TI 2009-016/3
Tinbergen Institute Discussion Paper

The Health Penalty Of China's Rapid Urbanization

E. Van de Poel[a]
O. O'Donnell[b]
E. Van Doorslaer[a,c]

[a] *Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands;*
[b] *University of Macedonia, Thessaloniki, Greece;*
[c] *Tinbergen Institute.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

THE HEALTH PENALTY OF CHINA'S RAPID URBANIZATION

E Van de Poel[a*], O O'Donnell[b], E Van Doorslaer[a,c]

[a] *Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands. Tel: +31 10 408 1502, Fax: +31 10 408 91 41, E-mail: vandepoel@ese.eur.nl*

*(*corresponding author)*

[b] *University of Macedonia, 156 Egnatia Street, GR-540 06 Thessaloniki, Greece. E-mail: ood@uom.gr*

[c] *Tinbergen Institute, The Netherlands. E-mail: vandoorslaer@ese.eur.nl*

Abstract

Rapid urbanization could have positive and negative health effects, such that the net impact on population health is not obvious. It is, however, highly pertinent to the human welfare consequences of development. This paper uses community and individual level longitudinal data from the *China Health and Nutrition Survey* to estimate the net health impact of China's unprecedented urbanization. We construct an index of urbanicity from a broad set of community characteristics and define urbanization in terms of movements across the distribution of this index. We use difference-in-differences estimators to identify the treatment effect of urbanization on the self-assessed health of individuals. The results reveal important, and robust, negative causal effects of urbanization on health. Urbanization increases the probability of reporting fair or poor health by 5 to 15 percentage points, with a greater degree of urbanization having larger health effects. While people in more urbanized areas are, on average, in better health than their rural counterparts, the process of urbanization is damaging to health. Our measure of self-assessed health is highly correlated with subsequent mortality and the causal harmful effect of urbanization on health is confirmed using more objective (but also more specific) health indicators, such as physical impairments, disease symptoms and hypertension.

Keywords: urbanization, health, China, treatment effects, difference-in-differences

JEL: I12, I18, O18

Introduction

Urbanization and economic development are intimately related (Williamson, 1988). There is no better example of this than China in recent decades, where a remarkable rate of economic growth has been accompanied by a process of urbanization that is unprecedented in human history, both in scale and in speed. The proportion of the Chinese population living in urban areas increased from only 20% in 1980, to 27% in 1990, and reached 43% in 2005 (NBS, 2006; World Bank 2006). By the middle of this century, the country's urbanization rate has been forecast to reach 75% (Yusuf and Saich, 2008). In the space of just a few decades, China will complete the urbanization process that lasted hundreds of years in the West. The non-economic consequences of such rapid urbanization, including those for health, as well as more obviously for the environment, will determine the true welfare effects of development and the extent to which it is sustainable. The consequences for population health are not obvious. On the one hand, urban living offers improved access to modern medicine (particularly in China) and gains in income that can be invested in health. On the other, the health of city dwellers is threatened by air pollution, more sedentary and possibly more stressful work, social detachment, and Western, high-fat diets. This paper uses panel data from China covering the period 1991-2004 to estimate the net health impact of urbanization.

On average, health outcomes are found to be better in urban parts of the developing world (Van de Poel *et al*, 2007; Zimmer *et al*, 2007). This apparent urban health advantage contrasts with the historical evidence of urban populations suffering poorer health in Western Europe prior to and during its period of industrialization (Rosen, 1958; Woods, 1985, 2003). The most likely explanation for this difference in the urban-rural health disparity over time and space is the marked decline in the prevalence of infectious diseases, in low-income as well as high-income countries (Riley, 2005), prompted, in large part, by public health measures built on the germ theory of disease (Preston, 1975, 1980; Cutler and Miller, 2005) and the introduction of effective medicines, antibiotics and vaccinations (Davis, 1956; Cutler *et al*, 2006; Soares, 2007). In the past, the opportunities for material gain offered by cities had to be weighed against the dangers of infection. Today, while cities of the developing world continue to pose risks to health, the immediate threat to life through infection has receded. However, the overcrowding and pollution that accompany urbanization, particularly on the scale and speed with which it has occurred in China, may impose an urban health penalty. During the last decades, China's environment has deteriorated significantly as rapid urbanization and industrialization generate enormous volumes of air and water pollutants (World Bank, 1997; Wang and Smith, 2000; Brajer and Mead, 2003).[1] As other developing countries, most notably India, China relies very heavily on coal as a source of energy, with the result that levels of airborne pollution in Chinese cities are many times greater than those found in most US and European cities (Pandey *et al,* 2006).[2] A World Health Organization study has estimated that there are 300,000 premature deaths per year in Chinese cities attributable to outdoor air pollution (Cohen *et al,* 2004).[3]

[1] But the health effects of pollution from urbanization are not necessarily limited to urban areas. Rural areas rely more on unsafe water sources and are also affected by pollutants coming from urban areas (World Health Organization, 2001).

[2] Across Chinese cities each with a population of at least 100,000, the weighted average of estimated airborne particulate matter concentrations (PM10) is 87 $\mu g/m^3$ (Pandey *et al*, 2006). The equivalent figure for US cities is 25. It is 13 in Sweden, 15 in France, 19 in the UK and 22 in Germany. The WHO study (Cohen *et al,* 2004) predictions of premature deaths due to outdoor air pollution are based on these estimates.

[3] As pointed out in footnote 1, the health effects of pollution in rural areas should not be overlooked. The WHO study estimated that 420,000 deaths per year in all of China are caused by indoor air pollution created by the burning of solid fuels, which rural households rely on for 90% of their energy needs (Zhang and Smith, 2007).

Urbanization brings social and economic changes that can raise risk factors associated with chronic disease. Urban populations of middle-income countries are experiencing a rapid nutritional transition towards Western-style diets, dominated by more processed foods and a high fat content (Popkin, 2001; Popkin and Du, 2003). Urbanization inevitably implies a shift in work patterns from physical, agricultural labor towards more sedentary occupations (Monda *et al,* 2007). In China, it is claimed that these transitions have contributed to stark increases in the prevalence of obesity and hypertension (Liu *et al,* 2004; Wang, Mi *et al,* 2007; Weng *et al,* 2007).

But urbanization clearly has positive, as well as negative, consequences for population health. Closer proximity to health care facilities, particularly hospitals, equipped with modern technology and staffed by highly trained doctors is an obvious advantage of living in towns and cities. In China, urban-rural differences in access to health care, and in health insurance cover, have been marked and widening in recent decades (Liu *et al*, 1999). Access to schools and to health education initiatives confer a strong advantage on urban areas in the field of preventative health care. Urban populations can also use higher incomes to invest in health through health care, a nutritious diet or by reducing strenuous work effort (Moore *et al*, 2003).

In this paper, we estimate the *net* effect of urbanization on health using longitudinal data from the *China Health and Nutrition Survey* (CHNS). Besides being a household panel, this survey also collects data on the characteristics of communities, making it possible to identify what happens to individuals' health when the environment in which they live becomes more urbanized. This identification strategy avoids the selection biases that arise from comparisons between the health of urban and rural populations, or from monitoring the health of migrants, which is difficult or impossible in any case with most panel data.

A dichotomous urban-rural classification, most often done on the basis of population density, does not capture the variation in living and health conditions across areas at different stages of urbanization (McDade and Adair, 2001; Vlahov and Galea, 2002; Champion and Hugo, 2004; Dahly and Adair, 2007). In addition, there is a practical problem in that the categorization of an area as 'urban' or 'rural' is often fixed over waves of a longitudinal survey, as it is in the CHNS, and so this categorization does not capture the urbanization taking place. In order to identify communities at various stages of the urbanization process, and to track changes over time in the degree of urbanicity within each community, we exploit the CHNS data on the characteristics of communities to construct an index of urbanicity, which depends, for example, on population size, the proportion of the workforce engaged in agriculture, proximity to health and educational facilities, and the presence of paved roads, shops, restaurants, etc. This index has been shown to outperform the simple urban-rural classification that comes with the CHNS in detecting different degrees of urbanicity, measuring changes in urbanicity over time and being less prone to misclassification bias (Van de Poel *et al*, 2008). We define urbanization in terms of movement of a community up the distribution of this urbanicity index. We adopt a treatment effects framework and define treatment as movement from the bottom to the top half of the distribution of the index. To investigate whether the health impact varies with the degree of urbanization, we also define ordinal treatments in terms of movements up tertiles of the distribution and by standard deviation increases in the index. We use difference-in-differences estimators made robust to unobserved individual heterogeneity by exploiting the panel nature of the data (Blundell and Costa Dias, 2000; Wooldridge, 2002).

The main health outcome used in the paper is self-assessed health (SAH), reported on a four-point scale from *excellent* to *poor*. This general measure of adult health has repeatedly been shown to be highly predictive of mortality, even conditional on physiological measures of health (Idler and Benyamini, 1997). We show that SAH predicts mortality in the CHNS and demonstrate that it is highly correlated with more specific health outcomes such as obesity, hypertension, physical impairments and symptoms of illness. We also estimate the impact of urbanization on these narrower, but more objective, measures of health status.

To our knowledge, this is the first paper to estimate the causal effect of urbanization on health from longitudinal data on both individuals and communities. These data allow us to identify the effect of urbanization by comparing the health transitions of individuals living in areas that experience rapid transformations to an urban environment with those living in areas that remain rural. We find important, and robust, negative effects of urbanization on health. Urbanization increases the probability of reporting fair or poor health by 5 to 15 percentage points, with a greater degree of urbanization having larger health effects. While people in more urbanized areas are, on average, in better health than their rural counterparts, the process of urbanization is damaging to health. Urbanization raises the probability of suffering from physical impairments, disease symptoms and hypertension, but there is no significant impact on obesity or under-nutrition.

In the remainder, we first present the CHNS data, and explain construction of the urbanicity index. This is followed by an explanation of our identification strategy, estimation methods and the various definitions of urbanization used. In the fourth section, we first present the main results for the impact of urbanization on SAH, and then check their robustness, before examining the impact on other health outcomes. The concluding section provides an interpretation of the implications of the study and acknowledges its limitations.

Data

Sample

We use the *China Health and Nutrition Survey* (CHNS) panel data from 1991, 1993, 1997, 2000 and 2004[4]. The CHNS is a large scale longitudinal survey conducted in 9 provinces in China: Liaoning, Shandong, Jiangsu, Henan, Heilongjiang, Hubei, Hunan, Guangxi and Guizhou. Although the CHNS is not representative of all China, these provinces vary substantially in terms of geography, urbanization and economic development. While the CHNS provinces span some of the relatively more urbanized regions of China, Beijing and Shanghai, the two largest megacities in China, are not covered. Urbanization rates vary considerably within each province. There have been some changes in the composition of the CHNS sample across time. Liaoning province was added in 1997 when Heilongjiang Province was unable to participate. Heilongjiang returned to the study in 2000 (and Liaoning remained as well). New households in original communities were added to replace households no longer participating in the study in 1997 and in 2000. In 1997, new communities in original provinces were added to replace sites no longer participating in the survey.[5]

The CHNS collects information on a wide range of individual, household and community characteristics. A community, which is the primary sampling unit (PSU), is a government-designated administrative district. The community interview is held with the community head for questions related to public facilities and infrastructure, and with community health workers for questions related to health care provision. In total, there are about 200 communities in each wave (see Appendix – Table A1); an average of about 20

---

[4] In the 1989 survey, health and nutritional data were only collected from preschoolers and adults aged 20-45.
[5] More information on this survey can be found at the Carolina Population Center CHNS website: http://www.cpc.unc.edu/projects/china.

communities in each province. On average, there are about 15 households and a little less than 50 individuals interviewed within each community.

There are a total of 47418 person-wave observations across the five waves of the survey. After dropping observations with missing information on any of the individual or household level variables used in the regression analysis, or missing community characteristics used in construction of the urbanicity index, we are left with 31333 person-wave observations. 19% of respondents are only interviewed once in the survey, 25% twice, 26% three times, 20% four times and 9% are interviewed in all waves. The panel dynamics and attrition rates are shown in Table 1. There is quite a high attrition rate, which is partly because Heilongjiang province was not interviewed in 1997. Individuals reporting poor health are more likely to drop out of the sample between the last two waves. We test for attrition bias in the analysis below.

Measurement of urbanization

In order to track the increasing urbanization that is taking place in communities across the survey waves, we construct an urbanicity index using factor analysis on a broad set of characteristics from the CHNS community level data pooled across all survey waves (Van de Poel *et al,* 2008). The urbanicity index captures information on population size, land use in the community, transportation facilities, economic activity and public services (see Appendix-Table A1). We have checked the validity of the urbanicity index in various dimensions and found that the factor loading of the community variables have intuitive signs; the time trend in the index indeed reveals increasing urbanization; the index correlates with a subjective classification of communities as urban, suburb, town or rural, that is available within the CHNS and with income (Van de Poel *et al*, 2008).[6] Although the urbanicity index is highly correlated with the administratively defined urban-rural classification available in the CHNS, it provides considerable additional information by displaying substantial variation within each category of the dichotomy.

Since the index is estimated from data on all communities in all waves, an increase in its value for a single community across time represents that community becoming more urbanized, in terms of reduced reliance on agriculture and increased availability of community infrastructure, services, etc, relative to the average over all communities within the whole period from 1990 to 2004.[7] If, within each wave, communities were homogeneous with respect to urbanicity, then the index would increase for all communities over time reflecting the general process of urbanization experienced commonly by all. Of course, in reality, communities differ greatly in their characteristics at each point in time and so changes in the index indicate not only the general process of urbanization but also the specific one experienced by a community relative to all others.

Since the index is constructed from factor analysis, it has no meaningful unit of measurement. We therefore identify the urbanization of a community through changes in its rank position in the (whole period) distribution of the index, conditioning on those that start off in the bottom part of the distribution. That is, we compare communities that move from the bottom to the top half of the distribution with those that remain in the bottom half. In 1991, 60% of the sample of communities were below the (all wave) median of the urbanicity index, while by 2004 61% of the sample was above the median. To investigate a dose-

[6] This subjective classification is not very useful for our purposes as there is not much variation across the survey waves. Van de Poel *et al* (2008) found that cities and towns have the highest average urbanicity index, followed by suburban and rural areas. This means that suburban areas do not come second on the continuum from city to rural.

[7] Similar, a decrease in the index points to deterioration in community infrastructure, meaning 'de-urbanization' has taken place. However, small changes in the index can also reflect reporting errors in the community survey. We return to this issue at the end of the Results section.

response effect, we also compare those that remain in the bottom third of the distribution with those that move from there to the middle and to the top third. The percentage of communities in the top (middle) third of the whole-period urbanicity distribution increased from 24% (30%) in the 1991 to 43% (35%) in 2004. To estimate the health effects of further urbanization in communities that are among the most urbanized even at the beginning of the panel, we also define treatment in terms of standard deviation increases in the index without conditioning on the initial degree of urbanicity.

Measurement of health

We use self-assessed health (SAH) as the principal measure of health. Respondents aged 18 years or over were asked to rate their health compared to that of people their own age on a four-point scale consisting of *excellent, good, fair* and *poor*. In the analysis, we mainly use a binary indicator of reporting *fair* or *poor* health (*poorhealth*), but in some specifications we exploit the information contained in the full ordinal scale.

SAH is a popular instrument for health status that is very widely used in research based on large scale household surveys. This is not just due to its availability, but because it provides a measure of general health status and numerous studies have demonstrated that it contains information on health over and above that which can be measured objectively by physiology-based instruments (Idler and Benyamini, 1997). Two potential limitations of the measure are, first, that its very generality means that it cannot reveal the dimensions of health that are most affected by a treatment, such as urbanization, and, second, that any heterogeneity in the reporting of health that is correlated with the treatment will bias the estimated effect. In the present context, reporting heterogeneity would affect our results if individuals living in communities that urbanize were to change their health expectations and therefore revise their SAH evaluation. To address both issues, we make use of the following more objective, but narrower, measures of health: mortality; obesity (Body Mass Index (BMI)>30); underweight (BMI<18.5); measured hypertension; reported physical impairments (goiter or angular stomatitis, loss of use of one or both arms or legs, blindness in one or both eyes); and, reported symptoms experienced in the four weeks preceding the survey (fever, headache, rash, diarrhea, joint pain, heart problems or others).

Table 2 shows the means of these more objective health indicators. Except for risk factors for chronic conditions, such as obesity and hypertension, most of the ill health indicators have a very low prevalence rate. Therefore, we create a binary variable that equals one if the respondent reported to suffer from at least one of the physical impairments. Also we use a binary variable to indicate whether the respondent reported any of the symptoms in the four weeks preceding the survey.

Table 3 confirms that SAH is correlated with each of the more objective measures of health. The first four columns show marginal effects from probit models explaining the probability of reporting *fair* or *poor* health. All of the more objective indicators of ill-health are significantly related with an increased probability of reporting *fair* or *poor* health, indicating that the latter binary measure captures at least some of the information contained in these more specific measures. The last three columns of Table 3 show marginal effects on the probability of dying by the subsequent wave. These results show that reporting *fair* or *poor* health at time t is predictive of mortality by the subsequent wave (column 5), increasing the baseline probability of dying by about one third, and that this predictive power remains after controlling for the set of more objective health indicators (column 6). This demonstrates that not only is the reporting *poor* or *fair* health strongly correlated with the other health indicators; but that it contains additional information relevant to predicting mortality. The last column of Table 3 illustrates that the marginal effect of reporting *fair* or *poor* health on mortality is an average of a smaller effect of *fair* health on the probability of dying by the next wave (0.004) and a much larger impact of *poor* health (0.04).

Control covariates

To identify the health effect of urbanization, we control for a set of individual and household level characteristics including demographics (age, sex, marital status, household size), socioeconomic status (education, income[8]) and household living conditions (availability of a flush toilet, use of solid fuels within the dwelling, water from a water plant and the presence of excreta around the household dwelling). Although the latter living conditions can also be correlated with a community's level of urbanization, we leave these variables out of the urbanicity index and include them separately in the models because they are not solely determined on the community level, but also by households' decisions. The exact definitions of all these variables are given in Table 4.

Item non-response is only substantial for the urbanicity index (24%) and household income (10%). The high proportion of missing information on the urbanicity index is due to the fact that it is constructed from a set of community variables, and so a missing value for any community characteristic causes the index to be missing for all individuals in that community.[9]

Table 5 shows summary statistics of the individual and household level health determinants across all 5 waves of the CHNS. The trends illustrate the rising (average)

---

[8] Household income is calculated by summing all market earnings across the household and then adding the total value of all other non-market goods and services produced within that household (see Liu *et al,* 2008, Fig. 1). Total household income is then deflated using a year/province/urban-rural specific consumer price index that was developed for use with the CHNS, and divided by the (square root of the) total number of household members to obtain real average household income per capita (Liu *et al*, 2008).

[9] Note that the community characteristics included in the urbanicity index have already been (partly) selected on the basis of their high response levels.

incomes in China in the period 1991-2004. Also household living conditions (water, sanitation, heating) seem to have improved substantially. The distribution of the sample across the provinces has remained quite stable, with Heilongjang entering the survey only in 1997. The last rows of Table 5 clearly illustrate the rapid urbanization taking place in China, with the urbanicity index rising from -0.27 in 1991 to 0.38 in 2004.

Identification strategy and estimation

As explained in the previous section, urbanization is defined in terms of movement of a community up the distribution of the urbanicity index, either from the bottom to the top half, or from the bottom to higher tertiles, and by standard deviation increases in the index. We identify the health impact of such urbanization by using difference-in-differences (DID) methods to compare the changes in health of those living in communities that experience urbanization with those that do not.

Model and estimation

We begin by restricting attention to individuals living in communities that are not urbanized at the beginning of the survey period, defined as those in the bottom half, or bottom third, of the distribution of the urbanicity index. A DID estimator of the treatment effect of urbanization is then obtained from the following logit model applied to the binary measure of SAH (and each of the other health outcomes examined) (Wooldridge, 2002; Blundell *et al*, 2004; Puhani, 2008; Böckerman and Ilmakunnas, 2009):

$$\begin{cases} y_{igt} = 1 \text{ if } \mathrm{y}^*_{\mathrm{igt}} > 0 \\ y^*_{igt} = \lambda_t \beta_1 + \alpha_g \beta_2 + x_{gt}\beta_3 + z_{igt}\beta_4 + \delta_{ig}\beta_5 + \varepsilon_{igt} \end{cases} \tag{1}$$

where i indexes individuals, g indexes treatment (urbanization) groups, defined at the level of the community, and t indexes time. y_{igt} equals one if the individual reports to have *fair* or *poor* health at time t. The model includes a full set of time dummies λ_t, which capture trends in reported health that are common across all individuals, and a set of treatment group dummies α_g, which capture time-invariant differences between those individuals living in communities that at some time experience a defined degree of urbanization and those that do not. The time varying group dummies, x_{gt}, equal one if the individual is exposed to a defined degree of urbanization at time t. Since we restrict the sample to those in the bottom part of the distribution of the urbanicity index at the beginning of the panel, these dummies are zero for all individuals at their first observation. The estimate of the average treatment effect of urbanization on the probability of experiencing *fair* or *poor* health is given by the marginal effect of these dummies. Further, we control for individual covariates z_{igt} (see Table 4) and a full set of both community and province dummies δ_{ig}.[10]

Although time-invariant differences between treatment and control communities are taken into account, this DID estimator does not exploit the panel nature of the data and so is potentially rendered inconsistent by any individual level unobserved heterogeneity that is correlated with any of the right-hand-side variables in (1). We deal with this by applying the conditional logit estimator to a model like (1), but including a fixed unobservable individual level effect and, consequently, no time invariant regressors. This comes at the cost of smaller sample size, as the fixed effects logit model only uses those observations for which there is variation in the dependent variable.

[10] In order to avoid the introduction of other indices to denote communities and provinces, we define δ_{ig} to be a set of dummies that for a given treatment group g, which indicates whether or not urbanization is ever experienced, varies across individuals according to the precise community and province in which they are located.

With a third estimator, we exploit more of the information in the ordinal SAH variable by taking the approach of Ferrer-i-Carbonell and Frijters (2004), who have shown that an ordered logit model with fixed effects can be estimated as a fixed effects logit model, where the ordered data are collapsed to binary data and the model allows individual-specific thresholds.[11] This involves creating a binary health indicator ($worsehealth$) that equals one if the individual reports worse health at time t than the average he/she reports across all waves and then using this as the dependent variable in a fixed effects variant of (1) estimated by conditional logit (Böckerman and Ilmakunnas, 2009). In the remainder of the paper, we will refer to this as the fixed effects ordinal logit.

Using Verbeek and Nijman's (1992) test, we found some evidence of attrition bias in the simple logit model.[12] However, once fixed effects are taken into account, attrition can only induce inconsistency when selection is related to the idiosyncratic errors. We tested this by adding the lagged selection indicator to the fixed effects logit model and the fixed effects ordered logit model (estimated on the total panel), and doing a t-test for the significance of the selection indicator (Jones *et al*, 2006).[13] The null of no effect was not rejected in both models (p-value=0.781 and 0.199 respectively), indicating that our fixed effects estimators are not biased by attrition, providing further reason for focusing on them.

Throughout, standard errors are corrected for clustering at the individual (and so any higher) level.

[11] We have also estimated ordered probit models on the ordinal SAH variable and these results confirmed the ones with the binary health indicator.
[12] This involves testing the significance of a count variable of the number of waves that are observed for the individual in the model explaining *poorhealth*. Under the null hypothesis, the error is uncorrelated with attrition for all t, and so attrition in the previous time period should not be significant in the equation at time t.
[13] Note that this method loses the first time period for all observations.

Definition of urbanization

In a first instance, we define the treatment of urbanization as a community moving from below the median (across all waves) of the urbanicity index to above it. We only use those individuals living in communities that fall below the median of the urbanicity index when they are first interviewed. It is important to emphasize that this median is defined on the sample of communities *pooled across all waves*; which means that in principle every community could start off below the median and end up above it. In reality, at each wave, some communities will have crossed the median, and other will not. In this setting, model (1) will consist of only one treatment group dummy α_g, equal to one in every wave if the individual's community ever rises above the median, and only one treatment dummy x_{gt}, which is unity only in the periods when the individual's community is above the median. It is possible that communities experience a drop in their urbanicity index, which could cause them to be above the median in one wave and fall below it in the next. We keep these observations in the sample, and hereby treat urbanization as potentially reversible. At the end of the Results section, we will return to this issue.

We can also use model (1), and its variants that take account of fixed effects, to investigate whether the health effects vary with the intensity of urbanization by defining treatment indicators that distinguish between smaller and larger movements up the distribution of the urbanicity index. We consider the sample of individuals whose communities start off in the lowest third of the urbanicity index and define two treatments: a move to the middle third of the urbanicity index by any subsequent wave and a move to the upper third of the index. This model has two time invariant group dummies in α_g and two time varying group dummies in x_{gt}, and the marginal effects of the latter are the estimated treatment effects of the two intensities of urbanization.[14]

Finally, in order to investigate the health effects of increased urbanization from any level, and not only from originally non-urban environments, we estimate the effects of varying magnitudes of increase in the urbanicity index from one wave to the next. We examine increases of i) 0.25-0.5 standard deviations (sd), ii) 0.5-1 sd, iii) 1-1.5 sd and iv) more than 1.5 sd between waves. Note that we are not restricting the starting level of urbanicity to any particular interval. The reference category is therefore communities that experience an increase in the urbanicity index smaller than 0.25 of a standard deviation (or a decrease). It should also be noted that with this definition, the treatment dummy is only switched on in the wave in which the change occurs. In subsequent periods it is turned off, unless an increase of the same magnitude is repeated. Therefore, the treatment effects estimated with this approach will reflect only the short term health impact of increased urbanization, unlike with the other approaches which identify the health effect that materializes over the whole period in which a community is exposed to a higher degree of urbanicity.

Results

Effects on self-assessed health

We first look at the health effect of a jump from below to above the median of the urbanicity index. After deleting those observations that start off in the upper half of the distribution, we are left with 17864 observations, of which 43% move to the upper half.

[14] Note that the treatment effect of first moving from the lowest to the middle third and then to the upper third is the same as moving to the upper third directly. We could not relax this assumption, because there are too few communities that actually jump from the lowest to the upper third from one wave to another.

Table 6 shows marginal effects obtained from the logit, fixed effects logit and fixed effects ordinal logit estimators. Sample sizes in the fixed effects models are substantially smaller because they only use observations that show variation in the dependent variable.[15]

All three models indicate a positive and significant treatment effect, indicating that urbanization increases the probability of reporting poorer health. The magnitude of the effect is about 5 to 6 percentage points, an increase of almost one-fifth in the baseline probability of reporting *fair* or *poor* health for those not originally living in urban environments. The estimates from the fixed effects logit (second column) indicate that urbanization raises the probability of reporting *fair* or *poor* health (6.5%) by slightly more than having excreta around the household dwelling (6%) or using solid fuels indoors (5%), and a little less than not obtaining water coming from a waterplant (10%) or not having a flush toilet (8%). Note that the treatment effect in the fifth column refers to the effect of urbanization on the ordinal SAH variable, and is therefore not directly comparable to the effects in the previous columns for the binary health variable. This marginal effect of 0.054 should be interpreted as the increase due to urbanization in the probability of an individual reporting worse health than he/she did on average across the panel.

The marginal effect of the *treatment group* dummy is negative and significant in the logit model (first column), indicating that those individuals that do experience urbanization are on average in better health than those who do not. This is consistent with the better average health outcomes that are usually found in more urban areas (Van de Poel *et al*, 2007; Zimmer *et al*, 2007). The combination of the positive effect of the time-varying *treatment* dummy and the negative effect of the time invariant *treatment group* dummy indicates that people living in areas that eventually become urbanized are originally in better health than

[15] The model using *worsehealth* as dependent variable exploits more of the variation in SAH and therefore uses more observations.

their counterparts living in areas that do not become urbanized, but the process of urbanization is itself harmful to health.

It is interesting that there appears to be an increasing trend in the probability to report poor health in China during the period 1991-2004. Model (1) imposes the restriction that urbanization has the same effect in every year, an assumption that may, to an extent, be justified by the fact that our treatment is defined in terms of crossing the median of the index computed from the data pooled across all waves. So, in terms of the index, moving from below to above the median in 1993 is not necessarily different from doing so in 2004. But, for a given value of the index, the degree and nature of urbanization may differ over time. To allow for this, we included interactions between the treatment variable and the wave dummies, but these were never found to be significant.

The estimates show the expected correlations of health with individual and household level determinants. Reporting *fair* or *poor* health is increasing with age, and decreasing with income and education. The education effect is not significant in the models including individual fixed effects, most likely due to its limited variation across time. Married individuals and females are more likely to report *poor* or *fair* health. As noted above, all of the household living conditions variables are significant in the expected directions.

By controlling for income and living conditions, we may have taken out any indirect effect that urbanization has on health through these factors. To investigate whether this is the case, we re-estimated the fixed effects logit model without these controls. As can be seen from the first column of Table 7, dropping income reduces the magnitude of the treatment effect of urbanization slightly (from 0.065 to 0.057), indicating there is a small, positive indirect effect from urbanization through income to health. Leaving the household living conditions variables out has no impact on the estimate (column 3). Finally, without control for income and living conditions (column 5), the estimated marginal effect of urbanization falls only marginally from 0.065 to 0.061. These results suggest there is a small indirect positive effect of urbanization on health operating through increasing household income, but not through household living conditions, which only very slightly offsets the direct negative effect.

Effects of varying intensities of urbanization

We now examine whether the health effect varies with the intensity of urbanization. 13409 individuals live in communities that start off in the lowest third of the distribution of the urbanicity index, 66% move to the middle third of the urbanicity index sometime in the period 1991-2004, and 12% to the upper third. Results are presented in Table 8 for the same three estimators (as in Table 6). Note that sample sizes are smaller as compared to Table 6, because the sample is restricted to those communities that start off in the lowest third (not lowest half) of the urbanicity index distribution. The results indicate that the treatment effect of moving from the lowest to the middle third of the urbanicity index is small and insignificant. However, moving from the bottom to the upper third of the index significantly increases the probability of reporting *fair* or *poor* health by about 6 percentage points in the logit model and 8 points in the fixed effects logit, which represents an increase of about one third in the baseline probability. The marginal effect estimated from the fixed effects ordered logit model implies that moving from the lower to the upper third of the index raises the probability of individuals reporting worse health than their average across survey waves by 0.12.

Next, we look at the estimated health effects of standard deviation changes in the urbanicity index. Because we use changes, the first observation is lost for each individual.

19% of the sample experiences an increase in the urbanicity index of 0.25-0.5 standard deviations, 18% an increase of 0.5-1 sd, 6% an increase of 1-1.5 sd and 4% an increase of more than 1.5 sd. Table 9 shows the treatment effects of these different magnitudes of urbanization.[16] Individuals living in communities that undergo very small increases in urbanization (0.25-0.5 sd increase in the index) actually have a slightly reduced probability of reporting *fair* or *poor* health relative to those that experience no increase (or decrease) in urbanization. But larger increases in urbanization cause deterioration in reported health, with the probability to report *fair/poor* health rising by as much as 15 percentage points for those experiencing an increase of more than 1.5 standard deviations in the urbanicity index. It should be kept in mind that these are short run effects in the sense that they materialize in the period immediately following the increased urbanization. Note that the magnitude of the change in the index is – as would be expected – negatively correlated with its initial value. So, consistent with Table 8, these results indicate that it is individuals originally living in more rural settings that undergo the most rapid urbanization experience the greatest deterioration in health.

Effects on other health outcomes

To check whether the negative health effects of urbanization reported in the previous sub-sections are simply attributable to changes in health expectations that accompany urbanization and to obtain more insight into which aspects of health are most affected by increasing urbanization, we now turn to estimates of the impact of urbanization on a set of more objective and specific health outcomes. Treatment of urbanization is defined as moving from the lower to the upper half of the distribution of the urbanicity index (as in

[16] Estimates are presented only for the fixed effects models since the fact that individuals can belong to several treatment groups makes definition of the *treatment group* dummies rather complicated for the simple logit. In any case, the fixed effects estimators are preferred.

Table 6). Logit and fixed effects logit estimates of the treatment effects are presented in Table 10. Sample sizes for the fixed effects models are much smaller, as these require some variation across time in the dependent variable, which is considerably smaller than in the SAH variables. The results reveal that urbanization increases the probability of suffering from hypertension (although the effect decreases and loses significance in the fixed effects logit), physical impairments and ill-health symptoms, but has no significant impact on under- and over-nutrition.[17] This suggests that the health impact of urbanization does not operate through obesity, as a result of changes in diet and lifestyle, and there is only limited evidence of an effect through a cardiovascular disease risk factor, such as hypertension. Much more important are the effects on physical impairments and symptoms of illness and disease.[18] From the fixed effects logit models, we estimate that urbanization almost doubles the baseline probability of suffering from physical impairments, and increases the baseline probability of suffering from ill-health symptoms by more than half. While the impact on symptoms may, in part, be due to changes in reporting behavior, this is unlikely to be true for physical impairments, which refer to losses of (use of) arms, legs and sight, suggesting that the effect of urbanization on SAH is not solely reflecting a change in individuals' health expectations as their environment becomes more urbanized. Urbanization is also associated with an increased probability of dying, although the effect is not significant, which is perhaps not surprising given the low incidence of death.

[17] Estimating a fixed effects model on mortality did not prove useful, because of the small proportion of people dying and the fact that individuals drop out of the sample once they die.

[18] We also tried excluding goiter/angular stomatitis from the list of physical impairments, as this is quite a different condition than the loss of (use of) arms, legs and eyesight. This did not significantly change the treatment effect of urbanization. Goiter/angular stomatitis has been related to iodine deficiency, but also other factors such as contamination of water have been shown to play an important role (Kotwal *et al*, 2006).

Sensitivity to making urbanization irreversible

The urbanicity index is constructed using factor analysis on a broad set of community characteristics. Changes in the index therefore reflect actual increases or decreases in the presence or availability of community facilities and infrastructure. The strong increasing trend of the urbanicity index across the CHNS survey waves reflects the huge urbanization taking place in China. However, for some communities the index decreases from one wave to the next. These decreases are generally quite small and much less frequent than the increases in the index, and —in the context of China's urbanization— are more likely to reflect reporting errors in the recording of community characteristics rather than actual 'de-urbanization'. To test whether our results are influenced by these potential errors, we replicated the analysis excluding these negative-change observations. In the case of the first definition of urbanization, i.e. crossing the median of the index, we excluded observations from communities that had returned to the lower half of the index distribution, after having moved to the upper half in the previous wave (3% of the sample). With this restriction, the treatment of urbanization becomes irreversible, in the sense that once communities move to the upper half of the distribution they remain there. The treatment effects of urbanization on SAH based on this definition of 'irreversible treatment' and the restricted sample are presented in the Appendix – Table A2. The treatment effect remains positive and is significant for all but the fixed effects ordinal logit. Using the fixed effects logit, the estimated impact of urbanization on the probability of reporting *fair* or *poor* health falls from 0.065 with reversible treatment and the full sample to 0.053 with irreversible treatment and the restricted sample.

Conclusion

Urbanization is an important component of economic development. Indeed, it is difficult to imagine development occurring without a process of urbanization. The health consequences of urbanization not only represent a potentially important effect of development on human welfare, but may also act as a constraint on its sustainability. This paper investigates the net health effect of the tremendous urbanization taking place in China.

To identify communities at various stages of the urbanization process, and to track urbanization over time, we derive an urbanicity index from a broad set of community characteristics available in the CHNS. This, in combination with individual level panel data provides a rich source of variation from which to identify the health impact of urbanization. The results reveal substantial and significant negative effects of urbanization on health, with the probability of reporting *poor* or *fair* health increasing by 5 to 6.5 percentage points, an increase of almost one fifth in the baseline probability, when communities rise from the bottom to the top half of the distribution of urbanicity. This is comparable to the effects of household level living conditions such as excreta surrounding the household dwelling, use of solid fuels indoors, absence of a flush toilet and not obtaining water from a water plant. We find a small offsetting indirect effect of urbanization on health through income, but no indirect effects through household living conditions.

Larger degrees of urbanization have stronger health effects. Moving from the lowest to the top third of the distribution of urbanicity increases the probability of reporting *fair* or *poor* health by 6 to 8 percentage points, an increase of about a third in the baseline probability. An increase of more than 1.5 standard deviations in the urbanicity index is predicted to have severe and immediate adverse health effects, increasing the probability of reporting *fair*/*poor* health by 0.15. Our results confirm that people in urban areas are on average in better health than those in more rural areas, but the process of urbanization causes negative health effects.

While our panel estimators are robust to any time-invariant heterogeneity across individuals in the way they report their health, we cannot rule out the possibility that our results reflect across time variation in the reporting of health in response to the experience of urbanization. For example, people who experience urbanization, and are awakened to the potential of medical treatment for example, might raise their health expectations and therefore become more likely to report *fair* or *poor* health, given the same objective health. Deaton (2007) has found that, conditional on national income, recent economic growth makes people unhappier. If this phenomenon is present, our estimates reflect not only changes in objective medical conditions that respond to urbanization, but also the health consequences of the dissatisfaction individuals may derive from a changing environment. This is still a meaningful and relevant finding with respect to evaluation of the development process. But our results do not appear to derive only from an impact of urbanization on health expectations. Our SAH variable is a good predictor of mortality and correlates well with other more objective health outcomes such as hypertension, obesity, under nutrition, physical impairments and ill-health symptoms. The power of SAH to predict mortality remains after controlling for these more objective outcomes, indicating that it provides additional health information. Moreover, urbanization has a significant positive impact on the probability of suffering hypertension, physical impairments and symptoms of illness and disease. Moving from the bottom to top half of the distribution of urbanicity almost doubled the baseline probability of suffering from physical impairments and increased the probability of reporting any symptoms by about half.

In sum, we find that the Chinese are paying a health penalty for the tremendous urbanization they are experiencing, with larger urbanization causing worse health effects. This is a new and rather unexpected finding, as one typically associates urban populations with better health. Indeed, our analysis also found better average health in more urban areas. But given our finding that urbanization comes with negative net health consequences, it is questionable whether this urban health advantage will be sustained. To our knowledge, the net causal health effect of urbanization has gone unstudied, most likely because data were not available to measure changes in urbanization. Application of a composite index of urbanicity to panel data has allowed us to define various concepts of urbanization. The limitation of using such an index to identify the health effects of urbanization is that it is difficult to pinpoint which specific aspects of urban life have positive consequences for population health, and which are harmful to health. On the positive side, the closer proximity to health care, health insurance, health education, and economic opportunities are likely to benefit health (Liu *et al,* 1999). But on the other hand, rapid and uncontrolled urbanization is also associated with pollution, overcrowding, social isolation, changes in dietary and physical activity patterns, and inadequate service capacity for providing drinking water, sanitation and waste disposal, which will penalize population health (Popkin, 2001; World Health Organization, 2001; Moore *et al*, 2003). Our analysis suggests that currently in China these negative aspects dominate the positive ones. Given the importance of cities in national and global economies, and the inevitability of increasing urbanization in China, it is of utmost importance to turn this effect around and foster sustainable and healthy cities.

Acknowledgements

The data for this paper were supplied by the China Health and Nutrition Survey, funded by NIH (R01-HD30880, DK056350, and R01-HD38700) and the Carolina Population Center and the Chinese CD. The authors are grateful to the Institute for Housing and Urban Development Studies in Rotterdam for funding the project on "Urbanization, Health and Health Inequality", from which this paper derives. Also many thanks to Pilar Garcia Gomez and Robert Sparrow for useful comments on an earlier draft; and to Alex Fu, Gordon Liu, Lars Osberg and Kuan Xu for information on data issues.

References

Blundell R, Costa Dias M. Evaluation methods for non-experimental data. Fiscal Studies 2000; 21; 4; 427-468.

Blundell R, Costa Dias M, Meghir C, Van Reenen J. Evaluating the employment impact of a mandatory job search program. Journal of the European Economic Association 2004; 2; 569–606.

Böckerman P, Ilmakunnas P. Unemployment and self-assessed health: Evidence from panel data. Health Economics 2009; 18; 2; 161-179.

Brajer V, Mead R. Blue skies in Beijing? Looking at the Olympic Effect. The Journal of Environment Development 2003; 12; 239-263.

Champion AG, Hugo G. New forms of urbanization: Beyond the urban– rural dichotomy. Aldershot, Hants, England, Burlington, VT: Ashgate; 2004.

Cohen AJ, Anderson HR, Ostro B, Pandey KD, Krzyzanowski M, Künzil N, *et al*. Urban air pollution. In: Ezzati M, Rodgers A, Lopez A, Murry C (Eds), Comparative

Quantification of Health Risks, vol 2. Geneva: World Health Organization; 2004. p. 1353–1433.

Cutler D, Miller G. The role of public health improvements in health advances: The twentieth century United States. Demography 2005; 42; 1-22.

Cutler D, Deaton A, LLeras-Muney A. The determinants of mortality. The Journal of Economic Perspectives 2006; 20; 3; 97-120.

Dahly DL, Adair L. Quantifying the urban environment: a scale measure of urbanicity outperforms the urban-rural dichotomy. Social Science and Medicine 2007; 64; 1407-1419.

Davis K. The amazing decline of mortality in underdeveloped areas. The American Economic Review 1956; 46; 2; 305-318.

Deaton A. Income, aging, health and wellbeing around the world: Evidence from the Gallup World Poll. National Bureau of Economic Research working paper n°13317; Princeton University; 2007.

Ferrer-i-Carbonell A, Frijters P. How important is methodology for the estimates of the determinants of happiness? Economic Journal 2004; 114; 641-659.

Idler E, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. Journal of Health and Social Behavior 1997; 38; 1; 21-37.

Jones A, Koolman X, Rice N. Health-related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability-weighted estimators in non-linear models. Journal of the Royal Statistical Society 2006: Series A; 169; 543-569.

Kotwal A, Priya R, Qadeer I. Goiter and other iodine deficiency disorders: a systematic review of epidemiological studies to deconstruct the complex web. Archives of Medical Research 2006; 38; 1; 1-14.

Liu L, Ikeda K, Chen M, Yin W, Miushima S, Miki T, Nara Y, Yamori Y. Obesity, emerging risk in China: trend of increasing prevalence of obesity and its association with hypertension and hypercholesterolsemia among the Chinese. Clinical and Experimental Pharmacology and Physiology 2004; 31; S8-S10.

Liu GG, Dow WH, Fu AZ, Akin J, Lance P. Income productivity in China: on the role of health. Journal of Health Economics 2008; 27; 27-44.

Liu Y, Hsiao WC, Eggleston K. Equity in health and health care: the Chinese experience. Social Science and Medicine 1999; 49; 10; 1349-1356.

McDade TW, Adair LS. Defining the "urban" in urbanization and health: a factor analysis approach. Social science and Medicine 2001; 53; 1; 55-70.

Monda KL, Gordon-Larsen P, Stevens J, Popkin B. China's transition: The effect of rapid urbanization on adult occupational activity. Social Science and Medicine 2007; 64; 4; 858-870.

Moore M, Gould P, Keary S. Global urbanization and impact on health. International Journal of Hygiene and Environmental Health 2003; 206; 4-5; 269-278.

National Bureau of Statistics China. 2006. Available at http://www.stats.gov.cn/english/

Pandey KD, Wheeler D, Ostro B, Deichmann U, Hamilton K, Bolt K. 2006. Ambient particulate matter concentrations in residential and pollution hotspot areas of world cities: New Estimates based on the Global Model of Ambient Particulates (GMAPS), The World Bank Development Economics Research Group and the Environment

Department Working Paper; The World Bank; Washington DC. Available at: http://go.worldbank.org/3RDFO7T6M0

Popkin B. The nutrition transition and obesity in the developing world. The Journal of Nutrition 2001; 131; 871S-873S.

Popkin BM, Du S. Dynamics of the nutrition transition toward the animal foods sector in China and its implications: a worried perspective. The Journal of Nutrition 2003; 33; 3898S-3906.

Preston SH. The changing relation between mortality and level of economic development. Population Studies 1975; 29; 2; 231-248.

Preston SH. Causes and consequences of mortality declines in less developed countries during the 20th century. In Easterlin RA (Ed), Population and economic change in developing countries. Chicago: University of Chicago Press for National Bureau of Economic Research; 1980.

Puhani PA. The treatment effect, the cross difference, and the interaction term in nonlinear "difference-in-differences" models. IZA Working Paper n°3478; 2008.

Riley JC. The timing and pace of health transitions around the world. Population and Development Review 2005; 31; 4; 741-764.

Rosen G. A history of public health. Expanded Edition, Baltimore: Johns Hopkins University Press; 1958.

Soares RR. On the determinants of mortality reductions in the developing world. Population and Development Review 2007; 33; 2; 247-287.

Van de Poel E, O'Donnell O, Van Doorslaer E. Are urban children really healthier? Evidence from 47 developing countries. Social Science and Medicine 2007; 65; 1986–2003.

Van de Poel E, O'Donnell O, Van Doorslaer E. Urbanization and the spread of diseases of affluence in China. HEDG working paper WP 08/25; 2008.

Verbeek M, Nijman TE. Testing for selectivity bias in panel data models. International Economic Review 1992; 33; 681–703.

Vlahov D, Galea S. Urbanization, urbanicity and health. Journal of Urban Health: Bulletin of the New York Academy of Medicine 2002; 79; 4; Supplement 1.

Wang Y, Mi J, Shan XY, Wang QJ, Ge KY. Is China facing an obesity epidemic and the consequences? The trends in obesity and chronic disease in China. International Journal of Obesity 2007; 31; 177-188

Wang X, Smith KR. Near-term benefits of greenhouse gas reductions: health impacts in China. Environmental Science and Technology 2000; 33; 18; 3056–3061.

Weng X, Liu Y, Ma J, Wang W, Yang G, Caballero B. An urban–rural comparison of the prevalence of the metabolic syndrome in Eastern China. Public Health Nutrition 2007; 10; 131-136.

Williamson JG. Chapter 11: Migration and urbanization. In: Chenery H, Srinivasan TN (Eds), Handbook of Development Economics, vol 1. Elsevier; 1988. p. 425-465.

Woods R. The effects of population redistribution on the level of mortality in nineteenth-century England and Wales.The Journal of Economic History 1985; 45; 3; 645-651.

Woods R. Urban-rural mortality differentials: an unresolved debate. Population and Development Review 2003; 29; 1; 29-46.

Wooldridge JM. Econometric Analysis of Cross Section and Panel Data. Cambridge: MIT Press; 2002.

World Bank. Clean Water, Blue Skies: China's environment in the new Millenium. Washington DC; 1997. Available at:

http://siteresources.worldbank.org/INTEAPREGTOPENVIRONMENT/Resources/Clear_Water_Blue_Skies.pdf

World Bank. World Development Indicators. Washington DC; 2006.

World Health Organization. Environment and People's Health in China. Geneva; 2001. Available at http://www.wpro.who.int/NR/rdonlyres/FD5E0957-DC76-41F2-B207-21113406AE55/0/CHNEnvironmentalHealth.pdf

Yusuf F, Saich T. China urbanizes: consequences, strategies and policies. Washington DC: World Bank; 2008.

Zhang J, Smith KR. Household air pollution from coal and biomass fuels in China: Measurements, health impacts, and interventions. Environmental Health Perspectives 2007; 15; 6; 848-855.

Zimmer Z, Kaneda T, Spess L. An examination of urban versus rural mortality in China using community and individual data. Journal of Gerontology Series B: Psychological Sciences and Social Sciences 2007; 62; S349-357.

Tables and Figures

	attrition						raw drop out rate by self-assesed health category				
wave	# individuals	drop outs	rejoiners	survival rate	raw drop out rate (%)	net drop out rate (%)	excellent at t-1	good at t-1	fair at t-1	poor at t-1	later-joiners
1991	6685										
1993	3489	3196	0	0.52	0.48	0.48	0.52	0.46	0.47	0.56	3950
1997	2482	1961	954	0.37	0.56	0.29	0.48	0.57	0.57	0.58	5201
2000	2050	1309	877	0.31	0.53	0.17	0.54	0.52	0.55	0.51	2350
2004	2559	658	1167	0.38	0.32	-0.25	0.30	0.31	0.32	0.42	2567

Table 1: Attrition in the Chinese Health and Nutrition Survey.
Notes: The survival rate is the percentage of original sample members remaining at wave t. The drop-out rate is the difference in observations between waves $t-1$ and t relative to the number of observations at $t-1$. The raw drop-out rate excludes rejoiners, while the net drop-out rate includes them. Note that this table only considers these individuals present in the 1991 wave; late-joiners are presented in the last column.

Description of variables (1/0)	Mean
BMI>30	0.025
BMI<18.5	0.079
diagnosed hypertension: average of three systolic blood pressure measurements (at time of survey) was≥140mm Hg and/or average diastolic blood pressure was ≥90mm Hg and/or respondent was taking medication to lower blood pressure	0.185
physical impairments:	
goiter/angular stomatitis	0.010
loss of one arm or the use of 1 arm	0.002
loss of both arms or use of both arms	0.001
loss of one leg or the use of 1 leg	0.003
loss of both legs or use of both legs	0.001
blindness in one eye	0.002
blindness in both eyes	0.001
suffering from any of the above impairments	0.076
symptoms experienced in 4 weeks preceding the survey:	
fever, sore throat, cough	0.044
headache, dizziness	0.037
rash, dermatitis	0.003
diarrhea, stomachache	0.020
joint pain, muscle pain	0.026
heart disease/chest pain	0.009
other symptoms	0.020
suffering from any of the above symptoms	0.106
whether respondent dies by subsequent wave	0.021
Observations	31333

Table 2: Description and means (proportions) of ill health indicators.

	marginal effect on the probability of reporting fair or poor health				marginal effect on the probability of dying by next wave		
poorhealth (SAH=fair or poor)					0.007***	0.005***	
SAH=good							0.002
SAH=fair							0.004**
SAH=poor							0.037***
BMI>30	0.044**					-0.002	
BMI<18.5	0.090***					0.007***	
hypertension		0.051***				0.005***	
suffering from any impairments			0.069***			0.005**	
suffering from any ill-health symptoms				0.283***		0.005***	
Observations	29664	29707	31333	31001	31333	29598	31333

Table 3: Correlation between SAH and more objective health measures. Marginal effects from probit regression.
Notes: Models also include covariates as described in Table 4 and wave dummies. Standard errors are adjusted for clustering on individuals.
*** significant at 10%; ** significant at 5%; *** significant at 1%**

	variable	description
demographics	age	age (years)
	age squared	age squared
	male	whether respondent is male (1-0)
	married	whether respondent is married (1-0)
	size	number of household members
socioeconomic status	<u>edno</u>	whether respondent has had no education (1-0)
	edprim	whether respondent's highest education is primary education (1-0)
	edmid	whether respondent's highest education is secondary education (1-0)
	edhigh	whether respondent's highest education is higher education (1-0)
	log income	logarithm of household income (in Chinese Yuan)
household living conditions	flush	whether household has flush toilet (1-0)
	excreta	whether there is some or much excreta around the dwelling (1-0)
	waterplant	whether household has access to water that comes from a waterplant (1-0)
	fuel	whether household uses solid fuels within dwelling (1-0)

Table 4: Description of explanatory variables.
Notes: Underscored variables are used as reference category in regression models.

Variable	1991	1993	1997	2000	2004
age	41.16	41.59	42.06	44.37	47.90
age squared	1936.06	1962.75	2020.35	2190.70	2529.20
male	0.47	0.48	0.50	0.48	0.48
married	0.78	0.78	0.75	0.80	0.83
size	3.98	3.98	4.14	3.44	3.24
edno	0.36	0.32	0.28	0.23	0.21
edprim	0.20	0.22	0.23	0.23	0.23
edmid	0.28	0.29	0.31	0.32	0.31
edhigh	0.16	0.17	0.18	0.21	0.25
log income	6.92	7.20	7.27	7.64	8.03
flush	0.18	0.24	0.29	0.33	0.44
excreta	0.22	0.21	0.15	0.10	0.09
waterplant	0.44	0.45	0.49	0.47	0.50
fuel	0.85	0.79	0.69	0.61	0.57
Liaoning	0.12	0.12	0.00	0.10	0.10
Heilongjang	0.00	0.00	0.08	0.09	0.12
Jiangsu	0.13	0.16	0.16	0.18	0.13
Shandong	0.13	0.10	0.10	0.08	0.11
Henan	0.11	0.08	0.18	0.09	0.12
Hubei	0.13	0.13	0.10	0.12	0.10
Hunan	0.12	0.13	0.05	0.07	0.08
Guangxi	0.10	0.13	0.14	0.10	0.12
Guizhou	0.17	0.16	0.18	0.15	0.13
urbanicity index	-0.27	-0.14	-0.12	0.11	0.38
below (all-wave) median of urbanicity index (1/0)	0.57	0.51	0.54	0.49	0.38
above (all-wave) median of urbanicity index (1/0)	0.43	0.49	0.46	0.51	0.62
in lowest third of (all-wave) distribution of urbaniciy index (1/0)	0.44	0.39	0.37	0.27	0.20
in middle third of (all-wave) distribution of urbanicity index (1/0)	0.30	0.32	0.32	0.36	0.35
in upper third of (all-wave) distribution of urbanicity index (1/0)	0.26	0.29	0.31	0.37	0.45
Observations	6685	5298	6040	5339	7971

Table 5: Means of covariates by wave.

	logit		fixed effects logit		fixed effects ordinal logit	
	poorhealth[1]		poorhealth[1]		worsehealth[2]	
	marginal effect	standard error	marginal effect	standard error	marginal effect	standard error
treatment	0.0407***	0.014	0.065***	0.020	0.054***	0.017
treatment group	-0.371***	0.091				
log income	-0.020***	0.003	-0.018***	0.007	-0.019***	0.005
married	0.014	0.011	0.071**	0.029	-0.014	0.025
edprim	-0.035***	0.010	0.011	0.039	0.048	0.035
edmid	-0.044***	0.012	-0.056	0.058	-0.01	0.049
edhigh	-0.052***	0.014	-0.092	0.094	0.024	0.077
age	0.008***	0.002				
age squared	0.000	0.000				
male	-0.056***	0.008				
waterplant	-0.073***	0.013	-0.097**	0.024	-0.073	0.020
flush	-0.059***	0.014	-0.078**	0.028	-0.086***	0.023
excreta	0.052***	0.010	0.058**	0.015	0.049***	0.013
fuel	0.05***	0.012	0.045**	0.023	0.051***	0.019
size	-0.007**	0.003	0.000	0.006	0.000	0.005
1993	-0.030**	0.012	-0.005	0.018	0.015	0.015
1997	-0.01	0.013	0.065***	0.018	0.087***	0.016
2000	0.138***	0.015	0.249***	0.025	0.227***	0.018
2004	0.150***	0.015	0.292***	0.029	0.26***	0.020
Liaoning	-0.199***	0.047				
Heilongjang	-0.262***	0.039				
Jiangsu	-0.346***	0.036				
Shandong	-0.303***	0.026				
Henan	-0.234***	0.070				
Hubei	-0.255***	0.058				
Hunan	-0.26***	0.040				
Guangxi	0.034	0.077				
Observations	17864		8284		10994	

Table 6: Marginal effects of urbanization and covariates on self-assessed health.
Notes: *treatment* equals one if community is in the upper half of the urbanicity index at time *t*. *treatment group* equals one if community is ever in the upper half. All models include community dummies (δ_{ig}). Standard errors are adjusted for clustering on individuals.

[1] $poorhealth_{it}$=1 if SAH_{it}=fair or poor, 0 otherwise; [2] $worsehealth_{it}$=1 if SAH_{it}>$mean_i$(SAH), 0 otherwise
*** significant at 10%; ** significant at 5%; *** significant at 1%.**

	fixed effects logit poorhealth[1]					
	marginal effect	standard error	marginal effect	standard error	marginal effect	standard error
treatment	0.057***	0.017	0.065***	0.020	0.061***	0.018
log income			-0.020***	0.006		
married	0.060**	0.025	0.068**	0.029	0.060**	0.026
edprim	0.006	0.034	0.012**	0.041	0.008	0.037
edmid	-0.052	0.053	-0.060	0.058	-0.060	0.056
edhigh	-0.086	0.089	-0.114	0.091	-0.113	0.093
waterplant	-0.088***	0.023				
flush	-0.07***	0.026				
excreta	0.05***	0.013				
fuel	0.040**	0.020				
size	0.001	0.005				
1993	-0.009	0.015	-0.006	0.018	-0.011	0.017
1997	0.049***	0.015	0.043**	0.018	0.030*	0.016
2000	0.193***	0.020	0.244***	0.021	0.195***	0.016
2004	0.226***	0.023	0.278***	0.023	0.220***	0.018
observations	8184		8284		8284	

Table 7: Marginal effects of urbanization and covariates on self-assessed health – sensitivity to control for household income and living conditions.
Notes: *treatment* equals one if community is in the upper half of the urbanicity index at time *t*.
[1] poorhealth$_{it}$=1 if SAH$_{it}$=fair or poor, 0 otherwise; * significant at 10%; ** significant at 5%; * significant at 1%.**

	logit		fixed effects logit		fixed effects ordinal logit	
	poorhealth[1]		poorhealth[1]		worsehealth[2]	
	marginal effect	standard error	marginal effect	standard error	marginal effect	standard error
treatment (bottom to middle third urbanicity index)	0.012	0.012	0.02	0.018	0.004	0.015
treatment (bottom to top third urbanicity index)	0.056*	0.034	0.081*	0.044	0.116***	0.037
treatment group (bottom to middle)	-0.442***	0.114				
treatment group (bottom to top)	-0.092	0.075				
Observations	13409		6425		8505	

Table 8: Marginal effects of urbanization on self-assessed health – ordinal treatments.
Notes: All models include community dummies (δ_{ig}) and covariates as in Table 6 . Standard errors are adjusted for clustering on individuals.
[1] poorhealth$_{it}$=1 if SAH$_{it}$=fair or poor, 0 otherwise; [2]worsehealth$_{it}$=1 if SAH$_{it}$>mean$_i$(SAH), 0 otherwise
*** significant at 10%; ** significant at 5%; *** significant at 1%.**

	fixed effects logit		fixed effects ordinal logit	
	poorhealth[1]		worsehealth[2]	
SD increase in the urbanicity index	marginal effect	standard error	marginal effect	standard error
0.25-0.5	-0.034**	0.016	-0.029**	0.014
0.5-1	0.041**	0.017	0.023	0.015
1-1.5	0.034	0.026	0.020	0.023
>1.5	0.152***	0.031	0.132***	0.027
Observations	7806		10411	

Table 9: Marginal effects of degrees of urbanization on self-assessed health.
Notes: All models include community dummies (δ_{ig}), and covariates as in Table 6. Standard errors are adjusted for clustering on individuals.
[1] poorhealth$_{it}$=1 if SAH$_{it}$=fair or poor, 0 otherwise; [2] worsehealth$_{it}$=1 if SAH$_{it}$>mean$_i$(SAH), 0 otherwise
*** significant at 10%; ** significant at 5%; *** significant at 1%.**

	logit		fixed effects logit	
dependent variable	marginal effect of urbanization	standard error	effect of urbanizati on	standard error
hypertension	0.026***	0.008	0.017	0.026
observations	16734		3858	
BMI>30	0.003	0.003	0.078	0.093
observations	16708		427	
BMI<18.5	0.006	0.007	-0.041	0.040
observations	16708		1617	
any physical impairments	0.017***	0.006	0.095**	0.037
observations	17864		2974	
any ill-health symptoms	0.009	0.006	0.061*	0.031
observations	17864		3817	
dying by next wave	0.002	0.002		
observations	17864			

Table 10: Marginal effects of urbanization on probability of experiencing different health outcomes.
Notes: Urbanization is defined as crossing the median of the urbanicity index (similar as in Table 6). All models include community dummies (δ_{ig}) and covariates as in Table 6. Standard errors are adjusted for clustering on individuals.
*** significant at 10%; ** significant at 5%; *** significant at 1%.**

Appendix

Community variables	1991	1993	1997	2000	2004	factor loading
whether there is any farmland in the comm (1-0)	0.61	0.60	0.57	0.48	0.55	-0.64
% of workforce in comm that is working in agriculture (%)	47.27	42.70	42.48	41.04	34.07	-0.72
whether comm is near a bus station (1-0)	0.55	0.54	0.63	0.71	0.64	0.32
whether comm is near a train station (1-0)	0.19	0.19	0.19	0.19	0.23	0.25
whether dirt is main characteristic of roads in comm (1-0)	0.27	0.23	0.23	0.15	0.06	-0.50
whether stone/gravel is main characteristic of roads in comm (1-0)	0.23	0.23	0.22	0.22	0.26	-0.12
whether there is any paved road in comm (1-0)	0.86	0.88	0.87	0.92	0.99	0.41
distance to nearest paved road from comm (km)	0.61	0.27	0.29	0.07	0.03	-0.18
average distance to markets (average over different goods) (km)	1.03	0.86	0.94	0.23	0.23	-0.33
whether comm has convenient telephone service (1-0)	0.56	0.66	0.82	0.92	0.86	0.45
whether there is a post office in comm (1-0)	0.84	0.88	0.88	0.87	0.83	0.20
whether comm can receive daily newspaper on the day it is published (1-0)	0.31	0.39	0.47	0.47	0.59	0.49
whether there is a primary school in the comm (1-0)	0.66	0.73	0.75	0.77	0.71	-0.02
whether there is a (low) secondary school in de comm (1-0)	0.27	0.29	0.30	0.29	0.31	0.31
whether there is a (high) secondary school in de comm (1-0)	0.16	0.14	0.15	0.13	0.13	0.33
whether there is a vocational school in the comm (1-0)	0.07	0.09	0.09	0.08	0.09	0.31
average distance across all health facilities people in comm can go to (km)	5.35	3.22	4.52	4.12	2.95	-0.25
average number of days (a week) that electricity is cut off in comm (days)	1.20	0.87	0.61	0.42	0.31	-0.26
wether there is a child care center (children<3) in comm (1-0)	0.20	0.21	0.29	0.25	0.31	0.47
whether there is a child care center (children<6) in comm (1-0)	0.45	0.48	0.55	0.41	0.55	0.35
number of restaurants in comm	5.04	6.23	8.59	9.09	10.85	0.52
number of enterprises in comm	36.73	32.38	38.25	96.13	165.51	0.40
% of workforce that is working in enterprises with >20 people (%)	32.08	29.80	27.65	31.08	30.06	0.46
% of workforce that is working in enterprises with <20 people (%)	8.41	10.38	14.61	18.85	16.26	0.28
whether there is an open trade area in comm (1-0)	0.22	0.41	0.38	0.46	0.48	0.28
population of comm	2248	2747	2239	3587	5002	0.35
Observations	189	181	191	215	216	

Table A1: Description, means and factor loadings of community variables used in the urbanicity index.

	logit		fixed effects logit		fixed effects ordered logit	
	poorhealth*		poorhealth*		worsehealth**	
	marginal effect	standard error	marginal effect	standard error	marginal effect	standard error
treatment	0.039**	0.015	0.053**	0.021	0.026	0.019
treatment group	-0.312***	0.091				
Observations	17401		7966		10531	

Table A2: Marginal effects of urbanization on self-assessed health with irreversible definition of treatment.

Notes: *treatment* equals one if community is in the upper half of the urbanicity index at time *t*; *treatment group* equals one if community is ever in the upper half. Observations dropped from sample if living in community that experiences a move from above to below the median of the urbanicity index. All models include community dummies (δ_{ig}) and covariates as in Table 6. Standard errors are adjusted for clustering on individuals.

[1] $poorhealth_{it}$=1 if SAH_{it}=fair or poor, 0 otherwise; [2] $worsehealth_{it}$=1 if $SAH_{it}>mean_i(SAH)$, 0 otherwise

*** significant at 10%; ** significant at 5%; *** significant at 1%.**