewing and 40–60 min after ingesting a 4-mg nicotine tablet.

The four studies that investigated the effect of passive smoking employed exposure durations of 30–60 min [100–103]. Of these, only two studies reported significant effects. One found large increases to both SBP and DBP [100]; in the other study, however, the increases were only moderate and were confined to SBP in male participants [103].

5. Bladder distension
Three studies were included that reported the effect of a full bladder on BP (Supplementary Table 5, http://links.lww.com/HJH/A702) [104–106]. Effects ranged from small to large increases in both SBP and DBP (Table 2). In the study that found only small effects, participants drank as normal over a mean period of 7 h, with BP measured before and after voiding their full bladders [104]. The other studies found large increases in both SBP and DBP within 60 min after drinking 1200 ml of water [106], and after drinking water until the urge to void became irresistible [105].
6. Cold exposure
Nine studies that assessed the effect of patient exposure to cold were identified (Supplementary Table 6, http://links.lww.com/HJH/A702) [107–115]. Reported effects ranged from moderate to large increases in SBP and small to large increases in DBP (Table 2). These studies compared resting BP at varying temperatures. The majority compared air temperatures, whereas some used skin temperatures. All studies found that exposure to cold significantly increased SBP (and, where reported, DBP).

7. Paretic arm
Two studies were evaluated that compared BP measurements between the intact and paretic arms of stroke patients (Supplementary Table 7, http://links.lww.com/HJH/A702) [116,117]. One found that measurements from paretic arms yielded small but significant increases in both SBP and DBP relative to measurements from intact arms (Table 2) [117]. The other study found no significant effects for either SBP or DBP [116].

8. White-coat effect
The white-coat effect is the phenomenon in which a patient’s BP changes when in the presence of a clinician. Generally, this change is assumed to involve an increase in both SBP and DBP from normal daytime values [4]. Two systematic reviews with meta-analyses relating to the white-coat effect were identified in the literature search [118,119]. However, these focused on the differences in white-coat effect between age groups and ethnic groups, respectively, and not on the mean white-coat effect in general.

A total of 41 empirical studies were analysed in the current study, including some that had not been included in the aforementioned systematic reviews. A large range of estimates for the white-coat effect was apparent (Supplementary Table 8, http://links.lww.com/HJH/A702) [120–160]. Reported significant effects of the presence of a clinician ranged from moderate decreases to large increases in both SBP and DBP (Table 2). The most common method used to quantify the white-coat effect was to compare daytime ambulatory or home self-measurements with those made by a clinician in a clinical setting. By subtracting patients’ normal daytime BPs from clinical measurements, the impact of the clinician’s presence and the clinical setting could be quantified. A negative white-coat effect indicated that, on average, patients’ BP decreased in the presence of a clinician, whereas a positive value indicated that it increased. Of the studies that yielded significant results, all but four produced only positive mean white-coat effects. Of the four exceptions, one produced only negative mean effects [158], and three produced positive and negative effects for different sample groups [145,147,148]. Characteristics of the observers and patients (such as age, sex, ethnic group and the presence of hypertension) differed between studies and sample groups and may therefore have impacted the magnitude and direction of reported effects. The aforementioned systematic reviews addressed some of these differences, finding that older patients experienced greater positive white-coat effects than younger patients [118], and South Asian patients experienced smaller white-coat effects than patients of African or European descent, on average [119].

Further potential patient-related sources of inaccuracy
Other potential sources of inaccuracy were mentioned in the reviewed literature; however, they lacked valid empirical evidence for their effects in clinical settings. These included soft Korotkoff sounds (the audible cues heard from the stethoscope when performing manual sphygmomanometry) [13], and time of day [161].

Device-related sources of inaccuracy
There are two primary methods of BP measurement: invasive and noninvasive. Invasive, or intra-arterial, BP measurement through the use of a catheter is regarded as the gold standard of BP measurement [162,163]. However, due to its invasiveness, it is used less commonly than noninvasive alternatives. Most noninvasive BP measurements typically make use of an inflatable cuff to temporarily occlude blood flow in the upper arm. As the cuff is deflated, different methods can be employed to determine the SBP and DBP, depending on the type of device. These include the use of auditory or oscillometric cues.

9. Device model inaccuracy
A variety of sphygmomanometer devices can be used for noninvasive measurement of BP. Two primary types of clinical device were identified: manual and automated. Manual sphygmomanometers can be either mercury or aneroid. Due primarily to the toxicity of mercury, the use of mercury-free devices (both aneroid and automated) has increased in recent years [164].

Two methods of measuring the inaccuracy of different models were apparent in the literature: comparison with invasive measurement and comparison with noninvasive ‘standard manual measurement’. Reported inaccuracies from these two methods are discussed separately. The standard reporting of the inaccuracy of devices includes the mean bias (i.e. the average of the difference scores, which may be positive or negative) and 95% limits of agreement. For a particular study comparing a test device to a criterion standard (e.g. invasive measurement), these limits specify the lowest and highest difference scores that would fall within 2 SDs of the mean difference score for that particular sample, thus capturing approximately 95% of difference scores. The width of the limits of agreement indicates how much the difference score (or bias) varies from measurement-to-measurement. Wider limits of agreement indicate greater variation, whereas narrow limits indicate less variation. Therefore, when interpreting the accuracy of a device, both the mean bias and the width of the limits of agreement should be considered [165,166]. Supplementary Tables 9A–9E, http://links.lww.com/HJH/A702 include the mean biases and 95% limits of agreement for all included studies of device model inaccuracy.
Four studies were identified that examined the accuracy of mercury devices in comparison with the criterion standard of invasive measurement (Supplementary Table 9A, http://links.lww.com/HJH/A702) [167–170]. Three of these investigated standard mercury devices, whereas one investigated the accuracy of a random-zero device. Reported significant mean biases ranged from small to moderate underestimation of SBP, and small overestimation of DBP (Table 2), with moderately wide limits of agreement in each study (between ±13 and ±16.2 mmHg for SBP, and between ±6 and ±13 mmHg for DBP).

Aneroid devices: Three studies reported the bias of three different models of aneroid device compared with invasive measurement (Supplementary Table 9B, http://links.lww.com/HJH/A702) [162,171,172]. All studies reported underestimation of SBP (including two small-to-moderate significant effects) and overestimation of DBP (including one moderate significant effect) (Table 2). There was noticeable variation in the widths of the limits of agreement across studies, indicating the variability in accuracy between devices (between ±8.6 and ±27.1 mmHg for SBP, and between ±4.6 and ±14 mmHg for DBP).

Seven studies compared a total of seven different aneroid models to a noninvasive criterion standard manual measurement (Supplementary Table 9C, http://links.lww.com/HJH/A702) [173–179]. Significant reported biases were small underestimations of both SBP and DBP (Table 2). Again, there was noticeable variation in the widths of the limits of agreement (between ±4.4 and ±25 mmHg for SBP, and between ±5.2 and ±15.64 mmHg for DBP).

Automated devices: Two systematic reviews were identified that examined the accuracy of different models of automated devices compared with invasive measurement (Supplementary Table 9D, http://links.lww.com/HJH/A702) [162,167,168,171,182–189]. Reported significant mean biases ranged from large underestimation to moderate overestimation of SBP and small underestimation to moderate overestimation of DBP (Table 2). There was very marked variation in the width of the 95% limits of agreement with the invasive criterion, reflecting the diversity of automated devices that were evaluated (between ±10 and ±52.9 mmHg for SBP, and between ±8 and ±29.2 for DBP).

A total of 40 studies compared 33 different automated models with a noninvasive criterion standard (Supplementary Table 9E, http://links.lww.com/HJH/A702) [167,176,190–227]. The majority of studies used standard mercury measurement as the criterion, although some used aneroid measurement. Significant mean biases ranged from small underestimation to large overestimation for SBP and moderate underestimation to moderate overestimation for DBP. Again, there was marked variation in the width of the 95% limits of agreement (between ±5.2 and ±35.9 mmHg for SBP, and between ±3.8 and ±26.68 mmHg for DBP).

Various reasons why the discrepancy between noninvasive and invasive measures might be more apparent in some patients than others were also discussed in the literature. For example, patients with shock have been identified to exhibit lower noninvasive SBP than invasive SBP [169]. In addition, erroneously high BP values from noninvasive techniques, known as pseudo-hypertension, may be related to arterial stiffness [228].

10. Device integrity and calibration

Incorrect calibration or poor device integrity can affect all types of sphygmomanometers. This can lead to measurement error in either direction that is constant for a particular device (at least in the short-term). One review [164] and nineteen empirical studies were identified. Ten of these empirical studies assessed the calibration accuracy of mercury sphygmomanometers (Supplementary Table 10A, http://links.lww.com/HJH/A702) [229–238], 18 assessed aneroid sphygmomanometers (Supplementary Table 10B, http://links.lww.com/HJH/A702) [229–237,239–247] and three assessed automated devices (Supplementary Table 10C, http://links.lww.com/HJH/A702) [229–231]. The common method of analysing device inaccuracy was to measure the deviation of sphygmomanometer readings from a known accurate criterion device. The majority of studies categorized an individual sphygmomanometer as inaccurate if it showed a discrepancy of greater than ±3 mmHg from a criterion device. However, thresholds of ±4 and ±5 mmHg were also employed in some studies [242,247].

The prevalence of inaccurate devices varied greatly between studies and device types, with more than half of the tested devices showing inadequate accuracy in some studies, but fewer than 5% in other studies (Table 2). It should be noted that different studies evaluated the accuracy of sphygmomanometers at different pressures, and this may limit general comparisons between studies and device types. Specific issues with sphygmomanometers that were reported to decrease accuracy included blocked vents and leaky bulb valves [238,239,248].

11. Insufficient rest period

Few studies directly assessed the effect of rest period duration on BP. Of these, two were deemed to be eligible for inclusion in this review (Supplementary Table 11, http://links.lww.com/HJH/A702 [249,250]. The studies differed in methodology, including the rest periods that they compared (i.e. 5 vs. 10 and 0 vs. 16 min) but both found that patients had significantly higher SBP and DBP if they did not rest for a sufficient period of time prior to measurement. It was shown that resting for 10 or 16 min could decrease SBP by a small-to-moderate amount and DBP by a small amount (Table 2). However, it is unclear from these results whether longer rest periods are necessary to completely reverse the effects of prior physical activity.
12. Body position
Nine studies comparing patient body positions were included in this review (Supplementary Table 12, http://links.lww.com/HJH/A702). Of these, only two studies investigated the effects of standing, and they yielded bidirectional results. One study reported moderate increases for both SBP and DBP (in hypertensive patients) [251], whereas the other reported a small decrease in SBP and no significant effect on DBP, compared with sitting (in healthy young adults) [252].

All studies in this category compared supine posture with sitting posture [251–259]. The supine position generally yielded small-to-moderate significant increases in SBP relative to the sitting position [252,253,255,258,259]. Although one study showed a statistically significant moderate decrease in SBP, the researchers did not control for order effects; consequently, participants had been resting for a longer period by the time that their BP was measured in the supine position [254]. A further two studies found no statistically significant effect [251,256]. Similarly, DBP was shown to be affected bidirectionally by supine posture; most studies reported significant decreases [251,253,254,256,258] but two reported significant increases [255,259]. These effects ranged from moderate decreases to moderate increases (Table 2).

13. Legs crossed at knees
The effect of leg position was assessed by comparing legs crossed at the knee with feet flat on the floor. All seven studies reported an increase in SBP when legs were crossed compared with sitting with feet flat (Supplementary Table 13, http://links.lww.com/HJH/A702) [260–266]. Significant effects ranged from small to large increases in both SBP and DBP.

14. Unsupported back
A single study was evaluated that addressed the effect of back support on BP measurement (Supplementary Table 14, http://links.lww.com/HJH/A702) [267]. It was reported that sitting on a table with no back support increased DBP (but not SBP) by a moderate amount, compared with sitting on a chair with back support (Table 2). However, the results should not be overinterpreted as the study did not control for potential confounds between the table and the chair, such as height (which may have affected whether or not participants’ feet reached the floor) and level of comfort.

15. Unsupported arm
Three studies addressed measurement of BP in an unsupported arm compared with a supported arm (Supplementary Table 15, http://links.lww.com/HJH/A702) [268–270]. Small but significant increases were reported for both SBP and DBP when measured in the unsupported arm (Table 2).

16. Arm lower than heart level
The height of the measured arm relative to the heart was also found to affect BP. Ten studies included in this review compared BP when the measured arm was at heart level vs. lower than heart level (Supplementary Table 16, http://links.lww.com/HJH/A702) [255,269,271–278]. Both SBP and DBP were found to increase (by a small to large amount) when the arm was lower than heart level (Table 2). The majority of studies compared measurements of the arm at heart level with the arm hanging by the patient’s side. The magnitude of the increase may also have been affected by body position (which varied among standing, sitting and supine), or the absolute or relative distance of the arm from heart level (which was seldom reported).

17. Incorrect choice of cuff size
Five studies included in this review indicated that cuff width affected BP values (Supplementary Table 17, http://links.lww.com/HJH/A702) [279–283]. Generally, an erroneous reduction in cuff size significantly increased the measurements for SBP and DBP (to a small or moderate degree), whereas an increase in cuff size significantly decreased the measurements (to a small degree) (Table 2). Most studies compared specific cuff sizes with the recommended cuff size, which was dependent on the arm circumference of the patient (Supplementary Table 17, http://links.lww.com/HJH/A702 for details).

18. Cuff placed over clothing
Four studies were included that examined the effect of measuring BP with the cuff placed over clothing (Supplementary Table 18, http://links.lww.com/HJH/A702) [284–287]. However, all reported nonsignificant effects. Guidelines often specify that the sphygmomanometer cuff should be placed over a bare arm; however, the empirical evidence does not provide any clear support for this recommendation (Table 2).

19. Stethoscope under cuff
Two studies addressed the effect of placing the stethoscope under the cuff during auscultation (Supplementary Table 19, http://links.lww.com/HJH/A702) [288,289]. They reported small but significant increases in measured SBP and small-to-moderate decreases in measured DBP, when compared with placing the stethoscope in the recommended position outside the cuff (Table 2).

20. Talking during measurement
Six studies on the effect of the patient talking during BP measurement were evaluated, all of which found increases in both SBP and DBP (Supplementary Table 20, http://links.lww.com/HJH/A702) [290–295]. Significant effects ranged from small to large increases for SBP, and small-to-moderate increases for DBP, compared with not talking (Table 2). However, variation in methodologies between studies, including the verbal content, may have affected the magnitude of BP increases.

21. Use of stethoscope bell (vs. diaphragm)
Five studies were identified that investigated the effect of using the bell rather than the diaphragm of the stethoscope to measure Korotkoff sounds (Supplementary Table 21, http://links.lww.com/HJH/A702) [267,296–299]. Two studies reported significant effects of using the bell [297,299]. It reduced the SBP measurement by a small amount in trauma patients and increased it by a small amount in young women (Table 2). In these studies, the effect on DBP ranged
from no difference to a small decrease (for young women). Other similar studies found no significant difference between bell and diaphragm.

22. Excessive pressure on stethoscope head
One study reported moderate decreases in the value of DBP when excessive force was applied to the head of the stethoscope during auscultation, with greater force yielding a greater decrease (Table 2, Supplementary Table 22, http://links.lww.com/HJH/A702 [300]).

23. Fast cuff deflation rate
Three studies examined the effect of cuff deflation faster than the recommended 2–3 mmHg/s (Supplementary Table 23, http://links.lww.com/HJH/A702) [301–303]. Reported significant effects ranged from small to moderate decreases in SBP and small to moderate increases in DBP (Table 2). The magnitude of effects was found to depend on the speed of deflation and the patient’s heart rate.

24. Short interval between measurements
Three studies were identified that investigated the effect of a shorter-than-usual interval between subsequent measures (Supplementary Table 24, http://links.lww.com/HJH/A702) [304–306]. Of these, only two reported the mean differences in SBP and DBP between different intervals (specifically 15 s vs. 1 min and 1 vs. 2 min), but they were not statistically significant. The third study did not report mean differences as it was focused specifically on variation (i.e. fluctuation) in BP. Results indicated that, when a standard mercury sphygmomanometer (but not an electronic device) was used, SBP (but not DBP) measures were significantly less variable when consecutive measurements were taken with no interval to separate them vs. a 1-min interval.

25. Reliance on a single measurement
Six studies were identified that investigated potential variability between consecutive measurements of BP (Supplementary Table 25, http://links.lww.com/HJH/A702) [147,258,259,275,307,308]. The intervals between compared measurements ranged from 1 to 45 min. Studies reporting statistically significant differences consistently found that, on average, initial measurements of SBP were higher by small-to-moderate amounts, compared with measurements taken 1–25 min later. However, of the two studies that reported a significant mean difference in SBP (both of which employed a 2-min interval and yielded a small effect), the initial measurement was higher in one study [147], and lower in the other [259] (Table 2). These findings were interpreted as indicating a need for multiple consecutive measurements to minimize inaccuracy.

26. Interarm variability
Two systematic reviews on the differences in BP between the left and right arms of individuals were identified [309,310]. Fifteen empirical studies, some already included in these reviews, were analysed (Supplementary Table 26, http://links.lww.com/HJH/A702) [311–325].

Because individuals can differ in terms of which arm (i.e. right or left) will yield higher BP readings, some studies reported absolute interarm differences (i.e. calculated for each participant by deducting the lower value from the higher value, irrespective of which arm each was measured at, before averaging). These studies reported small-to-moderate significant mean absolute interarm differences for both SBP and DBP (Table 2). For a patient whose BP is only measured in one arm, these values can be regarded as estimates of the average amount by which their apparent BP may have been higher or lower if it had been measured using the other arm.

Among studies that instead deducted the values for one specific arm from the values for the other arm before averaging, small-to-moderate significant mean interarm differences were also reported for SBP (+1.6 to +5.1 mmHg) and DBP (–1.4 to +3.73 mmHg) (standardized as right minus left). However, these latter figures represent estimates of the extent to which one specific arm tends to yield a higher apparent BP in the study population (e.g. positive values indicate that the right arm is higher), rather than the likely magnitude of interarm differences for patients within that population. This is because, in averaging over participants, values from those who have higher right-arm BP will, to some extent, ‘cancel out’ values from those who have higher left-arm BP (and vice-versa).

Further potential procedure-related sources of inaccuracy
Other unsubstantiated procedure-related sources of inaccuracy that were mentioned in the reviewed literature included the temperature of the stethoscope or the examiner’s hands, environmental noise and initial inflation into an auscultatory gap [13]. This gap exists in some patients and is a range between SBP and DBP in which there are no audible Korotkoff sounds [326,327].

Observer-related sources of inaccuracy
Sources of inaccuracy related to the observer are primarily apparent during manual sphygmomanometry (either mercury or aneroid). Empirically evaluated sources of inaccuracy of this kind included observer hearing deficit, incorrect Korotkoff sound interpretation and terminal digit bias.

27. Observer hearing deficit
One study was identified that assessed the impact of hearing loss on the manual measurement of BP (Supplementary Table 27, http://links.lww.com/HJH/A702) [328]. When hearing loss of 5, 10, 15, 20 and 25 dB was simulated across all frequencies, small decreases in SBP and small increases in DBP values were reported (Table 2).

However, the flat decrease of 5–25 dB across all frequencies does not accurately represent common forms of hearing loss, such as presbyacusis (age-related). Presbyacusis primarily affects the detection of higher frequency sounds, particularly those with frequency bands greater than 1000 Hz, whereas Korotkoff sounds have frequencies of 20–300 Hz [329–331]. Therefore, conclusions from this study should not be generalized to actual hearing loss.

28. Korotkoff sound interpretation
The pressures at which Korotkoff sounds appear and disappear in manual sphygmomanometry correspond to the
SBP and DBP, respectively. Differing interpretation of these sounds can have a bidirectional effect on measurement. There has been debate regarding which phase of Korotkoff sounds best corresponds to the DBP – phase IV (muffling of sounds) or phase V (disappearance of sounds) [332,333]; however, phase V has been recommended for most circumstances [4]. Two studies were evaluated that quantified the effect of different Korotkoff sound interpretations (Supplementary Table 29, http://links.lww.com/HJH/A702) [334,335]. Measuring DBP at phase IV (rather than phase V) was reported to significantly increase the DBP measurement by a moderate amount (Table 2).

29. Terminal digit preference for zero

Terminal digit bias refers to the preference to round BP readings to a specific end-digit, usually zero. An observer who biases zero as an end digit may, for example, record 120 mmHg when the sphygmomanometer shows a value between 117 and 122 mmHg. Thirty-four studies were identified that reported over-representation of terminal zeros in BP records, indicative of such a bias (Supplementary Table 29, http://links.lww.com/HJH/A702) [8,275,334,336–336]. Given that the expected frequency of terminal zero is 20% (because measurements are generally recorded to the nearest even number), approximately 79% of all SBP and DBP values analysed may have included erroneous terminal zeros in the study that reported the highest prevalence (i.e. 99%) (Table 2) [350].

Further potential observer-related sources of inaccuracy

Other observer-related sources of inaccuracy were mentioned in the reviewed literature but found to lack sufficient empirical evidence. These included value bias and visual impairment [4]. Value bias is related to, but distinct from, terminal digit bias. Although terminal digit bias is a preference for a specific final digit of the reading, value bias relates to the entire value of the reading. For example, in one study, there was a reported preference for recording 88 mmHg for DBP when a value of 90 mmHg would lead to treatment consideration [358]. In addition, one article examined ‘gauge reading errors’ relating to the reading of a mercury column in conjunction with Korotkoff sounds [367]. The researchers assessed the readings of groups of novice and experienced individuals in a single-trial simulation. They reported under-reading for SBP and over-reading for DBP, compared with the true value. In addition, there was notable variability between observers, as indicated by a moderately sized SD. However, due to the highly simplified representation of the mercury column as a rectangular bar presented on a computer screen, the single-trial design and a lack of statistical significance testing, the conclusions that can be drawn from this particular study are extremely limited.

DISCUSSION

The current systematic review has described 29 empirically evaluated potential sources of inaccuracy in the measurement of adults’ resting BP at the upper arm in clinical settings, such as wards and offices (Table 2). Each of these was categorized as relating to either: the patient, the device, the procedure or the observer. Significant directional effects on SBP and/or DBP were found for 27 of the 29 potential sources of inaccuracy, and the ranges of these effects were presented for each source. The significant effects of individual sources were highly variable and ranged from a mean underestimation of 24 mmHg to a mean overestimation of 33 mmHg for SBP, and a mean underestimation of 14 mmHg to a mean overestimation of 23 mmHg for DBP. Differences of this magnitude between ‘true’ resting BP and measured BP can have important implications in several clinical domains, including the physiological monitoring of inpatients, and the diagnosis and surveillance of hypertension.

Clinical impact

Patient monitoring

The majority of patient monitoring protocols, such as track-and-trigger systems, incorporate routine BP measurement, obtained through either manual or automated methods [1]. Typically, the patient’s vital sign values are compared against predefined ranges to determine whether they are ‘normal’ or ‘abnormal’ and, in some cases, to categorize the extent of any deviation from the normal range. Monitoring protocols also vary in terms of the mechanisms by which clinical responses are triggered [368,369]. In some systems, a criterion breach in relation to just one individual vital sign – such as SBP – can be enough to trigger a clinical response, whereas other systems require multiple vital signs to reach predefined thresholds before a response is triggered. More complex monitoring protocols utilize both of these mechanisms in different circumstances.

The current systematic review has presented evidence indicating that up to 25 factors may affect the accuracy and interpretability of any particular SBP measurement (and up to 27 for DBP). The magnitude of these potential effects ranged from small to large in both the positive and negative directions, and some individual sources of inaccuracy had potentially bidirectional effects. Therefore, the net effect of multiple sources may be negative, positive or (in some cases) neutral. However, in practice, it will often be impossible to know how many sources of inaccuracy influenced a particular BP measurement recorded on an observation chart. The results of this systematic review therefore call into question the use of patient monitoring protocols in which a single criterion breach in relation to SBP alone triggers a clinical response. Such trigger protocols are more vulnerable to overestimation or underestimation of SBP response, whereas other systems require multiple vital signs to reach predefined thresholds before a response is triggered. More complex monitoring protocols could be improved by more appropriate interpretation of BP values.

One potential suggestion to discourage unjustified reliance on precise BP values is to use observation charts in which observations are simply recorded as dots (or, in the case of BP, arrowheads) in a series of range rows, rather than recording the precise values numerically. Recent experimental results suggest an additional advantage of recording vital sign data (including BP) as dots/arrowheads in a graphical format. Specifically, novice chart-users who were tested using a range of observation chart designs
identified abnormal vital signs faster and more accurately when recorded this way rather than as numbers, even when those numbers were presented in otherwise identical quasi- graphs [370]. That is, in addition to their limited interpret- ability, the presence of numerical vital signs on a chart can actually impair user-performance.

An important caveat to the findings of this systematic review is that hypotensive patients were seldom studied in the reviewed literature, with most participants being normotensive or hypertensive. The extent to which the identified sources of inaccuracy generalize to hypotensive deteriorating patients is not apparent, and replications utilizing varying populations would help to clarify this.

**Goal-directed treatment and blood pressure maintenance**

Recommendations for goal-directed therapy often specify target BP values [371,372]. However, given potential measurement inaccuracies and transient changes in patient-related and procedure-related factors, a difference between two subsequent BP readings (in some cases, even a seemingly substantial difference) may not indicate a meaningful change. This should be considered when attempting to maintain BP within a narrow range of values. If a single reading deviates slightly from the desired range, it is unlikely that immediate action, such as titration of vasoactive drugs, would be required; and it may even be counterproductive.

It is important to recognize that many of the identified sources of inaccuracy may not apply to invasive measurement, which may be employed in acute care settings. This is particularly true of sources relating to the device and the procedure. However, although the sources of inaccuracy may vary between contexts, the general principle that BP measurement can be inaccurate, and is seldom likely to be precise, applies to all known methods of measurement.

**Hypertension diagnosis and treatment**

Errors such as terminal digit bias can have significant impacts on clinical outcomes. For example, in one study population, simply changing the definition of hypertension from SBP of ‘140 mmHg or greater’ to ‘greater than 140 mmHg’ would have decreased the prevalence of diagnosed hypertension from 25.9 to 13.3%, as the cut-off value no longer contained a terminal zero [362]. This potential decrease was the result of minimizing the effect of a single source of inaccuracy, and there is also a need to consider the potential cumulative effects of multiple sources before making clinical diagnoses. Repeated measurement of BP at multiple clinic visits is strongly advised before treatment for hypertension is initiated or adjusted.

**Conclusion and limitations**

It was not the purpose of this review to critique current clinical guidelines for BP measurement, particularly given the international readership of the journal and the variation in guidelines across jurisdictions. Rather, our goal was to supplement existing guideline recommendations with additional evidence-based information (which may also be of use to those charged with reviewing the guidelines or commissioning further empirical research designed to inform them). In considering the evidence, it should be noted that standardized procedures can be used to target many of the identified sources of inaccuracy but lack the ability to negate issues such as the white-coat effect, device model bias, and most observer-related factors. However, the impact of some of these sources of inaccuracy may be decreased by employing multiple observers and through training programmes focused on minimizing observer biases.

Although many of the identified sources of inaccuracy may not individually affect a reading of BP enough to alter a clinical decision, the cumulative effect of multiple patient, procedure, device and/or observer factors could have a substantial impact – as could some of the more influential sources of inaccuracy in isolation. To make sound interpretations of BP measurements, clinicians should recognize the extent of variability that can exist between any two measurements. The values in Table 2 act as a reminder that any single BP measurement includes the potential for substantial inaccuracy and should not be treated as a meaningful clinical indicator on its own.

Furthermore, by considering which sources of inaccuracy may be introduced at each stage in the process of measuring BP, clinicians may be better placed to recognize and mitigate their effects. To this end, Table 3 presents all of the empirically evaluated potential sources of inaccuracy rearranged into six sequential categories and indicates the direction of all reported significant effects on SBP and DBP. On average, potential sources of inaccuracy to which the plus sign has been assigned tend to increase measurements of SBP or DBP (as the case may be), and sources to which the minus sign has been assigned tend to result in a decrease. These monodirectional sources of inaccuracy include most of the more transitory patient-related factors (i.e. acute meal ingestion, acute caffeine use, acute nicotine use or exposure, bladder distension and cold exposure), the influence of which can only be eliminated by repeating the measurement once the relevant factor is no longer current. For the other sources with monodirectional effects on SBP and/or DBP (such as legs crossed at knees and unsupported arm), repeated measurements cannot eliminate their influence, irrespective of the interval between measurements (with the obvious exceptions of insufficient rest period and reliability on a single measurement). Hence, it is important that procedures are in place to address these factors.

Plus-or-minus signs have been used to indicate bidirectional results. In the case of device model bias, this is because the direction of the effect depends on the specific device type and model being used. However, for any given device, it is also important to consider variation from measurement-to-measurement (as indicated by the width of the 95% limits of agreement), the influence of which can be reduced by taking multiple measurements. For the other potential sources of inaccuracy, there are three general possibilities. First, some are inherently bidirectional (e.g. device calibration error can add a constant error in either direction, and terminal digit preference for zero can result in rounding up or rounding down). Second, for some potential sources of inaccuracy, the direction of the effect may be moderated by another factor (e.g. the directional effect of interarm variability depends on which arm has been
selected for measurement, the effect of acute alcohol use may depend on the time elapsed since ingestion, and the white-coat effect may vary according to factors such as age and race). Third, for the remaining factors with bidirectional effects in relation to SBP and/or DBP (i.e. standing body position, supine body position, use of stethoscope bell and reliance on a single measurement), there is limited and somewhat conflicting evidence, as outlined in the ‘RESULTS’ section.

Although this review fulfilled the goal of identifying and listing empirically evaluated sources of inaccuracy in the measurement of BP, the breadth of the topic limited the level of detail with which each source of inaccuracy could be investigated. Future reviews should therefore focus on specific individual sources of inaccuracy and their causes. Further, sources of inaccuracy that have been investigated by few high quality empirical studies (such as bladder distension, unsupported back and fast cuff deflation rate) require stronger empirical evidence and should be prioritized in future research on BP measurement. In addition, the focus on studies relevant to upper arm BP measurement in clinical settings meant that we did not evaluate studies that were specific to home and ambulatory BP measurement, or measurement at locations other than the upper arm. Nevertheless, many of the reported sources of inaccuracy would also be relevant to these contexts.

**Recommendations**

As demonstrated by these studies, the accuracy of BP measurement can be affected by many factors. A single BP value outside the expected range should therefore be interpreted with caution and should not be seen as a definitive indicator of clinical deterioration or lead to an alteration in clinical care without the support of other evidence. Where a BP measurement indicates an abnormally high or low level, further measurements should therefore be taken and the average used (instead of the original reading) before deciding on appropriate clinical interventions.

Wherever possible, BP values should be recorded graphically within ranges, rather than numerically; or else an agreed-upon standardized rounding method should be adopted. This may reduce the impact of some sources of inaccuracy. For example, given the substantial variability in

---

**TABLE 3. Empirically-evaluated potential sources of inaccuracy that may be introduced at each stage in the process of measuring an adult’s resting blood pressure, and the reported direction of significant effects (note that numbers in parentheses correspond to the numbering of potential sources of inaccuracy in the ‘RESULTS’ section, Table 2 and the Supplementary Tables, http://links.lww.com/HJH/A702)**

<table>
<thead>
<tr>
<th>Stage in the process of blood pressure measurement</th>
<th>Potential sources of inaccuracy introduced</th>
<th>Direction of significant effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before measurement</td>
<td>Acute meal ingestion (1)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Acute alcohol use (2)</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>Acute caffeine use (3)</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>Acute nicotine use or exposure (4)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Bladder distension (5)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Cold exposure (6)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Insufficient rest period (11)</td>
<td>+</td>
</tr>
<tr>
<td>Selecting the device</td>
<td>Device model bias* (9)</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>Device calibration error (10)</td>
<td>±</td>
</tr>
<tr>
<td>Positioning the patient</td>
<td>Standing body position (vs. sitting) (12)</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>Supine body position (vs. sitting) (12)</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>Legs crossed at knees (13)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Unsupported back (14)</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Unsupported arm (15)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Arm lower than heart level (16)</td>
<td>+</td>
</tr>
<tr>
<td>Attaching the device</td>
<td>Paretic arm (7)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Smaller cuff (vs. correct size) (17)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Larger cuff (vs. correct size) (17)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Cuff placed over clothing (18)</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Stethoscope under cuff (19)</td>
<td>+</td>
</tr>
<tr>
<td>Taking the measurement</td>
<td>White-coat effect (8)</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>Taking during measurement (20)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Use of stethoscope bell (vs. diaphragm) (21)</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>Excessive pressure on stethoscope head (22)</td>
<td>ns</td>
</tr>
<tr>
<td></td>
<td>Fast cuff deflation rate (23)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Observer hearing deficit (27)</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Korotkoff phase IV (vs. V) for DBP (28)</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>Short interval between measurements (24)</td>
<td>ns</td>
</tr>
<tr>
<td>Interpreting the measurement</td>
<td>Reliance on a single measurement (25)</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Interarm variability* (26)</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>Terminal digit preference for zero (29)</td>
<td>±</td>
</tr>
</tbody>
</table>

Symbols indicate whether significant increases (+), significant decreases (−) or both (+) were reported. Sources for which all reported directional effects were nonsignificant are marked ns.

*Direction depends on the device type and model (Table 2 and Supplementary Tables 9A–9E, http://links.lww.com/HJH/A702).

*Direction depends on which arm was selected for measurement (individuals differ in which arm will yield the higher values).
the reported frequency of terminal digit preference, it could eliminate some of the variation between individual measurements recorded by different observers. Significantly, it would also reduce the scope for misinterpretation of small, likely erroneous or misleading, changes in BP values. In line with the recommended change to graphing or rounding, protocols using BP to guide clinical interventions need to be reviewed.

**ACKNOWLEDGEMENTS**

The University of Queensland received research consultancy fees from the Clinical Skills Development Service, Metro North Hospital and Health Service for some of N.K. and M.S.H.’s work.

**Conflicts of interest**

M.O.W. has two relevant patents issued (method and means of physiological monitoring and method and means of blood pressure monitoring), but none of the articles cited in this review relate directly to these patents. For the remaining authors, none were declared.

**REFERENCES**


Acute effects of passive smoking on peripheral blood pressure allow for a better assessment of the white-coat effect in treated hypertensive patients. J Hypertens 2013; 11:1799–1805.
Kallioinen et al.


Blood pressure measurement inaccuracy


Sphygmomanometers in use in general practice: an over-


Coozantis DA, Jones CJ. The extent of inaccurate aneroid sphygmomano-


