Supplementary Material

Impact of community based screening for hypertension on blood pressure after two years: regression discontinuity analysis in a national cohort of older adults in China

In this supplementary material to the *BMJ* paper "Impact of community based screening for hypertension on blood pressure after two years: regression discontinuity analysis in a national cohort of older adults in China", we provide additional details related to our study.

A: Data manipulation test

We performed the McCrary test[1] for data manipulation separately for systolic and diastolic blood pressure. We used the optimal bandwidth. We used the integrated mean-squared error (IMSE)-optimal quantile-spaced method for bin width selection.[2] Log Difference in Frequency Bins was -0.11 (standard error=0.15) for systolic blood pressure and -0.19 (standard error=0.16) for diastolic blood pressure. We failed to reject the null hypothesis of a smooth density across the threshold, providing further support for the assumption that the field workers did not manipulate the running variable.

B: Continuity of participants' observed characteristics around the threshold

We tested whether participants' observed characteristics are continuous near the threshold following the method of Lee et. al. (2004)[3] and Imbens and Lemieux (2008).[4] As is shown in the main paper, most of the variables are balanced around the threshold. Two variables—years of education and sugar consumption—are slightly imbalanced when using diastolic blood pressure

(BP) as running variable. After narrowing the bandwidth (from 6 mmHg to 4 mmHg), these two variables are balanced, as is shown in Table S1. For the imbalanced variables when using systolic BP as running variable, we further used these variables as outcomes. Table S2 shows there is no significant jump of these variables at the threshold. Thus, these results show that all the covariates are continuous near the threshold.

Table S1. Covariate means above and below the diastolic blood pressure discontinuity threshold within a narrower bandwidth

Dependent variables	**Baseline diastolic** BP below 90	**Baseline diastolic** BP above 90	**Difference in means** (p-value)
Years of education	2.30	2.78	0.48 (0.105)
Sugar consumption	1.84	1.86	0.02 (0.624)

Table S2. Using the control variables as outcomes to estimate the continuity of observed characteristics around the threshold

Dependent variables	**Systolic BP as running variable**	
	Coefficient (95% CI)	P value
Age in years	1.43 (-1.47 to 4.34)	0.333
Urban	-0.19 (-0.45 to 0.06)	0.136
Exercise	-0.17 (-0.37 to 0.03)	0.099
Smoke	0.06 (-0.15 to 0.28)	0.562
Salted preserved vegetable consumption	0.02 (-0.30 to 0.35)	0.895

C: Heterogeneity among groups

We estimated whether there is heterogeneity among different groups regarding the impact of blood pressure screening in 2011-2012 on blood pressure in 2014 by adding a dummy variable indicating this group and an interaction term.

Specifically, we fitted the following regression models within the optimal bandwidth to implement the local linear regression.[4]

For systolic BP:

$$Y_i = \alpha_0 + \alpha_1 Above_i + \alpha_2(SBP_i - 140) + \alpha_3 Above_i(SBP_i - 140) + \alpha_4 G_i + \alpha_5 G_i Above_i + \nu_i$$

For diastolic BP:

$$Z_i = \beta_0 + \beta_1 Above_i + \beta_2(DBP_i - 90) + \beta_3 Above_i(DBP_i - 90) + \beta_4 G_i + \beta_5 G_i Above_i + \varepsilon_i$$

where Y_i is a measure of individual i's systolic blood pressure and Z_i is a measure of individual i's diastolic blood pressure in 2014; $Above_i$ is an indicator variable equal to 1 for people who have systolic blood pressure of at least 140 mm Hg (in the first equation) or have diastolic blood pressure of at least 90 mm Hg (in the second equation) in 2011-12; SBP_i is a measure of individual i's systolic blood pressure in 2011-12 and DBP_i is a measure of individual i's diastolic blood pressure in 2011-12; G_i is the main impact of group. α_5 and β_5 shows whether there is heterogeneity among groups. Table S3 shows the coefficients of α_5 and β_5, respectively for different groups and we did not find significant heterogeneity among groups.

Table S3. Regression discontinuity estimates of the impact of household blood pressure screening in 2011-2012 on blood pressure in 2014

Interaction terms	**Impact of screening on systolic BP using baseline systolic BP as running variable (mm Hg)** (95% CI)	**Impact of screening on diastolic BP using baseline diastolic BP as running variable (mm Hg)** (95% CI)
Age in years	0.07 (-0.21 to 0.35)	-0.008 (-0.17 to 0.16)
Male	2.07 (-3.38 to 7.51)	-0.83 (-4.38 to 2.71)
Urban	2.29 (-9.80 to 14.38)	-5.52 (-12.22 to 1.18)
Married	1.61 (-3.84 to 7.06)	0.09 (-3.43 to 3.61)
Number of children	0.22 (-1.09 to 1.52)	-0.27 (-1.20 to 0.67)
Years of education	0.07 (-0.65 to 0.81)	0.10 (-0.41 to 0.61)
Self-reported relative economic status	-0.25 (-4.20 to 3.71)	-1.80 (-4.38 to 0.78)

Note: The sample is restricted to systolic BP or diastolic BP within optimal bandwidth of the hypertension diagnose threshold at 140/90 mm/Hg. In all regressions, we used a triangular kernel function, which gives more weight to

observations closer to the threshold. Each cell in the table represents the coefficient from a separate regression. Sources: China Longitudinal Healthy Longevity Survey (2011-2015).

D: Robustness test

The results in Table S4 show that while the impact sizes fluctuate as the bandwidth changes, they are still consistent with our baseline estimates, indicating a strong and precise causal impact of community based screening on systolic BP two years later and a weaker and insignificant impact on diastolic BP two years later. Table S5 shows that our main results remain essentially the same when we adjusted for age using broad age categories (65-79, 80-89, 90-99, and 100+ years of age) instead of continuous functional forms to control for age.

Table S4. Using the hypertension diagnostic threshold to estimate the impact of screening on BP using different bandwidth

Bandwidth selection	**Impact of screening on systolic BP using baseline systolic BP as running variable** (mm Hg) (95% CI)	**Impact of screening on diastolic BP using baseline diastolic BP as running variable** (mm Hg) (95% CI)
50% of optimal bandwidth	-8.53 (-17.17 to 0.11)	-1.03 (-7.50 to 5.44)
P value	0.053	0.755
80% of optimal bandwidth	-7.89 (-14.28 to -1.50)	-1.79 (-6.74 to 3.16)
P value	0.015	0.478
Optimal bandwidth	-6.26 (-11.23 to -1.29)	-2.22 (-5.94 to 1.50)
P value	0.014	0.241
120% of optimal bandwidth	-5.25 (-9.82 to -0.68)	-2.43 (-5.82 to 0.96)
P value	0.024	0.161
150% of optimal bandwidth	-4.07 (-8.08 to -0.06)	-2.79 (-5.63 to 0.06)
P value	0.047	0.055

Note: The sample is restricted to systolic BP or diastolic BP within different bandwidth ranges from 50% to 150% of the optimal bandwidth the hypertension diagnose threshold at 140/90 mm/Hg for robustness check. In all regressions, we used a triangular kernel function, which gives more weight to observations closer to the threshold. Each cell in the table represents the coefficient from a separate regression. Models do not control for covariates.
Sources: Chinese Longitudinal Healthy Longevity Survey (2011-2015).

Table S5. Regression discontinuity estimates (95% confidence intervals) of the impact of community based hypertension screening in 2011-12 on blood pressure in 2014 (Chinese Longitudinal Healthy Longevity Survey)

Impact of screening on blood pressure (mm Hg)*	**Only with age group covariate**	**With demographic covariates**	**With demographic and social covariates**	**With demographic, social, and behavioural covariates**
Systolic blood pressure (local linear)	-6.4 (-11.4 to -1.4)	-7.6 (-12.8 to -2.4)	-8.2 (-13.4 to -2.9)	-8.5 (-13.7 to 3.2)
P value	0.01	0.004	0.002	0.002
Diastolic blood pressure (local linear)	-2.5 (-6.2 to 1.2)	-2.4 (-6.4 to 1.7)	-2.6 (-6.6 to 1.4)	-2.4 (-6.5 to 1.6)
P value	0.19	0.25	0.21	0.25

Note: The sample comprises people within the optimal bandwidth of the threshold of 140 or 90 mm Hg. Each cell represents the coefficient from a separate regression. In all regressions, we used a triangular kernel function, which gives more weight to observations closer to the threshold. Models in the second column do not control for any covariates; models in the third column additionally include demographic covariates (age, sex, urban residence, marital status, and number of children); models in the fourth column additionally include social covariates (education and self reported relative economic status); models in the fifth column additionally include behavioural covariates (such as exercise, smoking, alcohol consumption, and diet).

*Baseline systolic blood pressure or diastolic blood pressure used as assignment variable.

E: Placebo test

In the placebo test, we analyzed the impact of hypertension screening in 2011-2012 wave on systolic and diastolic BP in 2008 instead of BP in 2014. Table S5 shows that there was no impact in the placebo test, further supporting our finding of a strong causal impact of community based blood pressure screening.

Table S6. Placebo test using systolic and diastolic BP in 2008 as outcomes

Bandwidth selection	**Impact of screening on systolic blood pressure using baseline systolic blood pressure as running variable (mm Hg)**	**Impact of screening on diastolic blood pressure using baseline diastolic blood pressure as running variable (mm Hg)**
50% of optimal bandwidth	2.43 (1.43)	0.12 (2.56)
80% of optimal bandwidth	1.82 (1.57)	0.59 (1.76)
Optimal bandwidth	1.88 (1.43)	1.11 (1.55)
120% of optimal bandwidth	1.58 (1.27)	1.25 (1.38)
150% of optimal bandwidth	1.48 (1.11)	1.31 (1.15)

Note: The sample is restricted to systolic blood pressure or diastolic blood pressure within different bandwidth ranges from 50% to 150% of the optimal bandwidth the hypertension diagnose threshold at 140/90 mmHg for robustness check. Each cell in the table represents the coefficient from a separate regression, and standard errors are reported in parentheses. Models do not control for covariates. *Significant at 10%, **significant at 5%, ***significant at 1%. Sources: Chinese Longitudinal Healthy Longevity Survey (2011-2015).

References

1. McCrary J. Manipulation of the running variable in the regression discontinuity design: A density test. *J Econometrics* 2008;142(2):698-714.

2. Jacob R, Zhu P, Somers M-A, et al. A Practical Guide to Regression Discontinuity. *MDRC* 2012

3. Lee DS, Moretti E, Butler MJ. Do voters affect or elect policies? Evidence from the US House. *The Quarterly Journal of Economics* 2004;119(3):807-59.

4. Imbens GW, Lemieux T. Regression discontinuity designs: A guide to practice. *J Econometrics* 2008;142(2):615-35.