Evaluating Measurement Error in Readings of Blood Pressure among Adolescents and Young Adults in a Developing Country

Shawn Bauldry

University of Alabama at Birmingham

Kenneth A. Bollen

Linda S. Adair

University of North Carolina at Chapel Hill

Abstract

Biomarkers are increasingly being included in surveys used by population researchers and prominent among these biomarkers are blood pressure readings. To help address random fluctuations in blood pressure, it has long been known that multiple readings of blood pressure are preferable to a single reading. In this study, we use a Structural Equation Modeling approach to evaluate measurement error in blood pressure readings at three different time points spanning six years among adolescents and young adults who were part of a longitudinal epidemiological study based in Cebu, Philippines. Our results indicate that there are no systematic differences in the measurement properties of the first, second, and third readings for both females and males across each of the three waves of data. There are, however, differences in the measurement properties across waves and for females and males. Finally, we find that simple linear combinations of

the blood pressure readings have quite high validity, and therefore do a good job of reflecting the underlying "true" blood pressure.

Introduction

Biomarkers are increasingly being included in surveys used by population researchers. Prominent among these biomarkers are blood pressure readings. This is done because high blood pressure is related to cardiovascular disease, a leading cause of death around the world, strokes, and kidney disease. As population researchers incorporate blood pressure readings into their analyses it is important to understand the quality of the measurements.

To help address random fluctuations in blood pressure, it has long been known that multiple readings of blood pressure are preferable to a single reading (Souchek, Stamler, Dyer, Oglesby, and Lepper 1979). In addition to random fluctuations, however, numerous studies have demonstrated that blood pressure readings are influenced by a number contextual factors, including the device used for measurement (Bassein, Borghi, Costa, Strocchi, Mussi, and Ambrosioni 1985; Niyonsenga, Vanasse, Courteau, and Cloutier 2008), the time of year (Andersen, Henriksen, Jense, and The Copenhagen City Heart Study Group 2002), and potential sources of stress such as the "white coat" effect or the timing of measurement (Bodegard, Erikssen, Sandvik, Kjeldsen, Bhørnhold, and Erikssen 2002) among others. Furthermore, blood pressure readings are subject to recording errors with digit preference the most frequently studied source (Bennett 1994; Hessel 1986; Keary, Atkins, and O'Brien 1998; Niyonsenga, Vanasse, Courteau, and Cloutier 2008).

To evaluate measurement error in readings of blood pressure, Batista-Foguet, Coenders, and Ferragud (2001) adopted a structural equation modeling framework and used multiple group multi-trait multi-method (MTMM) models to decompose the variance in readings of blood pressure into components representing “true” blood pressure, random fluctuations, and systematic error. Using data from elderly patients in Spain they found that the second blood pressure reading had the best relationship with “true” blood pressure and that a linear combination of the readings using factor score weights had better measurement properties than a simple average of the readings.

We adopt a similar analytic approach as Batista-Foguet et al. (2001) to evaluate measurement error in blood pressure readings at three different time points spanning six years among adolescents and young adults who were part of a longitudinal epidemiological study based in Cebu, Philippines. Our analysis is guided by four research questions concerning measurement error. First, are there any systematic differences in the measurement properties of the first, second, or third readings done at approximately the same time? Second, are there any differences in measurement properties of the three readings across the three waves of data? Third, are there any differences in measurement properties of the three readings across females and males? Finally, are there any differences in the measurement properties of a simple average of the three readings as compared with a linear combination based on factor scores?

To our knowledge, this is the first study to evaluate measurement error in readings of blood pressure (a) among adolescents and young adults, (b) with a sample of respondents from a developing country, and (c) across three time waves of data. Given the centrality of blood pressure as a measure of adult health, this study will contribute to our understanding of the measurement properties of blood pressure readings across a range of contexts and how best to operationalize blood pressure for analysis.

Data

The data for our analysis are drawn from the Cebu Longitudinal Health and Nutrition Survey (CLHNS) (Adair, Popkin, Akin, Guilkey, Gultiano, Borja, Perez, Kuzawa, McDade, and Hindin 2011). The CLHNS began with an initial survey in 1983-1984 of 3,327 expectant mothers in 33 randomly selected communities located in the Cebu, Philippines metropolitan area. The mothers and their children were periodically resurveyed to capture the process of infant and adolescent development as well as changing family circumstances. Beginning in the 1998-1999 wave and continuing in the 2002 and 2005 waves of the survey blood pressure measurements of the adolescents were collected. During these waves the adolescents were respectively aged 14 to 16, 16 to 18, and 20 to 22 in the final wave.

A standard procedure was used for obtaining blood pressure measurements from each of the respondents. Respondents were measured after a 10 minute seated rest, during home visits. Trained interviewers using a mercury sphygmomanometer, and appropriate cuff sizes took three measurements.

Analysis Sample

The sample for this analysis consists of 2,127 cases (1,015 women and 1,112 men) with blood pressure readings for at least one of the three waves of data. Over 80 percent of the cases have blood pressure readings for all three waves. The sample sizes for the individual waves range from 2,087 at wave 1 to 1,966 at wave 2 and 1,812 at wave 3.

Blood Pressure Readings

Figures 1 and 2 provide box plots for the three readings of systolic and diastolic blood pressure across the three waves of data separately for females and males. For systolic blood pressure among both females and males we see similar distributions across the three readings within each wave. Across waves mean systolic blood pressure appears to be slightly increasing for females and males and the variance is increasing for females. For diastolic blood pressure we also observe similar distributions across readings for females and males within waves. Once again, across waves mean diastolic blood pressure appears to be slightly increasing for females and males, particularly by wave 3, and the variance appears to be increasing for females. Blood pressure increases with height as well as weight in children and adolescents, and would therefore be expected to increase over the time period covered by the study.

-- Figures 1 and 2 about here --

Analytic Approach

Following Batista-Foguet et al. (2001), we rely on a structural equation modeling approach to address our research questions concerning the measurement properties of the blood pressure readings. For our first analysis we specify separate multi-trait multi-method (MTMM) models for females and males and for each wave of blood pressure readings. The two traits in our MTMM models are systolic and diastolic blood pressure. The three methods in our MTMM models are the three readings. These first, second, and third method factors permit us to capture systematic error that covers over the systolic and diastolic readings for each measurement occasion.   These models allow us to decompose the variance in each of the individual blood pressure readings into components attributable to “true” systolic or diastolic blood pressure, systematic error associated with each reading occasion, and random error (sometimes referred to as unique factor) associated with each individual reading.

Our MTMM models can be written as

$$x_{ijk} = \alpha_{jk} + \lambda_{Tjk}\xi_{Tik} + \lambda_{Mjk}\xi_{Mij} + \delta_{ijk}, \tag{1}$$

where *x* is the blood pressure reading for trait *k* (systolic or diastolic blood pressure) with method *j* (reading 1, 2, or 3) on the *i*th subject. The ξ_{Tik} are the latent trait variables representing “true” systolic and diastolic blood pressure. The factor loadings, λ_{Tjk}, give the effects of actual underlying blood pressure on the readings. The ξ_{Mij} are the latent method variables representing the shared variance for the three reading occasions and the factor loadings, λ_{Mjk}, give the effects of the reading occasions on the readings. The α_{jk} are intercepts that capture any systematic differences in the means of the blood pressure readings. The δ_{ijk} are the random error terms for the blood pressure readings that we assume have means of zero and are uncorrelated with the ξs.

To ensure the model is identified we constrain the factor loadings for the methods factors to equal 1 and we scale the trait factors to the second reading of systolic and diastolic blood pressure respectively by setting these factor loadings equal to 1. We chose the second reading of blood pressure because it had been found to be more reliable than the first or third readings (Batista-Foguet, Coenders, and Ferragud 2001). Finally, we constrain the method factors to be uncorrelated with each other and with the latent traits. This set of constraints is consistent with a MTMM model where the number of traits does not equal the number of methods (Bollen and Paxton 1998). In the following analyses, we refer to this specification as the initial model.

We impose additional constraints to test for relative bias across the readings. The first set of additional constraints involves setting the remaining factor loadings for systolic and diastolic blood pressure to equal 1. The second additional set of constraints involves setting the intercepts, α_{jk}, equal to 0. These restrictions imply that the intercepts and slopes relating the blood pressure reading to the latent blood pressure are same across the three occasions. They represent restricted versions of the initial model and can thus be assessed with chi-square difference tests as well as changes in the BIC.

Our second research question concerns testing for measurement invariance across waves. To conduct these tests we specify a confirmatory factor analysis (CFA) model that combines the preferred MTMM models from the first analysis from each of the waves separately for females and males. In the CFA model we allow all of the latent trait variables for systolic and diastolic blood pressure across the waves to be correlated, but we maintain the restriction that the method factors at each wave are uncorrelated with each other, with the method factors across waves, and with all of the latent trait variables. We refer to this specification as the initial CFA model.

For this analysis we maintain all of the cases by using a casewise ML estimator which only requires Missing At Random (MAR) rather than Missing Completely at Random (MCAR) data (Arbuckle 1996). To test for measurement invariance across waves we consider two sets of constraints. The first set constrains the random error variances for the respective blood pressure readings to be equal across waves. The second set constrains the variances of the method factors to be equal across waves. These sets of constraints also represent restricted versions of the initial CFA model and can thus be assessed with chi-square difference tests as well as changes in the BIC.

A useful means of assessing the measurement properties of each reading is to examine the standardized validity coefficient (Bollen 1989). In this setting, the standardized validity coefficient is equal to the standardized factor loading for the latent trait variables and is given by

$$\lambda_{Tjk}^{s} = \sqrt{\frac{\lambda_{Tjk}^{2}\phi_{Tkk}^{2}}{\sigma^{2}(x_{ijk})}}, \tag{2}$$

where ϕ_{Tkk}^{2} is the variance of the latent trait variable and $\sigma^{2}(x_{ijk})$ is the variance of the blood pressure reading. Batista-Foguet et al. (2001) refer to this quantity as a measure of "measurement quality," but we use the earlier terminology.

To test for measurement invariance across females and males, we place the preferred CFAs from our second analysis into a multiple group (MG) framework with groups defined by sex. We continue to use a Casewise ML estimator to maintain all of the cases in this analysis. The initial MG CFA model allows for all of the free parameters to vary by sex. We consider a similar set of constraints as with the analysis of measurement invariance across waves. First, we test whether the random error variances are equal for females and males. Second, we test whether the method factor variances are equal for females and males.

Our final research questions involves assessing the measurement properties of a simple average of the three readings as compared with a weighted averaged based on factor scores from the best MTMM models from the first analysis. To assess the two approaches to constructing linear combinations of the readings we rely on a measure of validity given by

$$Q = 1 - \frac{\sum w_{jk}^2 \theta_{jk} + \sum w_{jk}^2 \phi_{Mj} + \sum 2 w_{jk} w_{k'j} \phi_{Mj}}{var(\sum w_{jk} x_{ijk})}, \tag{3}$$

where w_{jk} are weights, either the estimated factor scores or 1/3 and 0 for the simple average, θ_{jk} are the error variances for each reading, and ϕ_{Mj} are the variances of the method factors (Batista-Foguet, Coenders, and Ferragud 2001).

Results

The first research question concerns whether there are any differences in the measurement properties of the first, second, and third readings. We begin by testing for relative bias across the readings by first constraining the factor loadings for the latent systolic and diastolic blood pressure variables (the latent trait variables) to all equal 1 and then constraining the intercepts for each of the readings to equal 0. Table 1 provides the model fit statistics for the initial MTMM model and then the two restricted versions of the initial model separately for females and males and for each of the waves. To partially take account of the multiple tests performed, we use the Holm (1979) multiple testing correction to take account of the number of tests performed within each table.

-- Insert Table 1 about here --

We find that on balance the initial MTMM models have a good fit with the data for both females and males across all three waves. Fit is assessed using the chi-square test of whether the model exactly predicts the covariance matrix and means of the observed variables. Because of the potential of excessive statistical power of the chi-square tests in large samples, we supplement them with fit indices. Specifically, we use the BIC (Schwarz 1978), CFI (Bentler 1990), TLI (Tucker and Lewis 1973), and RMSEA (Steiger and Lind 1980). For the BIC [chi-square – df* ln(N)], large negative values signify a good fit (Raftery 1995) while values at or close to 1 represent the best fit for the CFI and TLI. RMSEA values of less than 0.05 are desired while values greater than 0.10 suggest a poor fit in large samples (Browne and Cudeck 1993).

The chi-square tests are all non-significant, a result that is particularly impressive given the large sample size. The BICs are all negative, which indicate that the model is preferred over the saturated model. The CFIs and TLIs are all close to 1 and the RMSEAs are all less than 0.05.

Turning to the restricted versions of the initial model, we find that the difference in BICs strongly prefer the restricted models over the initial model for both females and males across all waves. The chi-square difference tests also largely prefer the restricted models. The one exception is for males at wave 1. The chi-square different test indicates a statistically significant deterioration in model fit with restricting the factor loadings to equal 1 and with the additional constraint of restricting the intercepts to equal 0. We examined the estimates for the factor loadings and intercepts for all of the initial models and found all factor loadings to lie between 0.98 and 1.00 and no consistent pattern among the estimates for the intercepts that would suggest systematic bias. Thus we conclude that there is little evidence of substantively important bias for the loadings and intercepts across the blood pressure readings for females and males across the three waves of data. We adopt the MTMM models with factor loadings constrained to 1 and intercepts constrained to 0 in the following analyses.

The second research questions concerns whether the measurement properties of the three readings differ across waves. For this analysis we specify separate CFAs for females and males that combine the restricted version of the MTMMs from each wave. Table 2 reports the model fit statistics for the initial CFA MTMM models for females and males as well as two restricted versions of the initial model. The first restricted version constrains the error variances for each of the respective readings to be equal across waves. The second restricted model relaxes the error variance constrain and instead constrains the variances of the respective method factors to be equal across waves.

-- Insert Table 2 about here --

As with the individual MTMM models we find that the initial CFA MTMM models have a good fit with the data. The BICs are negative, the CFIs and TLIs are close to 1, the RMSEAs are both below 0.02, and the chi-square statistic for females is non-significant. The chi-square statistic for males is significant, but once again is not that large given the sample sizes and the complexity of the model. We find, however, that model fit deteriorates quite substantially with constraints either to the error variances or to the method factor variances. For both females and males, the chi-square difference tests reject both versions of restricted models as do the BICs. These results indicate that the variances of the error terms and the variances of the method factors are not stable over time (presented below), which suggests that the measurement properties of the readings differ across waves.

Our third research question concerns whether the measurement properties of the readings vary by sex. We address this question by specifying a multiple group version of the CFA MTMM models discussed above with the groups defined by sex. Table 2 includes the model fit statistics for the initial multiple group CFA MTMM and for two restricted versions of the initial model. The first restricted version constrains the error variances for the respective readings to be equal for females and males. The second restricted version of the model includes the constraint that the method factor variances are equal for females and males.

As with the CFA MTMM models we find that the initial model has a reasonable fit, though it does have a statistically significant chi-square statistic. Both of the restricted models, however, result in substantially worse model fits and are rejected by the chi-square difference tests. Thus the results indicate that the variances of the error terms and the variances of the method factors are not equivalent for females and males. There is no clear pattern to the variation between females and males. At some waves and for some readings the method factor variance is greater for females, while at some waves and for some readings the method factor variance is greater for males.

Table 3 reports estimates for the standardized trait factor loadings, the method factor variances, and the error variances from the initial CFA MTMM models. These estimates provide additional information about the measurement properties of the blood pressure readings. The estimates for the standardized loadings are all quite close to 1 for both females and males across all three waves. There are some slight differences in the estimates across the different readings. At wave 1, the validities of the three readings all virtually identical for female and male, systolic and diastolic blood pressure. At wave 2, the second reading of both systolic and diastolic blood pressure has slightly higher validity than the first and third readings for both females and males with the one exception that validities of the second and third readings for systolic blood pressure for females are equal. Finally, at wave 3, the third reading of systolic and diastolic blood pressure has slightly higher validity for both females and males than the first and second readings, with the exception that the second reading has the highest validity for diastolic blood pressure for females. Nonetheless, these slight differences in estimates indicate that the readings are all roughly equally valid and thus, in contrast to past studies, there is little evidence to prefer one reading over the others in our sample.

-- Insert Table 3 about here --

We find quite small estimated variances for the method factors for both females and males across all waves. In all cases, the method variances are less than one, which are at least two orders of magnitude less than the variances of the latent traits that range from 78 to 117 for systolic blood pressure and 58 to 86 for diastolic blood pressure. The estimated method variances are also smaller than the estimated variances of the random errors. This suggests that the systematic errors in the readings in this data are minimal and in fact even less than the random errors. It is particularly notable that the systematic errors in readings are virtually zero for females at wave 1. It is possible that this is due to a tendency for the people taking the readings to fill in the results from the first reading for the subsequent readings. It is not clear why this would be the case for females more so than for males.

We also find relatively small estimated error variances for each of the blood pressure readings when compared with the variances of the latent traits. The error variances for the blood pressure readings at wave 1 are roughly an order of magnitude less than the error variance for the readings at waves 2 and 3, though the error variances at wave 2 are generally greater than the variances at wave 3. At wave 1 for females and males, we find the third reading of systolic blood pressure and the second reading of diastolic blood pressure have the smallest error variances. At wave 2, the second reading for systolic blood pressure for females and males, the second reading of diastolic blood pressure for females, and the third reading of diastolic blood pressure for males have the smallest error variances. While at the third wave, the third reading has the smallest error variance with the exception of diastolic blood pressure for females in which the second reading has the smallest error variance. Thus, as with the validities, there does not appear to be a systematic pattern of lower error variances for any specific reading across females and males across the three waves.

Our final research questions concerns whether there are differences in the measurement properties of a simple average of the readings as compared with a linear combination of the readings using factor scores. Table 4 presents the factors scores from the individual MTMM models that constrain the trait factor loadings to 1 and the intercepts to 0. Table 5 presents the validities for the simple averages and the weighted averages using the factor scores reported in Table 4 for systolic and diastolic blood pressure among females and males across the three waves.

-- Insert Tables 4 and 5 about here --

We observe variation in the factor score weights that suggest the readings contribute unequally to predicting the underlying latent systolic and diastolic blood pressure variables. The variation, however, does not appear to be systematic. For instance, among females at wave 1, the third reading has the highest weight for systolic blood pressure, but the second reading has the highest weight for diastolic blood pressure. In contrast, among females at wave 2, the weights are pretty similar across the three readings for both systolic and diastolic blood pressure. The lack of a systematic pattern suggests that no particular blood pressure reading is uniformly preferred for females and males across waves. This is further underscored when we consider the validity indices reported in Table 5. As one would expect, the indices for the factor scores are all equal to or greater than the indices for the simple averages, but the differences are substantively nil. These results suggest that among the adolescents and young adults in the CLHNS a simple average of the three blood pressure readings for females and males across the three waves of data provides a valid measure of blood pressure that performs essentially as well as linear combination using factor score weights.

Discussion

Our research was primarily motivated by four research questions: (1) are there any differences in the measurement properties of the first, second, or third readings done at approximately the same time? (2) are there any differences in measurement properties of the three readings across the three waves of data? (3) are there any differences in measurement properties of the three readings across females and males?, and, (4) are there any differences in the measurement properties of a simple average of the three readings as compared with a linear combination based on factor scores?

With respect to the first question, we do not observe any systematic differences in the measurement properties of the first, second, and third readings for both females and males across each of the three waves of data. This contrasts with past studies that suggest the second reading is the most reliable (e.g., Batista-Foguet, Coenders, and Ferragud 2001). The contrast in findings may be due to two sources: (1) our analysis estimates measurement properties on a younger population than in past studies, which could have less variance in blood pressure readings and (2) the blood pressure readings in our analysis have such high validity that it's not easy to distinguish a best reading.

We do find, however, that the there are differences in the measurement properties across waves and for females and males. In particular, we observed larger method factor variances at waves 2 and 3 than at wave 1, but otherwise few systematic patterns among the method factor variances. We also observed larger error variances at waves 2 and 3 than at wave 1, particularly at wave 2. Furthermore, in general, males had larger error variances than females. Thus, our results suggest that it is important to attend to potential differences in measurement properties over time and by sex.

Our final research question concerned how well linear combinations of the readings capture underlying “true” blood pressure and whether there are any differences in using a simple average as opposed to a linear combination based on factor score weights. We find that the linear combinations have quite high validity, and therefore do a good job of reflecting the underlying “true” blood pressure. In addition, we find that despite the unequal factor score weights, the simple average of the readings performs essentially as well as the linear combinations based on factor score weights. This result is also different than what Batista-Foguet et al. (2001) found in their analysis of an elderly population in Spain.

References

Adair, Linda S., Barry M. Popkin, John S. Akin, David K. Guilkey, Socorro Gultiano, Judith Borja, Lorna Perez, Christopher W. Kuzawa, Thomas McDade, and Michelle J. Hindin. 2011. "Cohort Profile: The Cebu Longitudinal Health and Nutrition Survey." *International Journal of Epidemiology* 40:619-625.

Andersen, Ulla Overgaard, Jens H. Henriksen, Gorm Jense, and The Copenhagen City Heart Study Group. 2002. "Sources of Measurement Variation in Blood Pressure in Large-scale Epidemiological Surveys with Follow-up." *Blood Pressure* 11:357-365.

Arbuckle, James L. 1996. "Full Information Estimation in the Presence of Incomplete Data." Pp. 243-277 in *Advanced Structural Equation Modeling: Issues and Techniques*, edited by G. A. Marcoulides and R. E. Schumacker. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bassein, Leona, Claudio Borghi, Francesco Vittorio Costa, Enrico Strocchi, Alessandra Mussi, and Ettore Ambrosioni. 1985. "Comparison of Three Devices for Measuring Blood Pressure." *Statistics in Medicine* 4.

Batista-Foguet, J. M., G. Coenders, and M. Artes Ferragud. 2001. "Using Structural Equation Models to Evaluate the Magnitude of Measurement Error in Blood Pressure." *Statistics in Medicine* 20:2351-2368.

Bennett, Stan. 1994. "Blood Pressure Measurement Error: Its Effect on Cross-Sectional and Trend Analyses." *Journal of Clinical Epidemiology* 47:293-301.

Bentler, Peter M. 1990. "Compariative Fit Indices in Structural Models." *Psychological Bulletin* 107:238-246.

Bodegard, Johan, Gunnar Erikssen, Leiv Sandvik, Sverre E. Kjeldsen, Jørgen Bhørnhold, and Jan E. Erikssen. 2002. "Early Versus Late Morning Measurement of Blood Pressure in Healthy Men. A Potential Source of Measurement Bias?" *Blood Pressure* 11:366-370.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, Kenneth A. and Pamela Paxton. 1998. "Detection and Determinants of Bias in Subjective Measures." *American Sociological Review* 63:465-478.

Browne, Michael W. and Robert Cudeck. 1993. "Alternative Ways of Assessing Model Fit." Pp. 136-162 in *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long. Newbury Park, CA: Sage.

Hessel, Patrick A. 1986. "Terminal Digit Preference in Blood Pressure Measurements: Effects on Epidemiological Associations." *International Journal of Epidemiology* 15:122-125.

Holm, S. . 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics* 6:65-70.

Keary, L., N. Atkins, and E. T. O'Brien. 1998. "Terminal Digit Preference and Heaping in Office Blood Pressure Measurements." *Journal of Human Hypertension* 12:787-788.

Niyonsenga, Theophile, Alain Vanasse, Josiane Courteau, and Lyne Cloutier. 2008. "Impact of Terminal Digit Preference by Family Physicians and Sphygmomanometer Calibration Errors on Blood Pressure Value: Implication for Hypertension Screening." *Journal of Clinical Hypertension* 10:341-347.

Raftery, Adrian. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25:111-163.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6:461-464.

Souchek, Julianne, Jeremiah Stamler, Alan R. Dyer, Paul Oglesby, and Mark H. Lepper. 1979. "The Value of Two or Three Versus a Single Reading of Blood Pressure at a First Visit." *Journal of Chronic Diseases* 32:197-210.

Steiger, James H. and John C. Lind. 1980. "Statistically-Based Tests for the Number of Common Factors." in *Annual Meeting of the Psychometric Society*. Iowa City, IA.

Tucker, Ledyard R. and Charles Lewis. 1973. "A Reliability Coefficient for Maximum Likelihood Factor Analysis." *Psychometrika* 38:1-10.

Tables and Figures

Figure 1. Box plots for systolic blood pressure readings across three waves for females and males.

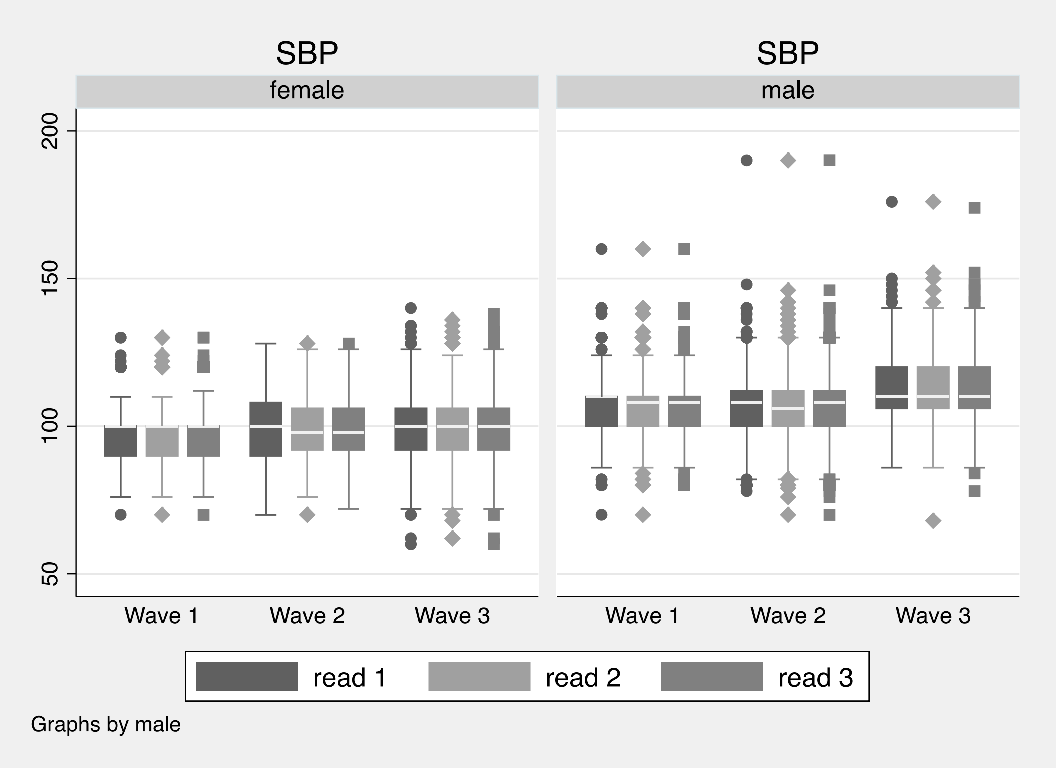

Figure 2. Box plots for diastolic blood pressure readings across three waves for females and males.

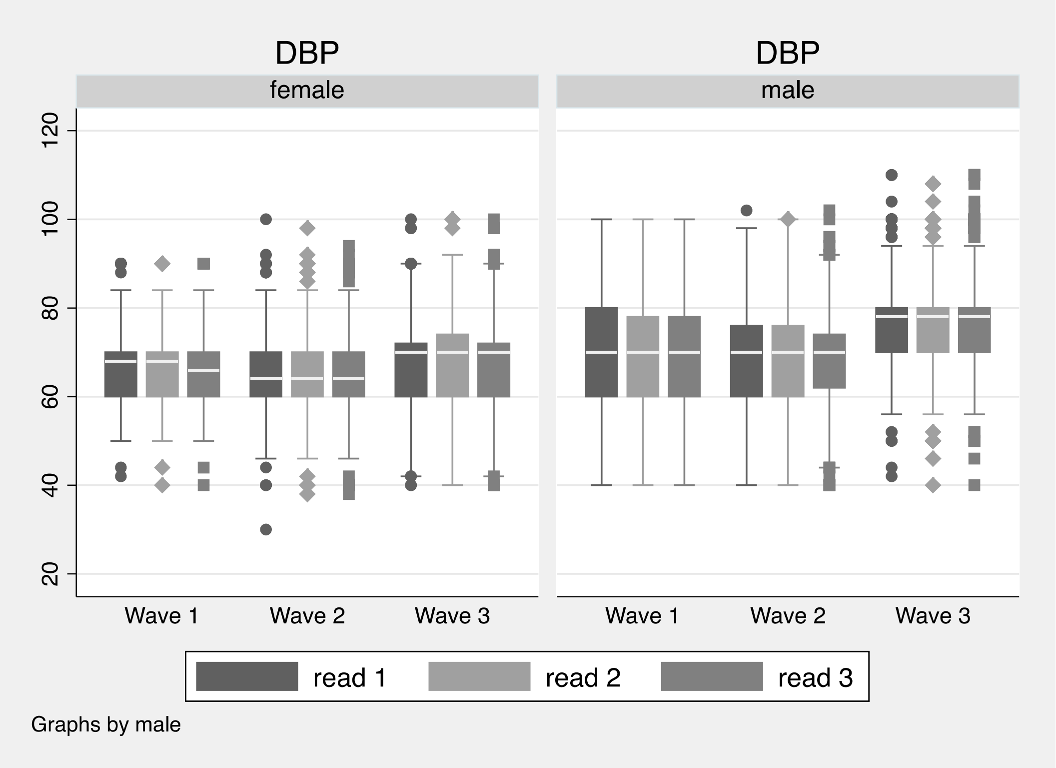

Table 1. Model fit statistics for separate MTMM models by sex and wave.

	N	χ^2	df	Adj p-val	BIC	CFI	TLI	RMSEA
Wave 1 Female								
M1: initial model	997	2.121	5	0.832	-32.403	1.000	1.000	0.000
M2: trait loadings constrained to 1	997	8.147	9	1.000	-53.996	1.000	1.000	0.000
M3: M2 + intercepts constrained to 0	997	13.402	13	1.000	-76.360	1.000	1.000	0.006
Difference M2 - M1		6.026	4	1.000	-21.593			
Difference M3 - M1		11.281	8	1.000	-43.957			
Wave 2 Female								
M1: initial model	896	12.776	5	0.486	-21.214	0.999	0.998	0.042
M2: trait loadings constrained to 1	896	17.392	9	0.687	-43.789	0.999	0.999	0.032
M3: M2 + intercepts constrained to 0	896	31.694	13	0.072	-56.679	0.998	0.998	0.040
Difference M2 - M1		4.616	4	1.000	-22.576			
Difference M3 - M1		18.918	8	0.337	-35.466			
Wave 3 Female								
M1: initial model	822	4.650	5	1.000	-28.909	1.000	1.000	0.000
M2: trait loadings constrained to 1	822	12.535	9	1.000	-47.871	1.000	0.999	0.022
M3: M2 + intercepts constrained to 0	822	17.829	13	1.000	-69.424	1.000	0.999	0.021
Difference M2 - M1		7.885	4	1.000	-18.962			
Difference M3 - M1		13.179	8	1.000	-40.515			
Wave 1 Male								
M1: initial model	1090	17.084	5	0.104	-17.886	1.000	0.999	0.047
M2: trait loadings constrained to 1	1090	27.600	9	0.031	-35.345	0.999	0.999	0.044
M3: M2 + intercepts constrained to 0	1090	43.213	13	0.001	-47.708	0.999	0.999	0.046
Difference M2 - M1		10.516	4	0.554	-17.460			
Difference M3 - M1		26.129	8	0.030	-29.822			
Wave 2 Male								
M1: initial model	1070	7.615	5	1.000	-27.262	1.000	0.999	0.022
M2: trait loadings constrained to 1	1070	18.726	9	0.497	-44.053	0.999	0.999	0.032
M3: M2 + intercepts constrained to 0	1070	31.120	13	0.081	-59.560	0.999	0.998	0.036
Difference M2 - M1		11.111	4	0.507	-16.791			
Difference M3 - M1		23.505	8	0.072	-32.298			
Wave 3 Male								
M1: initial model	990	8.760	5	1.000	-25.729	1.000	0.999	0.028
M2: trait loadings constrained to 1	990	11.906	9	1.000	-50.173	1.000	1.000	0.018
M3: M2 + intercepts constrained to 0	990	27.556	13	0.240	-62.114	0.999	0.999	0.034
Difference M2 - M1		3.146	4	1.000	-24.445			
Difference M3 - M1		18.796	8	0.336	-36.386			

Notes: Adj. p-val are adjusted p-values using Holm's sequential procedure.

Table 2. Model fit statistics for CFA MTMM models incorporating all waves.

	N	χ^2	df	Adj. p-val	BIC	CFI	TLI	RMSEA
Female								
M1: initial CFA model	1015	152.153	135	0.149	-782.404	1.000	1.000	0.011
M2: equal error variances across waves	1015	4306.216	147	0.000	3288.587	0.908	0.904	0.167
M3: equal method variances across waves	1015	289.695	141	0.000	-686.398	0.997	0.996	0.032
Difference M2 - M1		4154.063	12	0.000	4070.991			
Difference M3 - M1		137.542	6	0.000	96.006			
Male								
M1: initial CFA model	1112	199.804	135	0.001	-747.075	0.999	0.999	0.021
M2: equal error variances across waves	1112	4170.735	147	0.000	3139.689	0.922	0.918	0.157
M3: equal method variances across waves	1112	256.192	141	0.000	-732.770	0.998	0.998	0.027
Difference M2 - M1		3970.931	12	0.000	3886.764			
Difference M3 - M1		56.388	6	0.000	14.305			
Multiple Group Male & Female								
M1: initial MG CFA model	2127	351.957	270	0.001	-1716.909	0.999	0.999	0.017
M2: equal error variances across sexes	2127	1463.565	288	0.000	-743.226	0.988	0.987	0.062
M3: equal method variances across sexes	2127	570.974	279	0.000	-1566.855	0.997	0.997	0.031
Difference M2 - M1		1111.608	18	0.000	973.684			
Difference M3 - M1		219.017	9	0.000	150.055			

Notes: Adj. p-val are adjusted p-values using Holm’s sequential procedure.

Table 3: Measurement parameters from initial CFA MTMM models combining waves.

	Female			Male		
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
Std Trait Loadings						
SBP1	1.000	0.969	0.990	0.999	0.975	0.988
SBP2	0.999	0.982	0.986	0.999	0.989	0.989
SBP3	1.000	0.982	0.992	0.999	0.979	0.992
DBP1	0.997	0.968	0.982	0.997	0.969	0.988
DBP2	0.998	0.983	0.993	1.000	0.985	0.990
DBP3	0.995	0.976	0.981	0.999	0.984	0.994
Method variances						
M1	0.000	0.632	0.739	0.067	0.676	0.277
M2	0.000	0.543	0.287	0.005	0.000*	0.262
M3	0.000	0.320	0.513	0.121	0.895	0.339
Error variances						
SBP1	0.024	4.722	1.284	0.266	5.216	2.459
SBP2	0.100	2.556	2.551	0.192	2.772	2.329
SBP3	0.004	2.731	1.173	0.113	4.006	1.540
DBP1	0.274	3.957	1.929	0.464	4.879	1.847
DBP2	0.211	1.887	0.758	0.034	2.904	1.448
DBP3	0.517	3.087	2.223	0.068	1.913	0.663

Notes:

*The estimate for the variance is -0.206 with a standard error of 0.162. Given that this estimate is non-significant and a negative variance is nonsensical, we treat this estimate as 0.

Table 4. Factor score coefficients for blood pressure readings.

		systolic blood pressure			diastolic blood pressure		
Group	Trait	read 1	read 2	read 3	read 1	read 2	read 3
Female							
Wave 1	SBP	0.140	0.034	0.826	0.000	0.000	0.000
	DBP	0.000	-0.001	0.003	0.353	0.457	0.187
Wave 2	SBP	0.217	0.380	0.382	0.009	-0.014	0.018
	DBP	0.006	-0.022	0.027	0.229	0.439	0.308
Wave 3	SBP	0.351	0.231	0.408	-0.044	0.073	-0.023
	DBP	-0.010	-0.006	0.022	0.223	0.553	0.211
Male							
Wave 1	SBP	0.218	0.361	0.420	-0.008	0.223	-0.215
	DBP	0.021	0.034	-0.055	0.053	0.756	0.190
Wave 2	SBP	0.214	0.506	0.265	-0.011	0.073	-0.053
	DBP	-0.008	0.069	-0.054	0.191	0.398	0.394
Wave 3	SBP	0.281	0.296	0.413	0.017	0.018	-0.030
	DBP	0.019	0.014	-0.029	0.225	0.279	0.489

Table 5. Comparison of validity measures for simple average of readings and linear combination constructed using factor scores.

	Simple Average		Factor Score Linear Combination	
	SBP	DBP	SBP	DBP
Female				
Wave 1	1.000	0.998	1.000	0.998
Wave 2	0.985	0.983	0.986	0.985
Wave 3	0.993	0.990	0.993	0.992
Male				
Wave 1	0.999	0.999	1.000	1.000
Wave 2	0.987	0.986	0.989	0.988
Wave 3	0.993	0.994	0.993	0.994