IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10481

Social Networks and Mental Health Problems: Evidence from Rural-to-Urban Migrants in China

Xin Meng
Sen Xue

JANUARY 2017

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10481

Social Networks and Mental Health Problems: Evidence from Rural-to-Urban Migrants in China

Xin Meng
Australian National University and IZA

Sen Xue
Jinan University

JANUARY 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Social Networks and Mental Health Problems: Evidence from Rural-to-Urban Migrants in China

Over the past two decades, more than 160 million rural residents have migrated to cities in China. They are usually separated from their rural families and work in an unfamiliar, and sometimes hostile, city environment. This paper investigates to what extent city social networks alleviate mental health problems among these migrants. Using the longitudinal migrant survey from the Rural-to-Urban Migration in China (RUMiC) project, we find that larger social networks are significantly correlated with fewer mental health problems in both OLS and fixed effect estimates. To mitigate the endogeneity issue, we use past rainfall in the home county and the distance between home village and the closest transportation centre as the instrument variables for city social networks. The instrument variable estimates and fixed effect instrumental variable estimates suggest that an additional person in the city social networks of migrants reduces GHQ 12 by 0.12 to 0.16 Likert points. The results are robust for migrants who are less educated, who work long hours and who do not have access to social insurances in the city.

JEL Classification: I12, I18, J61

Keywords: mental health, social networks, migration, China

Corresponding author:
Xin Meng
Research School of Economics
College of Business and Economics
Australian National University
HW Arndt Building 25a
ACT 0200
Australia

E-mail: Xin.Meng@anu.edu.au

1 Introduction

Migration is a process which separates individuals from their familiar social networks and engages them in unfamiliar surroundings. During this process, migrants have to adjust to new and sometimes hostile environments, navigate new social systems and cultures, and cope with various stresses. All these circumstances can influence migrants' health in general and their mental health in particular. In a society where social networks play a crucial role in one's life, such as China, migration process can be a very disruptive experience.

In the past twenty years, more than 160 million rural Chinese have moved to cities to work. Unlike internal migrants in other countries, rural migrants are treated as 'guest workers' in Chinese cities due largely to an institution — the household registration system (*'hukou'*) which was established in the early 1950s. In its extreme form, the *hukou* system confined individuals to live and work in their birth place; no mobility was allowed. One justification for the system was that China could not afford to provide the social services and welfare that urban residents could access to all citizens. In the 1990s, as China speeded up its economic growth, the sudden inflow of foreign direct investment substantially increased the demand for unskilled labour. It was then the *hukou* restrictions began to ease. Twenty years on, moving from rural areas to urban cities to work has become a normal way of life for rural people. However, the essence of the *hukou* system has not changed: migrants are not urban citizens and are not entitled to the services and social welfare available to their local urban counterparts (Meng, 2012). This institutionalised discrimination, together with local urban residents' natural prejudice against rural people, means that migrants in Chinese cities not only face the normal pressure of migration in most societies, namely losing their familiar social networks and social environment, but also face significant discriminatory treatment. In addition, migrants in China often engage in manual labour jobs and they work extremely long hours. Due to high working pressure, cultural shock and discrimination, it is not surprising that studies have found that, on average, migrants are less mentally healthy than both rural and local residents (Li et al., 2009; Chen, 2011) and that more than 20 percent of the them had clinically relevant mental problems (Mou et al., 2011; Qiu et al., 2011).

Mental health is an important form of human capital, affecting productivity and wellbeing. Mental illness can trap individuals into disadvantageous positions in labour and marriage markets (e.g. Bartel and Taubman, 1979, 1986; Ettner et al., 1997; Frijters et al., 2010), and impose huge costs on the society (Olesen et al., 2012). Given these negative effects, it is important to understand how to improve people's mental health.

The relationship between social networks and mental health has long been studied, mostly by psychologists or epidemiologiests. In theory, social networks have both beneficial and harmful effects on mental health. They may improve mental health by creating a sense of social integration and buffering stress; but participation in social networks can also induce psychological costs in terms of indebtedness and obligation (Kawachi and Berkman, 2001). A large body of psychological research empirically examines the association between social networks and mental health for different populations (see review by Kawachi and Berkman, 2001; Smith and Christakis, 2008; Cohen and Janicki-Deverts, 2009). However, many of these studies use samples based on mental health patients (Kawachi and Berkman, 2001; Smith and Christakis, 2008). In addition, in the studies focusing on causation, the type of 'social networks' examined is often abnormal, such as having regular conversations with nurses or social workers. The conclusions of these studies may not be generalised to the effect of natural networks, such as friends and acquaintances (Cohen, 2004; Cohen and Janicki-Deverts, 2009; Ertel et al., 2009).

In this study, we examine whether migrants' social networks in their newly settled cities can protect them from experiencing mental health problems, using a unique large-scale survey from the Rural-Urban Migration in China project (RUMiC). The survey collects mental health, social networks, and other detailed socio-economic information from a random sample of rural-urban migrants. The RUMiC dataset has three key advantages for the current study. First, it allows us to assess the effect of the *natural* social networks on migrants' mental health condition. Second, it contains a large longitudinal component, using which we are able to estimate a fixed effects model. The key challenge to identifying the impact of social networks on mental health is the endogeneity issue arising from the omitted individual unobservable characteristics. The fixed-effect model permits us to tease out the time-invariant unobservable characteristics and hence reduce the scope of the endogeneity issue. Finally, RUMiC offers a rich variation in the hometowns of migrants (around 1600 counties/districts), which enables us to find plausibly exogenous variation to further correct the endogeneity issue. In particular, we use past rainfall in the home county and the distance between home village and the closest transportation centre as the instrument variables for city social networks. We discuss the validity of these instruments in details in Section 3.

Our results suggest that an increase in one's city contacts significantly reduces migrant mental health problem: an additional person in one's city social networks reduces GHQ score by 0.12-0.16 points. For a migrant with average network size (13 contacts in the city), the effect of city social networks is 1.6 to 2.1 Likert score, equivalent to 35% to 47% of the standard deviation of the Likert score. These results can also be found in subsamples of migrants who are less educated, who work long hours and who do not have access to social insurances in the city.

The structure of this paper is as follows. Section 2 provides background information on rural-urban migration and mental health condition of migrants in China. Section 3 presents the empirical strategy. Section 4 introduces the data. Section 5 discusses the results, and Section 6 concludes with discussion.

2 Background

There is institutionalised discrimination against internal rural-urban migrants in Chinese cities, primarily due to the household registration system, or *hukou*. During the pre-reform era (1949-1978), labour mobility was strictly controlled by the government and rural to urban migration was forbidden. While urban locals were covered by a cradle-to-grave social welfare system, rural people (then more than 80% of the total population) were left to cover their own education, health, pension and other social services within their small communities (villages or communes). This rural-urban segregation was insured by a food rationing system where to purchase any food in cities one needed food coupons, which were only available to urban dwellers. In addition, the income gap between rural and urban households was very large. Prior to the economic reforms of 1978, an average urban household was making 2.5 times the income of their rural counterparts, not to mention the gap in social services and social welfare. With such a large income gap plus other benefits urban dwellers were receiving, the motivation for rural workers to move to cities were very strong. However, in the 1980s and early 1990s, governments in all urban cities tightly controlled rural-urban migration, despite that the food rationing system was no longer at work due to a significant increase in agriculture productivity during the initial economic reform period. In those days, some rural workers moved to the city to work in the service sector, but the number was very limited. In 1990, there were fewer than 30 million rural workers who migrated to cities, and the number increased to 38 million by 1997 (World Bank, 2009). During this period, many cities implemented various policies to restrict local firms to hire migrant workers. Many cities even published a long list of occupations from which rural migrants were prohibited from being hired (Meng, 2000; Cai et al., 2001; Meng, 2012).

It was not until the late 1990s when foreign direct investment increased the demand for unskilled workers that the Chinese government began to loosen controls on rural-urban migration. In the past 20 years, the number of rural-urban migrants has increased from 38 million in 1997 to 168 million in 2015. However, allowing migrants to work in cities did not change the institutionalised discrimination against them. Migrants still work much longer hours, receive lower wages, and the majority of them are not covered by social health, unemployment and pension insurances. Their children do not have equal access to attend child care and schooling in cities. As a result, many children and spouses are left behind in the home town when migrants move to cities to work (Mu and van de Walle, 2011; Meng, 2012; All China Women's Federation, 2013).

Migration always imposes significant mental stress on those who leaves a familiar environment to embrace a new and unfamiliar surrounding. Migration to a hostile environment, such as the one facing rural-urban migrants in China, is even more stressful. Several studies document the high prevalence of mental problems among Chinese rural migrants. Qiu et al. (2011) find 23.7% of migrant workers in Chengdu had clinically relevant depression symptoms, and 12.8% were consistent with a clinical diagnosis of depression. In Shenzhen, 21.4% of migrant workers are found to have clinically relevant depression symptoms (Mou et al., 2011). He and Wong (2013)'s survey of female migrants in Shenzhen, Kunshan, Dongguan and Shanghai finds that 24% of migrants could be classified as having poor mental health. Similarly, Wong et al. (2008) find that 25% of male migrants and 6% of female migrants in Shanghai are mentally unhealthy. The existing literature also indicates that the mental condition of migrants is worse than their urban and rural counterparts. Li et al. (2009) compare migrants with urban residents in Beijing and rural residents at emigrating region and find that both the urban and rural residents are mentally healthier than migrants. Chen (2011) shows that the psychological distress of migrants in Beijing is worse than that of urban locals, using multivariate regression techniques. There is only one exception to this pattern of results. Li et al. (2007) find that the migrants in Hangzhou are mentally healthier than urban locals, but their mental condition is still worse than that of rural people in Western Zhejiang, which is the origin of many migrants. All these studies suggest that the mental health condition of migrants in China is serious and deserves our attention.

3 Empirical Strategy

The role of social networks in shaping mental health has been widely discussed in psychological and epidemiology studies. The literature proposes several mechanisms describing how social networks affect mental health. Cohen and Wills (1985) develop two of the most prominent models in psychology, the main effect model and the stress-buffering model. The main effect model proposes that social networks could be beneficial to mental health, regardless of whether the individual is experiencing difficulty or not. Specifically, this model predicts that the social interaction provided by an individual's network could generate positive psychological states by increasing his/her sense of security, social belonging and self-worth. The stress-buffering model focuses on individuals in crisis. This model posits that, before a crisis, the individual expects that his/her networks will provide help, and this helps him/her confidently address the future crisis and mitigate stress. During the crisis the networks can also directly reduce stress by providing material and emotional support.[1] In their review paper, Kawachi and Berkman (2001) point out the negative side of social networks. The reciprocal nature of social networks can impose mental cost on individuals who find it difficult to respond to the needs of their network members. This negative effect can be particularly large for people with limited social and economic resources, such as rural migrants. Given the theoretical ambiguity, whether social networks improve or worsen mental health is largely an empirical question.

To understand whether social networks improve migrant mental health, we first estimate the following OLS regression equation:

$$MHP_{ijct} = \beta_1 * SN_{ijct} + X_{ijct} * \beta_2 + W_{ct} * \beta_3 + \tau_t + \rho_j + \varepsilon_{ijct}, \tag{1}$$

where subscripts i, j, c, and t denote individual, destination city, sending county, and survey year, respectively; MHP_{ijct} is the measure of mental health condition; SN_{ijct} is the measure of migrants' social networks in cities, X_{ijct} represents the vector of individual characteristics; W_{ct} is a vector of sending community characteristics; while τ_t is the year fixed effect; ρ_j is the destination city fixed effect; and ε_{ijct} is the unobserved factor.

We estimate two specifications of Equation 1: the baseline model and the extended model. The variables included in X_{ijct} for the baseline model are individual's age, gender, education, marital status, number of children, height, self-reported health, years since the first migration,

[1] Please refer to Thoits (2011) for the detailed channels through which social networks improve mental health in the main effect model and stress-buffering model.

self-employment indicator, and the number of people over 16 years old in the household. These are commonly included variables in other economic studies of mental health (e.g. Stillman et al., 2009; Bjorklund, 1985; Akay et al., 2012). We also include a variable to indicate whether any family members had passed away in the past 12 months in X, as this type of family tragedy can greatly affect an individual's mental health. W_{ic} is a vector of sending community controls which may affect mental health. The variables included are home village geographic location (whether it is a mountainous village and its distance to the closest county), availability of medical services (whether the home village has a medical centre or not), cost of hiring a day-labourer and the long-term rainfall in the home county.

In the extended model, we also control for individual occupation and industry affiliation in X_{ijct} to reduce the omitted variable problem. Although Equation 1 already controls for destination city fixed effects, some time-varying city characteristics may still affect individual's mental health. To this end, we include two more control variables: the growth rates of GDP and minimum wage in the destination city.

Our goal is to identify β_1 in Equation 1. If social networks are exogenous, the OLS estimate of β_1 indicates the impact of social networks on mental heath. The negative value of β_1 implies that social networks help reduce mental health problems; otherwise, social networks do not have a beneficial effect. However, there are three reasons why social networks may not be exogenous. First, reverse causality between social networks and mental health may exist. A person's mental health may affect his/her relationship with others and thereby his/her social networks. Second, certain unobserved personal attributes can be correlated with both social networks and mental health. For example, introverted people may have fewer friends and be more likely to have mental health problems (Kawachi and Berkman, 2001; McKenzie et al., 2002). Third, our social networks measure is self-reported and retrospective, so it may contain large measurement error. An indication of measurement error in the data is that 52% of respondents rounded their answers on network size to a multiple of five. All of these factors indicate that the OLS estimation of β_1 from Equation 1 could be inconsistent. It is important to know that the direction of the OLS bias is unknown here. For example, while it is intuitive to think that reverse causality might cause negative bias, since people with fewer mental problems are more attractive and thereby tend to have a larger networks, people with more mental health problems could also choose a destination with larger existing networks to relieve their stress.

One way to reduce the endogeneity bias, in particular, omitted unobservable personal attributes problem, is to use the fixed effect (FE) model as follows:

$$MHP_{ijct} = \beta_1 * SN_{ijct} + X_{ijct} * \beta_2 + W_{ct} * \beta_3 + \tau_t + n_i + \varepsilon_{ijct}, \tag{2}$$

where n_i is the individual fixed effect and X_{ijt} and W_{ct} include only the time-variant characteristics. The main advantage of the FE model is that it explicitly controls for the individual fixed effect n_i, which removes the bias caused by the time-invariant individual characteristics. However, the FE model has two limitations. First, it cannot resolve the endogeneity bias caused by unobserved time-variant factors. Second, measurement error can induce large attenuation bias in the FE estimator. Pischke (2007) summarized that if the correctly measured variables are persistent and the measurement errors are uncorrelated with each other across waves, then the attenuation bias would be particularly large in the FE estimator. Given the nature of misreporting in the social networks measure, this is a real possibility.[2]

An alternative way to mitigate endogeneity bias and circumvent the disadvantages of the FE estimator is to adopt the instrumental variable approach. The instrumental variable approach jointly estimates Equation 1 and an equation of social networks:

$$SN_{ijct} = \gamma_1 * Z_{ct} + X_{ijct} * \gamma_2 + W_{ct} * \gamma_3 + \tau_t + \rho_j + \epsilon_{ijct}, \tag{3}$$

where Z_{ct} is the instrumental variable which identifies the effect of social networks. A valid instrumental variable should satisfy two conditions. First, the instrument(s) must be correlated with the endogenous variable(s) (relevance condition); and second, the instrument(s) cannot be correlated with the error term of the mental health problem (exclusion restriction). In this paper, we use two variables as potential instrumental variables. One is the average spring and summer rainfall (i.e., from April to August) two years ago in the home county, and the other is the distance between the home village and its closest transport station. We discuss whether these variables could be strong predictors for migrant social networks and whether they satisfy the exclusion restriction below.

Rainfall in the home county

Several studies use rainfall to instrument migrant networks (e.g. Munshi, 2003; Giles and Yoo, 2007). Rainfall and migrant social networks are correlated, because rainfall is related to agricultural income. In general, rainfall increases agricultural production, which, in turn, reduces migration from the region by weakening the migration push effect. Migrants usually form city social networks with other migrants from the same sending region, so rainfall should be a strong predictor for migrant city social networks.

To test this argument, we use the rural household survey data from the Rural-Urban Migration in China project, matched with rainfall data from the Meteorological Information Centre in China to estimate the relationship between rainfall during spring and summer, agricultural income and migration intentions.[3] Table 2 presents the results. We find that rainfall indeed increases agricultural income and consequently reduces rural people's migration intentions. As migrants tend to move to destinations where they have existing networks (Bao et al., 2007), and migrants also tend to form networks with people from the same hometown, the impact of rainfall on migration intention could eventually translate into an impact on network size. It takes time to migrate and form social networks, so we use the spring and summer rainfall two years ago as the instrumental variable. The first-stage results are shown in Section 5 and as can be seen there that this is a strong IV.

[2]We discuss the measure of social networks in Section 4.

[3]For detailed discussion of these data, see Appendix A.

However, it is not enough to simply have strong instruments. A valid instrumental variable should have no direct effect on the outcome variable. In our case, rainfall may be directly related to mental health, for the following reasons. First, gloomy weather can depress people. For this reason, we include average daily rainfall for the previous 10 years and its squared term as measures for long-term rainfall in the W_{ct} vector in Equation 1. We think it is reasonable to assume that the direct effect of rainfall on mental health is mainly shaped by long-term rainfall.[4] We expect that once the long-run rainfall is controlled for, transitory rainfall in migrants' home counties two years ago (the instrumental variable) should not have a direct effect on the current mental health of migrants in cities.[5]

Second, rainfall may affect migrants' mental health via its impact on agricultural income in their hometowns. This potential violation is likely to underestimate the beneficial effect of social networks (i.e. a positive bias). As shown in Table 2 rainfall in rural China is positively correlated with agricultural income. We also know that hometown agricultural income increase is likely to be negatively associated with mental health problems of migrants. To mitigate this potential direct income effect from rainfall to mental health, we include the daily wages of day-labour in migrant sending villages in Equation 1. Further, the literature also provides evidence that individuals can adapt to external income shocks (Tella et al., 2010), and in some cases income shock can be fully adapted within one year (Brickman et al., 1978). Given these, we expect that any remaining bias through agricultural income should be small.

Third, for new migrants, rainfall in home counties may be correlated with an unobserved preference for city life. In home regions with good rainfall, only the most adaptive people choose to move; but when the home region is in drought, many people choose to migrate, regardless of whether they can adapt well to city life. We follow Munshi's (2003) strategy of using the fixed effect instrumental variable model (FEIV) to control for this unobserved preference, which assumes that this unobserved preference for city life is largely time-invariant. Thus, the FEIV estimates should be internally valid.[6]

Distance between the home village and its closest traffic station

The second instrumental variable is the distance between a migrant's home village and its closest transportation station. This information is sourced from the survey question which asked respondents to estimate the "distance between your home village and the nearest transportation station (coach, train or dock)". This variable may be correlated with the size of migrant networks, through two channels. First, as a factor determining the cost of migration, distance may affect villagers' intention to migrate. Second, transport stations are usually built in populated areas. Villages closer

[4] It is still debated in the health literature whether long-term rainfall affects mental condition for one particular population (See e.g. Henríquez-Sánchez et al., 2014; OHare et al., 2016).

[5] Connolly (2013) uses rainfall during the day of the interview as well as the rainfall one-day before the interview in a subjective well-being regression and find that the one-day before rainfall variable is not statistically significantly.

[6] Note that the last potential violation of exclusion restriction is likely to bias our IV estimate downward, because rainfall is likely to be negatively correlated with mental health problems via a preference for city life, and it is also negatively correlated with social networks.

to these stations often have larger populations than those further away. These two channels can affect how many people migrate from a source village and thereby influence the potential network size of the migrants in the destination city. Specifically, these two channels predict that the greater the distance between migrant home village and the closest transportation station, the smaller the social networks in the destination city.

The validity of this instrumental variable relies on the assumption that the distance between the home village and its closest trasportation station is not correlated with the error term in Equation 1. This assumption may not hold if there are omitted variables which are correlated with both error terms in Equations 1 and 3. The distance between the home village and the closest transport station is usually correlated with the level of regional economic development and other geographic factors, and these variables may affect the mental health of villagers. Thus, we include the characteristics of home village (W_{ict}) directly in the regression to avoid the potential omitted variable problem. We assume that, conditional on these variables, this instrumental variable does not directly affect the mental health problems of migrants.

In the analysis below, we use these two instrumental variables both separately and jointly. However, in the FEIV estimation we only use the rainfall instrumental variable, because the distance variable does not vary over time.

4 Data

4.1 RUMiC Survey and Other Data Sources

The main data used in this paper are from the Rural-to-Urban Migration in China (RUMiC) project. The RUMiC project aims to provide a longitudinal dataset to document the socio-economic impact of internal rural-urban migration in China. The project comprises three independent surveys: the Migrant Household survey (MHS), the Urban Household survey (UHS) and the Rural Household survey (RHS). However, the UHS and RHS were terminated in 2012 due to funding constraints. The MHS is still on-going. In this paper, we use the 2008, 2009, 2011 and 2012 waves of the migrant household survey to obtain information on mental health, social networks and other individual and sending community characteristics.[7]

The RUMiC migrant household survey is currently the largest longitudinal survey of rural migrants in China. It covers 15 cities in 9 provinces or municipalities: Guangzhou, Dongguan, Shenzhen, Zhengzhou, Luoyang, Hefei, Bengbu, Chongqing, Shanghai, Nanjing, Wuxi, Hangzhou, Ningbo, Wuhan and Chengdu. These 15 cities represent cities in both largest migration sending and destination provinces (Gong et al., 2008), and cover the coastal, central and western regions of China. Each wave contains around 5000 households.

The RUMiC survey was designed to be a longitudinal survey. However, due to the high attrition rate, each year the survey team drew a representative refreshment sample to restore the sample size to around 5,000 households. Thus, in addition to the longitudinal component, the sample in the

[7] The 2010 wave does not include information on migrants' mental health, so it is not included in the analysis.

baseline wave and the subsequent random refreshments constitute a representative sample.[8] This special design offers us two precious opportunities. First, we can use the longitudinal component to control for individual fixed effects and provide more internally valid estimates. Second, we are able to use the representative sample to provide estimates which are free of attrition bias and also relatively representative of the general migrant population.

In addition to RUMiC survey data, we also utilise data from *City Statistical Yearbook of China* and weather condition data from the Meteorological Information Centre of China. We discuss these data sources later in this section.

4.2 Main Variables

Mental health problems

Mental health information is obtained from the MHS's General Health Questionnaire (GHQ) 12. GHQ is widely used to screen for psychiatric disorders in psychological and medical studies. In the economics literature the abbreviated version (GHQ 12) is frequently used to measure mental health conditions or subjective well-being (e.g. Clark and Oswald, 1994; Gardner and Oswald, 2007; Kuroda and Yamamoto, 2016; Cornaglia et al., 2015). GHQ 12 consists of 12 questions, which focus on "two main classes of phenomena: inability to carry out one's normal 'healthy' functions, and emergence of new phenomena that are distressing" (Graetz, 1991). The answer to each question has a 4-point score, generally denoting not stressed (1), slightly stressed (2), fairly stressed (3) and highly stressed (4). The RUMiC survey asked respondents, who were 16 years or older and present at the time of the interview, to answer these questions.

There are several ways to measure mental health problems using GHQ 12. In our main analysis, we use the Likert score, which is the sum of all the answers to the questions in GHQ 12 and then subtract 12.[9] Thus, the Likert score ranges from 0 to 36. The larger the Likert score, the worse the mental health condition. This measure of mental health is widely used in the literature (e.g. Gardner and Oswald, 2006, 2007; Akay et al., 2013).

Social networks

The RUMiC MHS contains a module on social networks. Ideally, social networks should be

[8]See the discussion on the sample representativeness of the first wave survey in Gong et al. (2008). For more discussion on attrition see Xue (2015)

[9]In the robustness check, we also consider the GHQ score, another measure of mental health problems that is used in the literature (e.g. Clark and Oswald, 1994). The GHQ score counts the number of items for which the respondent reported "fairly" or "highly" stressed. It ranges from 0 to 12. Similar to the Likert score, a larger GHQ score indicates worse mental health condition. We choose the Likert score in the main analysis because it has better distributional properties (i.e., with less skewness and kurtosis, Graetz, 1991), which may make the inference more reliable. Another way to measure mental health is to use factor analysis to measure different aspects of mental health (Cornaglia et al., 2015). However, although several studies have found that the 12 items in GHQ 12 could be attributed to different factors, the factor structure (especially for the number of factors) is different across populations. For example, Graetz (1991) finds that GHQ 12 can be modelled as three factors using Australian youth samples, but Kih et al. (1997) find only two factors in the Turkish sample. Doi and Minowa (2003) show that there are three factors for Japanese male adults, but only two factors for Japanese women. Given that there is no agreement on the factor structure of Chinese rural migrants and we do not find a robust factor structure in this dataset, we do not use this method.

measured in terms of both quantity (i.e., the number of network members) and the quality (i.e., the help which the network can offer). However, the social network module in the RUMiC survey only collects information on quantity. Its measure of network size comes from the following questions:

"During the period of the recent Chinese Lunar New Year, how many people in total did you send your greetings to in various ways (including visiting/phone call/mail/e-mail, etc.)?
Among them,
(1) approximately how many people are your relatives?
(2) how many are your friends and acquaintances?
(3) how many are currently living in the city?
(4) how many have city *hukou*?"

Because Chinese people have a tradition to greet friends or relatives during the Lunar New Year, we use the answer to these questions to measure the approximate size of social networks for migrant workers. In particular, as the purpose of the paper is to examine the impact of migrant city social networks on their current mental health condition, our measure of social networks uses the answer to question (3). This is similar to the way literature uses the number of people visited or sent cards to during Christmas period to measure network size in western society (Hill and Dunbar, 2003).[10]

<u>*Other control variables*</u>

All individual characteristic variables and most sending community characteristic variables used in this paper are drawn directly from the RUMiC Migrant Household Survey. The two economic variables at the destination cities and sending county rainfall related variables are from different sources.

The variable of destination city GDP growth rate is obtained from various years' *City Statistical Yearbook of China*. The minimum wage information is collected from online material (see Appendix A).

The rainfall data is obtained from the Meteorological Information Centre, which collects daily rainfall information from 824 national climatological base stations. These base stations collect accurate and representative weather information for analysing climate change in China. We match the migrant home counties with the nearest weather stations and take the rainfall information from the closest weather station to proxy the rainfall condition in the home county of the migrant. Based on this matching, we generate rainfall-related variables — the two-year lagged spring and summer rainfall in the home county, and the 10-year average rainfall at the home county.[11]

[10]RUMiC data also allow us to construct network size in rural areas. In the main analysis, we do not control for migrant hometown network size as it may be endogenous and bias the coefficients of other variables. But in the robustness check we check whether the results are sensitive to controlling for it. We show that such an inclusion does not change the results (see Panels 1 and 2 in Table 9).

[11]Around 12.2% of respondents did not provide accurate information on home counties. For these respondents, we match the closest weather station to the location of their home prefecture. Also, 0.2% of respondents did not

4.3 Sample Restrictions and Summary Statistics

The RUMiC migrant survey consists of around 5,000 households and 8,000-10,000 individuals for each wave. The social network module and home village information, however, were answered only by the household heads or spouses, whereas the GHQ 12 questions were answered only by individuals who were present at the time of the survey. Thus, the sample used for the analysis below is restricted to these respondents.[12]

In the four waves data we use, there are 23,894 people reported GHQ 12. Among them, 19,560 household heads or spouses provided information on social networks and home village information. Excluding respondents aged below 16 or above 65 leaves us with 19,512 observations. Further excluding individuals with missing values on other covariates leaves 17,704 observations.[13] Finally, to reduce measurement error, we remove 58 observations who reported having more than 150 city contacts. This is consistent with Dunbar (1993) and Hill and Dunbar (2003) who find that the maximum network size of human beings is approximately 150 people, due to cognitive limit. The remaining final sample is 17,646 observations.

In the following analysis, we construct three samples: a pooled sample which includes all the observations across waves, a representative sample which consists of the initial wave and the random refreshment samples in each follow-up wave; and a panel sample which contains the observations which appeared in at least two waves. We use the pooled sample to provide the efficient estimates, the representative sample to provide estimates free of attrition bias, and the longitudinal sample to provide fixed effect estimates.

Table 1 presents summary statistics for the three samples, respectively. Panel A shows that, in the pooled sample, the average Likert score is 7.7, and the average city network size is 13.5 people. The Liker score seems to be similar across the three samples but the panel individuals seem to have slightly more connections relative to the other two samples. Figure 1 presents the unconditional relationship between mental health problems and size of social networks for the pooled sample. It shows that the Likert score decreases as the network size increases, and the variance becomes larger when networks expand as the sample size shrinks (only around 4% sample has the size of social networks greater than 50). We also present this unconditional relationship for the three samples separately in Figure A.1, which suggests that the relationship is similar across different samples. It is interesting to see that the downward trend of the curve mainly concentrates in the region

---

provide precise information on home prefecture. These observations are excluded from the analysis. The average distance between the location of the local county/prefecture government and the nearest weather station is around 35 kilometres. We also tested the robustness if we include the observations which do not have precise information on home prefecture and use the nearest weather station to the province government to proxy their hometown rainfall information. The results are very similar to the results in Section 5 and available upon request.

[12]Potentially, we could assume that the rest of the household members have the same social network as the head or spouse. In the robustness check section we provide results based on the sample under this assumption.

[13]Of all the variables, the home village daily wage for day labour has the most missing values, accounting for more than 60% or 1,130 observations. In the unreported results, we tried to use the average village daily wage from the same home county/district to impute the missing values. The results are very similar, except the coefficients of FEIV drop slightly but are still statistically significant at 10% level in the baseline model and 5% level in the extended model. The results are available upon request.

where the network size does not exceed 50, which indicates that the effect of networks is possibly non-linear. In the main analysis, we use the whole sample to provide a complete picture, avoiding potential sample selection problem, and make the first stage of the IV estimation strong enough as well.[14]

Table 1 also shows that the average age of respondents in the pooled, representative, and panel samples is 32, 30, and 33 years, respectively. Around 64% of the total respondents are males and around two thirds of respondents have junior high school or below education. The panel sample are more likely to be married than the representative sample (64% vs. 52%). On average, the respondents first migrated 9 years ago, but for the representative sample this number is slightly less than 8 years. 27% of the panel sample are self-employed, while for the representative sample the proportion is lower at 18%. In terms of household characteristics, the average migrant family in the city comprises 1.6 adults. The number of children is 0.76 to 0.94 across the three samples, including those who live in the cities with our respondents and those left-behind in rural villages.

The next group of variables presented in Table 1 are related to migrant home village/county and destination city information. They show that the average distance between home village and the closest county centre is around 25 kilometres, home villages of nearly 90% of respondents have a medical centre, home villages of around 20% are located in the mountainous areas, and the daily wage for an unskilled labourer is around 50 to 59 yuan. Finally, the average daily rainfall for the last 10 years is around 3 mm.

Finally, Table 1 also shows the summary statistics for the two instruments. The average two-year lagged spring to autumn daily rainfall is around 5 mm, and the distance between home village and the closest transport station is around 16km. Figure 2 presents the unconditional relationship between the social networks and the two IVs, separately. It is clear that both the lagged rainfall and the distance from home village to the nearest transport station are negatively related to migrant social networks in cities. Thus, the higher the rainfall two years ago at the home county, the smaller the social networks in the destination cities. The further away the village is from a transport station, the higher the transport cost, and the smaller the social networks in the destination cities. Moreover, Panel B in Figure 2 suggests that there is an inverse relationship between distance and network size, so in the following analysis we use the inverse of 1+distance as the instrumental variable.[15]

[14] We also check the robustness by excluding the observations whose contacts are more than 50, and the results are shown in Panel 4 of Table 9. An alternative way to account for potential non-linearity in the effect is to add square or inverse terms of social networks in the regression specification. However, we choose not to do so because of the weak instrumental variable problem when multiple endogenous variables are used.

[15] Around 5% - 6% of respondents reported the distance as zero kilometres. To include these observations, we add one kilometre to the distance to make the inverse feasible. We also tried to add the logarithm of 1 kilometer plus distance or the square term of distance in the regression. But the first stages are weak.

5 Main Results

In this section, we estimate Equations 1 and 2 using OLS and instrumental variable approaches. The standard errors in the cross-sectional estimations are clustered at the home county level to avoid potential common factors related to home counties in the error term. The standard errors in the fixed effect estimations are clustered at the individual level.

5.1 OLS and Fixed Effect Results

Table 3 presents the OLS estimates for the pooled (Columns 1 and 2) and representative samples (Columns 3 and 4). We observe that migrant city networks are significantly and negatively correlated with mental health problems in both samples. The estimates of the baseline model in the two samples suggest that one additional person greeted in the Chinese Lunar New Year while currently living in urban areas is associated with a reduction in the Likert score by 0.016, which is equivalent to 0.4% of the standard deviation of the Likert score. Comparing the estimates between the baseline models and extended models, we find that adding additional variables does not significantly change the estimates. This suggests that the mechanism for social networks affecting mental health does not depend strongly on individual occupation and industry affiliation or city level time-varying factors.

The associations for other control variables are also interesting. Given that the results are largely consistent, our discussion below focuses on the pooled sample in Columns 1 and 2 of Table 3. Men have fewer self-reported mental health problems than women, which is also found by the literature (Akay et al., 2013). The years since the first migration has an inverted U-shape relationship with mental health problems. This suggests that as migrants spend more time in cities, mental health problem first increase and then decrease. This non-linear relationship seems to be intuitive and perhaps indicates an assimilation process. When migrants first arrived, the excitement of city life might curb their homesickness. But as time goes by, life and work pressures begin to bite and migrants feel depressed about the unfamiliar and, in many cases, hostile environment. However, if they stay long enough, they may get used to the city environment. Those who cannot adapt might decide to return home village as well. Thus, the result presented here is likely a combination of these two factors.

Table 3 also shows that more educated migrants tend to be mentally healthier. Relative to individuals with primary school education, those with junior high school education has fewer mental health problems (0.4 to 0.5 of a Likert score), while for those with senior high school degrees or above, the Likert score decreases by 0.8 or more. Married people have better mental health than single people, whereas divorced people have worse mental health. This is perhaps because spouses usually provide emotional support for each other (Smith and Christakis, 2008), or because mentally healthier people are selected for marriage. The self-rated health level is strongly correlated with the Likert score, which is similar to Akay et al.'s (2012) findings.

Columns 5 to 8 of Table 3 shows the OLS and FE results for the panel sample. Although the panel sample has a much smaller sample size, hence subject to sample selection problem, the magnitude of the OLS estimate (0.013 Likert score) is quite similar to those obtained using both the pooled and representative samples. However, once individual fixed effects are controlled for, the estimated coefficient becomes much smaller (-0.009 Likert score, which is equivalent to 0.3% of the standard deviation of the Likert score), even thought it is still statistically significant and negatively correlated with mental health problems. As discussed in Section 3, this could be due to two reasons. First is that when the independent variable (social network) is measured with errors, the fixed-effects estimation leads to a larger attenuation bias than the bias in the OLS estimation. The second possibility is related to omitted unobservable individual attributes problem. Using introversion as an example, introverted people have fewer contacts and more likely to have mental health problems. Thus, without controlling for it, we would observed larger impact of networks on mental health. Fixed-effects estimation may mitigate this problem and lead to a smaller estimated coefficient.

To mitigate the attenuation bias due to measurement error problem and the remaining time-varying omitted variable problem, which are unable to be resolved by the fixed-effects model, we adopt the instrumental variable approach in the next section.

5.2 IV Results

Before presenting IV results, we would like to discuss the validity of the rainfall instrument further. To this end, we conduct a falsification test, which examines the impact of the rainfall instrument on the number of contacts with city *hukou*. The idea is that the two-year lagged rainfall in hometown is supposed to be unrelated to city *hukou* networks. This is because urban hukou people are permanent residents in the city the migrant migrated to, the number of local people the migrant be-friended with should not be affected by his/her home town rainfall situation. If the instrumental variable is correlated with city *hukou* network, the lagged rainfall back home may somehow be related to migrants' city life through the channel other than social networks. Hence, it is likely to be associated with the error terms in Equations 1 and 2 and violate the exclusion restriction. The results are presented in Table 4. They show that such a relationship does not exist in any of the three samples (see Panel 1). For comparison, we show the estimated impact of the rainfall IV on the number of contacts living in the city, which includes both the people with and without city *hukou* (Panel 2). The stark difference between the two estimates indicates that it is not likely that the lagged rainfall variable directly affects migrants' city life through the channel other than social networks.

Table 5 presents the first stage estimations (Equation 3). The results here indicate the impact of rainfall and distance variation on migrants' city social networks, separately and jointly. The Kleibergen-Paap F statistic presented at the bottom of each panel indicate the strength of the instruments (Kleibergen and Paap, 2006). The results suggest that both instrumental variables are strongly correlated with social network size. In particular, Panel 1 shows that a 1 mm increase in the two-years lagged average daily rainfall between April and August reduces the social network size by 0.44-0.56 person. This means that, an average 4.7mm lagged daily rainfall contributes 2.1-2.6 persons to a network, accounting for around 16%-20% of migrant average networks.[16] Panel 2 presents the results of the impact of the distance between home village and the closest transport station on migrant city social networks. Note that the distance variable used here is the inverse of the distance plus 1 km to capture the non-linear relationship between distance and social networks. The results show that the closer the transport station to the home village, the larger the network (low cost on migration induces more people from migrant home town to migrate). The test statistics of these two individual instrumental variables all exceed 10, indicating that the instrumental variables are strong. Panel 3 shows the first-stage results using the two IVs jointly. Compared with the first two panels, both the magnitudes and statistical significances of the coefficients in Panel 3 remain similar, and the Kleibergen-Paap F statistics are also above 10. Thus, these two instrumental variables are not correlated with each other, and each of them provides different identification information to the second stage regression. Hence, jointly using these two instrumental variables makes the 2SLS estimation more efficient.

Table 6 shows the results of the second-stage regressions. For comparison, we present the OLS results in Panel 1. The IV estimates using different IV combinations are presented in Panels 2 to 4. All the IV results show that social networks help relieve mental health problems and that the magnitudes obtained from the IV estimations are larger than those obtained from the OLS estimations. This is to be expected and we will discuss the reasons later in this section. The results presented in Panel 2 indicate that each additional friend the migrant has reduces his/her mental health Likert score by 0.104 to 0.108 points for the pooled sample, and 0.024 to 0.033 points for the representative sample. However, the results for the representative sample are not precisely estimated. When using distance as the instrumental variable (in Panel 3), the effect of social networks on mental health are estimated with precision in all samples and all specifications. Here, we observe that each additional friend reduces mental health problems by 0.14 to 0.25 Likert points. Since different instrumental variables give different local average treatment effects, it is natural that the magnitudes of the estimates differ between Panels 2 and 3. The most important message from these two panels is that all the estimates have negative signs, indicating that social networks help reduce mental health problems.[17] Panel 4 presents the results using the two instrumental variables jointly. All the results are statistically significant at the 1% or 5% level. The magnitudes of the estimates indicate that an additional network member reduces the Likert score by -0.12 to -0.15 points, which is equivalent to 3% of the standard deviation of the Likert score. Since jointly using the two instrumental variables makes the estimation more efficient and covers a larger complier group in the setting of local average treatment effect, the results in Panel 4 are preferred.

Table 7 provides the FEIV estimates. Since the distance instrumental variable does not vary

[16] These calculations are based on the fact that the mean value of the rainfall instrumental variable is 4.7 mm and the average size of migrant network is 13, as shown in Table 1.

[17] It is a puzzle to us that the coefficients decrease in Panel 2 but increase in Panel 3 from the pooled sample to the representative sample. However, these changes may be induced by noise in the data, given the large standard error.

across different waves of the survey, the FEIV estimation employs only lagged rainfall as the instrumental variable. FEIV estimation has two advantages over the IV cross-sectional estimation described above. First, as the individual heterogeneity of mental health is potentially large, controlling for the individual fixed effects could enhance the efficiency of the estimation. Second, controlling for the individual fixed effects could also reduce the bias caused by any individual unobserved time-invariant characteristics, including individual preferences for city life. Thus, the FEIV estimates are more internally valid. However, we do realise that the respondents in the longitudinal sample may not be representative. The migrants in the longitudinal sample tend to be more socio-economically advantaged, better established in the city, and less mobile than those in the representative sample (Xue, 2015). We will keep this caveat in mind when interpreting the results.

Panel 2 of Table 7 indicates that rainfall is a strong instrumental variable in the FEIV estimations. The coefficients suggest that a 1 mm increase in the rainfall IV reduces the number of contacted people living in the city by 0.6 to 0.7 persons, which is slightly larger than the cross-sectional IV results. The Kleibergen-Paap F statistics are greater than 10. Panel 3 again suggests that social networks significantly reduce mental health problems for the longitudinal sample. The FEIV estimates are around -0.16 to -0.19, which is equivalent to around 4% of the standard deviation of the Likert score. The FEIV estimates are larger than the IV estimates in Panel 2 of Table 6. This is perhaps partly because FEIV estimation removes the bias caused by unobserved individual characteristics, such as preference for city life (see discussion in Section 3), and partly because the sample used is a selected group of migrants.

In summary, Tables 6 and 7 suggest that social networks reduce migrant mental health problems. The effect is statistically significant except when only the rainfall instrumental variable is used in the estimation in the representative sample. In terms of the magnitude of the effect, the IV cross-sectional estimations using two IVs jointly suggest that the effect for an migrant with the average network size of 13 is between 35% and 42% of the standard deviation of the Likert score for the sample. For the panel sample using FEIV estimations, the effect for an migrant with the average network size is between 47% and 55% of the standard deviation of the Likert score. These are very large and economically meaningful effects.[18]

Relative to the OLS and fixed effect results, the magnitudes of IV results are larger. This is consistent with Munshi's (2003) study, which finds that the IV estimates of network effect on employment and obtaining higher paid jobs are larger than the OLS and fixed effect associations. The larger IV estimates may be due to two reasons. First, mentally unhealthy people probably endogenously move to cities where they have larger potential networks. Migration is a stressful process, and potential migrants probably realise this, so it is possible that migrants with more mental health problems choose cities with larger potential networks in case they need help. Second, measurement error in social networks is large which creates substantial attenuation bias in OLS and fixed effect estimates. As mentioned in Section 3, 52% of respondents rounded their answers

[18]Unfortunately, we did not find any existing work that uses similar measures for mental health problems and social networks, so we are unable to compare these estimates with the literature.

on social network to a multiple of five. Given these two possibilities, the OLS and FE estimates can be seriously biased downwards.

5.3 Analysis of Sub-samples

The above analysis suggests that social networks are beneficial to the mental health of an average migrant worker. In this section, we investigate whether this beneficial effect still exists in mentally vulnerable migrants. In particular, we analyse three sub-samples: low-educated migrants (individuals with junior high school education or below), migrants who work long hours (weekly working hours>50) and migrants with no access to social welfare in cities. Table 8 shows the results.

Lowly educated migrants is more likely to be prone to poor mental condition (see the previous results in Table 3 and Miech et al. (1999)). Kawachi and Berkman (2001) stresses that social networks can impose huge psychological burdens on low SES people. Is the network effect on mental health we find for the full sample also applicable to the lowly educated migrants? We examine this issue in Panel 1 and restrict our our attention to those migrants with junior high school education or below, who account for 64% of the sample. The results suggest that social networks are still significantly and negatively correlated with mental health problems, and the IV and FEIV estimates are significantly negative as well.

One important feature of migrants is that they tend to work very long hours. In our sample, a migrant works an average of 62 hours per week, which is far more than the usual working hours of urban residents (40 hours per week). Long hours not only tend to worsen mental health (Sparks et al., 1997; Kuroda and Yamamoto, 2016), but may also crowd out social time and render social networks less helpful. In Panel 2 we examine whether social networks are still protective of migrants who work long hours. Specifically, we restrict the sample to respondents who work more than 50 hours per week.[19] The results indicate that social networks are still protective for the migrants who work long hours. The coefficients of the IV and FEIV estimation are even larger than those in the main analysis, suggesting that the beneficial effect of social networks is more effective for these migrants.

In Panel 3, we investigate whether the effect exists among migrants with no access to social insurance in the city. The "guest worker" system in China prevents rural migrants from accessing social welfare in the city (Meng, 2012). In the representative sample, 70% of migrants have no access to unemployment, pensions or health insurance.[20] This makes them vulnerable to shocks in life, which, in turn, could cause mental health problems. For the representative sample, for example, those migrants who have no access to social insurances have an average Likert scores of 7.9, while for those with access to social insurances the average Likert score is 7.24 and the difference is statistically significant at the 1% level. Thus, it is important to understand whether

[19]We also tried to restrict the sample to those who work more than 60 hours per week. But in this case the IVs are weak in the representative sample, though the estimate is still significant at 5% level and the first stage is strong in FEIV estimation.

[20]The pooled sample and longitudinal sample have similar proportions of migrants who do not have access to social welfare.

social networks can mitigate the mental health problems for migrants who do not have access to social insurances. We test this in Panel 3 by restricting the sample to migrants who have no access to pension, medical insurance and unemployment insurance. Although the OLS estimates are similar to the estimate obtained from the full sample, the FE, IV and FEIV estimates in Panel 3 are all larger. This suggests that social networks are more responsive among migrants with no access to social welfare.

5.4 Robustness Check

In this subsection we conduct several sensitivity tests. The results are presented in Table 9.

First, we examine whether our results are sensitive to additional control variables. To this end, we add to the existing controls in the extended model the following variables: 'weekly hours worked', 'monthly income', 'whether the children or spouse are left-behind in the rural village', 'the remittance to income ratio', and 'network size in rural areas'. The reasons that these variables may affect migrant mental health are as follows. Regarding working hours, studies have found that long working hours are highly associated with mental health problems (Sparks et al., 1997; Kuroda and Yamamoto, 2016) and this is true for Chinese migrants as well (Frijters et al., 2009). It is also found that income is a significant predictor of migrants' mental health problems (Akay et al., 2012). Separating from spouse or children can often affect migrant mental health problems (Li et al., 2007; Mou et al., 2011). Regarding remittances, we think that migrants who need to remit a larger proportion of their income back home may be under pressure to work harder than otherwise. Finally, strong home networks may to some extent offset the lack of the city social networks. The reason that these variables were not included in the main specifications is because they may cause endogeneity problem. Panel 1 of Table 9 presents the results of this sensitivity test. It shows that including these variables does not reduce the estimated effect of social networks on mental health. If anything, the IV and FEIV results are slightly larger than before.

Second, there may be a concern that people from different regions could have different social norms which affect the mean level of mental health in the region. Although in our main specification we have included many home village characteristics variables, it may not fully account for the difference in regional norms. Panel 2 reports results from the estimations with home-county fixed effects. They show that adding county fixed effects do not change our results.

Third, instead of using Likert score, we try to use the GHQ score of GHQ 12 as the dependent variable. As discussed in Section 4, GHQ score is also a frequently used measure in the literature (e.g. Clark and Oswald, 1994). We test whether the results are sensitive to the choice of dependent variable. The results in Panel 3 confirm our main findings in Tables 6 and 7. Thus, the choice of dependent variable does not substantively alter the results.

Fourth, as Figure 1 shows that the downward trend between mental health problems and network size mainly concentrates on observations whose contacts do not exceed 50, we restrict the sample to those observations with 50 contacts or less to test the sensitivity of our results. Panel 4 shows that the magnitudes of the estimates become larger, but the instruments become weaker, probably due to the reduction in sample size. We conduct Anderson-Rubin Wald test, which is robust to weak IV problem (Anderson and Rubin, 1949; Stock and Wright, 2000). It suggests that the IV and FEIV are still significant at 5% level. Hence, our results are robust to this restricted sample.

Finally, we test whether the results are sensitive to including other household members who also answered GHQ questions, rather than just the heads or spouses who answered all the GHQ, social network and home village questions. Here we assume that all household members have the same social network and clustered the standard errors at the home county level in the OLS and the cross-section IV estimates. Panel 4 presents these results, which remain similar.

6 Conclusion

This paper investigates whether and to what extent social networks can mitigate mental health problems among Chinese rural-urban migrants, who constitute more than one third of the Chinese urban labour force, produce most of the goods exported from China to the rest of the world, and yet are institutionally discriminated against in the Chinese urban labour market. They work much longer hours, are paid less, and are denied the social services and social welfare available to their urban local counterparts. Very often, they are separated from their immediate families. Because of such a discriminatory, sometimes even hostile treatment, migrants in China have exceptionally high rates of mental health problems. Although China has embarked on institutional reforms to reduce or eliminate unfair treatment of migrant workers, such change may take a long time to have real effect. In the meantime, understanding other informal channels which may help alleviate migrant mental health problems is of significant policy relevance.

Our results from the IV estimates suggest that having one additional city contact reduces the GHQ Likert score by 0.12 to 0.15 points, and the FEIV estimates indicate the effect is 0.16 Likert points. Given that the average network size is around 13, the effect of average network size is 1.6 to 2.1 Likert score, which accounts for 35% to 47% of the standard deviation of the Likert score among our sample. These are large effects. Knowing the potential impact of social networks on mental health among migrants may be important for the government and NGOs in China to find ways to foster and facilitate migrant social networks.

References

Akay, A., O. Bargain, and K. F. Zimmermann (2012). Relative concerns of rural-to-urban migrants in China. *Journal of Economic Behavior & Organization 81*(2), 421 – 441.

Akay, A., C. Giulietti, J. Robalino, and K. Zimmermann (2013). Remittances and well-being among rural-to-urban migrants in China. *Review of Economics of the Household*, 1–30.

All China Women's Federation (2013). *Report on rural left-behind children and rural-urban migrated children in China (wo guo nong cun liu shou er tong cheng xiang liu dong er tong zhuang kuang yan jiu bao gao)*. All China Women's Federation.

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, 46–63.

Bao, S., O. B. Bodvarsson, J. W. Hou, and Y. Zhao (2007). Interprovincial migration in china: The effects of investment and migrant networks. IZA Discussion Papers 2924, Institute for the Study of Labor (IZA).

Bartel, A. and P. Taubman (1979). Health and Labor Market Success: The Role of Various Diseases. *The Review of Economics and Statistics 61*(1), 1–8.

Bartel, A. and P. Taubman (1986). Some economic and demographic consequences of mental illness. *Journal of Labor Economics*, 243–256.

Bjorklund, A. (1985). Unemployment and mental health: Some evidence from panel data. *The Journal of Human Resources 20*(4), pp. 469–483.

Brickman, P., D. Coates, and R. Janoff-Bulman (1978). Lottery winners and accident victims: Is happiness relative? *Journal of personality and social psychology 36*(8), 917.

Cai, F., Y. Du, and M. Wang (2001). Hukou system and labour market protection in china. *Economic Research (jing ji yan jiu, in Chinese)*, 41–49.

Chen, J. (2011). Internal migration and health: Re-examining the healthy migrant phenomenon in china. *Social Science & Medicine 72*(8), 1294 – 1301.

Clark, A. E. and A. J. Oswald (1994). Unhappiness and unemployment. *The Economic Journal 104*(424), pp. 648–659.

Cohen, S. (2004). Social relationships and health. *American psychologist 59*(8), 676.

Cohen, S. and D. Janicki-Deverts (2009). Can we improve our physical health by altering our social networks? *Perspectives on Psychological Science 4*(4), 375–378.

Cohen, S. and T. A. Wills (1985). Stress, social support, and the buffering hypothesis. *Psychological Bulletin 98*, 310–357.

Connolly, M. (2013). Some like it mild and not too wet: The influence of weather on subjective well-being. *Journal of Happiness Studies 14*(2), 457–473.

Cornaglia, F., E. Crivellaro, and S. McNally (2015). Mental health and education decisions. *Labour Economics 33*, 1 – 12.

Doi, Y. and M. Minowa (2003). Factor structure of the 12-item general health questionnaire in the japanese general adult population. *Psychiatry and Clinical Neurosciences 57*(4), 379–383.

Dunbar, R. I. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and brain sciences 16*(04), 681–694.

Ertel, K. A., M. M. Glymour, and L. F. Berkman (2009). Social networks and health: A life course perspective integrating observational and experimental evidence. *Journal of Social and Personal Relationships 26*(1), 73–92.

Ettner, S. L., R. G. Frank, and R. C. Kessler (1997). The impacr of psychiatric disorders on labor market outcomes. *Industrial and Labor Relations Review 51*, 64.

Frijters, P., D. W. Johnston, and X. Meng (2009). The mental health cost of long working hours: the case of rural chinese migrants. working paper.

Frijters, P., D. W. Johnston, and M. A. Shields (2010). Mental health and labour market participation: Evidence from iv panel data models. *IZA Discussion Papers, No. 4883*.

Gardner, J. and A. J. Oswald (2006). Do divorcing couples become happier by breaking up? *Journal of the Royal Statistical Society: Series A (Statistics in Society) 169*(2), 319–336.

Gardner, J. and A. J. Oswald (2007). Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. *Journal of Health Economics 26*(1), 49–60.

Giles, J. and K. Yoo (2007). Precautionary Behavior, Migrant Networks, and Household Consumption Decisions: An Empirical Analysis Using Household Panel Data from Rural China. *The Review of Economics and Statistics 89*(3), 534–551.

Gong, X., S. T. Kong, S. Li, and X. Meng (2008). Rural-urban migrants: A driving force for growth. *China's dilemma: economic growth, the environment and climate change, Edited by Ligang Song and Wing Thye Woo, ANU Press*.

Graetz, B. (1991). Multidimensional properties of the general health questionnaire. *Social Psychiatry and Psychiatric Epidemiology 26*, 132–138.

He, X. and D. F. K. Wong (2013). A comparison of female migrant workers mental health in four cities in china. *International Journal of Social Psychiatry 59*(2), 114–122.

Henríquez-Sánchez, P., J. Doreste-Alonso, M. A. Martínez-González, M. Bes-Rastrollo, A. Gea, and A. Sánchez-Villegas (2014). Geographical and climatic factors and depression risk in the sun project. *The European Journal of Public Health 24*(4), 626–631.

Hill, R. A. and R. I. Dunbar (2003). Social network size in humans. *Human nature 14*(1), 53–72.

Kawachi, I. and L. Berkman (2001). Social ties and mental health. *Journal of Urban Health 78*, 458–467.

Kih, C., M. Rezaki, B. Rezaki, I. Kaplan, G. zgen, A. Sagduyu, and M. O. zturk (1997). General health questionnaire (GHQ12 & GHQ28): psychometric properties and factor structure of the scales in a turkish primary care sample. *Social Psychiatry and Psychiatric Epidemiology 32*, 327–331.

Kleibergen, F. and R. Paap (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics 133*(1), 97–126.

Kuroda, S. and I. Yamamoto (2016). Workers' mental health, long work hours, and workplace management: Evidence from workers' longitudinal data in japan. Technical report, Research Institute of Economy, Trade and Industry (RIETI).

Li, L., H.-m. Wang, X.-j. Ye, M.-m. Jiang, Q.-y. Lou, and T. Hesketh (2007). The mental health status of chinese ruralurban migrant workers. *Social Psychiatry and Psychiatric Epidemiology 42*, 716–722.

Li, X., B. Stanton, X. Fang, Q. Xiong, S. Yu, D. Lin, Y. Hong, L. Zhang, X. Chen, and B. Wang (2009). Mental health symptoms among rural-to-urban migrants in china: a comparison with their urban and rural counterparts. *World Health and Population 11*(1), 24 – 38.

McKenzie, K., R. Whitley, and S. Weich (2002). Social capital and mental health. *The British Journal of Psychiatry 181*(4), 280–283.

Meng, X. (2000). *Labour Market Reform in China.* Cambridge: Cambridge University Press.

Meng, X. (2012). Labor market outcomes and reforms in china. *The Journal of Economic Perspectives 26*(4), 75–101.

Miech, R. A., A. Caspi, T. E. Moffitt, B. R. E. Wright, and P. A. Silva (1999). Low socioeconomic status and mental disorders: a longitudinal study of selection and causation during young adulthood 1. *American journal of Sociology 104*(4), 1096–1131.

Mou, J., J. Cheng, S. M. Griffiths, S. Y. Wong, S. Hillier, and D. Zhang (2011). Internal migration and depressive symptoms among migrant factory workers in shenzhen, china. *Journal of Community Psychology 39*(2), 212–230.

Mu, R. and D. van de Walle (2011). Left behind to farm? women's labor re-allocation in rural china. *Labour Economics 18, Supplement 1*, S83 – S97.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics 118*(2), 549–599.

Olesen, J., A. Gustavsson, M. Svensson, H.-U. Wittchen, and B. Jönsson (2012). The economic cost of brain disorders in europe. *European Journal of Neurology 19*(1), 155–162.

OHare, C., V. OSullivan, S. Flood, and R. A. Kenny (2016). Seasonal and meteorological associations with depressive symptoms in older adults: A geo-epidemiological study. *Journal of affective disorders 191*, 172–179.

Pischke, S. (2007). Lecture notes on measurement error. Technical report.

Qiu, P., E. Caine, Y. Yang, Q. Chen, J. Li, and X. Ma (2011). Depression and associated factors in internal migrant workers in china. *Journal of Affective Disorders 134*(13), 198 – 207.

Smith, K. P. and N. A. Christakis (2008). Social networks and health. *Annual Review of Sociology 34*(4), 405–429.

Sparks, K., C. Cooper, Y. Fried, and A. Shirom (1997). The effects of hours of work on health: A meta-analytic review. *Journal of Occupational and Organizational Psychology 70*(4), 391–408.

Stillman, S., D. McKenzie, and J. Gibson (2009). Migration and mental health: Evidence from a natural experiment. *Journal of Health Economics 28*(3), 677–687.

Stock, J. H. and J. H. Wright (2000). Gmm with weak identification. *Econometrica 68*(5), 1055–1096.

Tella, R. D., J. H.-D. New, and R. MacCulloch (2010). Happiness adaptation to income and to status in an individual panel. *Journal of Economic Behavior & Organization 76*(3), 834 – 852.

Thoits, P. A. (2011). Mechanisms linking social ties and support to physical and mental health. *Journal of Health and Social Behavior 52*(2), 145–161.

Wong, D., X. He, G. Leung, Y. Lau, and Y. Chang (2008). Mental health of migrant workers in china: prevalence and correlates. *Social Psychiatry and Psychiatric Epidemiology 43*, 483–489.

World Bank (2009). *From Poor Areas to Poor People: China's Evolving Poverty Reduction Agenda : an Assessment of Poverty and Inequality in China.* World Bank.

Xue, S. (2015). Who are the movers and who are the stayers? attrition in the migrant household survey of the rural-to-urban migration in china project: 2008-2013. Working paper.

Table 1: Summary Statistics

	Panel A: Pooled Sample		Panel B: Represent Sample		Panel C: Longitudinal Sample	
VARIABLE	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Likert score	7.71	4.53	7.64	4.50	7.68	4.53
# of contacted people living in urban areas	13.48	17.58	13.05	17.60	13.82	17.55
Age	31.62	10.38	30.38	10.33	33.12	10.12
Male	0.64	0.48	0.64	0.48	0.65	0.48
Education:						
Primary school or below	0.12		0.12		0.13	
Junior high school	0.52		0.52		0.51	
Senior high school or equivalent	0.29		0.30		0.29	
College education or above	0.07		0.06		0.07	
Height (cm)	166.46	7.21	166.57	7.16	166.46	7.23
Self-rated health:						
Excellent health	0.37		0.39		0.34	
Good health	0.48		0.46		0.49	
Average health	0.14		0.13		0.15	
Poor or very poor health	0.01		0.01		0.02	
Married	0.57		0.52		0.64	
Divorced	0.02		0.02		0.02	
Years since the first migration	8.89	6.89	7.83	6.66	10.00	6.84
Self-employment	0.22		0.18		0.27	
# of family members over 16 years old	1.57	0.79	1.45	0.71	1.68	0.83
# of children	0.84	0.90	0.76	0.90	0.94	0.89
Death of family member	0.03		0.02		0.03	
Daily wage of unskilled labour at home village (yuan)	56.49	25.24	50.24	22.85	59.96	25.50
Distance bw home village and the closest county (km)	24.96	35.69	25.81	36.70	24.92	33.42
Home village has medical station	0.89		0.89		0.89	
Home village is in a mountainous area	0.21		0.23		0.20	
Average daily rainfall from t-10 to t-1 (1mm)	2.98	0.87	3.01	0.89	2.94	0.83
Growth of GDP in destination cities (%)	11.67	2.71	11.66	2.59	11.91	2.76
Growth of real minimum wage in destination cities (%)	5.96	9.38	4.97	8.57	6.36	9.84
Average daily rainfall btw Apr and Aug at t-2 (1mm)	4.75	1.81	4.71	1.89	4.74	1.72
Distance btw home village and the closest traffic centre (km)	16.25	32.76	16.54	33.90	15.72	30.81
Observations	17646		10951		8745	

Note: The pooled sample consists of all the observations across waves. The representative sample consists of 2008 wave and new households in each wave after 2008. The longitudinal sample consists of individuals who appeared in two or more waves.
Source: 2008, 2009, 2011 and 2012 migrant surveys in the RUMiC project.

Table 2: OLS estimates of impact of rainfall on migration choice

	(1) Log household net agricultural income per capita in previous year	(2) Intention to migrate in 3 months	(3) Intention to migrate in 12 months	(4) Migration for business over 12 months in the previous year
Daily rainfall between Apr and Aug in year t-1		-0.005**	-0.005**	
		(0.002)	(0.003)	
Daily rainfall between Apr and Aug in year t-2	0.016***	-0.009***	-0.009***	-0.007***
	(0.005)	(0.002)	(0.003)	(0.002)
Daily rainfall between Apr and Aug in year t-3	0.015***	-0.004*	-0.004	-0.004***
	(0.006)	(0.002)	(0.003)	(0.001)
Daily rainfall between Apr and Aug in year t-4	0.003			-0.000
	(0.004)			(0.001)
Observations	12991	31621	31621	34092
Adjusted R-squared	0.278	0.134	0.142	0.065

Note: Standard errors are clustered at household level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. Number of family members over 16 years old, male dummy, age, squared age, education dummies, dummies of marriage status, number of children, height, dummies of self-rated health and dummy of death of family member in the last 12 months, year dummies, county dummies and constant are included. The reference group is unmarried females with education below junior high school and excellent health. In the first column each household is an observation, and in the last three columns each household member is an observation. The daily rainfall variable is in 1 mm. Source: 2008 and 2009 rural surveys in the RUMiC project.

Table 3: OLS and fixed effect estimates of network effect on mental health problems

	Panel A: Pooled Sample-OLS		Panel B: Represent Sample-OLS		Panel C: Panel Sample-OLS		Panel D: Panel Sample-FE	
	Baseline	Extended	Baseline	Extended	Baseline	Extended	Baseline	Extended
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# of contacted people living in cities	-0.016***	-0.016***	-0.016***	-0.016***	-0.013***	-0.012***	-0.009**	-0.009**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)
Age	0.054*	0.052*	0.046	0.042	0.054	0.046		
	(0.030)	(0.030)	(0.034)	(0.034)	(0.044)	(0.044)		
Squared age *10ˆ(-2)	-0.057	-0.055	-0.053	-0.049	-0.054	-0.044		
	(0.040)	(0.040)	(0.046)	(0.046)	(0.058)	(0.057)		
Male	-0.596***	-0.581***	-0.731***	-0.702***	-0.462***	-0.481***		
	(0.100)	(0.101)	(0.122)	(0.123)	(0.149)	(0.153)		
Years since the first migration	0.019	0.021	0.017	0.023	0.025	0.028		
	(0.018)	(0.019)	(0.022)	(0.023)	(0.026)	(0.026)		
Squared year since the first migration *10ˆ(-2)	-0.093	-0.093	-0.085	-0.101	-0.085	-0.084		
	(0.062)	(0.063)	(0.080)	(0.081)	(0.085)	(0.085)		
junior high school	-0.504***	-0.518***	-0.440***	-0.429***	-0.636***	-0.649***		
	(0.133)	(0.132)	(0.164)	(0.162)	(0.194)	(0.193)		
senior high school or equivalent	-0.878***	-0.849***	-0.866***	-0.813***	-0.888***	-0.832***		
	(0.150)	(0.148)	(0.172)	(0.173)	(0.219)	(0.218)		
college education or above	-1.282***	-1.142***	-1.460***	-1.296***	-1.248***	-1.024***		
	(0.196)	(0.195)	(0.236)	(0.236)	(0.279)	(0.280)		
Married	-0.499***	-0.477***	-0.503***	-0.494***	-0.369**	-0.350*	0.052	0.065
	(0.131)	(0.131)	(0.161)	(0.161)	(0.180)	(0.181)	(0.303)	(0.304)
Divorced	0.654**	0.631**	0.188	0.223	1.213***	1.164***	0.539	0.516
	(0.275)	(0.272)	(0.351)	(0.350)	(0.374)	(0.367)	(0.629)	(0.635)
# of kids	-0.067	-0.055	-0.063	-0.043	-0.155*	-0.146	0.020	0.028
	(0.067)	(0.066)	(0.082)	(0.082)	(0.092)	(0.092)	(0.178)	(0.176)
Height (cm)	-0.012*	-0.013*	-0.003	-0.005	-0.016	-0.015	-0.026	-0.028
	(0.007)	(0.007)	(0.008)	(0.008)	(0.010)	(0.010)	(0.029)	(0.029)
Good health	1.272***	1.275***	1.515***	1.520***	1.119***	1.127***	0.915***	0.924***
	(0.084)	(0.083)	(0.095)	(0.095)	(0.128)	(0.124)	(0.136)	(0.136)
Average health	2.926***	2.932***	2.978***	2.986***	2.730***	2.726***	2.193***	2.231***
	(0.125)	(0.125)	(0.151)	(0.151)	(0.176)	(0.174)	(0.199)	(0.199)
Poor or very poor health	5.272***	5.250***	5.202***	5.228***	5.158***	5.112***	3.849***	3.819***
	(0.348)	(0.349)	(0.451)	(0.450)	(0.434)	(0.438)	(0.558)	(0.562)
Self-employment	-0.137	-0.072	-0.155	-0.034	-0.061	-0.090	-0.447	-0.549*
	(0.102)	(0.137)	(0.124)	(0.171)	(0.138)	(0.172)	(0.288)	(0.306)
# of family members over 16 years old	-0.035	-0.048	-0.059	-0.081	-0.027	-0.022	-0.072	-0.074
	(0.059)	(0.059)	(0.071)	(0.071)	(0.082)	(0.082)	(0.129)	(0.128)
Death of family member	0.135	0.178	0.357	0.373	0.029	0.067	0.317	0.325
	(0.199)	(0.197)	(0.264)	(0.266)	(0.276)	(0.271)	(0.333)	(0.332)
Daily wage of unskilled labour in home	-0.005**	-0.004**	-0.001	-0.000	-0.008***	-0.008***	-0.010***	-0.010***
village (yuan)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Home village is in a mountainous area	0.093	0.098	0.030	0.048	0.283**	0.271**		
	(0.093)	(0.092)	(0.109)	(0.109)	(0.135)	(0.134)		
Inverse of 1 + the distance btw home village	-0.216	-0.261	-0.234	-0.267	-0.351	-0.379		
and the closest county	(0.190)	(0.190)	(0.242)	(0.243)	(0.261)	(0.258)		
Home village has medical centre	-0.262**	-0.254**	-0.523***	-0.543***	0.082	0.094		
	(0.114)	(0.114)	(0.145)	(0.144)	(0.157)	(0.157)		
Average daily rainfall from t-10 to t-1	-0.343	-0.379*	-0.307	-0.335	-0.602*	-0.627*		
	(0.227)	(0.226)	(0.264)	(0.263)	(0.348)	(0.344)		
Squared average daily rainfall from t-10 to t-1	0.044	0.051	0.052	0.055	0.057	0.065		
	(0.034)	(0.034)	(0.040)	(0.039)	(0.052)	(0.051)		
Growth of GDP in destination		0.137***		0.125***		0.157***		0.110***
cities (%)		(0.024)		(0.029)		(0.033)		(0.038)
Growth of real minimum wage in destination		-0.005		-0.004		-0.004		-0.006
cities (%)		(0.004)		(0.006)		(0.005)		(0.005)
Industry and occupation dummies		Yes		Yes		Yes		Yes
City and year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17646	17646	10951	10951	8745	8745	8745	8745
Adjusted R-squared	0.106	0.111	0.121	0.125	0.094	0.100	0.047	0.056

Note: Standard errors are clustered at home county level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. All the regressions include year dummies, destination city dummies and constant. The representative sample consists of 2008 wave and new samples in each wave after 2008. The pooled sample consists of all the observations across waves. Longitudinal sample consists of individuals who appeared in two or more waves. The daily rainfall variable is in 1 mm. The reference group is unmarried females with primary school education or below and excellent health.
Source: 2008, 2009, 2011 and 2012 migrant surveys in the RUMIC project.

Table 4: A falsification test on rainfall instrumental variable

	Cross-sectional model		FE model
	pooled sample	represent sample	longitudinal sample
Panel 1: # of contacted people with city Hukou			
Average daily rainfall bw Apr. and Aug. at t-2	0.042	0.099	-0.004
	(0.056)	(0.066)	(0.094)
Observations	17636	10945	8739
Adjusted R-squared	0.055	0.054	0.016
Panel 2: # of contacted people living in the city			
Average daily rainfall bw Apr. and Aug. at t-2	-0.556***	-0.441***	-0.683***
	(0.124)	(0.133)	(0.183)
Observations	17646	10951	8745
Adjusted R-squared	0.085	0.096	0.030

Note: Standard errors are clustered at the home county level in the cross-sectional model and at the individual level in the FE model. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. In the cross-sectional model, the other control variables are the same as those in extended model in Panels A to C of Table 3. In the FE model, the other control variables are the same as those in the extended model in Panel D of Table 3. The pooled sample consists of all the observations across waves. The representative sample consists of 2008 waves and new households for each waves after 2008. The longitudinal sample consists of the individuals appearing in two or more than two waves. The daily rainfall variable is in 1 mm.
Source: 2008, 2009, 2011 and 2012 migrant surveys in the RUMiC project.

Table 5: The first-stage results of the cross-sectional IV estimation

	Pooled sample		Represent sample	
	Baseline	Extended	Baseline	Extended
	(1)	(2)	(3)	(4)
Panel 1: Using average daily rainfall btw Apr and Aug at t-2 as IV				
Rainfall	-0.563***	-0.556***	-0.485***	-0.441***
	(0.125)	(0.124)	(0.134)	(0.133)
Weak IV test statistics	20.457	20.126	12.991	11.056
Panel 2: Using distance to the closest transportation centre as IV				
Distance	3.012***	3.033***	2.588***	2.532***
	(0.845)	(0.859)	(0.759)	(0.751)
Weak IV test statistics	12.714	12.456	11.639	11.376
Panel 3: Using the two IVs				
Distance	2.941***	2.965***	2.535***	2.481***
	(0.841)	(0.855)	(0.758)	(0.751)
Rainfall	-0.550***	-0.543***	-0.475***	-0.431***
	(0.123)	(0.123)	(0.134)	(0.132)
Weak IV test statistics	13.508	13.183	11.457	10.345
Observations	17646	17646	10951	10951

Note: Standard errors are clustered at home county level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. The regression specifications are the same as Table 3. The pooled sample consists of all the observations across waves. The representative sample consists of 2008 wave and new samples in each wave after 2008. The daily rainfall variable is in 0.1 mm. KP test statistics are Kleibergen-Paap rk Wald F statistic. The daily rainfall variable is in 1 mm.
Source: 2008, 2009, 2011 and 2012 migrant surveys in the RUMiC project.

Table 6: The second-stage results of the cross-sectional IV estimation

	Pooled sample		Represent sample	
	Baseline	Extended	Baseline	Extended
	(1)	(2)	(3)	(4)
Panel 1: OLS				
# of contacted people living in cities	-0.016***	-0.016***	-0.016***	-0.016***
	(0.002)	(0.002)	(0.002)	(0.002)
Panel 2: Using average daily rainfall bw Apr. and Aug. at t-2 as IV				
# of contacted people living in cities	-0.108*	-0.104*	-0.033	-0.024
	(0.058)	(0.058)	(0.077)	(0.084)
Panel 3: Using distance to the closest transportation centre as IV				
# of contacted people living in cities	-0.144**	-0.140**	-0.253**	-0.252**
	(0.060)	(0.059)	(0.099)	(0.100)
Panel 4: Using the two IVs				
# of contacted people living in cities	-0.126***	-0.122***	-0.142**	-0.145**
	(0.043)	(0.042)	(0.058)	(0.061)
Observations	17646	17646	10951	10951

Note: Standard errors are clustered at home county level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. The regression specifications are the same as Table 3. The pooled sample consists of all the observations across waves. The representative sample consists of 2008 wave and new samples in each wave after 2008. The daily rainfall variable is in 0.1 mm. KP test statistics are Kleibergen-Paap rk Wald F statistic. The daily rainfall variable is in 1 mm.
Source: 2008, 2009, 2011 and 2012 migrant surveys in the RUMiC project.

Table 7: FEIV estimates of network effect on mental health problems

	Baseline (1)	Extended (2)
Panel 1: Fixed effect		
# of contacted people living in cities	-0.009**	-0.009**
	(0.004)	(0.004)
Panel 2: 1st-stage results of FEIV		
Average daily rainfall btw Apr and Aug at t-2	-0.603***	-0.683***
	(0.181)	(0.183)
weak IV test statistics	11.036	13.953
Panel 3: 2nd-stage results of FEIV		
# of contacted people living in cities	-0.192**	-0.164**
	(0.092)	(0.077)
Observations	8745	8745
Individuals	3646	3646

Note: Standard errors are clustered at individual level. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. The regression specifications are the same as Panel D of Table 3. Only the longitudinal sample is used in this table, which consists of individuals who appeared in two or more waves. The daily rainfall variable is in 1 mm. KP test statistics are Kleibergen-Paap rk Wald F statistic. Source: 2008, 2009, 2011 and 2012 migrant surveys in the RUMiC project.

Table 8: Analysis of disadvantaged migrant groups

	OLS	FE	IV	FEIV
Panel 1: Migrants with junior high school education or below				
# of contacted people living in cities	-0.021***	-0.010*	-0.174**	-0.158**
	(0.004)	(0.005)	(0.075)	(0.068)
Weak IV test statistics			9.040	23.960
Observations	6991	5207	6991	5207
Individuals				2558
Panel 2: Migrants with long working hours (weekly working hours>50)				
# of contacted people living in cities	-0.016***	-0.010*	-0.209**	-0.201**
	(0.003)	(0.005)	(0.096)	(0.088)
Weak IV test statistics			5.950	13.292
Observations	7298	4944	7298	4944
Individuals				2993
Panel 3: Migrants without access to welfare				
# of contacted people living in cities	-0.016***	-0.014***	-0.159***	-0.225**
	(0.003)	(0.005)	(0.054)	(0.093)
Weak IV test statistics			14.295	12.935
Observations	7678	4975	7678	4975
Individuals				2832

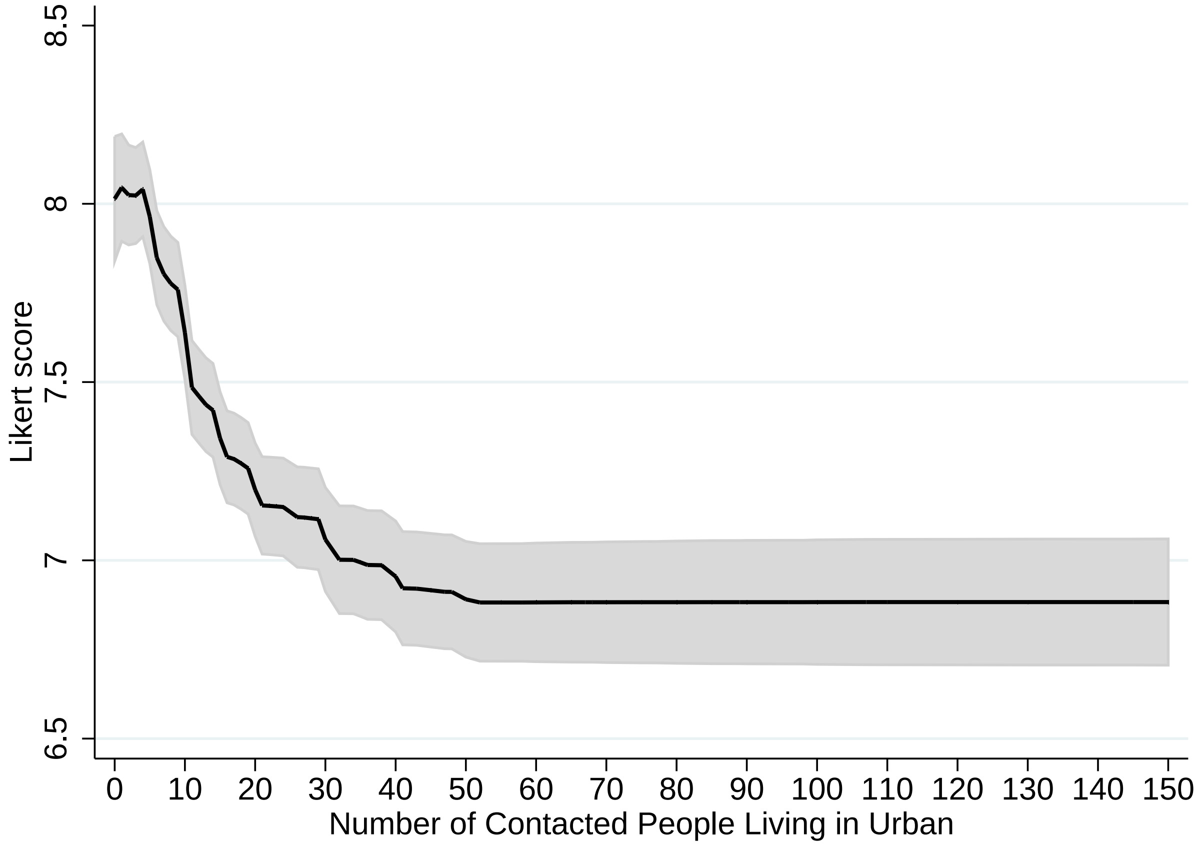
Note: Standard errors are clustered at home county level in OLS and IV estimates and at individual level at FE and FEIV estimates. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. The OLS and IV regression specifications are the same as the extended model in Panels A to C of Table 3, and the FE and FEIV regression specifications are the same as the extended model in Panel D of Table 3. All the regressions of OLS and IV estimations use the representative sample which consists of 2008 wave and new samples for each wave after 2008, and the regressions of FE and FEIV estimation use the longitudinal sample which consists of individuals who appeared two or more waves. The IV estimates are produced by rainfall and distance instruments jointly, and the FEIV restimates are produced by rainfall instrument only. The weak IV test statistics is the Kleibergen-Paap rk Wald F statistic.
Source: 2008, 2009, 2011 and 2012 migrant surveys in the RUMiC project.

Table 9: Robustness check

	OLS	FE	IV	FEIV
Panel 1: Adding additional control - without county fixed effects				
# of contacted people living in cities	-0.015***	-0.008**	-0.153**	-0.166**
	(0.002)	(0.004)	(0.073)	(0.082)
Weak IV test statistics			8.281	13.504
Observations	10808	8366	10808	8366
Individuals				3640
Panel 2: Adding additional control - with county fixed effects				
# of contacted people living in cities	-0.016***	-0.008**	-0.150**	-0.166**
	(0.003)	(0.004)	(0.075)	(0.082)
Weak IV test statistics			6.711	13.504
Observations	10808	8366	10808	8366
Individuals				3640
Panel 3: Mental health problem measure − GHQ score				
# of contacted people living in cities	-0.005***	-0.001	-0.038*	-0.064**
	(0.001)	(0.001)	(0.021)	(0.029)
Weak IV test statistics			10.345	13.953
Observations	10951	8745	10951	8745
				3646
Panel 4: Excluding respondents whose contacts exceed 50				
# of contacted people living in cities	-0.025***	-0.013**	-0.172	-0.284*
	(0.004)	(0.006)	(0.111)	(0.165)
Weak IV test statistics			7.951	8.818
Observations	10565	8308	10565	8308
				3489
Panel 5: All the household members				
# of contacted people living in cities	-0.015***	-0.009***	-0.162***	-0.228**
	(0.003)	(0.003)	(0.058)	(0.090)
Weak IV test statistics			11.405	13.587
Observations	12825	10212	12825	10212
Individuals				4234

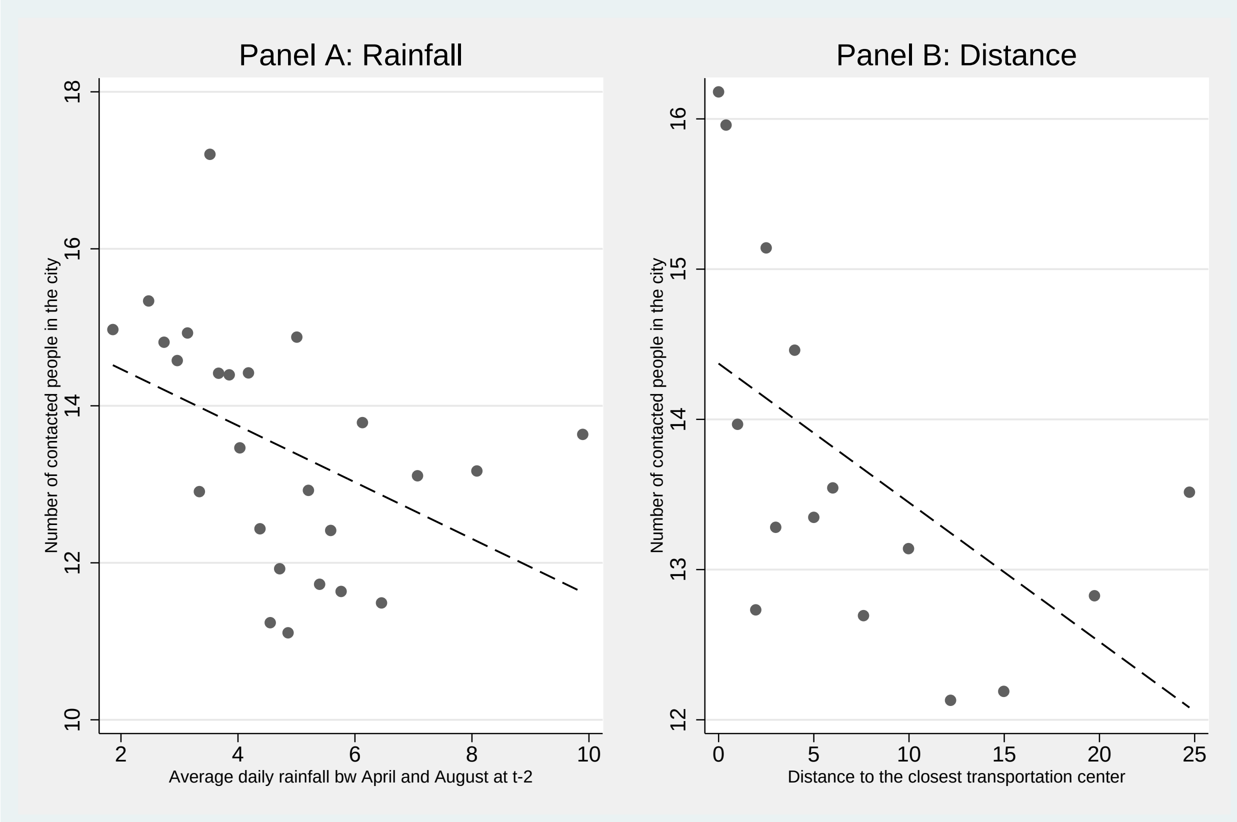
Note: Standard errors are clustered at home county level in OLS and IV estimates and at individual level at FE and FEIV estimates. * significant at 10% level; ** significant at 5% level; *** significant at 1% level. The OLS and IV regression specifications are the same as the extended model in Panels A to C of Table 3, and the FE and FEIV regression specifications are the same as the extended model in Panel D of Table 3. All the regressions of OLS and IV estimations use the representative sample which consists of 2008 wave and new samples for each wave after 2008, and the regressions of FE and FEIV estimation use the longitudinal sample which consists of individuals who appeared two or more waves. The IV estimates are produced by rainfall and distance instruments jointly, and the FEIV restimates are produced by rainfall instrument only. The weak IV test statistics is the Kleibergen-Paap rk Wald F statistic.
Source: 2008, 2009, 2011 and 2012 migrant surveys in the RUMiC project.

Figure 1: Unconditional relationship between social networks and mental health problems



Source: Pooled sample from the 2008, 2009, 2011 and 2012 waves of the migrant survey in RUMiC project. Each group represents 20% of the sample, and the groups are ranked based on network size.

Figure 2: Unconditional relationship between two IVs and size of networks



Note: We partition the sample into 25 equal-sized cells according to the magnitude of the instrument variable. Each dot in the graph represents average rainfall and network size within one cell.
Source: Pooled sample from the 2008, 2009, 2011 and 2012 waves of the migrant survey in RUMiC project.

A Data Appendix

A.1 General Health Questionnaire 12

The questions in General Health Questionnaire 12 are as follows: "In the last few weeks have you ever had the following feelings?

1. Have you been able to concentrate on whatever you're doing?
   (1) been able to concentrate; (2) attention occasionally diverted; (3) attention sometimes diverted; (4) attention frequently diverted, not been able to concentrate;
2. Have you lost much sleep over worry?
   (1) never; (2) occasionally; (3) fairly often; (4) very often;
3. Have you felt that you were playing a useful part in things?
   (1) true so; (2) to some extent; (3) rarely; (4) not at all;
4. Have you felt capable of making decisions about things?
   (1) very capable; (2) quite capable; (3) not quite capable; (4) not capable at all;
5. Have you felt constantly under strain?
   (1) never; (2) slightly; (3) considerably; (4) seriously;
6. Have you felt you couldn't overcome your difficulties?
   (1) never; (2) slightly; (3) considerably; (4) seriously;
7. Have you felt your normal day-to-day activities are interesting?
   (1) very interesting; (2) fairly interesting; (3) not very interesting; (4) not interesting at all;
8. Have you been able to face up to problems?
   (1) always; (2) most of the time; (3) sometimes; (4) rarely;
9. Have you been feeling unhappy or depressed?
   (1) never; (2) slightly; (3) considerably; (4) seriously;
10. Have you been losing confidence in yourself?
    (1) never; (2) slightly; (3) considerably; (4) seriously;
11. Have you been thinking of yourself as a worthless person?
    (1) never; (2) slightly; (3) considerably; (4) seriously;
12. Have you been feeling reasonably happy, all things considered?
    (1) very happy; (2) fairly happy; (3) not so happy; (4) not happy at all"

A.2 Data Details of Table 2

The sample used in Table 2 is extracted from the 2008 and 2009 waves of the RUMiC rural household survey. This survey covers 82 counties in 9 provinces in China. We restrict the sample to households whose agricultural income per household member in the previous year is not more than 50000 yuan to reduce the potential measurement error. We also exclude respondents who are younger than 16 or older than 65, because these respondents are unlikely to migrate. The rainfall data in Table 2 is constructed in the way described in Section 4.

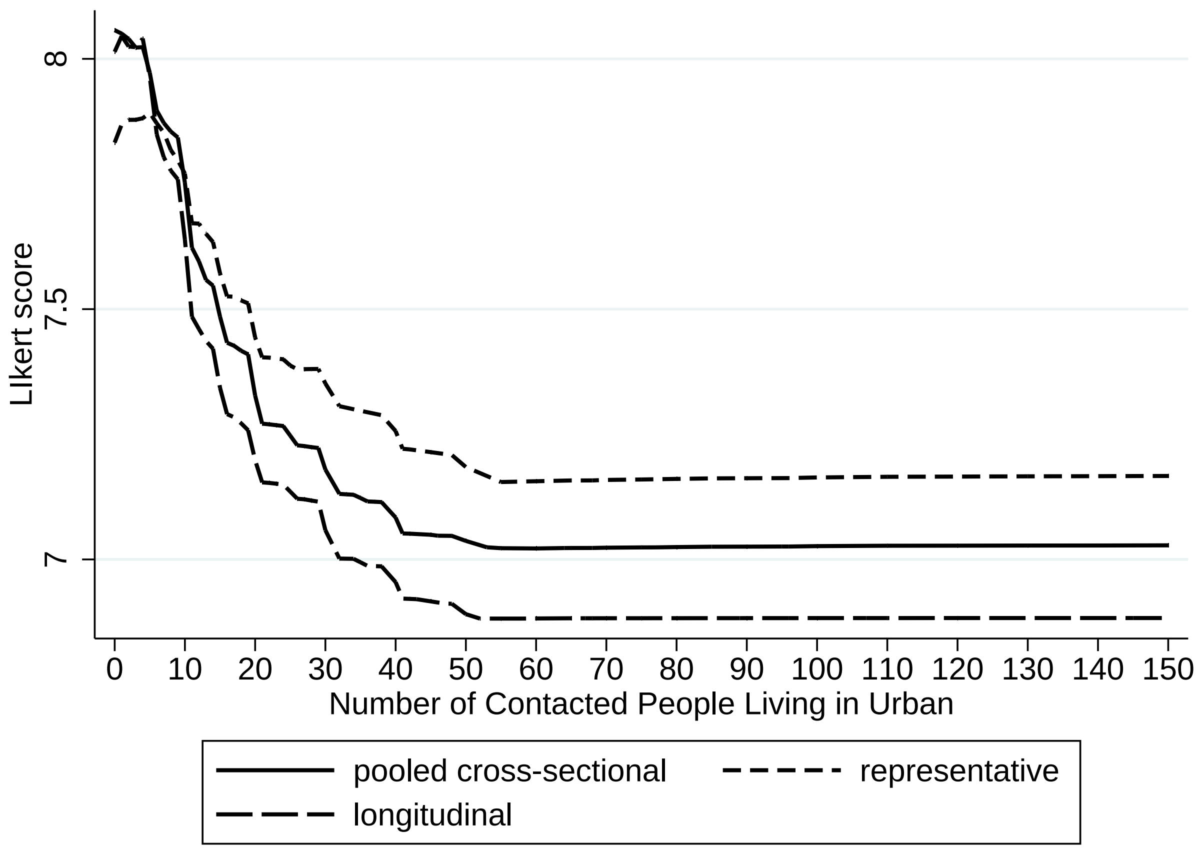
A.3 Data Source of Minimum Wage

The minimum wage data are extracted from the online websites. We browsed the following websites to obtain the minimum wage change.
http://www.btophr.com/v2/ic/city_low.shtml#ptop
http://www.hros.cn/zdgz/
http://www.labournet.com.cn/xinchou/zuidi/default.asp?number=gd#
http://www.51labour.com/zhuanti/0613/
http://www.360doc.com/content/11/0307/21/5079158 _ 99051974.shtml
http://news.hrloo.com/benzhan/19752.html
http://zx.cq.gov.cn/ydz/bmfw/26347bc2-d6d6-4964-a666-b612d6887135.shtml
http://www.cqhrss.gov.cn/u/cqhrss/cmd.shtml?action=search & keyword=
%D7%EE%B5%CD%B9%A4%D7%CA & submit.x=21 & submit.y=11
http://www.fl168.com/Lawyer9286/View/238686/
http://www.fl168.com/Lawyer9286/View/238681/
http://law.51labour.com/lawshow-11897.html
http://law.51labour.com/lawshow-36073.html
http://www.03964.com/read/15fc68bb0a70fc1600c5378b.html
http://www.ft22.com/shuju/2012-6/3462.html
http://www.03964.com/read/5ed5aeca628786848ab0215e.html
http://zhidao.baidu.com/question/394838325.html
http://www.updayday.com/guangzhou/79/0R43CO2012/
http://www.ycwb.com/ycwb/2006-09/01/content _ 1197676.htm
http://gz.bendibao.com/gzsi/2011919/si87494.shtml
http://365jia.cn/news/2011-07-06/555E4C5369162429.html
http://unn.people.com.cn/GB/14748/4860379.html
http://ah.cnpension.net/sbcx/2008-09-28/589372.html
http://ah.cnpension.net/sbzn/bb/gzdt/2008-11-08/656443.html
http://wenku.baidu.com/view/51a069ff910ef12d2af9e721.html
http://www.jshrss.gov.cn/sy/ldxxcx/200511/t20051113 _ 2486.htm
http://www.sundylawyer-aid.com/Item-132.aspx
http://www.njhrss.gov.cn/art/2012/8/28/art _ 2114 _ 54640.html
http://wgszq.blog.china.com/201006/6497348.html
http://wu-lawyer.blog.bokee.net/bloggermodule/blog _ viewblog.do?id=2725909
http://wgszq.blog.china.com/201006/6497350.html
http://www.zhlss.gov.cn/lder/zcfg/ldgz/20080916105011.htm
http://www.nbosta.org.cn/Html/pol _ zxzc/2010-04/02/1410010364643613100410021401.html
http://www.zh.gov.cn/zwgk/fggw/zcfg/201104/t20110408 _ 39427.htm
http://china.findlaw.cn/fagui/sh/23/40856.html
http://code.fabao365.com/law _ 313393.html
http://www.zhlss.gov.cn/lder/zcfg/ldgz/20060920102754.htm
http://www.zhlss.gov.cn/lder/zcfg/ldgz/20070929152727.htm
http://www.wenkudaquan.com/doc/20120621/289050.html
http://www.contracts.com.cn/news/page/14 _ 901 _ 2718.htm
http://www.china.com.cn/city/txt/2006-09/04/content _ 7130589.htm

http://sc.cnpension.net/shebao/chengdu/fuwu/620601.html
http://www.rc114.com/html/dispatch/2009/0624/3498.htm
http://cd.qq.com/a/20100127/001895.htm
http://www.ft22.com/shuju/2012-4/3296.html
http://edu.gongchang.com/f/zhichang-2011-09-30-22511.html
http://law.51labour.com/lawshow-6282.html
http://www.jinbw.com.cn/jinbw/xwzx/zzsx/201110213431.htm
http://panzhend.blog.163.com/blog/static/48903477201222444955449/
http://sz.bendibao.com/szsi/2008227/si62700.htm

B Appendix Figures

Figure B.1: Unconditional relationship between social networks and mental health problems for three samples



Source: 2008, 2009, 2011 and 2012 waves of the migrant survey in RUMiC project.