Minimally sufficient numbers of measurements for validation of 24-hour blood pressure monitoring in chronic kidney disease

Rajiv Agarwal, MD[1], **Wanzhu Tu, PhD**[2]

[1])Division of Nephrology Department of Medicine, Indiana University School of Medicine and Richard L. Roudebush Veterans Administration Medical Center, Indianapolis, IN, USA.

[2])Department of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, USA.

Abstract

Ambulatory blood pressure monitoring (ABPM) remains a reference standard, but the minimal number of ABPM readings required to diagnose hypertension has not been empirically validated. Among 360 patients with chronic kidney disease and 38 healthy controls, 24-hour blood pressure was recorded 2 times per hour during the night and 3 times per hour during the day. All subjects had at least 90% of the expected readings recorded. From this full set of ABPM recordings, we selected variable numbers of measurements and compared the performance of the selected readings against that of the full sample under either random or sequential sampling schemes. With 8 randomly selected systolic blood pressure readings, we were able to make diagnostic decisions in concordance with that from the full ABPM sample 91.0% of the time (kappa 0.804). With 15 randomly selected diastolic blood pressure readings, we made concordant decisions 96.3% of the time (kappa 0.810). A serial selection scheme generally required a greater number of readings to achieve the same levels of concordance with the full ABPM data. With a random selection scheme, 26 readings provided 95% confidence that the sample mean will be within 5 mmHg of the true systolic blood pressure mean, and within 3.5 mmHg of the true diastolic blood pressure mean.

Keywords

Cardiovascular disease; chronic kidney disease

Introduction

Ambulatory blood pressure (BP) monitoring (ABPM) over 24h is considered the reference standard for both hypertension diagnosis and BP control assessment [1]. ABPM is also frequently used to determine patient's BP response to specific antihypertensive therapies.

However, what constitutes an adequate number of BP recordings has never been empirically examined and is largely a matter of opinion. These opinions are embodied in various guidelines and some large studies.

The European Society of Hypertension Guidelines 2013 [2] recommend repeating ABPM for the following 3 reasons: (i) there are not at least 70% of the expected measurements over 24h; (ii) there are fewer than 20 readings during the awake period (0900 to 2100 h); and (iii) there are fewer than 7 recordings during the sleep period (0100 to 0600 h). Furthermore, for research purposes, the ESH guidelines suggest that the recording be repeated if there are fewer than 2 recordings per hour during the day and 1 per hour during the night.

The UK National Institute of Health and Clinical Excellence (NICE) guidelines 2011 [3] state, "When using ABPM to confirm a diagnosis of hypertension, ensure that at least two measurements per hour are taken during the person's usual waking hours (for example, between 08:00 and 22:00). Use the average value of at least 14 measurements taken during the person's usual waking hours to confirm a diagnosis of hypertension."

The American Heart Association guidelines have no explicitly stated position on how many measurements of ABPM are required for the recording to be considered adequate [4].

Other large studies published their own criteria. For example, The Coronary Artery Risk Development in Young Adults (CARDIA) Study [5] "defined nighttime as midnight to 0600 and daytime as 1000 to 2200. For a session to be deemed adequate, we required a minimum of 10 daytime measurements and 5 nighttime measurements during these specific intervals." The International Databases of Ambulatory Blood Pressure in relation to Cardiovascular Outcome (IDACO) required at least 10 daytime and at least 5 nighttime recordings to be considered further for analysis [6].

None of the above recommendations are based on empirical data. The ESH requirement of at least 70% of expected recordings, including 20 or more during the day and 7 or more during the night, are strict. Adherence to these guidelines often requires repeated ABPM, which the patients may not agree to perform. This also gives rise to missing data, which could bias the study results because those participants in a clinical trial who adhere to the ABPM protocol tend to have positive health behaviors; those who do not adhere to ABPM may have suboptimal adherence to their medications or to other positive health behaviors. As a result, the non-ignorable missingness and informative censoring of the study data could lead to biased findings.

The purpose of this research was to answer the question of how many recordings are needed for ABPM to be considered adequate. Adequacy was judged by answering two questions: first, how many recordings are needed to make a diagnosis that is consistent with that based on the full ABPM recording, and second, how many recordings are needed for achieving a reasonably tight 95% confidence interval around the mean of full ABPM data. Furthermore, besides performing these analyses on 24h recordings, we asked the question whether pre-specified thresholds are needed for daytime and nighttime recordings for the ABPM to be considered adequate.

Results

Of the 398 patients evaluated, 360 had hypertension and CKD, whereas 38 were healthy controls. For the entire cohort the average age (SD) was 68.6 ± 9.3 years, 389 (97.7%) were men, 320 (80.4%) were white, 65 (16.3%) black, 232 (58%) had diabetes mellitus, 67 (16.8%) were current tobacco users, 68 (17.1%) had a history of heart failure, 109 (27.4%) history of myocardial infarction, 44 (11.1%) stroke, and 76 (19.1%) peripheral vascular disease. The mean body mass index was 30.3 ± 4.7 kg/m^2 and their estimated GFR was 48 ± 20.2 mL/min/1.73m^2. Of those who had hypertension, the mean number of antihypertensive medications was 3.1 ± 1.4. Loop diuretics were used by 37%, thiazides 25%, ACE inhibitors 54%, ARBs 20%, β-blockers 68%, and dihydropyridine calcium-channel blockers by 43% of the study patients.

Actual ambulatory BP data from all 398 participants were analyzed. Patients had at least 23h of recording with a range of 46 to 73 measurements per patient. The mean ambulatory and seated clinic oscillometric BP are shown in Table 1.

Figure 1 presents the degree of observed agreement (concordance) and the kappa statistic against increasing number of BP measurements for making a diagnosis of hypertension control. Panel 1A shows the observed agreement to make a diagnosis of poorly controlled hypertension with just one randomly selected systolic BP with the gold-standard diagnosis was >75%, but yielded a modest kappa statistic of about 0.518. With 8 measurements, the observed agreement was 91.0% which means that that decisions were 91% concordant and 9% discordant. The kappa statistic was 0.804 (95% CI 0.798 to 0.809); kappa statistic over 0.8 indicates excellent agreement, after controlling for agreement by chance. Panel 1B shows at 29 serial readings (or about 10 h of recording) produced a concordance rate of 91.5% and a kappa value of 0.815 (95% CI 0.717 to 0.913). Panel 1C shows that for an accurate diagnosis of hypertension control by diastolic BP, at least 15 random recordings are needed (kappa 0.810 (95% CI 0.803 to 0.817), observed agreement 96.3%). Panel 1D shows that this number increases to 40 if diastolic BP recordings are serially collected (kappa 0.810 (95% CI 0.713 to 0.908), observed agreement 96.2%). Combining these observations, we conclude that 15 randomly selected recordings of either systolic or diastolic BP over a 24h period of ambulatory BP monitoring will allow physicians to reliably make a diagnosis of hypertension control.

When analysis was restricted to the diagnosis of daytime hypertension control, 6 random systolic recordings (concordance 92.8%, kappa 0.814) and 8 random diastolic BP recordings (concordance 97.5%, kappa 0.816) were needed to achieved desired level of accuracy (Figure 1S, **see supplementary appendix**). For serial recordings these requirements increased to 13 for systolic BP (concordance 92.7%, kappa 0.803) and 12 for diastolic BP (concordance 97.2%, kappa 0.810).

For the diagnosis of nighttime hypertension control, 3 random systolic recordings (concordance 90.5%, kappa 0.809) and 4 random diastolic BP recordings (concordance 92.2%, kappa 0.815) were needed (Figure 2S, **see supplementary appendix**). For serial

recordings these requirements increased to 4 for systolic (concordance 90.7%, kappa 0.814) and 5 for diastolic BP (concordance 93.5%, kappa 0.838).

The results of diagnostic performance of the random and serial selections of ABPM are shown in Figure 3S **(see supplementary appendix)**.

Figure 2 presents the widths of the 95% confidence intervals corresponding to different numbers of BP recordings. As expected, the width of the 95% confidence interval narrowed with increasing size of the random subsamples. The standard error of the mean (SEM) for the full data set which had up to 73 recordings was 1.60 mmHg systolic, which represented the maximal precision that one could reach with the fully observed ABPM data. With 10, 18, and 26 randomly selected observations, the SEMs were 4.0, 3.0, and 2.5 mmHg, respectively. (Table 2A). The corresponding SEMs for 24h diastolic BP were 1.15 for the full data-set, with increasing numbers of BP recordings needed for a greater precision (Table 2). Thus, the sample mean of 26 measurements would provide 95% confidence that it would be within 5 mmHg of the true mean systolic BP and approximately 3.5 mmHg of true diastolic BP mean.

In comparison to random selections where good asymptotic behaviors of the subsample means were guaranteed by the Central Limit Theorem (Figure 2 A and C), the performance of the serial samples was less well-behaved (Figure 2 B and D), despite the fact that larger sample sizes under both sampling plans led to narrower intervals (Table 2).

When data were limited to daytime recordings, SEM of the full daytime data-set was 1.98 mmHg for systolic and 1.40 mmHg for diastolic recordings. The SEM of 2 mmHg required 38 systolic recordings and 17 diastolic recordings (Figure 4S, **supplementary appendix and** Table 2).

For nighttime recordings, the number of recordings were limited, therefore SEM were wider. (Figure 5S, **supplementary appendix and** Table 2).

Discussion

ABPM is emerging from being predominantly a research tool to one that is increasingly adopted in clinical practice. Although still not used widely, the adoption of ABPM is on the rise in response to recent recommendations. In the UK, the National Institute of Health and Clinical Excellence (NICE) guidelines have endorsed its use prior to starting therapy for hypertension [3]. The United States Preventive Task Force (USPTF) also recommends its use for making a diagnosis of hypertension [7]. Since ABPM is the reference standard, it is somewhat surprising that no empirical study has investigated the minimally sufficient number of readings needed to achieve a valid assessment of 24h ABPM. Our study attempts to bridge this gap in the existing knowledge.

A main finding of the study suggested that surprisingly few recordings were needed for the diagnostic purposes. There was an excellent agreement between systolic BP control with the full-set gold standard if 8 randomly selected recordings or 29 serially collected recordings

were available over a 24h period. For diastolic BP control, these numbers increased to 15 random and 40 serial readings.

For daytime systolic hypertension, 6 random or 13 serial recordings were needed. For daytime diastolic hypertension, 8 random or 12 serial recordings were required. For nighttime systolic hypertension 3 random or 4 serial and for diastolic hypertension 4 random or 5 serial recordings were needed. Even with one available recording, whether random or serial, the AUC values of the ROC curves exceeded 0.80.

Assuming that data are missing at random, with 26 BP recordings one can be 95% confident that the subsample mean is within approximately 5 mmHg of the true systolic BP; with 40 recordings the sample mean will be within approximately 4 mmHg. In comparison, with the full data set, that is the best case scenario, we can be 95% confident that the mean systolic BP will be within 3.2 mmHg. With 27 BP recordings 95% confidence can be achieved that the sample mean will be within 3.5 mmHg diastolic and with 37 recordings within 3.0 mmHg diastolic. In comparison, with full data set, we can be 95% confident that the mean diastolic BP will be within 2.3 mmHg.

There are important clinical implications of our finding. The 2013 European Society of Hypertension Guidelines recommend 20 daytime and 7 nighttime recordings to be the minimal number needed to call 24h ABPM adequate [2]. Table 2 shows that with these recommendations, assuming that the data are missing at random, daytime systolic ABPM will be within 5 mmHg and nighttime systolic BP within 8 mmHg. In other words, the current recommendations of ESH are associated with at least 5 mmHg error. Overall, having 27 readings will allow the overall 24h ABPM to be within 5 mmHg and 3.5 mmHg diastolic. Our data therefore provide empirical support for the high certainty that can be achieved with the ESH recommendations. Requiring the presence of some daytime and some nighttime measurements as suggested by ESH will allow some assurance that the data are truly missing at random.

We believe that 15 recordings over a 24h period— exactly the number suggested by the CARDIA [5] and IDACO [6] studies— may be sufficient for a making an accurate diagnosis of systolic or diastolic hypertension. To be sure, this does not mean that ABPM should be abbreviated to just 5 h with three measurements per hour to provide 15 recordings. These would qualify as serial measurements and the number of measurements needed to provide the same certainty as 15 random recordings would be 40, or approximately 14 h of recordings performed three times an hour. The NICE guidelines recommend at least 14 recordings during the day to make a diagnosis of hypertension [3]. Our data confirm that 14 random recordings will provide a high degree of reliability in making a diagnosis of hypertension (kappa for systolic hypertension 0.860, 95% CI 0.856–0.865, kappa for diastolic hypertension 0.799, 95% CI 0.791–0.806). Specifically, for 14 systolic recordings the lower bound of CI of the area under ROC to will be 0.88 and for diastolic recordings 0.925.

These findings may be of particular relevance in the regulatory context. For example, the United States Food and Drug Administration suggests ABPM to assess pressor effects of

drugs [8]. The European Medicines Agency adopts the EMA guidelines and recommends that at least 2 readings per hour during the day and 1 reading per hour during the night should be available to call ABPM sufficient [9]. Considerably fewer readings would be needed given the findings of our study.

Some limitations are acknowledged. The participants were predominantly men. Although there is no *a priori* reason to believe that women would have ABPM test performance that would be significantly different from men, our study should be replicated in a larger group of women. Likewise, all participants with hypertension had CKD. CKD may be associated with increased BP variability and if so, the minimum number of readings recommended in this research would be more conservative, i.e., greater than in the general population. However, we had a group of 38 participants with no CKD or hypertension and the results in this group of patients were not substantially different from the overall cohort. Some strengths of the study included its prospective design, selection of ABPM based on adequate number of recordings and a reasonably large number of participants for a single site study.

In conclusion, in this first empirical study to our knowledge to determine the minimum number of BP recordings needed to validate ABPM, we provide evidence to validate the 2013 European Society of Hypertension guidelines [2] for what is considered an adequate ABPM and NICE guidelines [3] to confirm a diagnosis of hypertension. However, criteria derived from our empirical data are less stringent than the existing guidelines. For making a qualitative decision on making a diagnosis of hypertension or assessing its control a minimum of 13 daytime, 5 nighttime, or 40 24-hour serial measurements is sufficient. This would require 13 hours of measurement, thrice per hour instead of twice per hour from 8 AM to 10 PM proposed by NICE guidelines [3]. If a 24-hour recording is available and data are missing are random, only 15 measurements are needed to make a qualitative judgment regarding hypertension diagnosis or control; for daytime 8 and for nighttime 4 random measurements are sufficient. For quantitative analysis, the minimal sufficient number of 24h ambulatory BP is 26 random recordings. ESH requires 20 daytime recordings and 7 nighttime recordings to consider ABPM adequate. Given the stipulation of a certain number of recordings during the day and night makes us more confident that the data may be missing at random. Therefore, these recommendations made by ESH are for the first time supported by our empirical data [2]. We do not suggest that 24h ABPM be replaced with 8–15 random measurements over a 24h period. We simply provide empirical data when to consider 24h ABPM adequate.

Methods

Patients and BP measurements

Patients with chronic kidney disease (CKD) and hypertension were recruited from the renal clinic at Roudebush Veteran's Administration hospital in Indianapolis. Normotensive controls with no evidence of CKD or cardiovascular disease were recruited from the medicine clinic of the same hospital.

Ambulatory BP monitoring was performed over 24h using the SpaceLabs 90207 monitor, (SpaceLabs, Issaquah, WA) that has been validated [10]. The monitor was programmed to

record BP every 20 minutes from 0800h to 2200h and every 30 minutes from 2200h to 0800h as reported previously [11].

The study was approved by the Indiana University Institutional Review Board and the VA Research and Development Committee and all participants signed a written informed consent.

Diagnosis of hypertension

In this analysis, we included patients who had at least 23 hours of the 24-hour recordings. The diagnosis made based on the full sample were considered the gold standard in this analysis. A diagnostic decision was made for each individual patient based on following criteria: Patients with mean 24h ambulatory systolic BP   130 mmHg or diastolic BP   80 mm Hg were classified as hypertensive (if not on BP medications) or poorly controlled hypertensives (if receiving antihypertensive medications). Besides the 24h recordings, we also analyzed the data separately for daytime (0900 to 2100h) and nighttime (0100 to 0600). These times for definitions of day and night were chosen on the recommendation of the ESH guidelines. Daytime hypertension was diagnosed with BP   135 systolic or   85 mmHg diastolic, and nighttime hypertension with BP   120 systolic or   70 mmHg diastolic. For calculating the average systolic and diastolic BP, we first calculated the mean during a given hour and then averaged the means. Thus, a weighted average was used.

Sampling and analysis

In this research, we considered two different sampling schemes: (1) A sampling plan that randomly selected a predetermined number (m) of readings from each patient; (2) A serial sampling plan that randomly selected a series of sequential BP readings of size m.

Qualitative analysis of diagnosis of hypertension—We analyzed the selected readings and compared the diagnostic decision against that derived from the full data. Specifically, we calculate the mean systolic and diastolic BP. Diagnoses were made based on the calculated mean systolic and diastolic BP from the selected readings. We then compared the diagnosis based on the selected subsample to the gold standard; agreement and disagreement of the diagnostic decisions between the full sample and subsample were recorded. For each given m, we repeated the random experiment 1000 times. Rates of concordant and discordant decisions were calculated. A high proportion of concordant decisions suggested that the size of the subsample was sufficient to reach the same diagnostic decision as the full ABPM. We increased the number of selections from 3 to 73 to determine the optimal number of readings needed to reach diagnostic decisions that were concordant with that from the full dataset. In addition to the raw rates of agreement, we also reported kappa statistics, which accounts for both the observed accuracy and chance agreement [12]. A kappa statistic of >0.8 is typically considered near perfect agreement [13]. We evaluated the performance of serial samples similarly. We performed these analyses for the 24h data set and for daytime and nighttime recordings separately.

To assess diagnostic performance, we further calculated the area under the receiver operating characteristic curve associated with systolic and diastolic BP (see Appendix).

Quantitative analysis of agreement with full data set—To answer the second question posed by the study, we calculated 95% CIs for the mean systolic and diastolic BP. We then determine the average length of the calculated CIs at each given m. While empirical coverage probabilities should always approach the nominal level (95%), shorter interval lengths would indicate increased precision.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

References:

1. Pickering TG, Shimbo D, Haas D: Ambulatory blood-pressure monitoring. N Engl J Med 354:2368–2374, 2006 [PubMed: 16738273]

2. Mancia G, Fagard R, Narkiewicz K et al.: 2013 ESH/ESC Practice Guidelines for the Management of Arterial Hypertension. Blood Press 23:3–16, 2014 [PubMed: 24359485]

3. Anonymous: Hypertension: The Clinical Management of Primary Hypertension in Adults. Update of Clinical Guidelines 18 and 34. NICE Clinical Guidelines, No. 127. London, Royal College of Physicians (UK), 2011

4. Pickering TG, Hall JE, Appel LJ et al.: Recommendations for blood pressure measurement in humans and experimental animals: Part 1: blood pressure measurement in humans: a statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. Hypertension 45:142–161, 2005 [PubMed: 15611362]

5. Viera AJ, Lin FC, Hinderliter AL et al.: Nighttime blood pressure dipping in young adults and coronary artery calcium 10–15 years later: the coronary artery risk development in young adults study. Hypertension 59:1157–1163, 2012 [PubMed: 22547435]

6. Thijs L, Hansen TW, Kikuya M et al.: The International Database of Ambulatory Blood Pressure in relation to Cardiovascular Outcome (IDACO): protocol and research perspectives. Blood Press Monit 12:255–262, 2007 [PubMed: 17760218]

7. Siu AL: Screening for high blood pressure in adults: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 163:778–786, 2015 [PubMed: 26458123]

8. FDA guidance https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM609185.pdf accessed on 7-24-2018.

9. EMA guidance. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2016/07/WC500209943.pdf. accessed on 7-24-2018.

10. O'Brien E, Mee F, Atkins N, O'Malley K: Accuracy of the SpaceLabs 90207 determined by the British Hypertension Society protocol. J Hypertens 9:573–574, 1991 [PubMed: 1653299]

11. Agarwal R, Pappas MK, Sinha AD: Masked Uncontrolled Hypertension in CKD. J Am Soc Nephrol 27:924–932, 2016 [PubMed: 26163421]

12. Cohen J: A coefficient of agreement for nominal scales. Educ Psychol Meas 20:37–46, 1960

13. Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 33:159–174, 1977 [PubMed: 843571]
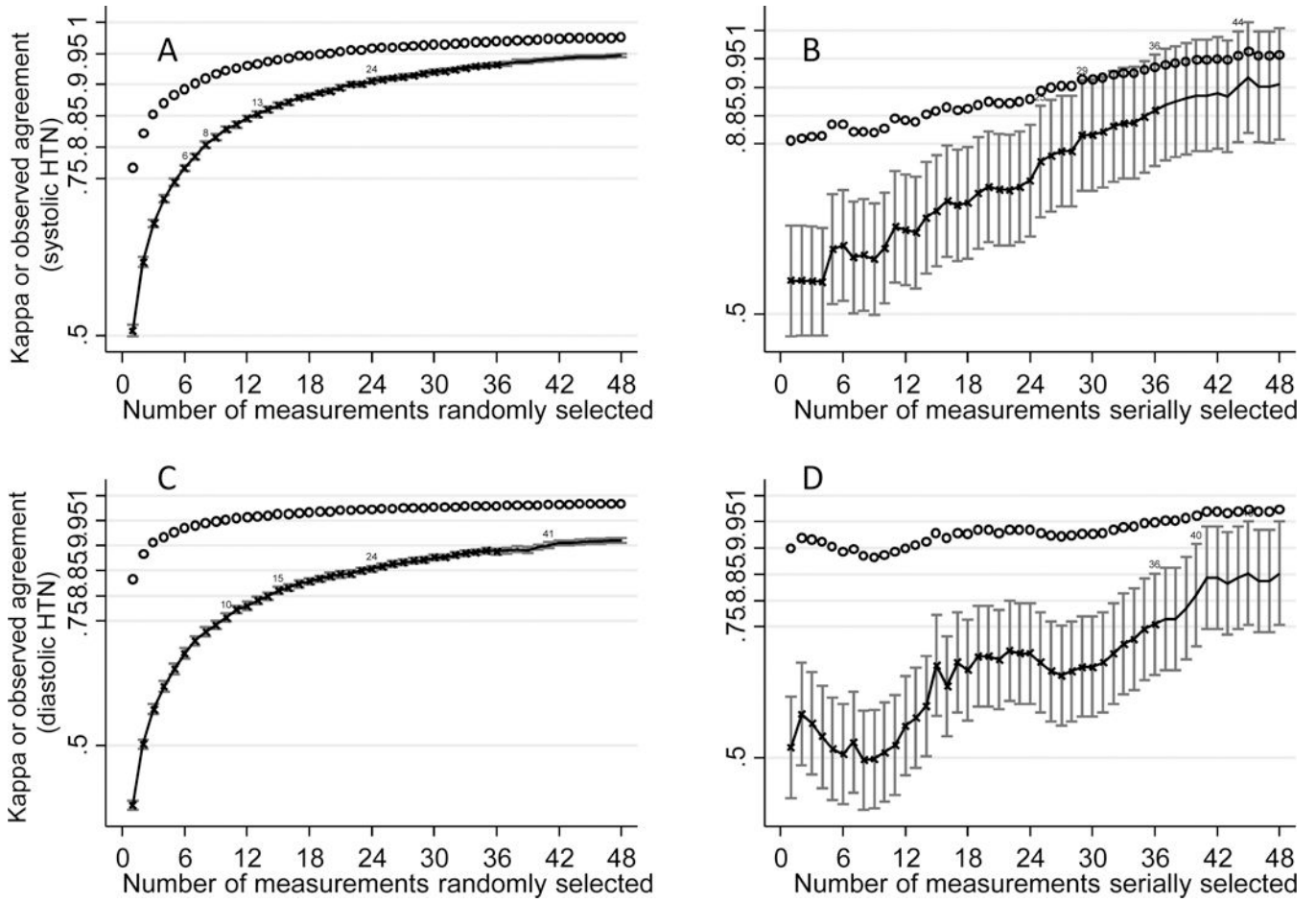
Figure 1:

Qualitative analysis of the diagnosis of hypertension control with increasing number of BP measurements: Each graph plots the observed agreement (concordance) and the kappa statistic (a measure of agreement) with increasing numbers of selected BP recordings. The reference set was the full set of 24h ambulatory BP recordings. Panel A and B show systolic and C and D diastolic recordings. The left panel shows readings selected at random from the full set. The right panel show serial selections of readings. As expected, increasing numbers of readings increased the confidence in making a diagnosis of hypertension control. Fewer recordings were needed when readings were selected at random, than when readings were selected serially. The numbers on the top of the kappa line correspond to the number of measurements for clarity of interpretation of the data. Error bars reflect the 95% confidence intervals of kappa.
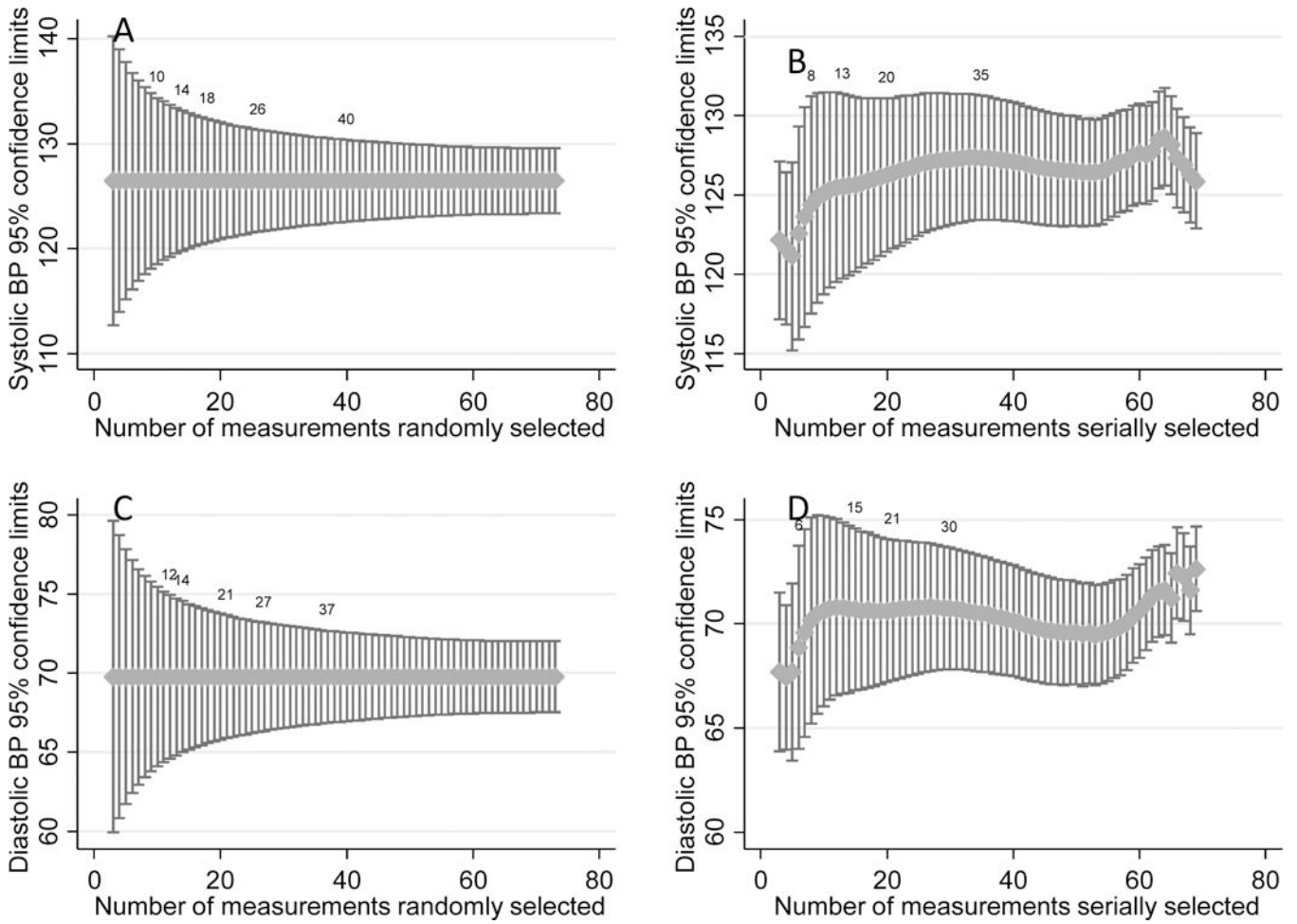
Figure 2:

Quantitative analysis of the relationship between precision of measurement with the number of BP recordings. Each graph plots the number of recordings against the 95th percentile of the estimate. Panel A and C are random measurement selections and Panel B and D serial selections. As in Figure 1, Panel A and B show systolic and C and D diastolic recordings. As expected, increasing numbers of readings increased the precision of the estimate. The numbers on the top of confidence intervals correspond to the number of measurements for clarity of interpretation of the data. See text for details.

Table 1:

Blood pressure measurements

Clinical characteristic	CKD	Healthy	p
n	360	38	
Age	69.2 ± 9.1	63.2 ± 9.3	<0.0001
Men	354 (98%)	35 (92%)	0.014
Ethnicity			0.40
White	287 (80%)	33 (87%)	
Black	60 (17%)	5 (13%)	
Other	13 (4%)	0 (0%)	
24h ABPM systolic (mmHg)	126.8 ± 13.5	121.7 ± 8.7	0.025
24h ABPM diastolic (mmHg)	69.2 ± 8.6	73.1 ± 6.0	0.006
Daytime systolic (mmHg)	127.9 ± 13.7	126.0 ± 8.9	0.40
Daytime diastolic (mmHg)	70.7 ± 8.9	76.4 ± 6.8	<0.001
Nighttime systolic (mmHg)	122.9 ± 16.1	113.7 ± 11.2	<0.001
Nighttime diastolic (mmHg)	65.4 ± 10.1	66.9 ± 7.1	0.37
Clinic BP systolic (mmHg)	119.8 ± 16.6	115.7 ± 11.0	0.13
Clinic BP diastolic (mmHg)	59.8 ± 10.7	64.7 ± 7.9	0.007
Clinic pulse rate (/min)	66.8 ± 11.8	69.7 ± 11.3	0.15

Table 2:

Minimum number of recordings needed to achieve a desired level of precision

A	Precision as measured by standard error of mean (SEM)					
Systolic BP Measure (mmHg)	**± 4**	**± 3.5**	**± 3**	**± 2.5**	**±2**	**Min SEM**
24h random	10	14	18	26	40	± 1.60
24h serial		8	13	20	35	± 1.54
Daytime random	9	12	16	22	38	± 1.98
Daytime serial		9	13	20	33	± 1.85
Nighttime random	7	9				± 3.1
Nighttime serial	5	7	10	11	11	± 1.53

B	Precision as measured by standard error of mean (SEM)					
Diastolic BP Measure (mmHg)	**2.75**	**2.5**	**2**	**1.75**	**1.5**	**Min SEM**
24h random	12	14	21	27	37	± 1.15
24h serial		6	15	21	30	± 1.07
Daytime random	10	11	17	23	31	± 1.40
Daytime serial		9	15	21	29	± 1.31
Nighttime random	9	11				± 2.5
Nighttime serial	7	9	11	11	11	± 1.19

Min SEM is the minimum standard error of the mean or the maximal precision observed for the full data set