Journal of Population Economics (2020) 33:155–195
https://doi.org/10.1007/s00148-019-00748-3

ORIGINAL PAPER

Social networks and mental health outcomes: Chinese rural–urban migrant experience

Xin Meng[1] · Sen Xue[2]

Received: 8 March 2018 / Accepted: 2 August 2019 / Published online: 24 August 2019

Abstract
Over the past two decades, more than 160 million Chinese rural workers have migrated to cities to work. They are separated from their familiar rural networks to work in an unfamiliar, and often hostile, environment. Many of them thus face significant mental health challenges. This paper is the first to investigate the extent to which migrant social networks in host cities can mitigate these adverse mental health effects. Using unique longitudinal survey data from Rural-to-Urban Migration in China (RUMiC), we find that network size matters significantly for migrant workers. Our preferred instrumental variable estimates suggest that a one standard deviation increase in migrant city networks, on average, reduces the measure of mental health problems by 0.47 to 0.66 of a standard deviation. Similar effects are found among the less educated, those working longer hours, and those without access to social insurance. The main channel of the network effect is through boosting migrants' confidence and reducing their anxiety.

Keywords Mental health · Social networks · Migration · China

JEL Classification I12 · I15 · J61

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

Responsible editor: Klaus F. Zimmermann

✉ Sen Xue
sen.xue@hotmail.com

1 Research School of Economics, College of Business and Economics, Australian National University, Canberra, Australia

2 Institute for Economic and Social Research, Jinan University, Guangzhou, China

and cultures, and cope with various stresses. All these circumstances can influence migrants' health in general and their mental health in particular. Psychologists have long been studying migration impact on mental health conditions of international migrants. They normally find that the cultural shock, social isolation, and racial discrimination often put significant stress on immigrants and lead to high prevalence of ill health and psychological problems (e.g. Wen 1976; Hull 1979; Bhugra 2004).[1]

In a society where social networks play a crucial role in one's life, such as China, the migration process can be a very disruptive experience. In the past 20 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 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 birthplace; no mobility was allowed. One justification for the system was that China could not afford to provide to all citizens the social services and welfare that were available to urban residents. In the 1990s, Chinese economic growth accelerated, and the sudden inflow of foreign direct investment substantially increased the demand for unskilled labour. It was then that hukou restrictions began to ease. Twenty years on, moving from rural areas to urban cities to work has become a common 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 that one faced in most societies, namely losing familiar social networks and social environment, but also face significant discriminatory treatment. In addition, migrants in China often engage in manual labour jobs and work extremely long hours. Due to high work 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 urban non-migrant residents (Li et al. 2009; Chen 2011) and that more than 20% percent of rural migrants have clinically relevant mental problems (Mou et al. 2011; Qiu et al. 2011).[2]

Mental health is an important form of human capital, affecting productivity and well-being. 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 large costs on 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 and epidemiologists. In theory, social networks have both beneficial and harmful effects on mental health. They may improve mental health

[1]While migration may adversely affect migrants' mental health condition, migrants are often found to be a selected group who are in general mentally stronger (see, for example, Lou and Beaujot 2005 for Canada; Janisch 2017 for Australia).

[2]Despite these potential mental health challenges, a large number of rural people migrated to the city to work due mainly to the large wage gaps between rural and urban regions.

by creating a sense of social integration and buffering stress, but participation in social networks can also induce psychological costs from indebtedness and obligation pressures (Kawachi and Berkman 2001). A large body of psychological research examines the association between social networks and mental health for different populations (see the 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 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 generalisable to the effect of more natural networks of friends and acquaintances (Cohen 2004; Cohen and Janicki-Deverts 2009; Ertel et al. 2009).

In this study, we examine whether migrants' social networks in host cities can help protect them from mental health problems. We analyse a unique large-scale longitudinal survey from the Rural-Urban Migration in China (RUMiC) project, which collects mental health, social networks, and other detailed socio-economic information from a random sample of rural–urban migrants. Studies of social network effects on mental health often face two challenges: the potential endogeneity between mental health and social networks, and the difficulty in measuring natural social networks. The RUMiC data have three key advantages to help overcome these challenges. First, it has a measure which we can use to proxy *natural* social networks, as opposed to man-made networks, such as nurses or social workers in clinical trials. In our study, social networks are measured as the number of friends and acquaintances who lived/worked mainly in cities and who the individual visited/called/sent a letter, a WeChat message, or an email to during the last Chinese New Year period. It includes both migrants and urban locals.[3] We use the term "migrant host city social networks" to describe our network measure. Second, our data contain a large longitudinal component, which allows us to estimate a fixed-effects model. The key challenge to identifying the impact of social networks on mental health is endogeneity arising from omitted individual unobservable characteristics. The fixed-effects model can tease out the time-invariant unobservable individual characteristics and hence reduce this source of bias. Finally, migrants observed in the RUMiC survey are from around 1600 counties.[4] This offers an opportunity to find plausible exogenous variation across sending counties as instruments to further mitigate the potential endogeneity problem. In particular, we use past rainfall in migrants' home county and the distance between the home village and the closest transportation hub as the instrumental variables for city social networks. We discuss the validity of these instruments in Section 3.

Our study is the first to describe the general mental health condition among a large sample of Chinese rural-to-urban migrants and the first to examine the causal link between the size of social networks and mental health problems. Our results suggest that social networks in host cities indeed mitigate migrant mental health problem: on

[3] Based on our data, the average share of urban locals among migrants' social networks is around 32–34%.
[4] There are 2861 county-level jurisdictions in China.

average, a one standard deviation increase in social networks reduces mental health problems (measured by a GHQ Likert score) by 0.47 to 0.66 of a standard deviation. This is a very large effect. Further, the impact is strong not only for average migrants but also for disadvantaged groups. For example, among migrants whose education level is 9 or fewer years, a one standard deviation increase in the social network measure reduces mental health problems by 0.56 to 0.68 of a standard deviation. Similar results are also found for migrants who work longer hours and who do not have access to city social insurance. Our results also suggest that the effects of social networks mainly work through boosting migrants' confidence and reducing their anxiety.

~~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, brigades, or communes). This rural–urban segregation was enforced 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. With such a large income gap, the motivation for rural workers to move to cities was very strong. In the 1980s and early 1990s, governments in all urban cities tightly controlled rural–urban migration, even though the food rationing system was no longer in place 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, for example, there were fewer than 30 million rural workers who had 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 the hiring of migrant workers. Many cities even published a long list of occupations into which rural migrants could not be hired (Meng 2000, 2012; Cai et al. 2001).

It was not until the late 1990s, when foreign direct investment increased the demand for unskilled workers in cities, that the Chinese government began to loosen controls on rural–urban migration. The number of rural–urban migrants 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. For example, the majority of migrants in cities are not covered by social health and unemployment insurance and social pensions. Their children do not have equal access to attending childcare and schools 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 leave a familiar environment to embrace a new and unfamiliar surrounding. Migration to a hostile environment, such as the one faced by rural–urban migrants in China, is even more stressful. Several studies document the high prevalence of mental health 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 health condition of migrants is worse than their urban and rural counterparts. Li et al. (2009) compare migrants with their rural non-migrant counterparts from their home town and with urban residents in Beijing and find that both the urban and rural non-migrant residents are mentally healthier than migrants. Chen (2011) shows that the psychological distress of migrants in Beijing is worse than that of urban locals. There is only one exception to this pattern of results. Li et al. (2007) find that 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 in his study. In contrast, Phillips et al. (2009) find that the prevalence rate of mental illness of China's general population is around 13.3%. For the younger population group (age 18–39), which is more comparable to the migrant population, the rate is 9.32%, much lower than the above-mentioned findings for migrant population.

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.[5] 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. Thus, 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:

$$MHP_{ijct} = \beta_1 * SN_{ijct} + X_{ijct} * \beta_2 + W_{ct} * \beta_3 + D_{jt} * \beta_4 + C_c + \tau_t + \rho_j + \epsilon_{ijct}, \quad (1)$$

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 problems; SN_{ijct} is the measure of migrants' social networks in cities, X_{ijct} is a vector of individual characteristics; W_{ct} is a vector of sending community characteristics; D_{jt} is a vector of destination city time-variant characteristics; C_c is the sending county fixed effects, τ_t is the year fixed effects; ρ_j is the destination city fixed effects; and ϵ_{ijct} is the unobserved factor.

The variables included in X_{ijct} are individual's age, gender, education, marital status dummy variable for married and divorced individuals (with singles as the omitted category), number of children, height, self-reported health, years since the first migration, self-employment indicator, the number of people over 16 years old in the household, and individual occupation and industry affiliation dummy variables. These are commonly included variables in other economic studies of mental health (e.g. Stillman et al. 2009; Bjorklund 1985; Akay et al. 2012). W_{ct} 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 centre), 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. While Eq. 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 in D_{jt}: the growth rates of GDP and minimum wage in the destination city.[6,7]

Our goal is to identify β_1 in Eq. 1. If social networks are exogenous, the OLS estimate of β_1 indicates the impact of social networks on mental health. 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

[5]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.

[6]We choose to include the growth of GDP and minimum wage rather than the level, given the rapid economic development in China and the importance of the adaption of one's subjective well-being (Brickman et al. 1978; Di Tella et al. 2010).

[7]Our results are robust if we do not control for individual occupation and industry affiliation dummy variables and D_{jt}. The results are available upon request.

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 a large measurement error. In a panel setting, the possibility that individuals may change their (mis)reporting behaviour can further induce a measurement error. One indication of measurement error in our data is that 52% of respondents rounded their answers on network size to a multiple of 5. The other is the low correlation of social network numbers overtime within an individual. Among our panel sample, the autocorrelation coefficient is quite low (0.25), although it is statistically significant at the 0.1 per cent level. All of these factors indicate that the OLS estimate of β_1 from Eq. 1 could be inconsistent. It is important to note 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 likely to make friends and thereby tend to have larger networks, people with more mental health problems could also choose a destination with larger existing networks to relieve their stress.[8]

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

$$MHP_{ijct} = \beta_1 * SN_{ijct} + X_{ijct} * \beta_2 + W_{ct} * \beta_3 + D_{jt} * \beta_4 + \tau_t + n_i + \epsilon_{ijct}, \quad (2)$$

where n_i is the individual fixed effect, and X_{ijct} and W_{ct} include only the time-variant characteristics. Because destination city and sending county fixed effects are collinear with the individual fixed effects, in the FE model, we do not include C_c and ρ_j. The main advantage of the FE model is that the individual fixed effect, n_i, 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 problem can induce large attenuation bias in the FE estimator, especially when the misreporting behaviour over time is inconsistent. 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.[9]

An alternative way to mitigate endogeneity bias and circumvent the disadvantages of the FE estimator is to adopt the instrumental variable approach. The cross-sectional

[8] In this paper, we mainly investigate migrant mental health condition in the destination cities, and we take the decision of migration as given here. However, how the migration decision correlates to mental health and social networks is important. Due to the data unavailability (i.e. RUMiC Rural Household Surveys do not include most rural home towns of migrants surveyed in RUMiC Migrant Household Surveys), we are unable to fully address this self-selection issue. Nevertheless, in Section 5, we will discuss how this potential self-selection bias could affect our results.

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

instrumental variable approach jointly estimates (1) and an equation of social networks:

$$SN_{ijct} = \gamma_1 * Z_{ct} + X_{ijct} * \gamma_2 + W_{ct} * \gamma_3 + D_{jt} * \gamma_4 + C_c + \tau_t + \rho_j + e_{ijct}, \quad (3)$$

and the fixed-effects instrumental variable approach jointly estimates (2) and the following equation of social networks:

$$SN_{ijct} = \gamma_1 * Z_{ct} + X_{ijct} * \gamma_2 + W_{ct} * \gamma_3 + D_{jt} * \gamma_4 + \tau_t + n_i + e_{ijct}, \quad (4)$$

where Z_{ct} is the instrumental variable which identifies the effect of social networks, and e_{ijct} is the error term of the social networks equation. 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).[10] 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) 2 years before the interview 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 the need for migration 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 choice.[11] The results are presented in Appendix Table 7. We find that rainfall indeed increases agricultural income and consequently reduces rural people's migration propensity. 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 home town, the impact of rainfall on migration choice could eventually translate into an impact on network size.[12] It takes time to migrate and form social networks, so we use the spring and summer

[10] For the measurement error issue, as long as the instrumental variable is uncorrelated with the measurement error of the network measure, the IV estimator is consistent. In the FEIV case, the consistency holds when the demeaned IV is uncorrelated with the measurement error of the demeaned network measure.

[11] For detailed discussion of these data, see Appendix B.

[12] These newly formed networks in cities do not have to coincide with migrants' pre-migration existing networks. Chinese culture normally regards people one meets in a foreign land (for example, destination cities for most rural migrants) from one's own home town as more reliable, trustworthy, and hence more likely to form social network with, regardless of whether the person one met was known pre-migration.

rainfall 2 years before as the instrumental variable. The first-stage results are shown in Section 5 and as can be seen there this is a strong IV.

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 10-year average daily rainfall between 3 and 13 years before the survey and its squared term as measures for long-term rainfall in the W_{ct} vector as well as the sending county fixed effects in Eq. 1. We think it is reasonable to assume that the direct effect of rainfall on mental health is mainly shaped by long-term rainfall.[13] We expect that once the long-run rainfall is controlled for, transitory rainfall in migrants' sending counties 2 years before (the instrumental variable) should not have a direct effect on the current mental health of migrants in cities.[14]

Second, rainfall may affect migrants' mental health via its impact on agricultural income in their home towns. Home town agricultural income increase is likely to be positively associated with mental health of migrants (or negatively associated with migrant mental health problem, which is our current dependent variable). To mitigate this potential direct income effect from rainfall, we include the daily wages of day-labour in migrant sending villages in Eqs. 1 and 2.[15] Further, the literature also provides evidence that individuals can adapt to external income shocks (Di Tella et al. 2010), and in some cases income shock can be fully adapted within 1 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-effects 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.[16]

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

[13]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; O'Hare et al. 2016).

[14]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 finds that the one-day before rainfall variable is not statistically significant.

[15]As the agricultural income can be seen as the opportunity cost of non-agricultural work, the daily wages of day-labour in villages should be positively correlated with agricultural income and thereby can be treated as its proxy.

[16]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.

transportation hub. 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 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 transportation station is not correlated with the error term in Eq. 1. This assumption may not hold if there are omitted variables which are correlated with both error terms in Eqs. 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_{ct}) 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 to mitigate reverse causality and omitted variable problem. As these two instrumental variables are from a separate data source or a different question in the survey, they are unlikely to be correlated with the measurement error in social networks, and hence they can resolve the measurement error issue as well. 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 longitudinal data to document the socio-economic impact of internal rural–urban migration in China. The project comprises three independent surveys: the Migrant Household Survey (MHS) which interviews migrant households in destination cities; the Urban Household Survey (UHS) that covers city local households; and the Rural Household Survey (RHS) that is conducted in major migrant-sending rural regions. The surveys began in 2008 and were conducted annually. The MHS has 9 completed waves (2008–2016), while the UHS and RHS were terminated in 2011 due to funding constraints. In this paper, we use the 2008, 2009, 2011, and 2012 waves of the MHS.[17] The household in the MHS survey refers to migrants from the same family moving to and living together in the destination city.

The RUMiC MHS is currently the largest longitudinal survey of rural migrants in China. It covers 15 cities in 9 provinces/municipalities: Guangzhou, Dongguan, Shenzhen, Zhengzhou, Luoyang, Hefei, Bengbu, Chongqing, Shanghai, Nanjing, Wuxi, Hangzhou, Ningbo, Wuhan, and Chengdu. These 15 cities represent cities in both the largest migration sending and receiving provinces and cover the coastal, central, and western regions of China (Gong et al. 2008). Each wave of the MHS contains around 5000 migrant households. The MHS adopts workplace-based sampling strategy, which enables RUMiC survey to obtain the most representative sample of migrant workers in China relative to other surveys.[18]

Due to high mobility of migrants, the attrition rate has been large. To ensure the sample size, each year, the survey team drew a random representative refreshment sample to restore the sample size to around 5000 households. Thus, in addition to the *panel sample*, the sample in the baseline wave and the subsequent random sample of refreshments constitute a repeated cross-section of a *representative sample*. 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.[19] 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.

[17] The 2010 wave does not include information on migrants' mental health, so it is not included in the analysis. We stop at 2012 because of the availability of the rainfall data. Up to now, the fully cleaned rainfall data are only available up to 2010.

[18] In China, there is a large fraction of migrant workers living in workplaces, such as dormitories of manufacturing factories and construction sites. These migrants are usually not sampled (or undersampled) in the other migrant household surveys, because the sampling frames of those surveys are all based on residential address. For example, the proportions of self-employed migrants in CHIP 1999 and 2002 and CULS 2002 are 64%, 66%, and 52%, respectively, and the proportions of production workers among migrants in CHIP 1999 and 2002 are around 7%. In contrast, the 2005 Mini Census shows that only 20% of migrant workers are self-employed, and 55% are production workers. Given these large differences, the RUMiC project conducted a census on workplaces in the aforementioned 15 cities to record the number of rural migrants in each workplace address. Based on this census information, the RUMiC project team then constructed the sampling frame. This sampling strategy leads to a more representative sample of rural–urban migrants. For example, in 2008 MHS, 22% of migrants are self-employed, and 40% are production workers, which is closer to the 2005 Mini Census than other migrant surveys. For more details of the sample representativeness and the RUMiC workplace-based census, see Gong et al. (2008). After being sampled in the workplaces, the home address and other contact details (e.g. telephone number and hometown address) of respondents are recorded during the interview for tracking purpose. For more details of tracking, please see Xue (2015).

[19] We are fully aware that due to attrition the panel sample is not representative. However, the panel sample still represents an important group of migrants and provides us with useful information. Our analysis shows that the attrition in the MHS survey is mainly caused by return migration or migrants moving to other cities. The stayers are those migrants who are more likely to be married, are better educated, and are willing to stay in cities permanently if policy allows them to. They are also likely to have better mental health conditions and to have larger economic gain from migration. Thus, the panel analysis is relevant to this group of migrants.

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. The GHQ is widely used to screen for psychiatric disorders in psychological and medical studies. In 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). The 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).[20] 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 the GHQ 12. In our main analysis, we use the Likert score, which is the sum of all the answers to the questions in the GHQ 12 and then subtract 12.[21] Thus, the Likert score ranges from 0 to 36. The higher 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 A social network module is included in every wave of the MHS. Ideally, social networks should be measured in terms of both quantity (i.e. the number of network members) and quality (i.e. the help which the network can offer). However, the social network module in the MHS 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/wechat messages, 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?"

[20] The details of the GHQ 12 questions are presented in Appendix B.1.

[21] In the robustness check, we also consider the Caseness GHQ score, another measure of mental health problem that is used in the literature (e.g. Clark and Oswald 1994). The Caseness 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 Caseness 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.

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 the literature uses the number of people visited or sent cards to during the Christmas period to measure network size in Western society (Hill and Dunbar 2003).[22]

Other control variables All individual characteristic variables and most sending community characteristic variables used in this paper are drawn directly from the RUMiC MHS. 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 B).

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 2-year lagged spring and summer rainfall in the home county, and the 10-year average rainfall at the home county.[23]

4.3 Sample restrictions and summary statistics

The RUMiC migrant survey consists of around 5000 households and 8000–10,000 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

[22] 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 panel A in Table 6).

[23] Around 10.6% 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, 0.1% of respondents did not 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 km. 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. In the sample, around 1.7% of observations are from home counties with only a single observation. After matching rainfall information, we also merge these counties with the neighbouring counties. The results remain similar if we do not merge these home counties and are available upon request.

of the survey. Thus, the sample used for the analysis below is restricted to these respondents.[24]

In the four waves of data we use, 23,894 people have a GHQ 12 score available. 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 17,711 observations.[25] Finally, to reduce measurement error, we remove 58 observations who reported having more than 150 city contacts. This restriction is based on Dunbar (1993) and Hill and Dunbar (2003), where they observe that the maximum network size of human beings is approximately 150 people due to cognitive limit. The remaining final sample is 17,653 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 respondents who 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 panel sample to provide fixed-effects estimates.

Table 1 presents summary statistics for the three samples. 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 Likert score is similar across three samples, but the panel individuals have slightly more connections relative to the other two samples. This may be due to a higher attrition rate for individuals with fewer social connections. 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 Fig. 4, 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 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 (i.e. respondents whose network size is equal to or smaller than 150) to provide a complete picture, avoiding potential sample selection problem, and make the first stage of the IV estimation strong enough as well.[26]

---

[24] 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.

[25] Of all the variables, the home village daily wage for day labour has the most missing values, accounting for more than 63% or 1130 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 and are available upon request.

[26] We also check the robustness by excluding the observations whose contacts are more than 50, and the results are shown in panel D of Table 6. 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.

Table 1 Summary statistics

Variable	Panel A: pooled sample		Panel B: representative sample		Panel C: panel sample	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Likert score	7.71	4.53	7.64	4.50	7.68	4.53
No. of contacted people living in urban areas	13.48	17.57	13.05	17.60	13.82	17.55
Age	31.63	10.39	30.38	10.33	33.12	10.13
Male	0.64		0.64		0.65	
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	
Single	0.41		0.47		0.34	
Married	0.57		0.52		0.64	
Divorced	0.02		0.02		0.02	
Years since the first migration	8.90	6.89	7.83	6.66	10.00	6.85
Self-employment	0.22		0.18		0.27	

Table 1 (continued)

Variable	Panel A: pooled sample		Panel B: representative sample		Panel C: panel sample	
	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
No. of family members over 16 years old	1.57	0.79	1.45	0.71	1.68	0.83
No. of children	0.84	0.90	0.76	0.90	0.94	0.89
Daily wage of unskilled labour at home village (yuan)	56.49	25.25	50.24	22.85	59.96	25.50
Distance between home village and the closest county (km)	24.96	35.69	25.81	36.69	24.91	33.41
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 $t-12$ to $t-3$ (mm)	2.98	0.90	3.02	0.94	2.94	0.85
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 between April and Aug at $t-2$ (mm)	4.75	1.81	4.71	1.89	4.74	1.72
Distance between home village and the closest traffic centre (km)	16.25	32.76	16.54	33.90	15.73	30.80
Observations	17653		10955		8756	
Average number of waves of the longitudinal sample					2.4	

The pooled sample consists of all the observations across waves. The representative sample consists of the 2008 wave and new households in each wave after 2008. The panel sample consists of individuals who appeared in two or more waves of the survey

Source: 2008, 2009, 2011, and 2012 Migrant Household Surveys of the RUMiC project

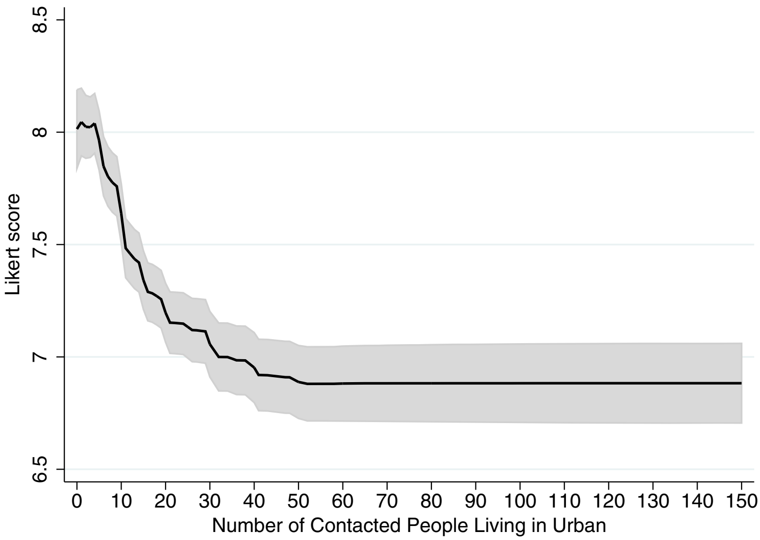

Fig. 1 Unconditional relationship between social networks and mental health problems. Source: Pooled sample from the 2008, 2009, 2011, and 2012 waves of the Migrant Household Survey in RUMiC project

Table 1 also shows that the average ages of respondents in the pooled, representative, and panel samples are 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. Migrants in the panel sample are more likely to be married than those in the representative sample (64% vs. 52%). On average, the respondents first migrated 9 years ago, but for the panel sample this number is slightly higher (10 years). A total of 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 1.5 to 1.7 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 km, home villages of 89% of respondents have a medical centre, home villages of more than 20% are located in the mountainous areas, and the average daily wage for an unskilled labourer is around 50 to 60 yuan.[27] The 10-year average daily rainfall is around 3 mm.

Finally, Table 1 also shows the summary statistics for the two instruments. The average 2-year lagged spring to summer daily rainfall is around 5 mm, and the

[27] One Chinese yuan is equal to about 0.14 to 0.16 US dollars during 2008 to 2012.

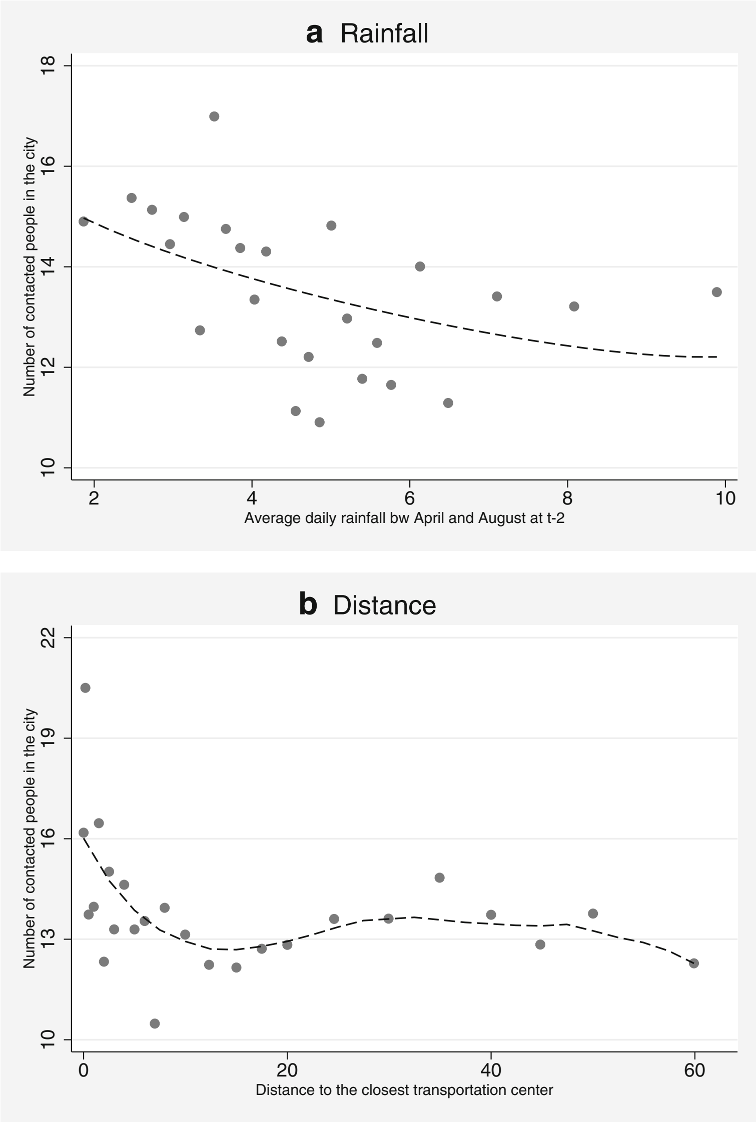

Fig. 2 Unconditional relationship between two IVs and size of networks. According to the magnitude of the instrumental variable, we partition the sample into 25 equal-sized cells. Each dot in the graph represents the average values of the instrumental variables and network size within one cell. Source: Pooled sample from the 2008, 2009, 2011, and 2012 waves of the Migrant Household Survey in RUMiC project

distance between home village and the closest transport station is around 16 km. Figure 2 presents the unconditional relationship between the social networks and the two IVs, separately. It is clear that the lagged rainfall is negatively related to migrant network size in cities, and there is an inverse relationship between the distance IV and network size, so in the following analysis we use the inverse of 1+distance as the instrumental variable.[28]

5 Results

In this section, we estimate (1) and (2) using OLS, fixed-effects, and instrumental variable approaches. The standard errors of the IV estimation, including both cross-sectional IV estimation and FEIV estimation, are clustered at the sending county level, as the rainfall instrument varies at sending-county-year cells. The standard errors in the non-IV estimations are clustered at the household level.

5.1 OLS and fixed-effects results

Table 2 presents two sets of the main results: the first set (panel A) directly uses the Likert score as the dependent variable and the number of friends and acquaintances in one's networks as the main independent variable. The second set (panel B) uses standardised mental health and social network size (with a mean of 0 and a standard deviation of 1) for ease of interpretation. Columns (1)–(3) report the OLS estimates for the three samples, pooled, representative, and panel, respectively, while column (4) reports the fixed-effects estimation using the panel sample.

We observe that migrant city networks are significantly and negatively correlated with mental health problems in all three samples. The estimate using the pooled sample suggests that one additional person in one's social networks in host cities is associated with a reduction in the Likert score by 0.016 for the pooled sample, which is equivalent to 0.4% of the standard deviation of the Likert score. The same results using standardised mental health and social networks (column (1) in panel B) suggest that one standard deviation increase in social networks is related to 0.06 standard deviation reduction in mental health problems.

The associations of other statistically significant control variables all have sensible signs and magnitudes (see Appendix Table 8). Male migrant workers have fewer mental problems than their female counterparts. More educated migrants tend to be mentally healthier. Relative to individuals with primary school education, those with junior high school education have fewer mental health problems (0.4 to 0.6 of a Likert score), while for those with senior high school degrees or above, the Likert score decreases by 0.7 or more. Married people have better mental health than singles, whereas divorced people have worse mental health than singles. This is perhaps because spouses usually provide emotional support for each other (Smith and Christakis 2008), or because mentally healthier people are selected into marriage. The self-rated health level is strongly correlated with the Likert score, which is similar to Akay et al.'s (2012) findings. In summary, the OLS results for the pooled

[28] Around 5–6% of respondents reported the distance as 0 km. To include these observations, we add 1 km to the distance to make the inverse feasible. We also tried to add the logarithm of 1 km plus distance or the square term of distance in the regression. But the first stages are weak.

Table 2 Selected OLS and fixed effect estimates of network effect on mental health problems

	OLS			FE
	Pooled sample	Representative sample	Panel sample	Panel sample
	(1)	(2)	(3)	(4)
Panel A: Estimates from raw network and mental health variables				
No. of contacted people living in cities	− 0.016***	− 0.017***	− 0.013***	− 0.009**
	(0.002)	(0.003)	(0.003)	(0.004)
Observations	17653	− 10955	8756	8756
Adjusted R^2	0.124	0.140	0.118	0.057
Panel B: Estimates from standardised network and mental health variables				
Standardised no. of contacted people living in cities	− 0.060***	–0.065***	–0.049***	–0.033**
	(0.008)	(0.010)	(0.012)	(0.014)
Observations	17653	10955	8756	8756
Adjusted R^2	0.124	0.140	0.118	0.057

Standard errors are clustered at household level. *Significant at 10% level; **significant at 5% level; ***significant at 1% level. In the OLS regressions, the other covariates included are age and its square, male, years since the first migration and its square, education attainment dummies, marriage and divorce dummies, number of children, height, self-reported health dummies, self-employment, household size, daily wage of unskilled labour at home village, inverse of 1+ the distance between home village and the closest county, home village has medical station, average daily rainfall from $t-12$ to $t-3$ and its square, growths of GDP and minimum wage in the destination cities, industry and occupation dummies, home county dummies, destination city dummies, year dummies, and a constant term. In the fixed-effects regression, the other covariates included are marriage and divorce dummies, number of children, height, self-reported health dummies, self-employment, household size, daily wage of unskilled labour at home village, growths of GDP and minimum wage in the destination cities, industry and occupation dummies, year dummies, individual fixed effects, and a constant term. The reference group is unmarried females with education below junior high school and excellent health. The pooled sample consists of all the observations across waves. The representative sample consists of the 2008 wave and new samples in each wave after 2008. The panel sample consists of individuals who appeared in two or more waves

Source: 2008, 2009, 2011, and 2012 Migrant Household Surveys of the RUMIC project

sample are largely consistent with other studies about internal migrants in China (e.g. Chen 2011; Akay et al. 2012, 2013), as well as international migrants in developed countries (e.g. Lou and Beaujot 2005; Chadwick and Collins 2015; Janisch 2017).

The OLS estimates using the representative sample (column (2)) and the panel sample (column (3)) are generally similar to those obtained using the pooled sample. While the estimate for the representative sample is slightly larger, and the one for the panel sample is smaller, the statistical significance levels are the same.

Once individual fixed effects are controlled for (column (4)), the estimated coefficient becomes even smaller (– 0.009 Likert score, which is equivalent to 0.2% of the standard deviation of the Likert score). As discussed in Section 3, this could be due to two reasons. First, when the independent variable (social network) is measured with error, 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. Using introversion as an example, introverted people have fewer contacts and are more likely to have mental health problems. Thus, without controlling for it, we would observe 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 further investigate the validity of the rainfall instrument. To this end, we conduct a falsification test, which examines the impact of the rainfall instrument on the number of contacts who have city hukou. The idea is that the 2-year lagged rainfall in hometown is supposed to be unrelated to city hukou networks. If the instrumental variable is correlated with city hukou networks, the lagged rainfall back home may somehow be related to migrants' city life through a channel other than social networks. Hence, it is likely to be associated with the error terms in Eqs. 1 and 2 and violate the exclusion restriction. The results are presented in Table 3. They show that such a relationship does not exist in either of the two samples (see panel A). 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 B). The stark difference between the two estimates indicates that it is not likely that the lagged rainfall variable directly affects migrants' city life through a channel other than social networks.

Table 4 presents main results of the cross-sectional IV and fixed-effects IV estimations using different instrumental variables.[29] The first column show the first-stage

[29]Because the IV estimates generated by the pooled sample and the representative sample are largely similar, in this section, we report the IV estimates based on the pooled sample and the panel sample. The results using the representative sample are available upon request from the authors.

Table 3 A falsification test on rainfall instrumental variable

	OLS model pooled sample (1)	FE model panel sample (2)
Panel A: No. of contacted people with city Hukou		
Average daily rainfall between April and Aug. at $t-2$	0.002	0.005
	(0.066)	(0.096)
Observations	17643	8750
Adjusted R^2	0.079	0.015
Panel B: No. of contacted people living in the city		
Average daily rainfall between April and Aug. at $t-2$	− 0.688***	− 0.690***
	(0.162)	(0.203)
Observations	17653	8756
Adjusted R^2	0.106	0.029

Standard errors are clustered at the home county level. *Significant at 10% level; **significant at 5% level; ***significant at 1% level. The other covariates included are the same as those in Table 2. The pooled sample consists of all the observations across waves. The panel sample consists of individuals who appeared in two or more waves. The daily rainfall variable is measured in mm

Source: 2008, 2009, 2011, and 2012 Migrant Household Surveys of the RUMiC project

results. The Kleibergen–Paap F statistic presented at the bottom of each panel indicates the strength of the instruments (Kleibergen and Paap 2006). Panels A to C show the results about the cross-sectional IV estimation. The results suggest that both instrumental variables are strongly correlated with social network size. Specifically, panel A shows that a 1-mm increase in the 2-year lagged average daily rainfall between April and August reduces the social network size by 0.69 person. This means that an average of 4.75-mm lagged daily rainfall contributes 3.27 persons to a network, accounting for around 24% of migrant average networks or 0.19 standard deviations of migrant networks.[30] Panel B 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 networks. The test statistics of these two individual instrumental variables both exceed 10, indicating that the instrumental variables are strong. Panel C shows the results using the two IVs jointly. Compared with the first two panels, both the magnitudes and statistical significances of the coefficients in panel C remain similar,

[30]These calculations are based on the fact that the mean value of the rainfall instrumental variable is 4.75 mm, and the average size and the standard deviation of migrant networks are 13.48 and 17.57, respectively, as shown in Table 1.

Table 4 IV estimates of network effect on mental health problems

	First stage	Second stage	
	Unstandardised measures (1)	Unstandardised measures (2)	Standardised measures (3)
Panel A: Cross-section IV estimation—only using rainfall IV			
Rainfall	– 0.688***		
	(0.162)		
No. of contacted people living in cities		– 0.089*	– 0.347*
		(0.053)	(0.207)
Kleibergen–Paap F statistics	18.033		
P value of robustified Durbin-Wu-Hausman test		0.144	0.144
Observations	17,653	17,653	17.653
Panel B: Cross-section IV estimation—only using distance as IV			
Distance	3.122***		
	(0.951)		
No. of contacted people living in cities		– 0.154**	– 0.597**
		(0.063)	(0.244)
Kleibergen-Paap F statistics	10.784		
P value of Robustified Durbin-Wu-Hausman test		0.003	0.003
Observations	17653	17653	17653
Panel C: Cross-section IV estimation—using the two IVs			
Distance	3.094***		
	(0.948)		
Rainfall	– 0.682***		
	(0.161)		
No. of contacted people living in cities		– 0.121***	– 0.471***
		(0.041)	(0.160)
Kleibergen-Paap F statistics	12.441		
P value of Robustified Durbin-Wu-Hausman test		0.002	0.002
Observations	17,653	17,653	17,653

Table 4 (continued)

	First stage	Second stage	
	Unstandardised measures (1)	Unstandardised measures (2)	Standardised measures (3)
Panel D: FEIV estimation—using rainfall IV			
Rainfall	− 0.690***		
	(0.203)		
No. of contacted people living in cities		− 0.171**	− 0.661**
		(0.081)	(0.313)
Kleibergen-Paap *F* statistics	13.819		
P value of Robustified Durbin-Wu-Hausman test		0.017	0.017
Observations	8756	8756	8756

Standard errors are clustered at home county level. *Significant at 10% level; **significant at 5% level; ***significant at 1% level. The other covariates included are the same as those in Table 2 (also see the description in Section 4 or Appendix Table 8). The results in panels A to C are obtained from the pooled sample which consists of all the observations across waves. The results in panel D are obtained from the panel sample which consists of the individuals appearing in two or more waves. Column 1 shows the first-stage results based on the unstandardised measures of the IVs and network variable. Columns 2 and 3 show the second-stage results based on the unstandardised and standardised measures of the network and mental health variables, respectively. The daily rainfall variable is measured in mm. The robustified Durbin-Wu-Hausman test is provided in the Stata package "ivreg2" and "xtivreg2" and derived from the C test

Source: 2008, 2009, 2011, and 2012 Migrant Household Surveys of the RUMiC project

and the Kleibergen–Paap F statistic is also above 10.[31] Thus, these two instrumental variables are not correlated with each other, and each of them provides different identification information to the second-stage regression.[32] Hence, for the cross-sectional analysis, jointly using these two instrumental variables makes the 2SLS estimation more efficient.

The first column of panel D presents the first-stage results of the fixed-effects IV estimation. The coefficient is similar to the cross-sectional IV results, suggesting that the rainfall instrument is largely orthogonal to the individual fixed effects. The Kleibergen–Paap *F* statistic again is greater than 10, indicating the rainfall instrument is strong enough to identify the effect of social networks in the fixed-effects IV estimation.

Columns 2 and 3 of Table 4 show the results of the second-stage regressions, based on the unstandardised and standardised measures of networks and mental health

[31]Note that Stock and Yogo (2005) suggest that in the two IV cases the critical value of the *F* test for the instruments being strong is 19.93 when standard error is not clustered and is not corrected for heteroskedasticity (i.e. the plain standard error). In our case, if we use the plain standard error, the *F* statistic is about 29.

[32]The correlation coefficient of the two instrumental variables is − 0.05.

problems, respectively. 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. Panels A and B show that when using the single IV, each additional friend the migrant has reduces his/her mental health Likert score by 0.089 to 0.154 points. In terms of standard deviation, this indicates that a standard deviation increase in social networks reduces the mental health Likert score by 0.35 to 0.6 standard deviations. Panel C, using the instruments jointly, produces the coefficient of − 0.121 Likert points, suggesting that a standard deviation increase in social networks decreases the mental health Likert score by 0.47 standard deviation. The estimates in panel C are statistically significant at the 1% level. 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 C are preferred in the cross-sectional analysis.

Panel D of Table 4 provides the FEIV estimates. As the distance instrumental variable does not vary across different waves of the survey, the FEIV estimation employs only lagged rainfall as the instrumental variable. The results again suggest that social networks help reduce mental health problems. In particular, the coefficients are around − 0.171 and − 0.661 for the unstandardised and standardised measures, respectively, and they are statistically significant at the 5% level. Compared with the cross-sectional IV estimation, FEIV estimation has two advantages. 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.[33] However, we do realise that the respondents in the panel sample may not be representative. The migrants in the panel sample are older, more likely to be self-employed, and have longer migration history than those in the representative sample. More importantly, they tend to be mentally healthier as respondents with fewer mental health problems are more likely to stay in the panel (see Appendix Fig. 5). In light of this caveat, we conjecture that our FE and FEIV results may be lower bound estimates as social networks are likely to have a stronger effect on those who are less mentally healthy.[34]

Relative to the OLS and fixed-effects results, the magnitudes of IV results are larger; and the Durbin-Wu-Hausman test shows that the OLS estimates are statistically significantly different from our preferred IV estimates (i.e. FEIV and the cross-section IV estimates jointly using the two instruments). This is similar to

[33] The FEIV estimates are larger than the IV estimates in panel A of Table 4. 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.

[34] The comparison of the results between the representative sample (column (2)) and the panel sample (column (3)) of Table 2 is indicative of the possibility that social networks have a stronger effect on those less mentally healthy people. The coefficient from the representative sample is larger than the panel sample, which has fewer people who have mental problems, due to attrition.

Munshi's (2003) study, where he finds that the IV estimates of network effects on employment and obtaining higher paid jobs are larger than the OLS and fixed-effects associations. The larger IV estimates may be due to three reasons. First, mentally unhealthy people probably 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.[35] Second, measurement error in social networks is large, which creates substantial attenuation bias in OLS and fixed-effects estimates. As mentioned in Section 3, 52% of respondents rounded their answers on social network to a multiple of 5. Last, the nature of the IV and FEIV estimates may also contribute to the difference. In general, IV estimation identifies the local average treatment effect (LATE) (Imbens and Angrist 1994). The effect identified by our instrument is for the group whose host city social networks are affected by the rainfall variation or distance to the closest transport centre at the rural hometown. Thus, the predicted social networks in the second stage of estimation only reflect the networks from the migrants' home town. It is likely, for example, that networks from one's home town play a more important role in generating mental stability than networks from urban local people or acquaintances acquired after migrants moved to the host city. If so, the LATE effect could be larger than the OLS estimates. Given these three possibilities, the OLS and FE estimates can be biased downwards and not completely comparable with the IV and FEIV estimates.[36]

5.3 Differential social network impacts on items of GHQ 12

Do social networks affect different items of the GHQ 12 differentially? In this subsection, we investigate this issue. Graetz (1991) finds that the twelve questions in the GHQ 12 can be grouped into three dimensions of mental health: anxiety (lost sleep, under strain, felt not able to overcome difficulty, depressed), loss of confidence (lack of self-confidence, felt worthless), and anhedonia and social dysfunction (lost concentration, felt not able to play a part in things, not able to make decisions, uninterested in day-to-day activities, could not face up to problems, were not happy). Following this classification, we put the items belonging to the same dimension

[35] Note that this possibility can only explain the difference between the OLS and cross-sectional IV estimates and cannot explain the difference between the FE and FEIV estimates.

[36] There is another issue which needs to be discussed before we turn to further analysis. In the empirical strategy section, we assumed that the migration decision is exogenous. This, however, may not be the case. If it is the case that mentally unhealthy people, with a lower possibility of establishing social networks in cities, are also less likely to migrate, the results obtained from our sample are likely to underestimate the true effect of social networks, as social networks seem to have a larger effect on mental health for less mentally healthy people (see the comparison between the representative and panel samples of Table 2). If on the contrary, mentally unhealthy people are more likely to migrate who also have smaller social networks in cities, our results could be overestimates of the true effect. This latter case is less intuitive though.

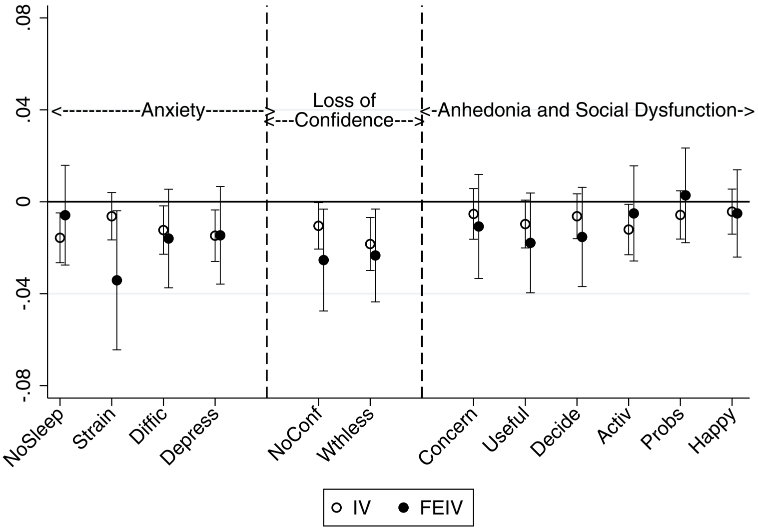

Fig. 3 The IV and FEIV estimates on GHQ 12 items. The point estimates and 95% confidence intervals of the effect of social networks on each GHQ item are presented. The regression specification is the same as those in Table 4, except that the dependent variable is one of the GHQ items. The IV estimates are produced by the pooled sample which consists of all the observations across waves, and the FEIV estimates are produced by the panel sample which consists of individuals who appeared two or more waves. The IV estimates are obtained by using rainfall and distance instruments jointly, and the FEIV estimates are obtained by using rainfall instrument only. Source: the 2008, 2009, 2011, and 2012 waves of the Migrant Household Survey in RUMiC project

together and present the IV and FEIV estimates in Fig. 3.[37,38] The results suggest that social networks mainly help boost confidence of migrants and reduce their anxiety. For items grouped in "anhedonia and social dysfunction", the effect of social networks is also beneficial, but the estimates are small in the magnitude and not precisely measured.

5.4 Analysis of subsamples

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 subsamples: low-educated migrants (individuals with junior high school education or below), migrants who work long hours (weekly working hour> 50), and migrants

[37] We also constructed the Likert score of each dimension by adding up the relevant items and estimated the effect of social networks. The results show that for IV estimation, the estimated effects for all three dimensions are negative and statistically significant. For FEIV, the results are precisely estimated for anxiety and confidence, but not for anhedonia and social dysfunction. The full results are available upon request from the authors.

[38] Cornaglia et al. (2015) also follow this classification to investigate the relationship between mental health and education decisions.

Table 5 Analysis of disadvantaged migrant groups

	OLS	FE	IV	FEIV
Panel A: Full sample				
No. of contacted people living in cities	− 0.016***	− 0.009**	− 0.121***	− 0.171**
	(0.002)	(0.004)	(0.041)	(0.081)
Kleibergen-Paap *F* statistics			12.441	13.819
Observations	17,653	8756	17,653	8756
Individuals		3651		3651
Panel B: Migrants with junior high school education or below				
No. of contacted people living in cities	− 0.018***	− 0.009*	− 0.144***	− 0.175**
	(0.003)	(0.005)	(0.055)	(0.080)
Kleibergen-Paap *F* statistics			10.234	18.868
Observations	11,312	5216	11,312	5216
Individuals		2188		2188
Panel C: Migrants with long working hours (weekly working hours >50)				
No. of contacted people living in cities	− 0.015***	− 0.010*	− 0.140***	− 0.178**
	(0.003)	(0.005)	(0.047)	(0.083)
Kleibergen-Paap *F* statistics			10.750	13.374
Observations	11,727	4949	11,727	4949
Individuals		2130		2130
Panel D: Migrants without access to welfare				
No. of contacted people living in cities	− 0.017***	− 0.014***	− 0.120***	− 0.259**
	(0.003)	(0.005)	(0.036)	(0.115)
Kleibergen-Paap *F* statistics			15.319	10.365
Observations	11,989	4983	11,989	4983
Individuals		2144		2144

Standard errors are clustered at household level for the non-IV estimates and home county level for the IV and FEIV estimates. *Significant at 10% level; **significant at 5% level; ***significant at 1% level. The other covariates included are the same as those in Table 2 (also see the description in Section 4 or Appendix Table 8). All the OLS and IV regressions use the pooled sample which consists of all the observations across waves, and the FE and FEIV regressions use the panel sample which consists of individuals who appeared in two or more waves. The IV estimates are produced by rainfall and distance instruments jointly, and the FEIV estimates are produced by rainfall instrument only

Source: 2008, 2009, 2011, and 2012 Migrant Household Surveys of the RUMiC project

with no access to social welfare in cities. Table 5 shows the results. For the ease of comparison, we include the results of the full sample in panel A.

Low-educated migrants are more likely to be prone to poor mental condition (see the previous results in Appendix Table 8 and Miech et al. 1999). Kawachi and Berkman (2001) stress 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 low-educated migrants? We examine this issue in panel B and restrict 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 helpful for the low-educated migrants.

One important feature of migrants is that they tend to work very long hours. In our sample, a migrant works an average of 62 h per week, which is far more than the usual working hours of urban residents (40 h 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 C, we examine whether social networks are still protective of migrants who work long hours. Restricting the sample to respondents who work more than 50 h per week, we find that social networks are still protective for the migrants who work long hours.

In panel D, 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 pooled sample, 68% of migrants have no access to unemployment and health insurance or pension.[39] This makes them vulnerable to shocks in life, which, in turn, could cause mental stress. For the pooled sample, for example, those migrants who have no access to social insurances have an average Likert score of 7.93, while for those with access to social insurances the average Likert score is 7.23, and the difference is statistically significant at the 1% level. Thus, it is important to understand whether social networks can mitigate the mental health problems for migrants who do not have access to social insurances. Restricting the sample to migrants who have no access to pension, medical and unemployment insurance revealed a larger effect of social networks on mental health for this group. This suggests that social networks are more responsive among migrants with no access to social welfare.

5.5 Robustness check

In this subsection, we conduct several sensitivity tests. The results are presented in Table 6.

First, we examine whether our results are sensitive to additional control variables. To this end, we add to the existing controls 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". These variables could affect migrant mental health and social networks at the same time and hence may be considered to generate omitted variable problem for us. Regarding working hours, as discussed in Section 5.4, studies have found that long working hours are associated with mental health problems (Sparks et al. 1997; Kuroda and Yamamoto 2016) and this is also the case for Chinese migrants as well (Frijters et al. 2009). Further, long working hours could impinge on

[39]The representative sample and panel sample have similar proportions of migrants who do not have access to social welfare.

Table 6 Robustness check

	OLS	FE	IV	FEIV
Panel A: Adding additional controls				
No. of contacted people living in cities	− 0.015***	− 0.008**	− 0.129***	− 0.166**
	(0.002)	(0.004)	(0.047)	(0.083)
Kleibergen-Paap F statistics			11.793	14.028
Observations	17,261	8375	17,261	8375
Individuals		3510		3510
Panel B: Removing subjective health controls				
No. of contacted people living in cities	− 0.014***	− 0.007*	− 0.109***	− 0.160**
	(0.002)	(0.004)	(0.042)	(0.079)
Kleibergen-Paap F statistics			12.302	14.136
Observations	17,653	8756	17,653	8756
Individuals		3651		3651
Panel C: Mental health problem measure—GHQ score				
No. of contacted people living in cities	− 0.004***	− 0.001	− 0.038***	− 0.068**
	(0.001)	(0.001)	(0.014)	(0.032)
Kleibergen-Paap F statistics			12.441	13.819
Observations	17,653	8756	17,653	8756
Individuals		3651		3651
Panel D: Excluding respondents whose contacts exceed 50				
No. of contacted people living in cities	− 0.024***	− 0.012**	− 0.236*	− 0.342
	(0.003)	(0.006)	(0.123)	(0.212)
Kleibergen-Paap F statistics			5.195	6.423
Observations	17042	8319	17042	8319
Individuals		3494		3494
Panel E: All household members				
No. of contacted people living in cities	− 0.015***	− 0.009***	− 0.138***	− 0.235**
	(0.002)	(0.003)	(0.045)	(0.100)
Kleibergen-Paap F statistics			11.365	12.047
Observations	20,575	10,210	20,575	10,210
Individuals		4233		4233

Standard errors are clustered at household level for the non-IV estimates and home county level for the IV and FEIV estimates. *Significant at 10% level; **significant at 5% level; ***significant at 1% level. The other covariates included are the same as those in Table 2 (also see the description in Section 4 or Appendix Table 8). All the OLS and IV regressions use the pooled sample which consists of all the observations across waves, and the FE and FEIV regressions use the panel sample which consists of individuals who appeared in two or more waves. The IV estimates are produced by rainfall and distance instruments jointly, and the FEIV estimates are produced by rainfall instrument only

Source: 2008, 2009, 2011, and 2012 Migrant Household Surveys of the RUMiC project

individuals' ability to socialise. It is also found that income is a significant predictor of migrants' mental health problems (Akay et al. 2012); and as socialising is costly, lack of income could also affect one's ability to develop social networks. Separating from a spouse or children can often affect migrant mental health problems (Li et al. 2007; Mou et al. 2011); at the same time, the loneliness brought about by the separation from one's family may encourage individuals to seek for other social connections. Migrants who need to remit their income back home may be under pressure to work harder than otherwise, and like the effect of working hours, they may also lack time and financial resource to socialise. Finally, strong home networks may be associated with better initial mental health conditions and also to some extent offset the lack of the city social networks. The reason that these variables were not included in the main specifications is that they may cause an additional endogeneity problem. Panel A of Table 6 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.[40]

Second, to be consistent with the literature (see, for example, Akay et al. 2012), we include self-reported health in our main specification to act as a proxy for physical health. However, it is possible that self-reported health (or the physical health) is affected by one's mental health, hence, endogenous. Panel B reports results from the specification excluding the self-reported health variable, and the results are robust to this exclusion.

Third, instead of using Likert score, we also adopted the Caseness GHQ score of GHQ 12 as the dependent variable. As discussed in Section 4, Caseness 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 our dependent variable. The results in panel C confirm that our main findings in Tables 2 and 4 are robust. The choice of dependent variable does not substantively alter the results.

Fourth, as Fig. 1 shows that the negative relationship 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 fewer to test the sensitivity of our results. Panel D shows that the magnitudes of the estimates become larger, but the instruments become weaker, probably due to the smaller variation in the endogenous variable and instruments. We conduct the Anderson–Rubin Wald test, which is robust to the weak IV problem (Anderson and Rubin 1949; Stock and Wright 2000). It suggests that the IV and FEIV results are still significant at the 5% level.

Finally, the sample we used so far includes individuals who answered GHQ 12 and the social network question. As the social network questions were asked only to one household member (either the household head or his/her spouse), our current sample excludes the rest of the family members. We now test whether the results are sensitive if we include other household members and assume that the household has the same social networks. Panel E 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

[40] We also further include the number of hukou friends to check the robustness, and the results remain similar and are available upon request.

one-third of the Chinese urban labour force and 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 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 a real effect. In the meantime, understanding other informal channels which may help alleviate migrant mental health problems is of significant policy relevance.

Our study reveals that the migrant social networks in cities are indeed helpful with regard to mitigating migrant mental health problems. Our preferred results from the IV estimates suggest that, on average, one standard deviation increase in social networks reduces the measured mental health problem by 0.47 to 0.66 standard deviation. The effect is not only found at the mean. The migrant social network is also an important factor in mitigating metal health problem among disadvantaged migrants: those with less education, those working long hours, and those without social insurances in cities. Further examinations of detailed GHQ questions suggest that social networks mainly help increase confidence and reduce anxiety of migrants.

Mental health is an important form of human capital, and mental health problems often impose substantial costs on society (Frijters et al. 2009; Olesen et al. 2012). Given that migrants form a large proportion of China's urban workforce, and contribute significantly to China's economic growth, it is important that city governments pay attention to migrant mental health. Fostering and facilitating migrant social networks should not be very costly. For example, city governments could encourage communities, trade unions, and employers to provide migrants with more opportunities to socialise and to form friendships. More efforts could be made to educate urban locals to treat migrants equally in their day-to-day dealings and to make them feel welcome. The reason city governments have not done so, perhaps, is partially related to the current institutional setting. Most migrants do not have city health insurance. Consequently, when they are unhealthy and are unable to work, they need to return to their home town to receive treatment. Thus, host city governments are not bearing the cost of migrants being physically or mentally unhealthy. To this end, the central government should speed up removal of institutionalised discrimination against migrants by providing equal job opportunities and equal access to social benefits in cities, which, in turn, would provide incentives for city governments to care about migrant workers' long-term physical and mental health. Better mental health and better health services in cities can increase the length of stay of migrants in cities, increasing their productivity and assimilation into the urban environment.

It is important to understand that the conclusions drawn from this paper only consider one aspect of mental health as it relates to rural–urban migration. To understand the full picture, additional research needs to be conducted on issues such as how migration may affect social networks for rural left-behind families and non-migrant households and how such an effect may in turn affect mental health conditions of these groups.

Acknowledgements We would like to thank Tue Gorgens, Bob Gregory, Jenny Williams, the Editor Klaus Zimmermann, and three anonymous referees and seminar and conference participants at Australian National University and the 2014 Australasian Econometric Society Annual Meeting for their invaluable comments.

Funding This study was funded by the Australian Research Council for RUMiC project (ARC grant number LP066972, LP140100514, and DP0988572) and Guangzhou Social Science "13-fives plan" Grant (2019GZYB25).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix A: Tables and figures

Table 7 OLS estimates of impact of rainfall on migration decision

	Log household net agricultural income per capita in previous year (1)	Migrated for work purpose over 12 months in the previous year (2)
Panel A: Standard error clustered at village level		
Daily rainfall between April and Aug in year $t-2$	0.016***	− 0.007***
	(0.006)	(0.002)
Daily rainfall between April and Aug in year $t-3$	0.015*	− 0.004**
	(0.008)	(0.002)
Daily rainfall between April and Aug in year $t-4$	0.003	− 0.000
	(0.005)	(0.001)
Observations	12991	34092
Adjusted R^2	0.278	0.065
Panel B: Standard error clustered at county level		
Daily rainfall between April and Aug in year $t-2$	0.016**	− 0.007**
	(0.007)	(0.003)
Daily rainfall between April and Aug in year $t-3$	0.015*	− 0.004
	(0.008)	(0.002)
Daily rainfall between Apr and Aug in year t-4	0.003	–0.000
	(0.004)	(0.001)
Observations	12991	34092
Adjusted R^2	0.278	0.065

*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 a constant are included in the regressions. The reference group is single females with education below junior high school and excellent health. In the first column, each household is an observation, and the characteristics of household head are used. In the second column, each household member is an observation. The daily rainfall variable is measured in mm

Source: 2008 and 2009 Rural Household Surveys of the RUMiC project

Table 8 OLS and fixed effect estimates of network effect on mental health problems

	OLS			FE
	Pooled sample (1)	Representative sample (2)	Panel sample (3)	Panel sample (4)
No. of contacted people living in cities	− 0.016***	− 0.017***	− 0.013***	− 0.009**
	(0.002)	(0.003)	(0.003)	(0.004)
Age	0.048	0.028	0.052	
	(0.031)	(0.036)	(0.052)	
Squared age/100	− 0.045	− 0.022	− 0.044	
	(0.041)	(0.048)	(0.068)	
Male	− 0.645***	− 0.751***	− 0.555***	
	(0.111)	(0.135)	(0.182)	
Years since the first migration	0.017	0.036	0.015	
	(0.018)	(0.023)	(0.029)	
Squared year since the first migration/100	− 0.072	− 0.152*	− 0.014	
	(0.062)	(0.084)	(0.095)	
Junior high school	− 0.453***	− 0.384**	− 0.585***	
	(0.130)	(0.159)	(0.211)	
Senior high school or equivalent	− 0.814***	− 0.829***	− 0.718***	
	(0.143)	(0.173)	(0.231)	
College education or above	− 1.085***	− 1.175***	− 1.044***	
	(0.195)	(0.248)	(0.303)	
Married	− 0.440***	− 0.448***	− 0.145	0.064
	(0.135)	(0.167)	(0.206)	(0.304)
Divorced	0.688**	0.362	1.332***	0.507
	(0.300)	(0.375)	(0.445)	(0.635)
No. of kids	− 0.068	− 0.062	− 0.280***	0.026
	(0.071)	(0.089)	(0.109)	(0.175)
Height (cm)	− 0.008	− 0.001	− 0.017	− 0.029
	(0.007)	(0.009)	(0.012)	(0.029)
Good health	1.231***	1.442***	1.066***	0.921***
	(0.076)	(0.096)	(0.117)	(0.135)
Average health	2.860***	2.900***	2.602***	2.235***
	(0.117)	(0.152)	(0.174)	(0.199)
Poor or very poor health	5.100***	4.939***	4.908***	3.818***
	(0.366)	(0.490)	(0.508)	(0.563)
Self-employment	− 0.065	− 0.050	− 0.132	− 0.585*
	(0.133)	(0.182)	(0.185)	(0.306)
No. of family members over 16 years old	− 0.068	− 0.127*	− 0.020	− 0.071
	(0.058)	(0.075)	(0.085)	(0.127)

Table 8 (continued)

	OLS			FE
	Pooled sample	Representative sample	Panel sample	Panel sample
	(1)	(2)	(3)	(4)
Daily wage of unskilled labour at home village (yuan)	− 0.003	0.002	− 0.008***	− 0.010***
	(0.002)	(0.003)	(0.003)	(0.003)
Home village is in a mountainous area	0.078	0.044	0.322*	
	(0.105)	(0.121)	(0.186)	
Inverse of 1 + the distance between home village and the closest county	− 0.409**	− 0.370	− 0.578*	
	(0.201)	(0.258)	(0.308)	
Home village has medical station	− 0.275**	− 0.560***	0.185	
	(0.117)	(0.152)	(0.178)	
Average daily rainfall from $t-12$ to $t-3$ (1 mm)	3.046***	5.599***	1.050	
	(1.148)	(1.746)	(2.559)	
Squared average daily rainfall from $t-12$ to $t-3$	− 0.371**	− 0.762***	− 0.030	
	(0.157)	(0.239)	(0.329)	
Growth of GDP in destination cities (%)	0.120***	0.108***	0.127***	0.109***
	(0.025)	(0.032)	(0.037)	(0.038)
Growth of real minimum wage in destination cities (%)	− 0.006	− 0.007	− 0.005	− 0.006
	(0.004)	(0.006)	(0.005)	(0.005)
Industry and occupation dummies	Yes	Yes	Yes	Yes
Home county dummies	Yes	Yes	Yes	No
Destination city dummies	Yes	Yes	Yes	No
Year dummies	Yes	Yes	Yes	Yes
Observations	17,653	10,955	8756	8756
Adjusted R^2	0.124	0.140	0.118	0.057

Standard errors are clustered at household level. *Significant at 10% level; **significant at 5% level; ***significant at 1% level. The estimates of the constant term are omitted. The reference group is single females with education below junior high school and excellent health. The pooled sample consists of all the observations across waves. The representative sample consists of the 2008 wave and new samples in each wave after 2008. The panel sample consists of individuals who appeared in two or more waves. The daily rainfall variable is measured in mm

Source: 2008, 2009, 2011, and 2012 Migrant Household Surveys of the RUMIC project

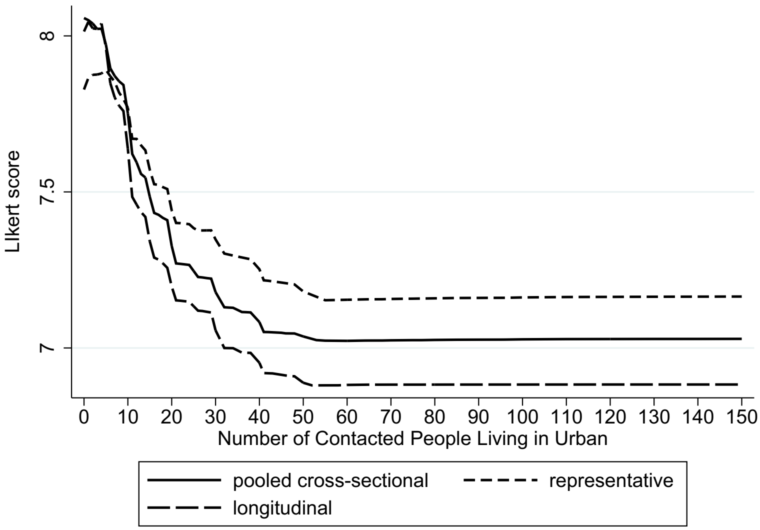

Fig. 4 Unconditional relationship between social networks and mental health problems for the three samples. Source: 2008, 2009, 2011, and 2012 waves of the Migrant Household Survey in RUMiC project

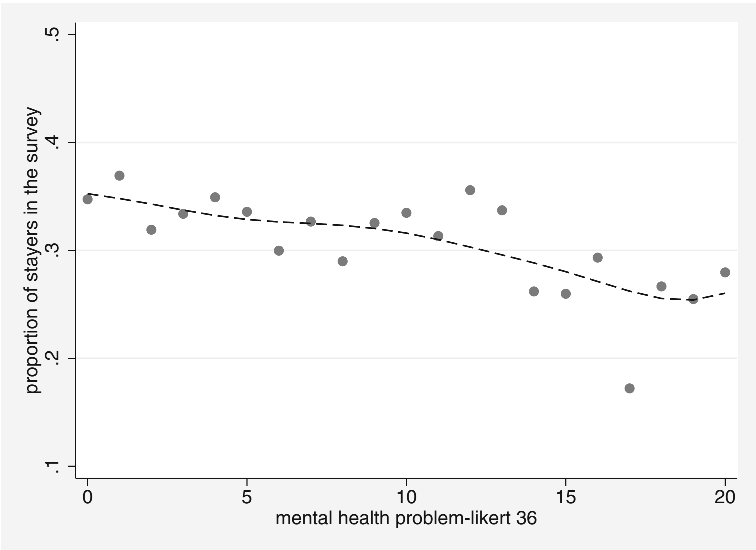

Fig. 5 Unconditional relationship between attrition and mental health problems. The x-axis shows the pre-determined mental health problem to attrition (i.e. the characteristics recorded when the respondent first participated the survey). Each dot represents a cell which is at a value of GHQ 12 Likert score, except the last dot represents a cell with the Likert score no less than 20 due to the small sample size. The y-axis shows the proportion of stayers in each cell. The stayers are those respondents who were tracked at least once after they participated the survey. Source: 2008, 2009, 2011, and 2012 waves of the Migrant Household Survey in RUMiC project

Appendix B: Data appendix

B.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. *Concen* 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. *NoSleep* Have you lost much sleep over worry?
   (1) never; (2) occasionally; (3) fairly often; (4) very often;
3. *Useful* 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. *Decide* 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. *Strain* Have you felt constantly under strain?
   (1) never; (2) slightly; (3) considerably; (4) seriously;
6. *Diffic* Have you felt you couldn’t overcome your difficulties?
   (1) never; (2) slightly; (3) considerably; (4) seriously;
7. *Activ* 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. *Probs* Have you been able to face up to problems?
   (1) always; (2) most of the time; (3) sometimes; (4) rarely;
9. *Depress* Have you been feeling unhappy or depressed?
   (1) never; (2) slightly; (3) considerably; (4) seriously;
10. *NoConf* Have you been losing confidence in yourself?
    (1) never; (2) slightly; (3) considerably; (4) seriously;
11. *Wthless* Have you been thinking of yourself as a worthless person?
    (1) never; (2) slightly; (3) considerably; (4) seriously;
12. *Happy* Have you been feeling reasonably happy, all things considered?
    (1) very happy; (2) fairly happy; (3) not so happy; (4) not happy at all”

B.2 Data details of appendix Table 7

The sample used in Appendix Table 7 is extracted from the 2008 and 2009 waves of the RUMiC rural household survey. We restrict the sample to households whose agricultural income per household member in the previous year is not more than 50,000 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 7 is constructed in the way described in Section 4.

B.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/2010-04/02/141001036464361310041002140l.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

References

Akay A, Bargain O, Zimmermann K (2012) Relative concerns of rural-to-urban migrants in China. Econ Behav Or 81(2):421–441

Akay A, Giulietti C, Robalino J, Zimmermann K (2013) Remittances and well-being among rural-to-urban migrants in China. Rev Econ Household 12(3):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 TW, Rubin H (1949) Estimation of the parameters of a single equation in a complete system of stochastic equations. Ann Math Stat, 46–63

Bao S, Bodvarsson OB, Hou JW, Zhao Y (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, Taubman P (1979) Health and labor market success: the role of various diseases. Rev Econ Stat 61(1):1–8

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

Bhugra D (2004) Migration and mental health. Acta Psychiatrica Scandinavica 109(4):243–258

Bjorklund A (1985) Unemployment and mental health: some evidence from panel data. J Human Resour 20(4):469–483

Brickman P, Coates D, Janoff-Bulman R (1978) Lottery winners and accident victims: is happiness relative? J Person Soc Psychol 36(8):917–927

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

Chadwick KA, Collins PA (2015) Examining the relationship between social support availability, urban center size, and self-perceived mental health of recent immigrants to Canada: a mixed-methods analysis. Soc Sci Med 128:220–230

Chen J (2011) Internal migration and health: re-examining the healthy migrant phenomenon in China. Soc Sci Med 72(8):1294–1301

Clark AE, Oswald AJ (1994) Unhappiness and unemployment. Econ J 104(424):648–659

Cohen S (2004) Social relationships and health. Am Psychol 59(8):676

Cohen S, Janicki-Deverts D (2009) Can we improve our physical health by altering our social networks? Perspect Psychol Sci 4(4):375–378

Cohen S, Wills TA (1985) Stress, social support, and the buffering hypothesis. Psychol Bull 98:310–357

Connolly M (2013) Some like it mild and not too wet: the influence of weather on subjective well-being. J Happiness Stud 14(2):457–473

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

Dunbar RI (1993) Coevolution of neocortical size, group size and language in humans. Behav Brain Sci 16(04):681–694

Di Tella R, Haisken-De New J, MacCulloch R (2010) Happiness adaptation to income and to status in an individual panel. J Econ Behav Org 76(3):834–852

Ettner SL, Frank RG, Kessler RC (1997) The impact of psychiatric disorders on labor market outcomes. Indus Labor Relat Rev 51:64

Ertel KA, Glymour MM, Berkman L (2009) Social networks and health: a life course perspective integrating observational and experimental evidence. J Soc Person Relat 26(1):73–92

Frijters P, Johnston DW, Meng X (2009) The mental health cost of long working hours: the case of rural Chinese migrants. Working paper

Frijters P, Johnston DW, Shields MA (2010) Mental health and labour market participation: evidence from IV panel data models. IZA Discussion Papers. No:4883

Gardner J, Oswald AJ (2006) Do divorcing couples become happier by breaking up? J R Stat Soc Series A (Stat Soc) 169(2):319–336

Gardner J, Oswald AJ (2007) Money and mental wellbeing: a longitudinal study of medium-sized lottery wins. J Health Econ 26(1):49–60

Giles J, Yoo K (2007) Precautionary behavior, migrant networks, and household consumption decisions: an empirical analysis using household panel data from rural China. Rev Econ Stat 89(3):534–551

Gong X, Kong ST, Li S, Meng X (2008) Rural-urban migrants: a driving force for growth. China's dilemma: economic growth, the environment and climate change. In: Song L, Woo WT (eds). ANU Press, pp 110–152

Graetz B (1991) Multidimensional properties of the general health questionnaire. Soc Psychiatry Psychiatric Epidemiol 26:132–138

He X, Wong DFK (2013) A comparison of female migrant workers mental health in four cities in China. Int J Soc Psychiatry 59(2):114–122

Henríquez-Sánchez P, Doreste-Alonso J, Henríquez-Sánchez MA, Bes-Rastrollo M, Gea A, Sánchez-Villegas A (2014) Geographical and climatic factors and depression risk in the sun project. Europ J Public Health 24(4):626–631

Hill RA, Dunbar RI (2003) Social network size in humans. Human Nat 14(1):53–72

Hull D (1979) Migration, adaptation, and illness: a review. Soc Sci Med 13(A):25–36

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. Econometrica 62(2):467–475

Janisch LM (2017) Mental health assimilation of Australian immigrants. Ruhr Economic Papers No. 728

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

Kleibergen F, Paap R (2006) Generalized reduced rank tests using the singular value decomposition. J Econometr 133(1):97–126

Kuroda S, Yamamoto I (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)

Lou Y, Beaujot R (2005) What happens to the 'Healthy Immigrant Effect': the mental health of immigrants to Canada. PSC Discussion Papers Series: 19(15), Article 1

Li L, Wang HM, Ye XJ, Jiang MM, Lou QY, Hesketh T (2007) The mental health status of Chinese rural urban migrant workers. Soc Psychiatry Psychiatric Epidemiol 42:716–722

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

McKenzie K, Whitley R, Weich S (2002) Social capital and mental health. British J Psychiatry 181(4):280–283

Meng X (2000) Labour market reform in China. Cambridge University Press, Cambridge

Meng X (2012) Labor market outcomes and reforms in China. J Econ Perspect 26(4):75–101

Miech RA, Caspi A, Moffitt TE, Wright BRE, Silva PA (1999) Low socioeconomic status and mental disorders: a longitudinal study of selection and causation during young adulthood. Am J Sociol 104(4):1096–1131

Mou J, Cheng J, Griffiths SM, Wong SY, Hillier S, Zhang D (2011) Internal migration and depressive symptoms among migrant factory workers in Shenzhen, China. J Commun Psychol 39(2):212–230

Mu R, van de Walle D (2011) Left behind to farm? Women's labor re-allocation in rural China. Labour Econ 18(Supplement 1):S83–S97

Munshi K (2003) Networks in the modern economy: Mexican migrants in the US labor market. Q J Econ 118(2):549–599

O'Hare C, OSullivan V, Flood S, Kenny RA (2016) Seasonal and meteorological associations with depressive symptoms in older adults: a geo-epidemiological study. J Affect Disorders 191:172–179

Olesen J, Gustavsson A, Svensson M, Wittchen H-U, Jönsson B (2012) The economic cost of brain disorders in Europe. Europ J Neurol 19(1):155–162

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

Phillips M, Zhang J, Shi Q et al (2009) Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001–05: an epidemiological survey. Lancet 373(9680):2041–2053

Qiu P, Caine E, Yang Y, Chen Q, Li J, Ma X (2011) Depression and associated factors in internal migrant workers in China. J Affect Disorders 134(13):198–207

Smith KP, Christakis NA (2008) Social networks and health. Ann Rev Sociol 34(4):405–429

Sparks K, Cooper C, Fried Y, Shirom A (1997) The effects of hours of work on health: a meta-analytic review. J Occup Org Psychol 70(4):391–408

Stillman S, McKenzie D, Gibson J (2009) Migration and mental health: evidence from a natural experiment. J Health Econ 28(3):677–687

Stock JH, Wright JH (2000) GMM with weak identification. Econometrica 68(5):1055–1096

Stock JH, Yogo M (2005) Testing for weak instruments in linear IV regression. In: Andrews DWK, Stock JH (eds) Identification and inference for econometric models: essays in honor of Thomas Rothenberg. Cambridge University Press, Cambridge, pp 80–108

Thoits PA (2011) Mechanisms linking social ties and support to physical and mental health. J Health Soc Behav 52(2):145–161

Wen K (1976) Theories of migration and mental health: an empirical testing on Chinese-Americans. Soc Sci Med 10(6):297–306

Wong D, He X, Leung G, Lau Y, Chang Y (2008) Mental health of migrant workers in China: prevalence and correlates. Soc Psychiatry Psychiatric Epidemiol 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. PhD thesis, Australian National University