

Accounting for Selectivity Bias and Correlation Across the Sequence From Elevated Blood Pressure to Hypertension Diagnosis and Treatment

Penny Gordon-Larsen,^{1,2} Samantha M. Attard,^{1,2} Annie Green Howard,³ Barry M. Popkin,^{1,2} Bing Zhang,⁴ Shufa Du,^{1,2} and David K. Guilkey PhD^{2,5}

BACKGROUND

It is unknown whether efforts to reduce hypertension burden in countries with very high prevalence, would be more effective if directed at hypertension diagnosis vs. treatment. Most analyses do not address bias and correlation across the sequence from elevated blood pressure (BP) to hypertension diagnosis and treatment, leading to potentially misleading findings.

METHODS

Using data spanning 18 years of the China Health and Nutrition Survey ($n = 18,926$; ages 18–75 years), we used an innovative 3-step, integrated system of equations to predict the sequence from: (i) elevated BP (systolic/diastolic BP $\geq 140/90$ mm Hg) to (ii) diagnosed hypertension conditional on elevated BP, and to (iii) treatment (medication use) conditional on diagnosis, accounting for measured and unmeasured individual- and community-level confounders at each of the 3 steps. We compared results to separate traditional logistic regression models without control for unmeasured confounding.

RESULTS

Using our 3-step model, elevated BP increased from 12.6% and 8.5% (1991) to 36.8% and 29% (2009) in men and women, respectively, but diagnosis remained under 50%. We found widening disparities in hypertension diagnosis (higher hypertension at lower vs. higher education (difference of 2% in 1991 that widened to 5% in 2009)) and narrowing disparities in education (difference of 6% in 1991 to 4% in 2009) and insurance status (difference of 7% in 1991 to 2% in 2009) for treatment.

CONCLUSIONS

Our 3-step model improved model fit over traditionally used models. Our findings highlight serious barriers to hypertension diagnosis in Chinese adults, particularly among men and individuals of low attained education.

Keywords: China; urbanization; health insurance; health infrastructure.

doi:10.1093/ajh/hpx137

One-third of global deaths result from cardiovascular diseases¹ with the vast majority of cardiovascular disease deaths in low-middle income countries.² China has recently experienced major increases in ischemic heart disease and stroke,³ concomitant with multiple health care system reforms.^{4,5} While 300 million individuals in China are estimated to have high blood pressure (BP),⁶ over half are undiagnosed and untreated,^{7–10} leading to substantial health and economic burden.¹¹ It is unknown whether efforts to reduce hypertension burden in China would be more effective if directed at hypertension diagnosis or treatment.

Much of the literature considers each step in the continuum from elevated BP to hypertension diagnosis and hypertension treatment as discrete unrelated steps, generally using standard logistic regression models.⁹ There are papers solely

on hypertension, others just on diagnosis, and yet others on treatment. No analyses consider the full sequence from having elevated BP, to hypertension diagnosis, to hypertension treatment, while addressing differential effects at each stage and taking each previous step into account (i.e., hypertension treatment conditional on diagnosis and elevated BP). It is likely that health insurance, health system availability, and socioeconomic factors may be differentially associated with elevated BP, diagnosis, and receiving treatment for hypertension. Yet, the vast majority of research uses separate logistic regression models and ignores these important sources of selectivity and correlation between elevated BP, hypertension diagnoses, and treatment, thus resulting in biased estimates.¹² Further, most studies do not account for underlying, unobservable factors influencing who gets diagnosed

Correspondence: Penny Gordon-Larsen (pglarsen@unc.edu).

Initially submitted May 29, 2017; date of first revision June 28, 2017; accepted for publication September 12, 2017; online publication September 26, 2017.

¹Department of Nutrition, Gillings School of Global Public Health & School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; ²Carolina Population Center, UNC-Chapel Hill, Chapel Hill, North Carolina, USA; ³Department of Biostatistics, Gillings School of Global Public Health, UNC-Chapel Hill, Chapel Hill, North Carolina, USA; ⁴National Institute of Nutrition and Health, Chinese Center for Disease Control and Prevention, Beijing, China; ⁵Department of Economics, UNC-Chapel Hill, Chapel Hill, North Carolina, USA.

© American Journal of Hypertension, Ltd 2017. All rights reserved. For Permissions, please email: journals.permissions@oup.com

and treated, despite the fact that health-consciousness, motivation, and demand for treatment clearly impact who visits the doctor and who receives treatment. Lack of such control limits understanding of the factors that differentially influence diagnosis and treatment.

We used a novel modeling strategy borrowing from techniques used in economics used to evaluate effects of health programs with control for demand-based health program placement^{13,14} (garnering James Heckman the Economics Nobel Prize in 2000). We modified Heckman's 2-step modeling strategy¹⁵ to include 3 steps to examine factors associated with each of 3 steps in the continuum from elevated BP (systolic/diastolic BP (SBP/DBP) \geq 140/90 mm Hg) to hypertension diagnosis and subsequent treatment using 18 years of longitudinal data from the population-based China Health and Nutrition Survey ($n = 19,509$ adults seen over 7 exams between 1991 and 2009), accounting for measured and unmeasured confounders (i.e., unobservable factors with differential influence on elevated BP, hypertension diagnosis, and treatment). In addition, we compared these results to separate logistic regression models, with the same measured confounders but without correction for unmeasured confounders.

METHODS

China Health and Nutrition Survey (CHNS)

The CHNS is a longitudinal study of 228 communities within 9 provinces of China. Surveys began in 1989, with subsequent surveys every 2 to 4 years, for a total of 8 rounds of surveys between 1989 and 2009. The CHNS was designed to provide representation of rural, urban, and suburban areas varying substantially in geography, economic development, public resources, and health indicators with a focus on examining changes in dietary behavior, physical activity, and health in the context of urbanization and economic change.¹⁶ It is the only large-scale, longitudinal study of its kind in China. The original 1989 survey used a multistage, random cluster design in 8 provinces (Liaoning, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi, and Guizhou) to select a stratified probability sample. Using this sampling strategy, 2 cities (1 large and 1 small city—usually the provincial capital and a lower income city) and 4 counties (stratified by income: 1 high, 1 low, and 2 middle income counties) were selected. Within cities, 2 urban and 2 suburban communities were selected; within counties, 1 community in the capital city and 3 rural villages were chosen. Twenty households per community were then randomly selected for participation. The study met the standards for the ethical treatment of participants and was approved by the Institutional Review Boards of the University of North Carolina at Chapel Hill and the Institute of Nutrition and Health, Chinese Center for Disease Control and Prevention. More detailed survey procedures have been described elsewhere.¹⁶

Sample

The eligible sample included 19,509 adults aged 18–75 years with SBP and DBP data for at least one of the 1991, 1993, 1997, 2000, 2004, 2006, and 2009 exams (60,289

observations). Of these individuals, 18,926 (56,843 observations [obs]) had complete covariate data on 1 to 7 measurement occasions and were included in the analysis. Subjects were excluded due to pregnancy at time of survey ($n = 105$) or missing hypertension diagnosis ($n = 150$), health insurance ($n = 31$), education ($n = 230$), or household income ($n = 67$). Some individuals had more than one covariate missing at more than one occasion. Relative to the full eligible sample ($n = 19,509$), the analytic sample ($n = 18,926$) was comprised of a greater proportion of males, older individuals, and those of higher education and income, who lived in areas of lower urbanization; individuals missing hypertension diagnosis or treatment information at one or more waves ($n = 1,471$) tended to be female, have higher SBP or DBP, lower attained education or income, and live in communities of lower urbanization.

Outcome measures

At each survey, experienced physicians measured SBP and DBP in triplicate on the right arm using an appropriately sized cuff from a standard mercury sphygmomanometer after 5-minute initial seated rest, with 30-second interval between measurements and mean SBP and DBP calculated. Hypertension was defined as having self-reported doctor diagnosis of hypertension, use of antihypertensive medication, SBP \geq 140 mm Hg, or DBP \geq 90 mm Hg¹⁷ and undiagnosed hypertension as SBP/DBP \geq 140/90 mm Hg without doctor diagnosis of hypertension. Diagnosed hypertension defined as self-reported doctor diagnosis of hypertension without reported hypertension medication use and treated hypertension as self-report of the use of antihypertension medication.

Individual-level covariates

Age, sex, highest attained education, health insurance availability, and type were self-reported at each survey using standard survey procedures. Income was reported at the household and individual levels at each survey and inflated to the 2009 yuan currency for comparability over time.

Community-level covariates

A validated community-level, multidimensional urbanization index¹⁸ comprised of 12 components (transportation infrastructure, health infrastructure, population density, economic activity, sanitation, education, education and income diversity, social services, sanitation infrastructure, housing infrastructure, communications, modern markets, and traditional markets) was derived from household and community survey questions at each wave. The community-level price for 100 pills of capoten, a commonly prescribed Angiotensin-Converting-Enzyme inhibitor, was calculated from prices in community drug stores in 2000 and 2004 and community pharmacies in 2006 and 2009. Provinces were categorized into 3 regions [northern (Liaoning, Heilongjiang), central (Jiangsu, Shandong, Henan), or southern (Guangxi, Guizhou, Hubei, Hunan)].

Analyses

Descriptive analyses were conducted in Stata 12 (Stata Corp, College Station, TX). Selection models were estimated in Absoft Pro Fortran 11.1 (Absoft Development Tools and Languages, Rochester Hills, MI). In descriptive analyses, we examined individual- and community-level characteristics according to survey year and by hypertension status (normotension, undiagnosed, diagnosed, treated) using analysis of variance for continuous variables and chi-squared tests for categorical variables.

We modified Heckman's method to involve joint estimation of 3 equations (rather than 2 in Heckman's original work) and the use of a semi-parametric estimator rather than assuming joint normality as was done in the original Heckman paper. The joint estimation controls for the correlation in unobserved variables that affect the selection and the resulting reduction in bias has been well documented in the literature.^{19,20} We used this 3-step Heckman-type selection model¹² to predict (step 1) elevated BP, (step 2) hypertension diagnosis, and (step 3) hypertension treatment, while accounting for measured and unmeasured confounding at the individual and community levels (including access to community-level health infrastructure) in each of the 3 model steps.^{12,13} Rather than assuming a specific distribution such as multivariate normality for the unobserved heterogeneity (i.e., confounding by unobservable factors) at the community and individual levels, we used a discrete factor approximation with 5 points of support for both distributions as a more robust¹⁹ extension of Heckman and Singer.¹⁵

To deal with selectivity in the sequence of elevated BP, diagnosis, and treatment, each step of the Heckman selection model included exogenous factors (i.e., observed confounders), and an unobserved heterogeneity component (i.e., correction for unmeasured confounders). We did not include statistical mediators, such as smoking, body mass index, and diet since adjusting for mediating variables can introduce bias to coefficient estimates.^{21,22}

The model for elevated BP included: time (years since baseline), age in 2000, squared age in 2000, interactions between time and age, household income and its interaction with time, health insurance and its interaction with time, highest attained education, region, and community-level urbanization covariates: access to communications infrastructure, health infrastructure, modern markets, and traditional markets; quality of housing infrastructure, social services, and transportation infrastructure and sanitation infrastructure; income and education diversity; and population density. The hypertension diagnosis model did not include region and certain community-level variables (housing infrastructure, sanitation infrastructure, presence of modern markets, and presence of traditional markets) that were in the elevated BP model on the basis of differences in predictors of elevated BP and diagnosis (e.g., availability of fast food may be associated with elevated BP, but not with diagnosis of hypertension), and because including different covariates for each step in the Heckman-type selection model aids identification of the system of equations. The hypertension treatment model included all covariates from the hypertension diagnosis model, with the addition of hypertension medication price

(low, medium, high, or missing) under the assumption that community-level medication prices would not be associated with elevated BP or diagnosis, but could be associated with hypertension medication use. We compared results from the 3-step model to 3 separate, traditional logistic regression models (Supplementary Table 1). We call these models naïve results as they controlled for the same measured confounders but did not include control for unmeasured confounding across each step in the continuum from undiagnosed hypertension to hypertension diagnosis to treated hypertension.

We examined model fit statistics to determine whether correction for unobservable factors was necessary (details and results presented in Supplementary Methods and Table).

Presentation of results

Using regression coefficients from the 3-step model system, we predicted: (i) elevated BP, (ii) hypertension diagnosis conditional on having elevated BP, and (iii) hypertension treatment conditional on hypertension diagnosis, according to sex, attained education, and health insurance. While we included the full analytic sample for model estimation, we present simulations for individuals aged 36–75 years in 2009 ($n = 7,135$) to hold observed characteristics of the same individuals constant over time. We used a parametric bootstrapping procedure with 1,000 replications to calculate SEs for predicted probabilities.

RESULTS

Descriptive statistics over time

The crude prevalence of hypertension in adults aged 18–75 years doubled from 13.2% in 1991 to 28.4% in 2009, coupled with an increase in BMI, income, and urbanization (Table 1). By 2009, almost 10% of the sample reported receiving treatment *via* the use of hypertension medication (Table 1). Unadjusted baseline characteristics suggest differences in hypertension status by education level, age, sex, urbanization, and health insurance (Table 2).

Model-based findings

Findings for the first step of the Heckman selection model (elevated BP) suggest higher odds of hypertension for men than women (Table 3). The odds ratios for hypertension were higher for individuals with secondary or greater attained education relative to individuals with primary attained education.

In the second step of the Heckman selection model (hypertension diagnosis among individuals with elevated BP), more women than men were diagnosed with hypertension; more individuals with primary or greater (vs. less than primary) attained education were diagnosed with hypertension (Table 3).

In the third step of the Heckman selection model (hypertension treatment among individuals diagnosed with hypertension), women were more likely than men to be treated for hypertension (Table 3).

Table 1. Analytic sample characteristics of Chinese adults (aged 18–75 years) according to CHNS survey year

	Total <i>n</i> = 56,843	1991 <i>n</i> = 7,612	1997 <i>n</i> = 7,885	2004 <i>n</i> = 8,573	2009 <i>n</i> = 8,748
Women, %	52.0	52.7	50.9	51.9	52.2
Age in years, mean (SD)	44.5 (14.4)	39.9 (14.4)	42.4 (14.4)	46.7 (13.9)	48.6 (13.7)
Household income (in hundreds) (2009 Yuan), (SD)	197.5 (323.4)	58.3 (66.5)	144.5 (127.7)	213.9 (221.8)	415.8 (603.2)
BMI kg/m ² , mean (SD)	22.7 (3.23)	21.7 (2.9)	22.3 (3.1)	23.1 (3.3)	23.4 (3.4)
SBP mm Hg, mean (SD)	119.1 (18.0)	114.0 (17.7)	118.3 (17.6)	121.7 (18.2)	123.6 (18.1)
DBP mm Hg, mean (SD)	77.5 (11.2)	74.1 (11.2)	77.0 (11.0)	78.6 (11.2)	80.4 (11.1)
Hypertension status, ^a %					
Normotension	79.6	86.8	81.0	76.7	71.6
Hypertension	20.4	13.2	19.0	23.3	28.4
Undiagnosed	13.3	8.9	15.1	14.9	16.5
Diagnosed	2.1	1.9	1.3	2.3	2.4
Treated	5.1	2.4	2.6	6.1	9.5
Community-level measures, ^b mean (SD)					
Total urbanization, ^c	57.8 (20.0)	46.8 (16.2)	53.0 (18.2)	62.7 (20.3)	67.1 (19.4)
Population density	5.8 (1.5)	5.8 (1.3)	5.6 (1.5)	5.8 (1.6)	5.9 (1.5)
Communication	5.3 (1.7)	3.9 (1.4)	4.7 (1.3)	5.7 (1.5)	6.8 (1.5)
Educational and economic diversity	4.6 (1.2)	3.8 (1.0)	4.2 (1.0)	4.7 (1.2)	5.4 (1.1)
Health infrastructure	5.6 (2.3)	5.7 (2.0)	5.8 (2.2)	5.3 (2.3)	5.9 (2.6)
Social services	1.7 (2.4)	– ^d	– ^d	2.9 (2.5)	3.6 (3.1)
Transportation infrastructure	5.6 (2.6)	4.9 (3.0)	5.4 (2.6)	5.9 (2.4)	5.9 (2.1)
Proportion of population with health insurance, %	40.1	34.0	26.4	27.5	91.4

Abbreviations: BMI, body mass index; CHNS, China Health and Nutrition Survey; DBP, diastolic blood pressure; SBP, systolic blood pressure.

^aNormotension defined as SBP/DBP <140/90 mm Hg. Undiagnosed hypertension defined as SBP/DBP ≥140/90 mm Hg without doctor diagnosis of hypertension. Diagnosed hypertension defined as self-reported doctor diagnosis of hypertension without reported hypertension medication use. Treated hypertension defined as self-report of the use of hypertension medication.

^bCommunity-level measures are derived from community-level surveys taking place every year in the CHNS. Each measure has a possible range of 0–10, with a higher score indicating higher urbanization.¹⁷

^cUrbanization based upon a 12-component urbanization scale (possible range: 0–120) formed specifically for the CHNS. Omitted from table: economic activity, sanitation, education, traditional markets, modern markets.

^dSocial services data collected from 2004 onward.

Given the complexity of interpreting these findings jointly across the 3 equations, with results conditional on the previous step, we used simulations based on our model coefficients to predict elevated BP, diagnosis, and treatment of hypertension across key covariates.

Our simulations by education (Figure 1) showed higher hypertension prevalence in individuals of lower (vs. higher) attained education (difference of 2% in 1991 that widened to 5% in 2009). However, diagnosis (conditional on having elevated BP) was higher in individuals of higher (vs. lower) education (difference of 7% in 1991 that widened to 10% in 2009). Yet, educational differences in treatment (conditional on having elevated BP and diagnosis of hypertension) narrowed over time in lower (vs. higher) attained education (from 6% in 1991 to 4% in 2009).

Similarly, our model-based simulations suggest a reversal in progression in men vs. women over time (Figure 2).

Men had higher predicted hypertension prevalence than women at all time points, with differences that widened over time (4% in 1991 to 8% in 2009). However, women had higher predicted hypertension diagnosis conditional on having hypertension and higher treatment conditional on diagnosis than men. These differences narrowed over time, from 4% in 1991 to 3% in 2009 for diagnosis and treatment.

Differences in predicted prevalence, diagnosis (conditional on having elevated BP), and treatment (conditional on having elevated BP and diagnosis of hypertension) were consistent in individuals with and without health insurance (Figure 3). These simulations suggest higher treatment among the insured (vs. uninsured), with narrowing differences in treatment (conditional on having elevated BP and diagnosis of hypertension) by insurance status over time (from 7% in 1991 to 2% in 2009).

Table 2. Analytic sample characteristics according to classification of hypertension in Chinese adults aged 18–75 years, CHNS 1991

	Normotension ^a	Undiagnosed hypertension ^a	Diagnosed hypertension ^a	Treated hypertension ^a	P value ^b
N	6,609	677	143	183	
Attained education, %					
<Primary	31.1	49.9	52.4	57.9	<0.0001
Primary	51.6	38.3	42.0 (4.1)	24.6	
Secondary	12.3	7.2	3.5 (1.5)	9.3	
>Secondary	5.0	4.6	2.1 (1.2)	8.2	
Age in years, mean (SD)	37.9 (13.4)	51.9 (14.0)	52.6 (12.8)	58.0 (10.5)	<0.0001
Urbanization, mean (SD) ^c	46.0 (16.2)	50.8 (15.6)	51.7 (15.3)	57.0 (13.4)	<0.0001
Women, %	53.6	41.8	52.4	60.1	<0.0001
Income in hundreds of 2009 Yuan, mean (SD)	58.8 (69.7)	53.7 (37.3)	62.2 (47.2)	56.4 (37.1)	0.23
Health insurance, %	32.4	41.1	42.7	57.9	<0.0001

Abbreviations: ANOVA, analysis of variance; CHNS, China Health and Nutrition Survey.

^aNormotension defined as systolic/diastolic blood pressure <140/90 mm Hg. Undiagnosed hypertension defined as systolic/diastolic blood pressure ≥140/90 mm Hg without doctor diagnosis of hypertension. Diagnosed hypertension defined as self-reported doctor diagnosis of hypertension without reported hypertension medication use. Treated hypertension defined as self-report of the use of hypertension medication.

^bP value for the difference in the explanatory variable across categories of normotension, undiagnosed hypertension, diagnosed hypertension, and treated hypertension using ANOVA tests for continuous variables or chi-squared test for categorical variables.

^cUrbanization based upon a 12-component urbanization scale (possible range: 0–120) formed specifically for the CHNS.

In contrast, naïve, logistic regression models (Table 4) suggest that the lack of control for unmeasured confounders resulted in comparatively stronger differences in treatment.

DISCUSSION

We aimed to determine whether efforts to reduce the burden of hypertension in China would be more effective if directed at hypertension diagnosis or at hypertension treatment. Using our 3-step model, elevated BP increased from 12.6% in 1991 to 36.8% in 2009 for men and from 8.5% in 1991 to 29% in 2009 for women, but hypertension diagnosis remained under 50% for men and women at all time points. While previous studies have examined crude prevalence, diagnosis, and treatment of hypertension in China^{7,8,9,10} and worldwide,^{23–25} few studies have focused on social and demographic characteristics that differentially predict elevated BP, hypertension diagnosis and treatment, accounting for each of the previous steps.^{7,10} We found lower hypertension prevalence in individuals of higher (vs. lower) attained education, but higher diagnosis (conditional on having elevated BP) in individuals of higher (vs. lower) education with narrowing of educational differences in treatment (conditional on having elevated BP and hypertension diagnosis). Treatment of hypertension among individuals with diagnosed hypertension was more equitably distributed (though differences were still present); almost 80% of individuals (across all SES levels) with diagnosed hypertension were treated.

Our model-based findings suggest that despite multiple health care reforms in China over recent decades, notable education and sex disparities in hypertension prevalence and diagnosis remain, and have widened. Thus, hypertension diagnosis remains the major barrier to improving BP control

in China, even after controlling for access to health infrastructure. Furthermore, population-level improvements in hypertension screening may hold a comparatively better advantage than intervening on treatment. In contrast, naïve logistic regression models showed smaller male–female differences in elevated BP, but larger absolute differences in hypertension diagnosis. If our modeling approach is correct, it would mean that, findings from simple logistic regressions would erroneously suggest that education and sex disparities in hypertension diagnosis have decreased or remained constant over time, while differences in treatment persisted. These naïve findings would inappropriately lead to recommendations to target hypertension treatment through availability or price reforms on hypertension treatment.

Previous cross-sectional examinations of hypertension prevalence and diagnosis in China suggest that individuals of low (vs. high) education are more likely to have hypertension and to be diagnosed.^{7,10} Similar to the United States, poorly educated Chinese adults are less able to afford health care and have more complications associated with uncontrolled, chronic disease than individuals of high education.⁵ We found that hypertension diagnosis and treatment among hypertensive individuals was lower in men than women, which is consistent with previous studies in China⁷ and worldwide.^{26,27}

Health insurance was more strongly associated with treatment than diagnosis particularly in early survey years, though model coefficients were not statistically significant. Universal health insurance reforms were implemented in 2006; by 2009, >90% of adults across China (and in the CHNS) were insured, though insurance quality varies greatly.^{5,28} Approximately 80% of individuals diagnosed with hypertension were treated, but less

Table 3. Model coefficients from 3-step Heckman selection model in Chinese adults aged 18–75 years, CHNS^a

Parameter	Hypertension prevalence,		Hypertension diagnosis,		Hypertension treatment,	
	coefficient (SE)	<i>P</i> value ^b	coefficient (SE)	<i>P</i> value ^b	coefficient (SE)	<i>P</i> value ^b
	Obs = 56,843		Obs = 11,615		Obs = 4,073	
Constant	-0.05 (0.26)	0.39	-2.41 (0.33)	<0.0001	-1.93 (0.43)	<0.0001
Time	0.11 (0.01)	<0.0001	0.12 (0.02)	<0.0001	0.11 (0.02)	<0.0001
Age in 2000	9.00 (0.43)	<0.0001	10.13 (1.43)	<0.0001	5.77 (2.38)	0.02
Age in 2000 * time	0.02 (0.03)	0.27	0.08 (0.08)	0.26	0.20 (0.14)	0.15
Age in 2000 ²	3.49 (1.80)	0.06	-2.46 (4.41)	0.34	4.05 (6.94)	0.34
Age in 2000 ² * time	-0.32 (0.13)	0.02	0.71 (0.30)	0.02	-0.69 (0.47)	0.13
Female	-0.58 (0.05)	<0.0001	0.24 (0.08)	<0.01	0.16 (0.10)	0.11
Medium income	-0.14 (0.09)	0.10	0.52 (0.19)	<0.01	-0.44 (0.26)	0.09
High income	-0.02 (0.09)	0.39	0.24 (0.19)	0.18	-0.56 (0.26)	0.04
Health insurance	-0.04 (0.08)	0.35	-0.11 (0.17)	0.32	0.35 (0.23)	0.12
Health insurance * time	0.01 (0.01)	0.08	0.02 (0.01)	0.11	-0.01 (0.02)	0.352
Primary education	-0.04 (0.05)	0.30	0.26 (0.09)	<0.01	0.19 (0.12)	0.112
Secondary education	-0.21 (0.08)	<0.01	0.37 (0.14)	<0.01	0.41 (0.19)	0.042
>Secondary education	-0.38 (0.09)	<0.0001	0.46 (0.15)	<0.01	0.32 (0.20)	0.11
Communication	0.00 (0.02)	0.40	0.10 (0.03)	<0.01	0.13 (0.04)	<0.01
Population density	0.03 (0.02)	0.03	-0.03 (0.03)	0.23	0.00 (0.03)	0.40
Educational and income diversity	0.03 (0.02)	0.18	0.14 (0.03)	0.0001	0.07 (0.05)	0.14
Health infrastructure	0.03 (0.01)	<0.01	0.03 (0.01)	0.07	0.07 (0.02)	<0.01
Housing infrastructure	0.06 (0.01)	<0.0001	–	–	–	–
Traditional markets	-0.01 (0.01)	0.11	–	–	–	–
Social services	-0.01 (0.01)	0.09	0.03 (0.01)	0.02	-0.02 (0.02)	0.26
Transportation infrastructure	0.01 (0.01)	0.18	0.03 (0.01)	0.04	0.04 (0.02)	0.06
Modern markets	0.02 (0.01)	<0.01	–	–	–	–
Sanitation infrastructure	-0.02 (0.01)	0.02	–	–	–	–
Central region	-0.25 (0.05)	<0.0001	–	–	–	–
South region	-0.79 (0.05)	<0.0001	–	–	–	–
Medium medication price ^c	–	–	–	–	0.03 (0.12)	0.39
High medication price ^c	–	–	–	–	0.15 (0.13)	0.21
Missing medication price ^c	–	–	–	–	0.07 (0.21)	0.37
Medium income * time	0.01 (0.01)	0.272	-0.03 (0.01)	0.02	0.02 (0.02)	0.21
High income * time	-0.00 (0.01)	0.372	-0.01 (0.01)	0.35	0.04 (0.02)	0.03

Abbreviations: BP, blood pressure; CHNS, China Health and Nutrition Survey.

^aThree-step Heckman selection model for sequential estimation of hypertension prevalence, hypertension diagnosis, and hypertension treatment, with 5 points of support, nonlinear error terms, and clustering at the individual and community levels.

^b*P* values for *z* test of model coefficient.

^cHypertension medication prices based upon pharmacy and drug-store prices for 100 pills of capoten, an ACE inhibitor, collected in 2000, 2004, 2006, and 2009. An indicator variable for missing hypertension medication price was used for 1991, 1993, and 1997, when hypertension medication prices were not collected.

than 50% of individuals with hypertension were diagnosed. Thus, increasing hypertension diagnosis through social or health insurance programs may hold a comparatively better health advantage than intervening at the treatment stage.

We compared our 3-step findings to 3 naïve separate logistic regressions, a modeling method typically found in the literature. Diagnostic testing indicated that unmeasured confounding was indeed present in the separate logistic models, yielding biased results. Accounting for unmeasured

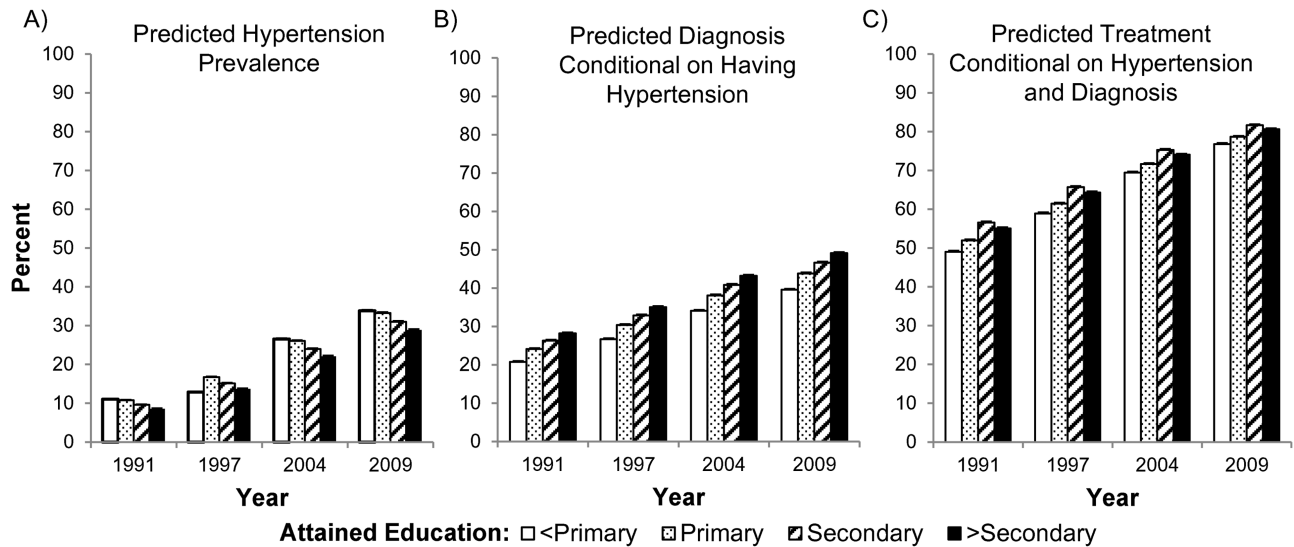


Figure 1. Predicted % from 3-step Heckman selection model according to attained education level from adults aged 18–75 years from China Health and Nutrition Survey 1991–2009. Three model outcomes were (a) hypertension prevalence (systolic/diastolic BP \geq 140/90 mm Hg), (b) hypertension diagnosis (self-reported doctor diagnosis), and (c) hypertension treatment (self-reported medication use). Simulations were based on characteristics of the 2009 sample to hold constant individual characteristics over time. Model for hypertension treatment controlled for age, sex, household income, education level, and community-level: communication, population density, educational and income diversity, health infrastructure, social services, transportation infrastructure, price of hypertension medication, and clustered at the community level. Model for hypertension diagnosis controlled for the same variables, except for price of hypertension medication. Model for hypertension prevalence controlled for the same variables as hypertension diagnosis, with additional control for region, and community level: housing infrastructure, traditional markets, modern markets, and sanitation infrastructure. Abbreviation: BP, blood pressure.

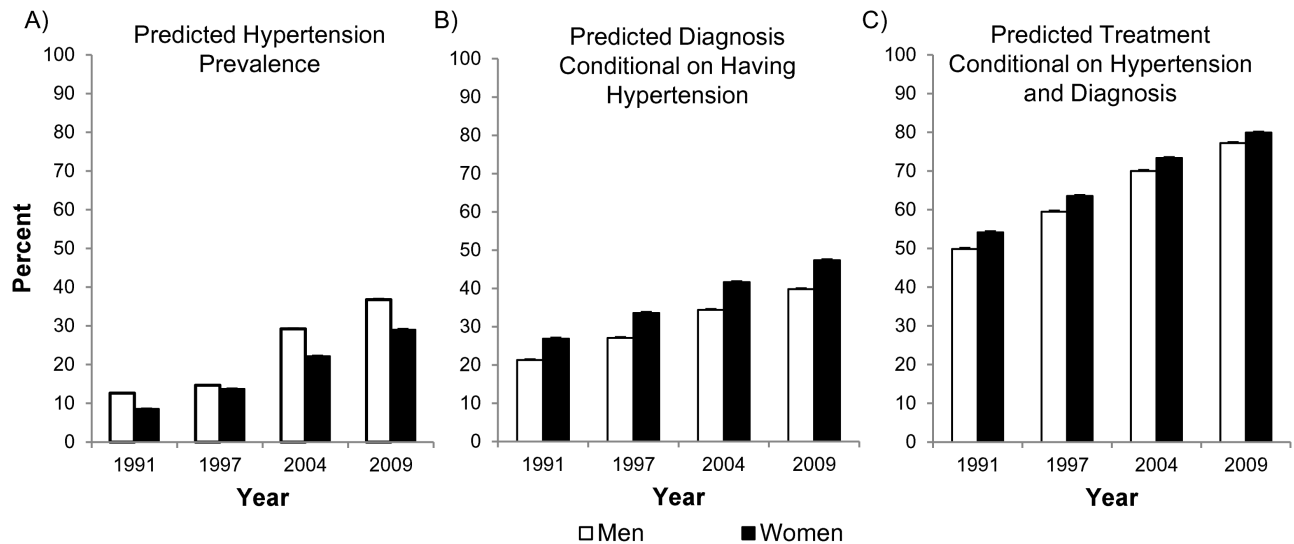


Figure 2. Predicted % from 3-step Heckman selection model according to sex from adults aged 18–75 years from China Health and Nutrition Survey 1991–2009. Three model outcomes were (a) hypertension prevalence (systolic/diastolic BP \geq 140/90 mm Hg), (b) hypertension diagnosis (self-reported doctor diagnosis), and (c) hypertension treatment (self-reported medication use). Simulations were based on characteristics of the 2009 sample to hold constant individual characteristics over time. Model for hypertension treatment controlled for age, sex, household income, education level, and community-level: communication, population density, educational and income diversity, health infrastructure, social services, transportation infrastructure, price of hypertension medication, and clustered at the community level. Model for hypertension diagnosis controlled for the same variables, except for price of hypertension medication. Model for hypertension prevalence controlled for the same variables as hypertension diagnosis, with additional control for region, and community level: housing infrastructure, traditional markets, modern markets, and sanitation infrastructure. Abbreviation: BP, blood pressure.

confounding in our 3-step model improved model fit and showed widening sex and education disparities in hypertension diagnosis. Separate models for the 3 steps amplified differences in treatment, likely because this model did not control for confounding by hypertension diagnosis.

There are some limitations of note. Diagnosis and treatment data were self-reported with misclassification a possibility as we do not know details regarding medication adherence, diagnosis misclassification due to respondents' health literacy, or lack of doctor visits or screening. It is

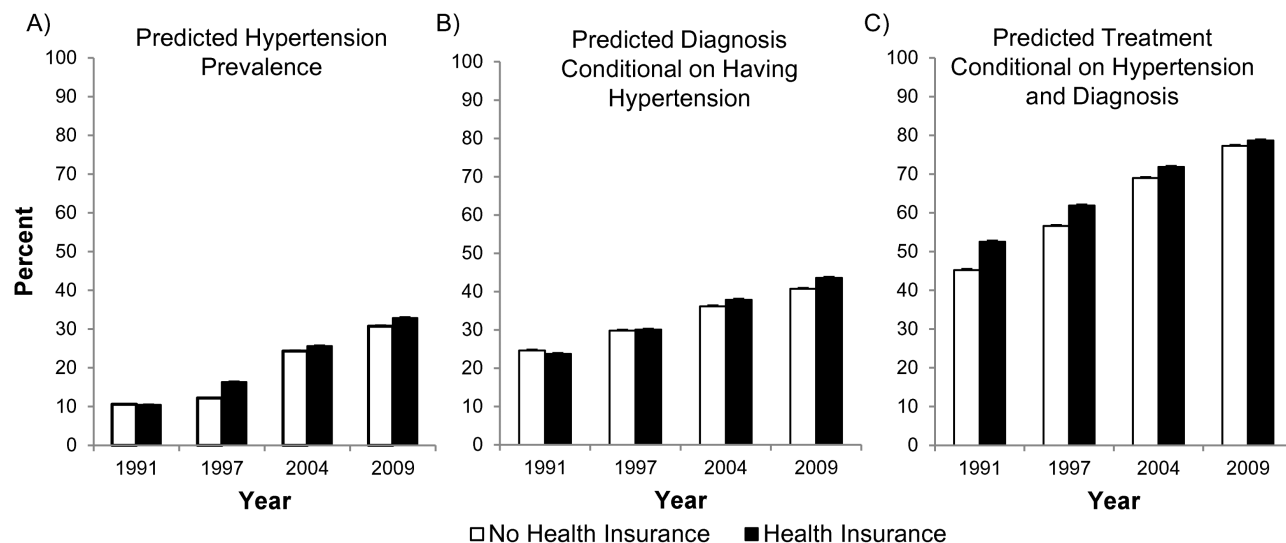


Figure 3. Predicted % from 3-step Heckman selection model according to health insurance status from adults aged 18–75 years from China Health and Nutrition Survey 1991–2009. Three model outcomes were (a) hypertension prevalence (systolic/diastolic BP \geq 140/90 mm Hg), (b) hypertension diagnosis (self-reported doctor diagnosis), and (c) hypertension treatment (self-reported medication use). Simulations were based on characteristics of the 2009 sample to hold constant individual characteristics over time. Model for hypertension treatment controlled for age, sex, household income, education level, and community-level: communication, population density, educational and income diversity, health infrastructure, social services, transportation infrastructure, price of hypertension medication, and clustered at the community level. Model for hypertension diagnosis controlled for the same variables, except for price of hypertension medication. Model for hypertension prevalence controlled for the same variables as hypertension diagnosis, with additional control for region, and community level: housing infrastructure, traditional markets, modern markets, and sanitation infrastructure. Abbreviation: BP, blood pressure.

Table 4. Comparing sign and *P* values of main exposure variables from 3-step Heckman Selection model vs. logistic regression models in Chinese adults aged 18–75 years, CHNS^a

	Hypertension prevalence		Hypertension diagnosis		Hypertension treatment	
	With correction	Without correction	With correction	Without correction	With correction	Without correction
Female	<0.0001 (– ^b)	<0.0001 (–)	<0.01 (+)	<0.0001 (+)	0.11 (+)	<0.01 (+)
Education						
Primary	0.30 (–)	0.17 (–)	<0.01 (+)	0.01 (+)	0.11 (+)	0.15 (+)
Secondary	<0.01 (–)	<0.01 (–)	<0.01 (+)	<0.01 (+)	0.04 (+)	0.09 (+)
>Secondary	<0.0001 (–)	<0.0001 (–)	<0.01 (+)	<0.001 (+)	0.11 (+)	0.13 (+)
Health insurance	0.35 (–)	0.20 (+)	0.32 (–)	0.17 (+)	0.12 (+)	0.03 (+)
Health insurance * time	0.08 (+)	0.40 (+)	0.11 (+)	0.35 (–)	0.36 (–)	0.17 (–)

Abbreviation: CHNS, China Health and Nutrition Survey.

^aThree-step Heckman selection model for sequential estimation of hypertension prevalence, hypertension diagnosis, and hypertension treatment had 5 points of support, nonlinear error terms, and clustering at the individual and household levels. Models additionally controlled for the confounders. The model for hypertension treatment controlled for age, household income, and community-level: communication, population density, educational and income diversity, health infrastructure, social services, transportation infrastructure, price of hypertension medication, and clustered at the community level. Model for hypertension diagnosis controlled for the same variables, except for price of hypertension medication. Model for hypertension prevalence controlled for the same variables as hypertension diagnosis, with additional control for region, and community level: housing infrastructure, traditional markets, modern markets, and sanitation infrastructure.

^bA single plus and minus denotes the sign of the beta coefficient from the regression model.

possible that health insurance may have had a greater association with hypertension treatment because diagnosed individuals are likely to be more actively engaged in the health care system than undiagnosed individuals, although we controlled for factors such as community-level health infrastructure. Future studies to assess external validity of the 3-step modeling approach in other data sets is a next step.

We used a novel application of a well-established econometric modeling strategy to examine how sex, education level, and health insurance are separately associated with hypertension prevalence, diagnosis, and treatment in Chinese adults. Our 3-step model had improved model fit over traditionally used models. Our findings highlight serious barriers to hypertension diagnosis in Chinese adults, particularly among men and individuals of low attained education. Given rising

hypertension rates across rural and urban parts of China, and the great health and economic costs of untreated hypertension, efficient diagnosis, and treatment of hypertension will only increase in importance over the coming years.

SUPPLEMENTARY MATERIAL

Supplementary data are available at *American Journal of Hypertension* online.

ACKNOWLEDGEMENTS

We are grateful to the participants in the China Health and Nutrition Survey. This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (R01-DK104371), the National Heart Lung and Blood Institute (R01-HL108427) and uses data from China Health and Nutrition Survey (CHNS), funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01-HD30880), although no direct support was received from grant for this analysis. We thank the National Institute of Nutrition and Health, China Center for Disease Control and Prevention, the NIH (DK056350 and R01-HD38700), and the NIH Fogarty grant 5 D43 TW009077 for financial support for the CHNS data collection and analysis files from 1989 to 2011 and future surveys, and the China-Japan Friendship Hospital, Ministry of Health for support for CHNS 2009. We also are thankful to general support from the Carolina Population Center (P2CHD050924), the University of North Carolina at Chapel Hill. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DISCLOSURE

The authors declared no conflict of interest.

REFERENCES

- Global Burden of Disease 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2014; 385:117–171.
- Alleyne G, Binagwaho A, Haines A, Jahan S, Nugent R, Rohhani A, Stuckler D; Lancet NCD Action Group. Embedding non-communicable diseases in the post-2015 development agenda. *Lancet* 2013; 381:566–574.
- Yang G, Wang Y, Zeng Y, Gao GF, Liang X, Zhou M, Wan X, Yu S, Jiang Y, Naghavi M, Vos T, Wang H, Lopez AD, Murray CJ. Rapid health transition in China, 1990–2010: findings from the Global Burden of Disease Study 2010. *Lancet* 2013; 381:1987–2015.
- Liu S, Griffiths SM. From economic development to public health improvement: China faces equity challenges. *Public Health* 2011; 125:669–674.
- Gong P, Liang S, Carlton EJ, Jiang Q, Wu J, Wang L, Remais JV. Urbanisation and health in China. *Lancet* 2012; 379:843–852.
- Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. *Lancet* 2005; 365:217–223.
- Wu Y, Huxley R, Li L, Anna V, Xie G, Yao C, Woodward M, Li X, Chalmers J, Gao R, Kong L, Yang X; China NNHS Steering Committee; China NNHS Working Group. Prevalence, awareness, treatment, and control of hypertension in China: data from the China National Nutrition and Health Survey 2002. *Circulation* 2008; 118:2679–2686.
- Gu D, Reynolds K, Wu X, Chen J, Duan X, Muntner P, Huang G, Reynolds RF, Su S, Whelton PK, He J; InterASIA Collaborative Group. The International Collaborative Study of Cardiovascular Disease in ASIA. Prevalence, awareness, treatment, and control of hypertension in China. *Hypertension* 2002; 40:920–927.
- Wang J, Ning X, Yang L, Lu H, Tu J, Jin W, Zhang W, Su TC. Trends of hypertension prevalence, awareness, treatment and control in rural areas of northern China during 1991–2011. *J Hum Hypertens* 2014; 28:25–31.
- Wang H, Zhang X, Zhang J, He Q, Hu R, Wang L, Su D, Xiao Y, Pan J, Ye Z. Factors associated with prevalence, awareness, treatment and control of hypertension among adults in Southern China: a community-based, cross-sectional survey. *PLoS One* 2013; 8:e62469.
- Alcoer L, Cueto L. Hypertension, a health economics perspective. *Ther Adv Cardiovasc Dis* 2008; 2:147–155.
- Dow WH, Norton EC. Choosing between and interpreting the Heckit and two-part models for corner solutions. *Health Serv Outcomes Res Methodol* 2003; 4:5–18.
- Guilkey DK, Hutchinson PL. Overcoming methodological challenges in evaluating health communication campaigns: evidence from rural Bangladesh. *Stud Fam Plann* 2011; 42:93–106.
- Rindfuss RR, Guilkey DK, Morgan SP, Kravdal Ø. Child-care availability and fertility in Norway. *Popul Dev Rev* 2010; 36:725–748.
- Heckman J, Singer B. A method for minimizing the impact of distributional assumptions in econometric-models for duration data. *Econometrica* 1984; 52:271–320.
- Zhang B, Zhai FY, Du SF, Popkin BM. The China Health and Nutrition Survey, 1989–2011. *Obes Rev* 2014; 15(Suppl 1):2–7.
- Whitworth JA; World Health Organization, International Society of Hypertension Writing Group. 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. *J Hypertens* 2003; 21:1983–1992.
- Jones-Smith JC, Popkin BM. Understanding community context and adult health changes in China: development of an urbanicity scale. *Soc Sci Med* 2010; 71:1436–1446.
- Mroz TA. Discrete factor approximations in simultaneous equation models: estimating the impact of a dummy endogenous variable on a continuous outcome. *J Econom* 1999; 92:233–274.
- Guilkey DK, Lance PM. Program impact estimation with binary outcome variables: Monte Carlo results for alternative estimators and empirical examples. In Sickles RC, Horrace WC (eds), *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*. Springer New York: New York, NY, 2014, pp. 5–46.
- Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol* 2013; 42:1511–1519.
- Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* 2009; 20:488–495.
- Egan BM, Zhao Y, Axon RN. US trends in prevalence, awareness, treatment, and control of hypertension, 1988–2008. *JAMA* 2010; 303:2043–2050.
- Kearney PM, Whelton M, Reynolds K, Whelton PK, He J. Worldwide prevalence of hypertension: a systematic review. *J Hypertens* 2004; 22:11–19.
- Devi P, Rao M, Sigamani A, Faruqui A, Jose M, Gupta R, Kerkar P, Jain RK, Joshi R, Chidambaram N, Rao DS, Thanikachalam S, Iyengar SS, Verghese K, Mohan V, Pais P, Xavier D. Prevalence, risk factors and awareness of hypertension in India: a systematic review. *J Hum Hypertens* 2013; 27:281–287.
- World Health Organization. *Global Status Report on Noncommunicable Diseases 2010*. World Health Organization: Geneva, Switzerland, 2011.
- Doumas M, Papademetriou V, Faselis C, Kokkinos P. Gender differences in hypertension: myths and reality. *Curr Hypertens Rep* 2013; 15:321–330.
- Yip WC, Hsiao WC, Chen W, Hu S, Ma J, Maynard A. Early appraisal of China's huge and complex health-care reforms. *Lancet* 2012; 379:833–842.