**Online Appendixes for Steinsaltz et al. Very short term Blood Pressure Variability and long-term Mortality.**

**Appendix 1. Addressing errors in blood pressure measurement**

The blood pressure errors were found in the at home measurements. While these did not destroy the usefulness of the home measurements, they did require some attention and decisions for how to work with these defects. We also consider them inherently interesting, and worth registering for future researchers working on these or similar data. In particular, the problem we have called “pseudo-repetition” was entirely unexpected, although not unprescedented, and is of particular concern to researchers trying to estimate individual variation in clinically relevant measures.

**Last digit preference.** Mild tendency for observers to prefer certain last digits in reporting BP measurements has been reported in other studies, though an analysis of the 1999 wave of NHANES reported no last-digit preference. [[1]](#endnote-1) The last-digit preference in NHANES III, on the other hand, is substantial, with about 26% of all the diastolic BP measurements ending in 0, but only about 30% ending in 4 or 6. A more complete analysis of this phenomenon is included in the appendix. Because the shifts due to last-digit preference are presumably small, we expect them to have little effect on the main effects that we are examining in this paper, but they do increase the probability of two measurements being rounded to the same value, something that needs to be taken into account in examining the problem of pseudo-repetition.

**Pseudo-replicates.**  While the protocol calls for each subject to have three independent BP measures taken, it is clear from looking at the distribution of the results that this is not what happened, in many cases, in the home measurements. Instead, one or two measurements were taken correctly, and the remaining “observations” were either copied in, or matched the earlier results because of the observer’s inclination to expect the same number to come up. We analyze this problem at length in the Appendix, showing that approximately 16% of all the subjects were corrupted by some amount of copying (whether intentional or inadvertent). This is a significant problem, because copying affects the variance of the reported measurements, shifting it consistently downward, to an extent that may be very large, and is impossible to reconstruct. The effect is to create a class of spurious low-variance individuals that dilute the effect of variance on survival. As we describe later, this phenomenon may have played a role in causing the negative result reported by Muntner *et al.*

**Missing or implausible measurements**: Some of the reported measures were extremely implausible, particularly for diastolic BP. 262 subjects had at least one diastolic BP measure recorded as 0, in addition to the 2108 subjects who were missing at least one measurement. We excluded all of these subjects, and indeed any subject who had at least one measurement recorded outside the ranges (40,140) for diastolic and (60,250) for systolic BP, as recommended by the CDC. There was just one subject with systolic BP measures that were too low, but 271 subjects with low diastolic BP (in addition to those with measures recorded as 0). One subject was excluded for diastolic BP 156, and three were excluded for systolic BP that was too high, with the maximum being 264. After excluding the subjects for whom there is no follow-up information, we are left with 15602 of the originally 18155 subjects. We refer to this as population B.

The empirical means of the home and clinic measures in population B are tabulated in Table 1. We note that the home measures are systematically higher than the clinic measures, within every demographic group, with greater differences for subjects who are white or Mexican, and female. The average difference is about 2.3 for diastolic and 2.7 for systolic, which is small compared with the general range of the differences, which have SD of 10.1 (diastolic) and 14.0 (systolic).

Appendix 1 Table 1 Mean BP (with SD of individual means in parentheses) by sex and ethnicity.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sex | Ethnicity | Home Systolic | Clinic Systolic | Home Diastolic | Clinic Diastolic |
| M | Black | 128.4 (18.9) | 127.4 (19.4) | 78.1 (13.1) | 77.4 (12.3) |
|  | White | 131.0 (19.2) | 127.8 (19.2) | 76.8 (11.2) | 75.2 (10.7) |
|  | Mexican | 126.7 (17.6) | 123.2 (17.8) | 76.1 (11.9) | 74.2 (11.7) |
|  | Other | 125.5 (18.4) | 122.5 (17.9) | 76.7 (11.8) | 75.3 (10.7) |
| F | Black | 122.7 (21.6) | 121.9 (22.6) | 74.5 (13.0) | 71.8 (12.7) |
|  | White | 126.9 (22.4) | 123.1 (22.5) | 73.4 (10.6) | 70.1 (10.2) |
|  | Mexican | 120.7 (20.7) | 117.4 (20.9) | 71.5 (10.9) | 68.6 (10.7) |
|  | Other | 119.0 (19.3) | 117.0 (21.0) | 72.5 (11.5) | 69.1 (10.2) |

**Last-digit preference**  One expects approximately 20% of the reported observations to end in each of the possible digits 0,2,4,6,8. These will be slightly modified by the overall distribution of blood pressure measurements. Since the true blood pressure values are spread over multiple decades – the range from first to ninth decile of diastolic BP measurements is (60,90), and for systolic BP is (104,156) – this effect will be very small.

The actual distributions of final digits are summarized in Table 2. The clinic measures seem to fit the expectations fairly well, but the home measures show a substantial preference for final digits 0 and 8. These distortions agree broadly with those previously reported for a later wave of NHANES blood pressure measurements, but they are more extreme.

Appendix 1 Table 2: Fraction of measurements with each final digit.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Site |  | 0 | 2 | 4 | 6 | 8 |
| Home | SBP | 0.239 | 0.198 | 0.159 | 0.170 | 0.234 |
| DBP | 0.267 | 0.187 | 0.160 | 0.159 | 0.227 |
| Clinic | SBP | 0.187 | 0.179 | 0.198 | 0.218 | 0.219 |
| DBP | 0.192 | 0.189 | 0.209 | 0.212 | 0.197 |

By increasing the clustering of the observations, this preference for reporting last digits 0 and 8 will somewhat reduce the variance of observations. The effect is probably small, and should not substantially affect our results. We have nonetheless taken it into account in our estimation of the empirical prior for the home and clinic variance distributions.

**Pseudo-replication** Looking at Figure 1, we see that the first bar – corresponding to individuals with 0 SD, hence three identical measurements – is about twice as high in the home measurements as in the clinic measurements. One possible explanation is that some of the examiners performing the home measurements may, either intentionally or inadvertently, have duplicated measurements, rather than recording three independent measurements. As we discussed, the prior work of Muntner *et al.* used a statistical technique that was very sensitive to the inclusion of a substantial number of these subjects with erroneously reported zero SD. Our methods are far more robust to the influence of these falsely reported zero SD measurements.

To clarify this issue, we considered several other features of the data. We would expect an individual’s having three identical SBP measurements to be approximately independent of having three identical DBP measurements. Not exactly, because variability in SBP is slightly correlated with variability in DBP, and lower variability increases the likelihood of three identical measurements. But the correlation should be small. In fact, whereas independence would predict only 0.064\*0.050=0.3% of subjects to have three identical measurements for SBP and DBP, we find that the actual fraction is 0.9%, so three times higher.

If each individual had three independent measurements, we would expect that when there are two identical measurements reported, the odd one is equally likely to be any one of the three. If there were a trend in the measurements – for example, a tendency of patients to relax and lower their BP, or the reverse – we would expect the second measurement to be the least likely to be distinct. Instead, we find the counts and proportions given in Table 3.

Appendix 1 Table 3 Number of individuals having two matching measurements whose unmatched measurement is in the given position. (DBP includes only individuals whose measurements are all at least 45.)

|  |  |  |  |
| --- | --- | --- | --- |
| Which one distinct | 1 | 2 | 3 |
| SBP | 3040 (41.9%) | 2461 (33.9%) | 1758 (24.2%) |
| DBP | 3272 (40.0%) | 2927 (35.8%) | 1980 (24.2%) |

The disparity is huge, and it is not easy to explain if the measurements are independent. On the other hand, it is very much consistent with what we would expect to see if some examiners at some times took two independent measurements, and then copied one of the first two measurements for the third. (Again, this could be inadvertent, if the examiner were influenced by the first two measurements to anticipate a repetition.) This pattern would also be strengthened if some examiners at some times saw two identical measurements for the first two measurements, and If we think of the 24.2% in category 3 (so with measurements 1 and 2 identical) as representing the true rate of matching, then would conclude that about 27% of the apparent triple matches are spurious, which agrees with the inference we drew from the simulations.

The last piece of evidence comes from looking at the results recorded by different examiners. There are 13 different examiners coded in the data. If the triple matches were an accurate reflection of the similarity of the subjects’ BP measurements, we would expect that all examiners would have approximately equal numbers of such results. Instead, we find the results in Figure 2. Three examiners seem to have much higher fractions of identical BP measures than all ten others, and it is the same three for both systolic and diastolic measures. If we exclude those three, the results for the other ten examiners are consistent (according to a chi-square test) with the hypothesis that they all have the identical frequency of three identical measurements, 3.9% for systolic and 5.4% for diastolic.

Appendix 1, Figure 2 Fraction of subjects with three identical BP measurements reported by each examiner, with 95% confidence intervals. Black dashed line is the average of all examiners other than 3001, 3002, and NA (not given, 0 or 88888).



There is very little variation among the examiners in the proportion of subjects reported with two matching measurements. But if we look at the position of the non-matching measurement there is considerable variation, as we see in the ternary plot Figure 3. Interestingly, every single examiner had fewer matches between the first and second measurements than between first and third or between second and third. The three examiners who stood out for their proportion of triple matches (marked green in Figure 3) also have the highest proportion of matches between the second and third measurements. Only one of the examiners has proportions that do not fail a chi-square test at the 0.05 level for equality of the proportions of the three different possible matches.

Appendix 1 Figure 3 Proportion of subjects who had matching results for the first two measurements (X3), last two (X1) or the first and third (X2). Top is SBP, bottom DBP. The black dot is the point (1/3,1/3,1/3). The green dots correspond to the two examiners who had exceptionally high proportions of triple matches. The red dots correspond to the other ten examiners. The blue dot includes all results where the examiner was either not listed, or had the “missing” code 88888.



1. Ostchega Y, Prineas R J, Paulose-Ram R, *et al.* National Health and Nutrition Examination Survey 1999-2000: Effect of observer training and protocol standardization on reducing blood pressure measurement error. *J Clin* Epidem. 2003;56: 768—774. [↑](#endnote-ref-1)