Validity and inter-observers reliability of blood pressure measurements using mercury sphygmomanometer in the PERSIAN Guilan cohort study

Farahnaz Joukar[a], Mohammad Reza Naghipour[a,b], Sara Yeganeh[b], Masood Sepehrimanesh[a,c], Abbasali Keshtkar[d], Mohammad Taghi Ashoobi[a], Soheil Hassanipour[e] and Fariborz Mansour-Ghanaei[a,e]

Downloaded from http://journals.lww.com/bpmonitoring by RzbJzYh7zLiwYJwFUlF+gJNtzYsbJlEdcgWVmVplfBF4dqDXJqXb48L6wS3W4KA6DYXDX3DP40keWofFPVszUBok40H51D8sYSl/3kLz+l49EBdJBJuO+do86x/1WS7uljZJ3Ag= on 03/06/2020

Objective: **Accurate measuring of blood pressure is a vital step in both clinical and para-clinical settings. The aims of the present study were to evaluate the validity and inter-observer reliability of measured blood pressures by two trained observers and one expert supervisor in the PERSIAN Guilan cohort study (PGCS).**

Participants and methods: **In a quasi-experimental study, two trained observers and one expert supervisor measured systolic and diastolic blood pressures (SBP and DBP) in 85 included participants. All measurements were done using Riester mercury sphygmomanometer as duplicate for each people.**

Results: **Lack of validity in the total SBP ($P = 0.017$), DBP in age <50 years ($P = 0.039$), and DBP in BMI >25 ($P=0.019$) of first observer and total SBP ($P = 0.045$), SBP of male ($P = 0.019$), both SBP and DBP in age >50 years ($P = 0.034$, $P = 0.012$) and DBP in BMI <25 ($P = 0.001$) of second observer were seen. In addition, total inter-rater reliability was found as 12.2% and 27.2% in SBP and DBP, respectively. Age, sex and BMI categorized inter-observer reliability were not more than 15% in SBP of BMI <25 kg/m^2 and 31.6% in DBP of female. The final inter-observer agreement after educational course was higher than 0.7 totally and in all categorical evaluations.**

Conclusion: **Based on lack of validity in some conditions and low level of reliability, education of all observers to measure both SBP and DBP accurately is needed. This is more necessary to done before performing the high population surveys. *Blood Press Monit* 25: 100–104 Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved.**

Blood Pressure Monitoring 2020, **25:**100–104

Keywords: diastolic pressure, systolic pressure, validity and reliability

[a]Gastrointestinal and Liver Diseases Research Center, Guilan University of Medical Sciences, [b]Caspian Digestive Disease Research Center, Guilan University of Medical Sciences, Rasht, Iran [c]NIH Clinical Center, Certified Protein and Proteomics specialist, Postdoctoral Fellowship at University of Louisiana at Lafayette, Louisiana, USA, [d]Department of Health Science Education Development, School of Public Health, Tehran University of Medical Sciences, Tehran and [e]GI Cancer Screening and Prevention Research Center, Guilan University of Medical Sciences, Rasht, Iran

Correspondence to Fariborz Mansour-Ghanaei, MD, Gastrointestinal and Liver Diseases Research Center, Guilan University of Medical Sciences, 41448-95655 Rasht, Iran
Tel: +98 1315535116; fax: +98 1315534951;
e-mail:fmansourghanaei@gmail.com

Received 7 July 2019 **Accepted** 20 October 2019

Introduction

Hypertension, also called persistent elevated arterial blood pressure, is an important risk factor for cardiovascular diseases, chronic kidney disease, stroke, and vision loss [1] and is one of the most important criteria of metabolic syndrome [2]. Therefore, accurate measurement of blood pressure is vital for prevention of such diseases [3,4]. Nowadays, with development of medical researchers, several events are estimated using questionnaire, educational tests or assessment of evaluators [5]. Therefore, the validity and reliability of such tools must be checked and should be in the acceptable level [6–9]. Although blood pressure measurement using mercury sphygmomanometer is a gold standard for assessment of blood pressure, this technique has certain limitations such as manmade or devise errors, patient anxiety, any movements during assessment, and korotkoff sounds which can lead to inaccurate measurements [10].

Since manmade error is one of the most important causes of wrong detection of blood pressure, therefore, the aims of the present study were to evaluate the validity of reported SBP and DBP by two expert observers and compare their agreement as inter-observer reliability assessed by mercury sphygmomanometer in the PERSIAN Guilan cohort study (PGCS).

Participants and methods

Participants

In a quasi-experimental study nested in PGCS, 85 Iranian participants consisted of 55 male and 30 female with age ranged from 35 to 70 years were enrolled. PGCS is a part of the Prospective Epidemiological Research Studies in Iran study [11–13] which conducted on the

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.bpmonitoring.com.

1359-5237 Copyright © 2019 Wolters Kluwer Health, Inc. All rights reserved. DOI: 10.1097/MBP.0000000000000420

Copyright © 2020 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

adult population of Sowme'eh Sara County, near the Rasht, capital of Guilan province, northern of Iran and surveys will perform every year for fifteen years. In the first phase of enrollment, 10 520 participants are included in the PGCS (for more information please see: http://www.gums.ac.ir/cohort).

For this nested study, the sample size was calculated based on the standard of the American Association for the Advancement of Medical Instrumentation (AAMI) [14]. The aims and protocol of the study explained for all participants and they were assured that this study did not induce any damage or danger. Local ethical committee of Guilan University of Medical Sciences approved the study protocol and all procedures were conducted based on Declaration of Helsinki revised in 2000.

Procedure and measurements

Since the stress is one of the factors, which alters blood pressure and detecting ability of observers, the participants, observers, and supervisor were asked to maintain their calm during the measurements. Heavy exercises, foods and coffee intake, and drugs consumption were forbidden before assessment. The bladder of the participants must be empty and they should not be fasted more than 14 hours. The SBP and DBP were measured with using mercury sphygmomanometer (Riester, Germany) from the right arm of each participant after 5 minutes sitting in the rest condition by the first trained observer. Others measurement with similar conditions from the same arm were done again by second trained observer and then by supervisor with 5 minutes rest-time interval. In the participants who had irregular heartbeat, the blood pressure was measured at least five times and the mean value was reported. All measuring was done duplicate and rounding the measured values was not allowed. In addition, talking of participant with observer or each of them separately was not permitted. The height and weight of the participants were also measured using the strip meter and digital scale, respectively, and BMI was calculated as weight (kg)/height (m)^2.

Statistical analysis

Data were expressed as mean and SD for quantitative variables and frequency and percentage for qualitative ones. Descriptive analysis of SBP, DBP, BMI, age, and sex were done using SPSS version 23. Agreement between each observer and supervisor (validity) plus inter-observer agreement (reliability) in total SBP and DBP and age, sex, and BMI-categorized SBP and DBP were assessed by Bland–Altman method and Cohen's kappa, respectively, using MedCalc version 15.8. In Bland–Altman method, we plotted the differences against supervisor values as a reference [15]. $P < 0.05$ was considered as significant difference. Figures were created using GraphPad Prism 7.0.

Results

Our participants were 55 male and 30 female with mean age of 53.2 ± 9.6 years (range: 37–70 years) and mean BMI of 27.7 ± 5.9 kg/m^2 (range: 19.7–63.6 kg/m^2). Descriptive analysis of SBP and DBP reported by two observers and one supervisor are presented in Table 1. As seen, very low levels of agreement were existed between two observers and between each observer and supervisor.

Validity of reported values about total, SBP and DBP is shown in Fig. 1 and for sex, age, and BMI categorized presented in supplementary file (Supplemental digital content 1, http://links.lww.com/BPMJ/A114). Moreover, related statistical analysis is presented in Table 2. As demonstrated, significant differences (disagreement) were seen between both observers 1 and 2 with supervisor in SBP. Indeed, first observer reported significant lower SBP (mean difference = –3.35 ± 12.64, P = 0.016) while second observer detected higher SBP (mean difference = 2.29 ± 10.39, 95% CI: 0.05–4.54, P = 0.045) in comparison to supervisor. When compared as categorized, first observer reported significantly lower DBP in age <50 years and SBP in BMI >25 kg/m^2 in comparison to supervisor. Although, second observer detected significantly higher SBP in male, SBP and DBP in age >50 years. Also second observer detected higher DBP in BMI <25 kg/m^2 (Table 2).

Inter-observer agreement (reliability) in total plus sex, age, and BMI categorized SBP and DBP is presented in Table 3. As shown, the reliability level in none of the parameters was not more than 31.6%, which related to the female categorized DBP. Moreover, the lowest reliability was related to the female categorized SBP, which was 7.9%.

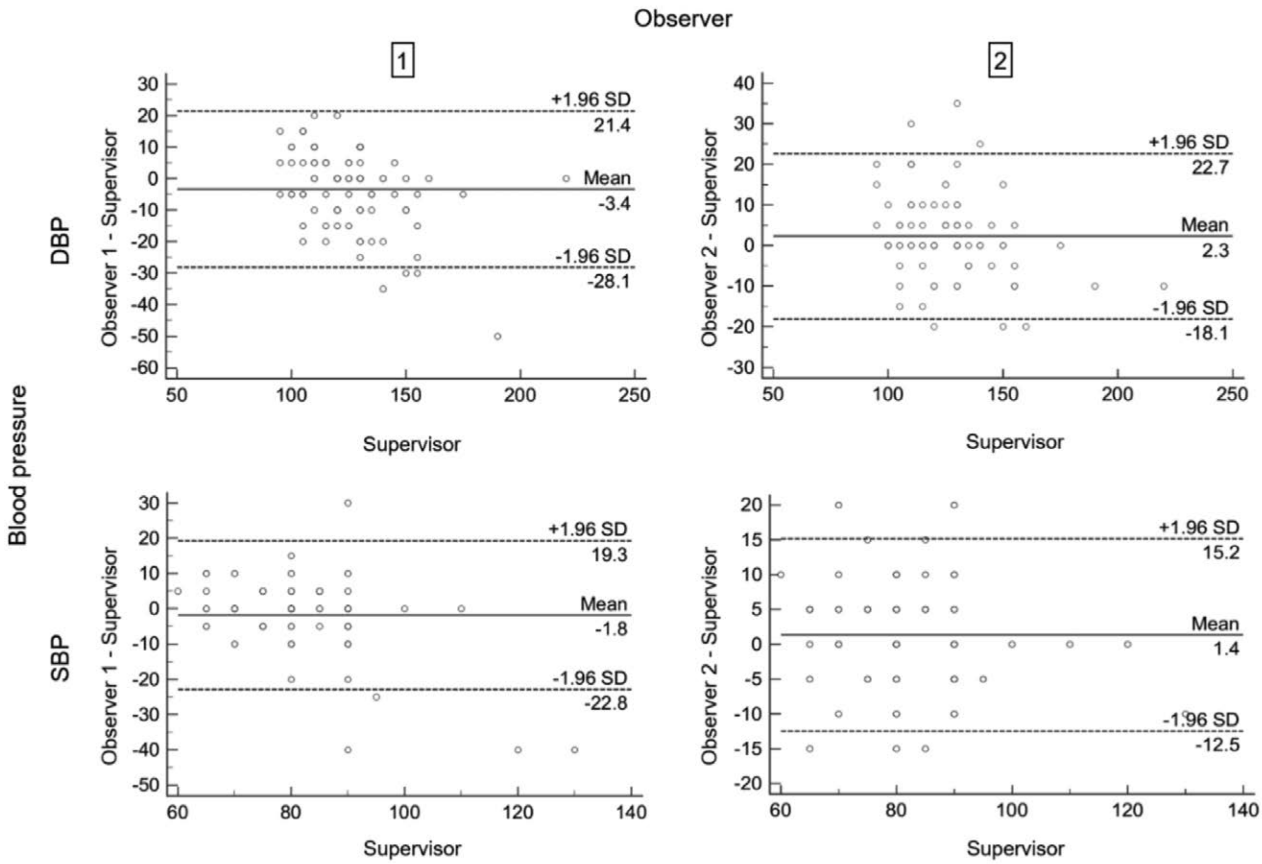
Immediately after seeing the low inter-observer reliabilities, the one-month educational course was applied for

Table 1 Descriptive statistics of SBP and DBP reported by two observers and one supervisor

Parameters	Observer 1	Observer 2	Supervisor
SBP			
Mean ± SD	122.65 ± 19.62	128.29 ± 20.65	126.00 ± 20.98
Median	120	130	125
Mode	120	130	130
Minimum	85	90	95
Maximum	220	210	220
Percentile			
10	100	100	103
25	110	112.5	110
75	132	140	135
90	140	150	152
DBP			
Mean ± SD	79.82 ± 11.45	82.94 ± 12.25	81.59 ± 11.42
Median	80	80	80
Mode	80	90	80
Minimum	50	50	60
Maximum	120	120	130
Percentile			
10	65	70	65
25	70	75	75
75	85	90	90
90	90	97	90

Copyright © 2020 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Fig. 1



Bland–Altman graph of agreement between two observers with supervisor in total SBP and DBP.

them to obtain higher reliabilities. The final inter-observer agreement after this educational course was higher than 0.7 totally and in all categorical evaluations.

Discussion

In the present study, the validity and reliability of reported SBP and DBP by two expert observers were evaluated and compared. We found that even expert observers had some mistakes included higher or lower reporting of blood pressure which affected by age, sex, and BMI. Furthermore, our detected inter-observer reliabilities were too low, which clarified the importance of training course before any assessments.

Despite of developments in tools of blood pressure measurement from mercury to aneroid sphygmomanometers and then digital oscillometric devices [16], this procedure is time-consuming, requires complex and approximately expensive training courses and accompanies by certain errors when using mercury sphygmomanometers [17]. However, mercury sphygmomanometer is widely used and available worldwide in hospitals, clinics, and homes. On the other hand, measurement of blood pressure with wrist devices showed several variations and leads to falsely elevated blood pressure [18]. Howbeit, exact measurement of blood pressure is one of the routine aspect of nursing practice [19], but manmade errors are unavoidable. Based on our best knowledge, there are no reports about evaluation of validity and inter-reliability of observers when measured blood pressure using mercury sphygmomanometers. However, several reports were existed about comparing this device with digital ones. For instance, Vera-Cala *et al.* [17] compared the accuracy of an automatic device with a mercury sphygmomanometer in 1084 participants and found that digital tool had similar accuracy and precision to the mercury sphygmomanometer. On the other hand, the variations between reported value by digital devices and mercury sphygmomanometers were also reported [20,21].

We found some variations in reported values by two observers in comparison to supervisor and together. These variations may be related to the human errors and measuring processes. In an observational study performed by Odagiri *et al.* [22], it has been reported that the physician and nurse had 78.8% agreement in the reading blood pressure values. The level of 69% inter-observer agreement was also reported by Ripollés Ortí *et al.* [23] which is

Copyright © 2020 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

Table 2 Statistical analysis of agreement between two observers with supervisor (validity) plus inter-observer agreement (reliability) in total and age, sex, and BMI categorized SBP and DBP

Parameters	Mean difference ± SD	95% CI	*P* value
SBP			
Observer 1 vs. supervisor	−3.35 ± 12.64	−6.08 to −0.63	0.016
Observer 2 vs. supervisor	2.29 ± 10.39	0.05 to 4.54	0.045
DBP			
Observer 1 vs. supervisor	−1.76 ± 10.74	−4.08 to 0.55	0.133
Observer 2 vs. supervisor	1.35 ± 7.04	−0.17 to 2.87	0.080
Sex categorized SBP			
Female			
Observer 1 vs. supervisor	−4.33 ± 12.09	−8.85 to 0.18	0.059
Observer 2 vs. supervisor	0.00 ± 9.28	−3.47 to 3.47	1.000
Male			
Observer 1 vs. supervisor	−2.82 ± 13.01	−6.33 to 0.70	0.114
Observer 2 vs. supervisor	3.54 ± 10.83	0.62 to 6.47	0.019
Sex categorized DBP			
Female			
Observer 1 vs. supervisor	−1.83 ± 9.60	−5.42 to 1.75	0.304
Observer 2 vs. supervisor	0.50 ± 7.11	−2.16 to 3.16	0.703
Male			
Observer 1 vs. supervisor	−1.73 ± 11.39	−4.81 to 1.35	0.266
Observer 2 vs. supervisor	1.82 ± 7.03	−0.08 to 3.72	0.060
Age categorized SBP			
<50 years			
Observer 1 vs. supervisor	−2.71 ± 10.24	−6.23 to 0.80	0.126
Observer 2 vs. supervisor	0.86 ± 9.96	−2.56 to 4.28	0.614
>50 years			
Observer 1 vs. supervisor	−3.80 ± 14.16	−7.82 to 0.22	0.064
Observer 2 vs. supervisor	3.30 ± 10.67	0.27 to 6.33	0.034
Age categorized DBP			
<50 years			
Observer 1 vs. supervisor	−3.57 ± 9.82	−6.94 to −0.20	0.039
Observer 2 vs. supervisor	0.00 ± 7.95	−2.73 to 2.73	1.000
>50 years			
Observer 1 vs. supervisor	−0.50 ± 11.26	−3.70 to 2.70	0.755
Observer 2 vs. supervisor	2.30 ± 6.24	0.52 to 4.07	0.012
BMI categorized SBP			
<25 kg/m^2			
Observer 1 vs. supervisor	−1.59 ± 0.53	−6.73 to 3.55	0.526
Observer 2 vs. supervisor	2.04 ± 6.66	−0.91 to 5.00	0.165
>25 kg/m^2			
Observer 1 vs. supervisor	−3.97 ± 13.02	−7.25 to −0.69	0.018
Observer 2 vs. supervisor	2.38 ± 11.46	−0.50 to 5.27	0.104
BMI categorized DBP			
<25 kg/m^2			
Observer 1 vs. supervisor	0.68 ± 6.60	−2.24 to 3.61	0.633
Observer 2 vs. supervisor	4.32 ± 5.19	2.02 to 6.62	0.001
>25 kg/m^2			
Observer 1 vs. supervisor	−2.62 ± 11.77	−5.58 to 0.35	0.082
Observer 2 vs. supervisor	0.32 ± 7.34	−1.53 to 2.17	0.733

CI, confidence interval.

lower than acceptable level of 75%. In opposite to above mentioned two studies, Montes Redondo *et al.* [24] by evaluation of 318 participants found well inter-observer agreement as more than 80% with no significant differences between observers. Although their level of agreement is higher than our highest agreement level (31.6%), our findings confirmed the importance of human-related effects. Furthermore, some age, sex, and BMI related significant differences were detected in reported values. Because these variations had no specific patterns, therefore their underlying cause is human-related errors.

Strengths and limitations

One of the strengths and novel part of the present study was to evaluate the validity of blood pressure measurements in a cohort study with a high sample size.

Table 3 Inter-observer agreement (reliability) in SBP and DBP

Parameters	Kappa	95% CI
SBP		
Total	0.122 ± 0.047	0.030 to 0.213
Sex categorized		
Female	0.079 ± 0.072	−0.062 to 0.219
Male	0.141 ± 0.061	0.022 to 0.260
Age categorized		
<50 years	0.141 ± 0.076	−0.008 to 0.290
>50 years	0.093 ± 0.058	−0.021 to 0.206
BMI categorized		
<25 kg/m^2	0.150 ± 0.093	−0.033 to 0.333
>25 kg/m^2	0.104 ± 0.054	−0.002 to 0.210
DBP		
Total	0.272 ± 0.059	0.157 to 0.388
Sex categorized		
Female	0.316 ± 0.097	0.126 to 0.507
Male	0.247 ± 0.071	0.107 to 0.387
Age categorized		
<50 years	0.282 ± 0.092	0.101 to 0.462
>50 years	0.263 ± 0.075	0.115 to 0.411
BMI categorized		
<25 kg/m^2	0.243 ± 0.118	0.011 to 0.475
>25 kg/m^2	0.273 ± 0.069	0.137 to 0.409

CI, confidence interval.

One limitation of our study is that sequential, not simultaneous, measurements were taken. When sequential readings are taken, bias may occur due to an ordering effect. For example, the initial reading may be spuriously high because of an alerting reaction. In addition, initially high readings my regress downwards and initially low readings, upwards. We note that, although we cannot rule out the possibility that these factors affected our results, the two observers differed in terms of the direction of bias relative to the supervisor. Therefore, an ordering effect cannot explain fully the findings of our study.

In conclusion, low level of validity and inter-observer reliability (between observers and the supervisor) in the measured SBP and DBP using mercury sphygmomanometer were detected in this study. The possible main cause of these disagreements is manmade errors, which can be resolved using training course for all observers as we performed for all of our observers who cooperated in PGCS immediately after seeing the low inter-observer reliabilities.

Acknowledgements

We wish to thanks all staffs of Gastrointestinal and Liver Diseases Research Center, Guilan University of Medical Sciences for their kindly help to perform Guilan Cohort study and this nested study.

A substantial part of this study was supported by the Research Council of Guilan University of Medical Sciences.

M.R.N. contributed to the study design and analysis. F.M.-G. contributed to the study design, analysis and interpretation of data, and critical revision of the manuscript for portent intellectual content. F.J. and S.H. contributed to the study design and analysis of data. S.Y.

Copyright © 2020 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

contributed to gather data and drafting of the manuscript. M.S. contributed to analysis and writing of numerous sections of the article. S.H., M.T.A. and F.M.-G. contributed to the study design and critical revision of the manuscript for important intellectual content. All authors read and approved the last version of the article.

Conflicts of interest

There are no conflicts of interest.

References

1 Lackland DT, Weber MA. Global burden of cardiovascular disease and stroke: hypertension at the core. *Can J Cardiol* 2015; **31**:569–571.

2 Fattahi MR, Niknam R, Safarpour A, Sepehrimanesh M, Lotfi M. The prevalence of metabolic syndrome in non-alcoholic fatty liver disease; a population-based study. *Middle East J Dig Dis* 2016; **8**:131–137.

3 Abbud L, Nzelu D, Salaria M, Kay P, Kametas NA. Validation of the Omron HBP-1300 in pregnancy for medium-arm and large-arm circumferences according to the British Hypertension Society Protocol. *Blood Press Monit* 2018; **23**:277–280.

4 Padwal R, Jalali A, McLean D, Anwar S, Smith K, Raggi P, Ringrose JS. Accuracy of oscillometric blood pressure algorithms in healthy adults and in adults with cardiovascular risk factors. *Blood Press Monit* 2019; **24**:33–37.

5 Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006; **119**:166. e7–166.16.

6 Golafshani N. Understanding reliability and validity in qualitative research. *The Qualitative Report* 2003; **8**:597–606.

7 Myers MG, Godwin M, Dawes M, Kiss A, Tobe SW, Grant FC, Kaczorowski J. Conventional versus automated measurement of blood pressure in primary care patients with systolic hypertension: randomised parallel design controlled trial. *BMJ* 2011; **342**:d286.

8 Drost EA. Validity and reliability in social science research. *Education Research and Perspectives* 2011; **38**:105–123.

9 Kuwabara M, Harada K, Hishiki Y, Kario K. Validation of an automatic device for the self-measurement of blood pressure in sitting and supine positions according to the ANSI/AAMI/ISO81060-2: 2013 guidelines: the Omron HEM-9700T. *Blood Press Monit* 2019; **24**:146–150.

10 Topouchian J, Agnoletti D, Blacher J, Youssef A, Ibanez I, Khabouth J, *et al*. Validation of four automatic devices for self-measurement of blood pressure according to the international protocol of the European Society of Hypertension. *Vasc Health Risk Manag* 2011; **7**:709–717.

11 Eghtesad S, Mohammadi Z, Shayanrad A, Faramarzi E, Joukar F, Hamzeh B, *et al*. The PERSIAN cohort: providing the evidence needed for healthcare reform. *Arch Iran Med* 2017; **20**:691–695.

12 Poustchi H, Eghtesad S, Kamangar F, Etemadi A, Keshtkar AA, Hekmatdoost A, *et al*. Prospective epidemiological research studies in Iran (the PERSIAN cohort study): rationale, objectives, and design. *Am J Epidemiol* 2018; **187**:647–655.

13 Mansour-Ghanaei F, Joukar F, Naghipour MR, Sepanlou SG, Poustchi H, Mojtahedi K, *et al*. The PERSIAN Guilan cohort study (PGCS). *Arch Iran Med* 2019; **22**:39–45.

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

15 Krouwer JS. Why Bland-Altman plots should use X, not (Y+X)/2 when X is a reference method. *Stat Med* 2008; **27**:778–780.

16 Turner MJ. Can we trust automatic sphygmomanometer validations? *J Hypertens* 2010; **28**:2353–2356.

17 Vera-Cala LM, Orostegui M, Valencia-Angel LI, López N, Bautista LE. Accuracy of the Omron HEM-705 CP for blood pressure measurement in large epidemiologic studies. *Arq Bras Cardiol* 2011; **96**:393–398.

18 Casiglia E, Tikhonoff V, Albertini F, Palatini P. Poor reliability of wrist blood pressure self-measurement at home: a population-based study. *Hypertension* 2016; **68**:896–903.

19 Heinemann M, Sellick K, Rickard C, Reynolds P, McGrail M. Automated versus manual blood pressure measurement: a randomized crossover trial. *Int J Nurs Pract* 2008; **14**:296–302.

20 Coleman A, Steel S, Freeman P, de Greeff A, Shennan A. Validation of the Omron M7 (HEM-780-E) oscillometric blood pressure monitoring device according to the British Hypertension Society protocol. *Blood Press Monit* 2008; **13**:49–54.

21 de Greeff A, Shennan A. Blood pressure measuring devices: ubiquitous, essential but imprecise. *Expert Rev Med Devices* 2008; **5**:573–579.

22 Odagiri T, Morita T, Yamauchi T, Imai K, Tei Y, Inoue S. Convenient measurement of systolic pressure: the reliability and validity of manual radial pulse pressure measurement. *J Palliat Med* 2014; **17**:1226–1230.

23 Ripollés Ortí M, Martín Rioboó E, Díaz Moreno A, Aranguren Baena B, Murcia Simón M, Toledano Medina A, Fonseca Del Pozo FJ. [Agreement in the measurement of blood pressure among different health professionals. Are mercury sphygmomanometers reliable?]. *Aten Primaria* 2001; **27**:234–243.

24 Montes Redondo G, Fernández García JA, Prada Vigil A, Polonio Enríquez R, Rodríguez Hornero D, Pérula de Torres LA. [Reliability of blood pressure measurement: the patient versus primary care workers]. *Aten Primaria* 2000; **25**:73–77.

Copyright © 2020 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.