PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Trends in end digit preference for blood pressure and associations with cardiovascular outcomes in Canadian and UK Primary Care: a retrospective observational study
AUTHORS	Greiver, Michelle; Kalia, Sumeet; Voruganti, R.; Aliarzadeh, Babak; Moineddin, Rahim; Hinton, William; Dawes, Martin; Sullivan, Frank; Syed, Saddaf; Williams, John; de Lusignan, Simon

VERSION 1 – REVIEW

REVIEWER	Romin Pajouheshnia University Medical Center Utrecht, The Netherlands
REVIEW RETURNED	09-Aug-2018

GENERAL COMMENTS	This study assessed the prevalence of end digit preference (EDP) when recording blood pressure (BP) in primary care, the effect of this on the average recorded blood pressure values and the (causal) association between EDP and three cardiovascular disease outcomes, using two electronic data bases. In addition, the study presents a survey of the prevalence and use of automated office blood pressure recording machines (AOBP) in Canadian primary care centres. The study addresses the important issue of how systematic errors in the recording of clinical measurements can impact on patient health outcomes and provides some evidence to support AOBP as a means to reduce recording error and improve future patient health outcomes. While I find this study is generally well-designed and a clear report, there are a number of issues that need to be addressed that currently limit the validity and interpretability of the report. Major issues 1. The sampling strategy needs clarification (p5 lines 10-20). It is unclear whether the sampling was repeated many times, or a single sample (obtained by stratified sampling) was bootstrapped. It is necessary that the analyses in the study be repeated in multiple randomly drawn samples of the data, to limit the chance that the findings are chance findings found only in a single sample of the whole data. Moreover, p5 lines 16-18 mention a regression model, which I assume is explained in reference #20. This needs more explicit explanation in the text. 2. Figure 2 shows the thresholds for the UK and Canada differ (0.1,0.25 and 0.15,0.3, respectively), likely due to the data-driven cluster analysis. Given that the probability of a BP value ending in 1,3,7 or 9 is 0.4, the thresholds provided by the cluster analysis seem too low, and somewhat arbitrary, and thus require further

	explanation and justification. In addition, this classification assumes that a proportion of 1,3,7,9-ending values lower than their thresholds is strictly due to EDP, whereas it may be simply due to chance (especially in smaller centres). 3. I am concerned about the conclusions drawn on the causal associations between EDP and CVD outcomes. As EDP (or the use of AOBP vs manual BP measurement) is not randomized across patients or centres, it could be that the associations (SMRs) are confounded. For example, better funding or hospital care standards may cause there to be less EDP in BP measurements (maybe due to more AOBP), whilst also leading to better clinical outcomes through other unobserved mechanisms- thus confounding the association between EDP and CVD outcome. To strengthen the claims made by this paper, the analysis should account for potential confounding, and there should be further discussion of the potential impact of residual (or unmeasured) confounding on the SMR values, beyond the discussion on p9 lines 33-37. 4. The authors tentatively assume that EDP leads to poor clinical decisions (such as failure to diagnose and treat hypertension), which in turn leads to more CVD outcomes. This line of reasoning would ideally be strengthened by evidence of differences in how patients were treated between the high, medium, low EDP groups, or AOBP vs manual BP groups. Without this evidence, the recommendation for widespread use of AOBP seems overstated. Minor issues 1. The paper focuses on EDP specifically in all the analyses, and thus it would be clearer to mention this in the title and discussion instead of “systematic recording errors”. In addition, “cardiovascular outcomes” would be more clear in the title. 2. In the results, it is not always clear which data were used for which analysis, and whether specific results/figures are based on the Canadian, Toronto only, UK or all databases. This requires explicit mentioning throughout, especially in figure legends. 3. The phrase “EDP rate” should be EDP proportion instead, as this value is not strictly speaking a rate. 4. The text in Figure 2 needs to be larger. 5. The abbreviation EMR needs to be stated in full. 6. P5 lines 27-28. I could not access the related supplements- were they submitted? 7. There is no mention of missing data. Is this because data on outcomes were complete or were individuals in the study included only if data were complete? Details on this need to be reported. 8. Page 8- there is some repetition in paragraphs 2, 3 and 4, and thus it would be clearer to compare and contrast findings from other studies with this study in a single paragraph. 9. The statement on P9, lines 40-41 about repeated measurement or recording bias needs further elaboration, as it is currently unclear to what this refers. 10. I miss a discussion of why EDP seems to have led to rounding down, and hence a decrease in recorded BP values, on average. This would help to explain/support the credibility of the findings.

REVIEWER	Ji-Guang WANG The Shanghai Institute of Hypertension, Shanghai, China
REVIEW RETURNED	23-Aug-2018

GENERAL COMMENTS	In the present manuscript, the authors combined data from Canada and UK and investigated effect of digit preference on health outcomes. The analysis was well done and the manuscript well written. There are several minor concerns. 1. The authors mainly classified the sites according to digit preference. Is it possible to know the details on the use of blood pressure monitors ? If yes, the authors should report such information or data. 2. In line with the above comment, the authors should also look at the digit preference of blood pressure readings according to the use of manual auscultatory and automated oscillometric devices to see whether the latter device also caused digit preference. Previous studies showed that some devices skipped some of the key digits or certain numbers. 3. With the current data, the authors should have the possibility to look that the association between blood pressure level or control and health outcomes. It would be very interesting to see whether the blood pressure level or control status really matters for the clinical outcomes in these primary care centres. 4. One of the major clinical implications is encouraging the use of automated devices. In the introduction as well as discussion, the authors should provide more information on the status of blood pressure measurement in the two countries.

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:

Reviewer: 1
Reviewer Name: Romin Pajouheshnia
Institution and Country: University Medical Center Utrecht, The Netherlands
Please state any competing interests or state ‘None declared’: None declared

Please leave your comments for the authors below
This study assessed the prevalence of end digit preference (EDP) when recording blood pressure (BP) in primary care, the effect of this on the average recorded blood pressure values and the (causal) association between EDP and three cardiovascular disease outcomes, using two electronic data bases. In addition, the study presents a survey of the prevalence and use of automated office blood pressure recording machines (AOBP) in Canadian primary care centres. The study addresses the important issue of how systematic errors in the recording of clinical measurements can impact on patient health outcomes and provides some evidence to support AOBP as a means to reduce recording error and improve future patient health outcomes.

While I find this study is generally well-designed and a clear report, there are a number of issues that need to be addressed that currently limit the validity and interpretability of the report.

Major issues
1. The sampling strategy needs clarification (p5 lines 10-20). It is unclear whether the sampling was repeated many times, or a single sample (obtained by stratified sampling) was bootstrapped. It is necessary that the analyses in the study be repeated in multiple randomly drawn samples of the data,

to limit the chance that the findings are chance findings found only in a single sample of the whole data.

The stratified sampling without replacement was performed multiple times on the original datasets. We clarified the explanation for bootstrap and added the following to the manuscript:

“Since many patients had BP recorded multiple times with irregular visit to primary care between Jan 2006 to Dec 2015, we chose to discard excess information using a sampling mechanism. In particular, we generated 1000 independent replicates using the stratified sampling without replacement where one BP measurement was randomly chosen for a given patient. Logistic regression was performed on 1000 independently sampled replicates of the CPCSSN and RCGP RSC database. The odds ratios were estimated using the mean and 95% confidence intervals were estimated using the 2.5% and 97.5% percentiles of one thousand bootstrap estimates. “

Moreover, p5 lines 16-18 mention a regression model, which I assume is explained in reference #20. This needs more explicit explanation in the text.

We have explained the logistic regression model in the text as follows:

“All covariates in the logistic regression model were held constant to their latest value for each patient with respect to the study follow-up. For example, the most recent information on BMI or the diagnosis of diabetes or hypertension medication was used for each patient.”

2. Figure 3 shows the thresholds for the UK and Canada differ (0.1,0.25 and 0.15,0.3, respectively), likely due to the data-driven cluster analysis. Given that the probability of a BP value ending in 1,3,7 or 9 is 0.4, the thresholds provided by the cluster analysis seem too low, and somewhat arbitrary, and thus require further explanation and justification. In addition, this classification assumes that a proportion of 1,3,7,9-ending values lower than their thresholds is strictly due to EDP, whereas it may be simply due to chance (especially in smaller centres).

The proportion of blood pressures ending in 1, 3, 7 or 9 is expected to be 0.4 if there is no end digit preference; in other words, 40% of readings should end in one of those four digits.

The cluster analysis shows that many practices had much less than the expected 40% of BPs ending in one of those four digits, thus providing evidence of EDP. The unsupervised cluster analysis grouped blood pressures that were more similar to each other; this is a technique that is commonly used in exploratory data mining. The decision boundaries for classifying practices into strong, some or no EDP group are based on the minimization of two-dimensional Euclidean distance with respect to cluster specific centroid. Hence, the decision boundaries are not arbitrary but rather deterministic using cluster analysis.

We excluded sites with fewer than 1,000 BP readings in any one year; the high number of readings per clinic per year would decrease the likelihood of findings being due to chance, which would be more likely with small sample size. As well, chance readings could also be associated with a proportion of greater than 40%, which is not the case as shown on the figure.

3. I am concerned about the conclusions drawn on the causal associations between EDP and CVD outcomes. As EDP (or the use of AOBP vs manual BP measurement) is not randomized across patients or centres, it could be that the associations (SMRs) are confounded. For example, better

funding or hospital care standards may cause there to be less EDP in BP measurements (maybe due to more AOBP), whilst also leading to better clinical outcomes through other unobserved mechanisms- thus confounding the association between EDP and CVD outcome. To strengthen the claims made by this paper, the analysis should account for potential confounding, and there should be further discussion of the potential impact of residual (or unmeasured) confounding on the SMR values, beyond the discussion on p9 lines 33-37.

Residual confounding could be a problem; the analysis cannot account for unmeasured confounders. Nonetheless, the association persisted over ten years and we are not aware of activities that could be associated with both more precise BP measurement in primary care and reduced cardiovascular outcomes concurrently over a large number of sites and wide geographic location. We also limited the conclusion to “association” between EDP and CVD frequency and also explicitly stated that this “association” does not imply “causal” relationship between EDP and CVD frequency. However, we hope that this association is important and will stimulate further research as suggested by the reviewer.

4. The authors tentatively assume that EDP leads to poor clinical decisions (such as failure to diagnose and treat hypertension), which in turn leads to more CVD outcomes. This line of reasoning would ideally be strengthened by evidence of differences in how patients were treated between the high, medium, low EDP groups, or AOBP vs manual BP groups. Without this evidence, the recommendation for widespread use of AOBP seems overstated.

Table 5 provides differences in mean BP for patients (which is the measurement of interest) with diabetes or hypertension between high EDP and low/no EDP groups. Patients with high EDP have consistently lower BPs for all subgroups, yet these patients also have higher rates of cardiovascular events. One would expect lower rates with better controlled BP in the high EDP group which presumably is not using AOBP.

The study points to systematic errors in measurement; better precision (less EDP) with AOBP has been well documented. Guidelines are now recommending more consistent use of AOBP, and this study strengthens current recommendation.

The current Hypertension Canada guideline states that: “Over the past century, auscultation has been the predominant blood pressure measurement method. If auscultatory blood pressure is performed properly (i.e., using standardized methodology), it correlates well with ambulatory measurements and can predict target organ changes (34-36). However, ‘real world’ routine office auscultatory measurement, when performed by both in nurses and physicians, is consistently inaccurate because standardized methodology is simply not followed (5,13-15,18,20,37-39). The BP obtained in routine clinical practice is on average 9/6 mm Hg higher than standardized measurements (40,41). Unfortunately repeated educational programs to improve blood pressure measurement do not produce sustainable improvements in technique (14,42-47). The widespread removal of mercury from clinics and hospitals has created an additional source of error, as replacement aneroid devices commonly used for auscultation are inaccurate unless regularly calibrated .

For these reasons, the Task Force strongly encourages the use of validated electronic digital oscillometric devices. These devices are pre-programmed to take either single measurements or an automated series of measurements with averaging of the results. Electronic oscillometric devices minimize or eliminate many auscultation-induced errors, including those related to provider hearing deficits, terminal digit preference (rounding the reading to 0 or 5) and rapid deflation (48,49). Many devices for both clinical and public use have been found to be accurate and reproducible when compared to research-quality OBP (www.dableducational.com).”

(https://guidelines.hypertension.ca/diagnosis-assessment/measuring-blood-pressure/)

Contrary to the guideline statement, we have found lower BP (rather than higher) with those likely using manual BP measurement rather than AOBP, perhaps due to clinical inertia2.

We are making an argument in support of no longer using manual methods where this can be avoided. Better methods of measurement exist and should routinely be used.

Minor issues
1. The paper focuses on EDP specifically in all the analyses, and thus it would be clearer to mention this in the title and discussion instead of “systematic recording errors”. In addition, “cardiovascular outcomes” would be more clear in the title.

This has been changed.

2. In the results, it is not always clear which data were used for which analysis, and whether specific results/figures are based on the Canadian, Toronto only, UK or all databases. This requires explicit mentioning throughout, especially in figure legends.

This has been added for Figures 2 and 4.

3. The phrase “EDP rate” should be EDP proportion instead, as this value is not strictly speaking a rate.

We are using EDP proportion per year. The time dimension makes this a rate.

4. The text in Figure 2 needs to be larger.

Image resolution for Figure 2 has been improved and font size is also increased.

5. The abbreviation EMR needs to be stated in full.

This has been changed

6. P5 lines 27-28. I could not access the related supplements- were they submitted?

The survey has been added.

7. There is no mention of missing data. Is this because data on outcomes were complete or were individuals in the study included only if data were complete? Details on this need to be reported.

Missing data for BMI are shown on table 2. Data were complete for age and gender. Missingness cannot be estimated for other elements, such as diabetes, hypertension or cardiovascular outcomes since these variables are derived directly from the information routinely collected in the RCGP RSC and CPCSSN databases.

8. Page 8- there is some repetition in paragraphs 2, 3 and 4, and thus it would be clearer to compare and contrast findings from other studies with this study in a single paragraph.

This has been modified in the text.

9. The statement on P9, lines 40-41 about repeated measurement or recording bias needs further elaboration, as it is currently unclear to what this refers.

We tested for the impact of repeated measurements of BP in RCGP RSC and CPCSSN databases on the outcome (i.e. EDP) using a random-effects model. The heterogeneity due to repeated measurements was found to be non-significant. This motivated us to rely on ordinary logistic regression using bootstrap method (1000 replicates) for statistical inference.

10. I miss a discussion of why EDP seems to have led to rounding down, and hence a decrease in recorded BP values, on average. This would help to explain/support the credibility of the findings.

A possible explanation for the observation of rounding down is provided by Prospect Theory, used in Behavioral Economics, which describes decisions under conditions of uncertainty. Negative perceptions about possible risks (or risk aversion) outweigh positive perceptions about possible gains.3 There may be a behavioral bias towards rounding down; this may avoid risks associated with adding more medications with less emphasis on gains from cardiovascular outcome prevention.

This explanation has been added to the manuscript.

Reviewer: 2
Reviewer Name: Ji-Guang WANG
Institution and Country: The Shanghai Institute of Hypertension, Shanghai, China
Please state any competing interests or state ‘None declared’: None declared.

Please leave your comments for the authors below
In the present manuscript, the authors combined data from Canada and UK and investigated effect of digit preference on health outcomes. The analysis was well done and the manuscript well written. There are several minor concerns.

1. The authors mainly classified the sites according to digit preference. Is it possible to know the details on the use of blood pressure monitors ？ If yes, the authors should report such information or data.

We have reported the survey on automated BP monitors. Unfortunately, we do not have information on which automated BP machines were used.

2. In line with the above comment, the authors should also look at the digit preference of blood pressure readings according to the use of manual auscultatory and automated oscillometric devices to see whether the latter device also caused digit preference. Previous studies showed that some devices skipped some of the key digits or certain numbers.

As mentioned in the introduction, studies have shown reduced EDP for AOBP compared to manual BP measurement. The CAMBO trial, for example, found that about 13% of AOBP readings ended in zero (expected 10%) compared to 50% with manual BP readings.4

3. With the current data, the authors should have the possibility to look that the association between blood pressure level or control and health outcomes. It would be very interesting to see whether the blood pressure level or control status really matters for the clinical outcomes in these primary care centres.

We agree; however, this study concentrated on methods of measuring BP, rather than outcomes of BP control.

4. One of the major clinical implications is encouraging the use of automated devices. In the introduction as well as discussion, the authors should provide more information on the status of blood pressure measurement in the two countries.

As mentioned in the article, in Canada, 43% of family physicians reported using automated BP machines. There was no information on the proportion of patients with a BP measured with AOBP, if the machine was present in the office; not all patients have BP measured this way. We estimated this using levels of EDP as a proxy.

We appreciate the thoughtful reviews and the opportunity to reply to the questions. We look forward to next steps.

1. Canzanello VJ, Jensen PL, Schwartz GL. Are aneroid sphygmomanometers accurate in hospital and clinic settings? Arch Intern Med. 2001;161(5):729-731.
2. Salisbury C, Fahey T. Overcoming clinical inertia in the management of hypertension. Cmaj. 2006;174(9):1285-1286.
3. Kahneman D, Tversky A. Prospect theory; an analysis of decision under risk. Economica. 1979;47:263-291.
4. Myers MG, Godwin M, Dawes M, et al. Conventional versus automated measurement of blood pressure in primary care patients with systolic hypertension: randomised parallel design controlled trial. BMJ. 2011;342.

VERSION 2 – REVIEW

REVIEWER	Romin Pajouheshnia University Medical Center Utrecht, The Netherlands
REVIEW RETURNED	26-Oct-2018

GENERAL COMMENTS	Dear authors, The revised version of the manuscript shows several improvements over the first version. Most of my concerns have been addressed now, and I only have a few minor comments. 1. I previously raised concerns that this manuscript does not provide any direct evidence an association between EDP and treatment/intervention rates. At several points in the manuscript the authors hypothesized over EDP potentially leading to under-

	treatment of hypertension- which would explain the association with cardiovascular outcomes. However, Table 3 shows odds ratios for the association (OR) between hypertension medications and EDP, which is ~1 for both cohorts. This goes against the authors' hypothesis. The authors need to discuss this result, and the implications on their study conclusions. 2. Table 3 provides "adjusted ORs". What variables are they adjusted for? Which variables, how they were adjusted for and why they were selected needs to be reported in the methods. In addition, the adjustment variables should be reported as a footnote for table 3.

REVIEWER	Jiguang WANG The Shanghai Institute of Hypertension, Shanghai, China
REVIEW RETURNED	28-Oct-2018

GENERAL COMMENTS	No further comment.

VERSION 2 – AUTHOR RESPONSE

We thank the reviewer for his comments and careful review. We are providing responses to the queries.

1. I previously raised concerns that this manuscript does not provide any direct evidence an association between EDP and treatment/intervention rates. At several points in the manuscript the authors hypothesized over EDP potentially leading to under-treatment of hypertension- which would explain the association with cardiovascular outcomes. However, Table 3 shows odds ratios for the association (OR) between hypertension medications and EDP, which is ~1 for both cohorts. This goes against the authors' hypothesis.

Response: The results showed that patients diagnosed with hypertension had less frequent EDP as compared to patients without hypertension in both Canadian and UK database. Interestingly however, ORs for last digit zero (both sBP and dBP) were close to 1 for patients on hypertension medications compared to those not on medications. Although this effect was statistically significant (due to large sample size in Canada and UK), we believe that this adjusted difference in odds ratio was not clinically significant. Furthermore, this result does not provide information on association between rounding and intensification of medication. We were not able to capture the data to confirm or refute this potential association using our data sources and research design; there were also some data limitations in terms of medication dose as well.

Table 5 shows that mean systolic BP for patients with hypertension is lower in practices with high EDP. Based on this information, a possible interpretation is that medications may not be intensified because of rounding down associated with less precise measurement of BP.

The authors need to discuss this result, and the implications on their study conclusions.

Response: We have added this in to the paper. The text now reads:

"While there was no clinically significant association between measurement precision and presence of BP lowering medication (ORs close to 1), our data does not permit us to determine whether more precise measurement was associated with medication intensification through increase in dosage or addition of more medications. This could benefit from additional research."

2. Table 3 provides "adjusted ORs". What variables are they adjusted for? Which variables, how they were adjusted for and why they were selected needs to be reported in the methods. In addition, the adjustment variables should be reported as a footnote for table 3.

Response: We adjusted for patient variables that may influence BP or its measurement: age; sex; presence of hypertension and/or diabetes; BMI; use of hypertensive medications. We also adjusted for the size of the practice panels, as this may influence quality of care. Finally, we adjusted for year of measurement as EDP levels changed over time.

A note to this effect was added to the text of the manuscript in the section on methods. The list of variables is also reported as a footnote to table 3.

The title for Table 3 was incorrectly described as recording zero for systolic BP. As described in the methods, the logistic regression model used zero for both systolic and diastolic BP as outcome. We have corrected this and apologize for our error.

Thank you once more for the opportunity to provide responses.

VERSION 3 – REVIEW

REVIEWER	Romin Pajouheshnia, Postdoctoral researcher Utrecht University, The Netherlands
REVIEW RETURNED	18-Dec-2018

GENERAL COMMENTS	I thank the authors for addressing all the concerns I previously raised and have no further comments.