Online data supplement

Systolic Blood Pressure, Socioeconomic Status, and Biobehavioral Risk Factors in a Nationally Representative U.S Young Adult Sample

Beverly H. Brummett, PhD[1], Michael A. Babyak, PhD,

Ilene C. Siegler, PhD, MPH [1], Michael Shanahan, PhD [2],

Kathleen Mullan Harris, PhD. [2,3],

Glen H. Elder, PhD[2, 3], & Redford B. Williams, M.D. [1]

[1] Department of Psychiatry and Behavioral Sciences, Duke University Medical Center

Durham, North Carolina

[2] Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

[3] Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

Corresponding Author: Beverly H. Brummett, Ph.D.

e-mail: brummett@duke.edu

Table of Contents:

Additional notes on study sample

The current study used data from a nationally representative sample of young adults—Add Health— or the National Longitudinal Study of Adolescent Health—that was designed to study health and developmental trajectories across the life course from adolescence to adulthood. The participants have been followed from grades 7 through 12 in 1995 through early adulthood in 2008. Participants were enrolled in one of 80 schools in the United States and data are also available from multiple sources, including parent interviews and assessments of community characteristics [1]. Study schools were selected to represent a stratified sample based on region, suburban/urban/rural, school type (public, private, or parochial), ethnic composition, and size. Wave I data collection occurred in 1994-95 and included a confidential in-school survey collected from 90,118 students grades 7 to 12. Approximately 80% of the schools that participated provided a roster from which students were randomly picked to participate in a 1 ½ hour in-home interview, resulting in a sample of 20,745 students aged 11 to 19. The in-home interview wave was conducted between April and December of 1995 and included the target child and one parent or parent-figure. Similarly, Wave II was conducted on the students who were in 7th to 11th grade during the first wave and occurred between April and August of 1996--yielding a sample of 14,738 adolescents ages 13 to 20. Wave III was conducted between August 2001 and April 2002 and included participants from Wave I, resulting in sample of 15,197 ages 18 to 26. SBP was assessed at Wave IV, which was conducted between April 2007 and February 2009 and included participants from Wave I, resulting in sample of 15,701 participants. Those without SBP measures or without sample weights were excluded from the present study, leaving 14,199 participants ages 24 to 34for the present analysis.

Additional notes on measures

The original biological sex variable was recoded as the variable “male” where 0=female and 1=male. Marital status was coded as 0= never married, and 1=ever married at Wave IV.

A single race variable was constructed from the responses to the following six questions asked at Wave I: 1) “Are you of Hispanic or Latino origin?”; 2) “What is your race? White”; 3) “What is your race? Black or African American”; 4) “What is your race? American Indian or Native American”; 5) “What is your race? Asian or Pacific Islander”; or 6) “What is your race? Other.”

Parental education was determined by two sources, from the adolescent in-home interview at Wave 4 and from the parent during an in-home parental interview conducted at Wave I. Self-reported education from the parental interview was typically supplied by the mother or female head of household. When data from both sources (adolescent and parent) were available, reports from the parent interview were used. Mother’s and father’s education level were combined into a variable that represented the highest level of education attained by either parent. This variable was coded into 5 classes that followed the approach used for adolescent level of education.

Self-reported financial strain was derived from a summary measure from 6 questions that assessed whether individuals reported the inability to pay bills, buy food, pay utilities, etc. Report of no inability across all items was scored 0, whereas a yes to one or more of the items was scored 1. Home Ownership was coded as 0= do not own a home, 1= own a home. Income, Financial Strain, and Home Ownership were assessed from questions at Wave IV. A “Built Environment” measure (rated by the study interviewer) was constructed by taking the sum of two items assessed at Wave I regarding: 1) how well the building in which the respondent was maintained, and 2) how well the surrounding buildings were maintained. These items were measured on an 4 point Likert-type response scale with higher scores reflecting poorer maintenance. The final summed score has a range of 2-8. To ascertain household income, participants were asked, “Thinking about your income and the income of everyone who lives in your household and contributes to the household budget, what was the total household income before taxes and deductions in (2006/2007/2008)? Include all sources of income, including non-legal sources.” Household income was represented by a variable that was originally in ordered categories. In order to approximate a continuous variable from the ordered categories, were assigned the midpoint of each category to a score in the given category (e.g., a value of 2,500 was assigned to individuals reporting between $0 and $5,000, a value of $125,000 to individuals reporting $100,000-149,999, etc. Income reported as > $150,000 was assigned the value $150,000). The resulted in the following possible values 2,500, 7,500, 12,500, 17,500, 22,500, 27,500, 35,000, 45,000, 62,500, 87,500, 12,5000, and 150,000. IN the regression analyses, we further rescaled these values by dividing by 50,000.

Exercise. The definition of regular exercise is measured using a standard physical activity behavior recall[2]. Lack of exercise, or no physical activity, is defined by self-reports of no bouts of moderate to vigorous physical activity (5-8 metabolic equivalents) across 7 groups of activities per week [3].

Cardiac medication. Participants were first asked, “Have you taken in medications in the last four weeks?” If the respondent answered yes, they were asked to collect their medications and the interviewer recorded a list of medications provided. If the respondent was not able to provide medications, medication status was documented based on the respondents recollection. We represented medication use with a yes/no variable. In addition to anti-hypertension medications we also included any other cardiovascular medication that would tend to lower blood pressure.

For our medication indicator variable, we included any medication in the cardiovascular class that was known to lower blood pressure. Participants taking one or more of these medications was coded on the Cardiac Medication variable as ‘1’ and the remainder were coded as 0. The medications for our sample appear in Table S1. Because a participant may have been taking more than one medication of this class the total number of medications exceeds the number of participants on at least one medication.

Notes on statistical analysis

All statistical models incorporated grand sample survey weights and also adjusted for individual school membership, thus adjusting for survey design effects of individuals clustered in the sampling unit of school and stratification of geographic region. Application of poststratification weights allows the results to be generalized to adults of similar age and background in the U.S. population.

Assessing nonlinearity: We used the rms package in R (http://cran.r-project.org) to conduct preliminary analyses of the linearity assumptions of the model. The rms package includes the option to model continuous covariates as nonlinear using a flexible, nonparametric algorithm called a restricted cubic spline [4]. Restricted cubic splines have many desirable properties, including imposing restrictions such that relatively few degrees of freedom are expended in determining the functional form of the regression line. In the case of BMI, waist circumference, and resting heart rate, we specified that the spline would have 3 knots. Unlike linear splines, however, the knots do not necessarily reflect inflection points (i.e. locations where the regression line turns). The knots in this case simply reflect locations where the piecewise cubic functions are joined. Stone [5] has shown that the location of the knot is not particularly important in terms of recovering the functional form of an association. We used the default knot locations in the rms package of the .10, .50, and .90 quantiles. Because spline coefficients can be hard to directly interpret, we simplified our model by developing a linear piecewise function for each of the above three variables, allowing the line to bend at the location suggested by the spline plots.

We also examined the assumption of additivity by testing a prespecified set of interactions between each of the biobehavioral variables (exercise, BMI, waist, heart rate, and smoking) and race, region, age, and gender. The inclusion of these terms explained only an additional 0.7% of the variance in SBP. A pooled test for this R-square difference between a model with the interactions and one without at 211-161 = 50 degrees of freedom is not statistically significant.

Missing Data: Missing data for the primary regression models were imputed using PROC MI in SAS with 50 imputations. The imputation model included all variables in full model, with binary variables rounded to nearest integer. We then estimated a general linear model using SAS PROC GLM for each of the imputed datasets. Regression estimates, and their 95% confidence intervals and p-values were then calculated from the GLM results using SAS PROC MIANALYZE.

Logistic Regression Predicting Systolic Hypertension: We supplemented the linear regression analyses with a logistic regression model in which the continuous SBP measure was replaced by a dichotomous variable reflecting systolic hypertension clinical categories (0 = SBP $\leq$ 140, 1 = SBP > 140). The results of the logistic model predicting systolic hypertensive status (SBP > 140) were similar to those of the linear regression model though, as would be expected, tests of the parameter estimates were associated with less statistical power compared to the linear regression. In this model, none of the SES predictors were statistically significant. Among the biobehavioral variables, the statistically significant predictors were alcohol use (OR = 2.45 for heavy drinking versus no drinking, 95% CI = 1.76, 3.41); BMI (OR = 1.31 per 5 kg/m^2 increase, 95% CI = 1.23, 1.41); heart rate (OR = 1.15 per 10 beats/minute increase, 95% CI = 1.09, 1.20); (OR = 1.16 per 10 cm increase, 95% CI = 1.01, 1.39); and smoking (OR = 1.15 for smoking versus not smoking, 95% CI = 1.00, 1.33).

Path models: We used Mplus (v 3.01) to estimate out path models. Alcohol, exercise, smoking, respondent and parent education were specified as categories (as these are on the “y” side of some equations). We used the theta parameterization. Code for the primary model appears below. The original exercise variable in the Add Health database was found to be incorrectly coded and was therefore reversed in the call. We also allowed an indirect path from respondent education to SBP by way of household income and the biobehavioral variables. Associations were modeled as linear. We used the weighted least-squares with mean and variance adjustment (WLSMV) estimation procedure which allows the path coefficients to be interpreted in the original metric of the variables. Missing data were managed in path models using the full information maximum likelihood procedure (FIML). Estimates were also weighted using grand sample weights, and the school identification code was specified as a clustering variable.

As was the case with Chaix et al., we also found that the strong correlation between waist and BMI created estimation problems when both variables were included in the same model. Whereas Chaix et al. combined these two variables into a single “body shape” index, we chose to use only BMI in the primary model so that the indirect effect estimate would be more straightforward to interpret. We estimated a second model, in which we replaced BMI with the waist measurement.

<u>Mplus code</u>

```
Mplus VERSION 3.01
MUTHEN & MUTHEN
INPUT INSTRUCTIONS
  TITLE:PRIMARY MEDIATION MODEL
  DATA:
    FILE IS "Z:\xx\xx\xx\xx\xxxx.csv";
  VARIABLE: NAMES ARE
  AID  male  sbp  hr  bmi  waist
  drink  edclass  finstrn2
  smoke  exer   married
```

```
ownhome  race PSID  region GSWGT
hhinc  edclassp
edp2  ed2  cardmed
builtenv  age10
bmi32  hr72  waist100
hisp  black  asian  nvam  other
edp3  edp4  edp5
ed3  ed4  ed5
drink1  drink2  drink3  drink4;
 weight is gswgt;
cluster is psid;
  missing are all (-999);
 categorical are edclass edclassp drink
  smoke rexer;
USEVARIABLES  sbp  edclass edclassp
hhinc age10 hisp black asian
 nvam other male cardmed
 drink smoke
 hr  bmi rexer;
   define:
    rexer = exer;
    if (exer == 0) then rexer = 1;
    if (exer == 1) then rexer = 0;
ANALYSIS:
PARAMETERIZATION=THETA;
 TYPE=COMPLEX MISSING;
    ITERATIONS = 10000;
  CONVERGENCE = 0.00005;
 MODEL: sbp  on edclass  edclassp hhinc drink
 rexer smoke bmi hr
 age10 hisp black asian nvam other male cardmed;
edclass on age10 hisp black asian nvam other male cardmed;
   edclassp on  hisp black asian nvam other male age10 cardmed;
   hhinc on edclass hisp black asian nvam other male age10 cardmed;

   drink on hisp black asian nvam other male age10 cardmed
   edclass hhinc;
   rexer on  hisp black asian nvam other male age10 cardmed
   edclass hhinc;
   smoke on  hisp black asian nvam other male age10 cardmed
   edclass hhinc;
   bmi on hisp black asian nvam other male age10 cardmed
   edclass hhinc;
   hr on hisp black asian nvam other male age10 cardmed
   edclass hhinc;
  MODEL INDIRECT:
```

```
   sbp ind edclass;
   sbp ind hhinc;
 output: sampstat standardized cint;
```

References

1. Harris CM, Halpern CT, Whitsel AE, Hussey DJ, Tabor JW, Entzel PP, Udry JR. The National Longitudinal Study of Adolescent Health: Research Design. 2009; http://www.cpc.unc.edu/projects/addhealth/design.

2. Andersen RE, Crespo CJ, Bartlett SJ, Cheskin LJ, Pratt M. Relationship of physical activity and television watching with body weight and level of fatness among children: results from the Third National Health and Nutrition Examination Survey. *JAMA.* 1998;279:938-942.

3. Harris KM. An integrative approach to health. *Demography.* 2010;47:1-22.

4. Stone CJ, Koo CY. Additive splines in statistics. Paper presented at: Proceedings of the Statistical Computing Section, American Statistical Association1985; Washington DC.

5. Pressman SD, Cohen S. Does positive affect influence health? *Psychol Bull.* 2005;131:925-971.

Table S1: Medication data

Medication class	N	% of total
ACE Inhibitor	148	20.4
Beta-Blockade-Cardioselective	118	16.3
Combination Antihypertensive	73	10.1
Calcium Channel Blocker	64	8.8
Diuretic-Thiazide	50	6.9
Beta-Blockade-Non Cardioselective	49	6.8
Diuretic-Potassium Sparing	44	6.1
ACE II Inhibitor	39	5.4
Aldosterone Receptor Antagonist	33	4.6
Antiarrhythmic Group II	27	3.7
Diuretic-Loop	19	2.6
Antiarrhythmic Group I	15	2.1
Antiadrenergic-Centrally Acting	15	2.1
Antiarrhythmic Group IV	12	1.7
Diuretic-Carbonic Anhydrase Inhibitor	4	0.6
Antiarrhythmic Group V	4	0.6
Antiadrenergic-Peripherally Acting	4	0.6
Agents for Pulmonary Hypertension	2	0.3
Vasodilator	2	0.3
Renin Inhibitor	1	0.1
Antianginal Agent	1	0.1
Angent for Hypertensive Emergency	1	0.1
Total	725	100.0

Table S2. Indirect effects from education and income to SBP

Mediating pathway from respondent education	Indirect Effect (unstandardized)	95% CI
Education→BMI→SBP	-0.50	-0.66, -0.34
Education→Alcohol→SBP	0.13	0.06, 0.20
Education→Heart Rate→SBP	-0.20	-0.26, -0.14
Education→Exercise→SBP	0.08	-0.02, 0.18
Education→Smoking→SBP	-0.08	-0.32, 0.15
Total indirect effect of Education	-0.91	-1.19, -0.63
Direct effect of Education	0.32	-0.10, 0.73
Total effect of Education	-0.59	-0.91, -0.26

Income→BMI→SBP	-0.14	-0.32, 0.04
Income→Alcohol→SBP	0.12	0.04, 0.19
Income→Heart Rate→SBP	-0.07	-0.13, -0.004
Income→Exercise→SBP	0.01	-0.02, 0.04
Income→Smoking→SBP	-0.04	-0.13, 0.06
Total indirect effect of Income	-0.13	-0.29, 0.04
Direct effect of Income	-0.61	-1.03, -0.20
Total effect of Income	-0.74	-1.19, -0.29