

# A National Cross-Sectional Survey Dataset on Common Disease and Influencing Factors of Chinese Internal Migrants in 2017

Shengfa Zhang<sup>1</sup>, Wei Luo<sup>2</sup>, Yuhuan Ma<sup>2</sup>, Zhili Wang<sup>3</sup>, Jing Chen<sup>4\*</sup>, Fang Zhou<sup>4</sup>, Xiaoyu Zhang<sup>2</sup>, Wei Zhou<sup>2\*</sup>

<sup>1</sup>Law School, Beijing Normal University, Beijing 100875, China

<sup>2</sup>National Population Health Data Center, Chinese Academy of Medical Sciences, Beijing 100730, China

<sup>3</sup>China Population and Development Research Center, National Health Commission, Beijing 100081, China

<sup>4</sup>Migrant Population Service Center, National Health Commission, Beijing 100088, China

## ABSTRACT

This data article describes the dataset from a national cross-sectional questionnaire survey on Chinese internal migrants in 2017. The survey was part of the Chinese Migrants Dynamic Survey, which is an annually conducted large-scale national questionnaire survey on internal migrants. The respondents in the described dataset were migrants and registered local residents aged over 15 years old. A multistage stratified probability-proportional-to-size (PPS) sampling method based on the 2016 annual report data of China Migrant Population was adopted. The questionnaire was designed to collect demographic information of respondents and their family members, intention of migration or settlement, health status, health services accessibility, social integration, and epidemic influencing factors of common diseases, including hypertension, type 2 diabetes, diarrhea, fever, rash, icterus, conjunctival redness, cold, and other illnesses or injuries. The sample population in this survey include 13,998 internal migrants and 14,000 registered local residents from eight domestic cities / prefecture / districts across China. It is the most widely covering and highly representative dataset on common diseases and influencing factors of internal migrants in China. The dataset can be used to study common diseases and influencing factors among floating Chinese population. It provides data support for government to improve healthcare accessibility and the equity of public health services for internal migrants in China.

**Key words:** health services; floating population; population survey; common diseases; influencing factors; China

## Specifications Table

Subject area	Common diseases; influencing factors; floating population
Type of data	Table, Raw data
Data format	*.dta
Parameters for data collection	The dataset consists of individual information, family members' information, intention of migration and settlement (for migrants), health status, health services, social integration, common diseases, and influencing factors.
Data source location	Suzhou, Jiangsu province; Qingdao, Shandong province; Zhengzhou, Henan province; Changsha, Hunan province; Guangzhou, Guangdong province; Jiulongpo District, Chongqing municipality; Xishuangbanna Dai autonomous prefecture, Yunnan province; Urumqi, Xinjiang Uygur autonomous region.
Time period of the data	May 1st 2017–May 31, 2017

Received July 13, 2022; accepted September 6, 2022; publish online september 23, 2022.

\*Corresponding authors: E-mail: Jing Chen: okokok\_cj@163.com; Wei Zhou: rkjkt@pumc.edu.cn

© The authors. Published by Chinese Academy of Medical Science. This is an open access article attributed under the CC-BY-NC license (<http://creativecommons.org/licenses/by-nc/4.01>).

Specifications Table (continued)

Subject area	Common diseases; influencing factors; floating population
Description of data collection	The national cross-sectional questionnaire survey was conducted by the former National Health and Family Planning Commission. The sampling was drawn based on the 2016 China Migrant Population Information System, and a multistage stratified probability proportional-to-size (PPS) cluster sampling strategy was adopted to select participants. There were 13,998 internal migrants and 14,000 registered residents from 8 cities/prefecture/districts participated in the survey. Questionnaire was completed face-to-face by well-trained interviewers.
Data accessibility	National Population Health Data Center (NPHDC) / Population Health Data Archive (PHDA) ( <a href="https://www.ncmi.cn/phda/dataDetails.do?id=CSTR:A0006.11.A000T.201906.000570-V1.0">https://www.ncmi.cn/phda/dataDetails.do?id=CSTR: A0006.11.A000T.201906.000570-V1.0</a> ) under CC0 license, or the China Migrant Population Data Platform ( <a href="https://www.chinaldrk.org.cn/wjw/#/home">https://www.chinaldrk.org.cn/wjw/#/home</a> ) after application and approval.
Related research article	<ol style="list-style-type: none"> <li>1. Liu LC. Will moving with family members promote health investment of the youth? on dynamic monitoring data of floating population in 2017. <i>Financ Trade Res</i> 2022; 33(8):1-16. (in Chinese). doi: 10.19337/j.cnki.34-1093/f.2022.08.001.</li> <li>2. Liao ZL, Wang DW. Acceptance of mental health education and its influencing factors among elderly migrants in China. <i>Chin J Public Health</i> 2022; 38(7):922-7. doi: 10.11847/zgggws1135395. (in Chinese)</li> <li>3. Zhu Z, Guo M, Dong T, et al. Do migrants receive tuberculosis education in China? evidence from the China Migrants Dynamic Survey. <i>Health Educ Behav</i> 2021;10901981211000309. doi: 10.1177/10901981211000309.</li> <li>4. Wang J, Bai Y, Zhu J, et al. Association between social integration, social exclusion, and vaccination behavior among internal migrants in China: a cross-sectional study. <i>Int J Environ Res Public Health</i> 2022; 19(13):7915. doi: 10.3390/ijerph19137915.</li> <li>5. Li X, Li Y. The impact of perceived air pollution on labour supply: evidence from China. <i>J Environ Manage</i> 2022; 306:114455. doi:10.1016/j.jenvman.2022.114455.</li> <li>6. Qiang Y, Liu CJ, Sun J. Inequality in health services for internal migrants in China: a national cross-sectional study on the role of fund location of social health insurance. <i>Int J Environ Res Public Health</i> 2020; 17(17): 6327. doi: 10.3390/ijerph17176327.</li> </ol>

## VALUES OF THE DATASET

The dataset is nationally representative on the common diseases in internal migrants in China, including hypertension, type 2 diabetes, diarrhea, fever, rash, icterus, conjunctival redness, cold, other illness or injury. The data comes from both internal migrants and registered residents. It could be used to analyze the impact of migration on health status, health service accessibility, and social integration, to compare the differences between internal migrants and registered residents on health status, health service utilization, and social integration. Moreover, this dataset provides profile of the common diseases in Chinese rural-urban and urban-urban migrants, which is important to investigate the influencing factors for common diseases in China.

This dataset can be used in investigations on the common diseases in Chinese in terms of potential influencing factors and associated topics. It also provides data support for the government to formulate health service programs and policies, improve the healthcare level for internal migrants and the accessibility of pri-

mary healthcare in China, thus increasing the equity of health services.

## DATA DESCRIPTION

### Sample size and geographic distribution of data

The dataset was derived from a national project of cross-sectional survey named the "Key Disease Epidemic Influencing Factors " on Chinese internal migrants in 2017. The sampling frame derived from the 2016 annual report data of China Migrant Population, and a multistage stratified probability proportional-to-size (PPS) cluster sampling strategy was adopted for participant selection<sup>[1]</sup>. The survey included 13,998 internal migrants and 14,000 registered local residents from 8 cities/prefecture/districts. The sample sizes of internal migrants and registered residents in each participated site are shown in **Table 1**.

### The survey data profile on internal migrants and registered local residents from 8 cities in 2017

The dataset consist of 2 data files in .dta format, which

are survey data of internal migrants and registered local residents, respectively, from 8 representative cities in China. The dataset contains data of 75 survey questions for internal migrants and 63 survey questions for registered local residents regarding basic information of the survey sites, and six investigated dimensions, which are individual information, family members' information, mobility and intention to settle down (only for migrants), healthcare and public health services, social integration, common diseases, and influencing factors. The question types included single choice, multiple choice, or list. During data cleaning and desensitization, information about income, expense, and employment situation were removed. The data coding,

number of questions and variables in the data files are illustrated in **Table 2**.

## STUDY DESIGN AND METHODS

### Study setting

The former National Health and Family Planning Commission (NHFP) sponsored and conducted a series of large-scale national surveys on internal migrants annually from 2009 to 2017<sup>[3-5]</sup>. The reported dataset is derived from the second part of the Chinese Migrants Dynamic Survey in 2017, which is a national cross-sectional questionnaire survey on common diseases and influencing factors of China's internal migrants.

**Table 1.** Geographic distribution of sampling of the national survey on Chinese internal migrants in 2017

Locations of data source	internal migrants		registered local residents	
	sample size	proportion (%)	sample size	proportion (%)
Suzhou, Jiangsu province	2,000	14.29	2,000	14.29
Qingdao, Shandong province	2,000	14.29	2,000	14.29
Zhengzhou, Henan province	1,999	14.28	2,000	14.29
Changsha, Hunan province	2,000	14.29	2,000	14.29
Guangzhou, Guangdong province	1,999	14.28	2,000	14.29
Jiulongpo district, Chongqing municipality	1,000	7.14	1,000	7.14
Xishuangbanna Dai autonomous prefecture, Yunnan province	1,000	7.14	1,000	7.14
Urumqi, Xinjiang Uygur autonomous region	2,000	14.29	2,000	14.29
Total	13,998	100	14,000	100

**Table 2.** Data dimension and size of the national survey on Chinese internal migrants and registered residents in 2017

Dimensions in the survey	Coding range		Questions in surveys		Questions retained in the dataset		Variables encoded into the dataset	
	internal migrants	registered residents	internal migrants (n=126)	registered residents (n=94)	internal migrants (n=75)	registered residents (n=63)	internal migrants (n=312)	registered residents (n=257)
Basic information of survey sample points	C1 to C14	C1 to C8	14	8	6	6	6	6
Individual and family members' information	Q100 to Q101	Q100 to Q101	15	11	15	11	121	81
Income and expenditure	Q102 to Q112	Q102 to Q112	11	11	0	0	0	0
Employment situation	Q201 to Q216	Q201 to Q216	16	16	0	0	0	0
Mobility and intention to settle down	Q301 to Q321	NA	21	NA	7	NA	17	NA
Health status, healthcare and public health services	Q401 to Q418	Q401 to Q418	18	18	18	18	88	88
Social integration	Q501 to Q506	Q501 to Q505	6	5	6	5	31	30
Influencing factors of common diseases	Q601 to Q621	Q601 to Q621	25	25	23	23	52	52

### Participants

In the survey on internal migrants, the participants were qualified if: 1) non-registered male or female residents aged 15 years or older; 2) live locally for more than one month. The internal migrants were defined as individuals who travelled from their registered residence place and live or work in another districts or counties for more than one month, not including people who temporarily departed for tourism, medical treatment, business trip, family visit, and students or soldiers who left hometown for study or military service. Of the survey on registered residents, participants were sampled from registered residents of the selected districts or counties. The other selecting requirements for participants are the same<sup>[6]</sup>.

### Sampling methods

In the survey on China's internal migrants, a multistage stratified PPS cluster sampling strategy was adopted in participant selection. Initially, Guangdong, Shandong, and Jiangsu provinces in the eastern region of China, Henan, and Hunan provinces in the middle region of China, and Yunnan, Chongqing municipality, Xinjiang Uygur autonomous region in the western region of China were selected by purposive sampling method. Secondly, PPS sampling method was used in townships or streets selection, and subsequent selection of villages or communities. In village or community, systematic sampling was used to select participants for the survey. It should be noted that the sampling frame of PPS method derived from the 2016 annual reported data of China Migrant Population. In the end, the survey included 13,998 internal migrants and 14,000 registered local residents in particular.

### Survey procedure

The questionnaires were administered through face-to-face interviews using computer-assisted personal interviewing (CAPI) technique, i.e., smart phone or iPad. Interviewers were rigorously recruited through examinations to ensure they have proper understanding of the questionnaire. They received training to follow the standard survey protocol. The interviews took place in the households of the respondents. Oral informed consent was obtained from each respondent prior to starting the interview. The individuals who were enrolled but could not take the survey were replaced by individuals with matched characteristics in terms of age, gender, location, and residential status, prefera-

bly from the 2016 annual report data of China Migrant Population.

### Quality control

Rigorous data verification processes were established and implemented. Logical errors were checked by the CAPI system, and once identified data logic errors, the questionnaire would be fed back to the interviewers for eliminating potential data deviation. A quality assurance officer checked the completed questionnaires at each sub-district (township/district/county). Additionally, there were survey supervisors at the provincial or national level who performed audits on the collected data through random telephone interviews or online platform review.

### Data preprocessing

The former NHFPC conducted data processing. Data coding and data standardization were completed according to the internal regulations. In order to facilitate data sharing, privacy information and sensitive data have been removed from the dataset.

## MEASURES IN THE SURVEY

In the project of Chinese Migrants Dynamic Survey in 2017, the questionnaires designed for the survey on internal migrant and the survey on registered residents were entitled as the volume C and volume D, respectively, both of which contained the following 6 sections:

### Basic information of survey sites

The basic information of the survey sites includes the location (province/municipality/autonomous region, city/prefecture/districts, and township/village/community of the respondents), type of residence (rural or urban), survey date, and interviewer code.

### Individual and family members' information

Demographic information of the respondents and their family members who live in the same household were collected, including gender, date of birth, ethnic type, educational level, household registration type, whether is a Party member or League member, location of domicile place, type of current residence, whether is immigrated from other province, city or county, reason of migration, and relationship of each family member with respondent.

### Mobility of migration and intention of settlement (only for migrants)

The replies to the following 7 questions in this section were retained in the data file: 1) When did you first leave the registered residence; 2) Why did you leave your registered residence for the first time; 3) What district/county did you first move to; 4) Who did you migrate with at that time; 5) Before your first move, had your parents ever gone to work elsewhere; 6) How many cities have you ever traveled in total; 7) How many cities have you ever migrated to in total.

### Health status, healthcare, and health-related public services

Health status of respondent was investigated by questions: 1) How do you think about your health status (healthy, almost healthy, unhealthy, but can self-help; can not self-care in daily life); 2) Have you ever suffered from high blood pressure, type 2 diabetes, diarrhea, fever, rash, icterus, conjunctival redness, cold, and/or other illness or injury in the past year; 3) Have you been sick (injured) or unwell in the past year.

The health-related behaviors of the respondent were investigated by questions: 1) In the past year, when you were sick (diarrhea, fever, etc.), did you go to see a doctor; 2) When you are sick or injured, where do you go first; 3) How long does it take to arrive the nearest healthcare service institution from your residence.

The public health services of the respondent was investigated by questions: 1) Have you ever received health follow-up or physical examination (for respondents with type 2 diabetes or hypertension); 2) In the past year, have you ever received health related education in your village/community; 3) What type of health education you have received in your current village/ community.

The family planning of respondent was investigated by questions: 1) How many children you have; 2) Are you planning to have a new baby in the next two years; 3) Do you use contraception currently; 4) What kind of contraceptive approaches do you use; 5) What is the reason if you do not use any contraception.

### Social integration

The social integration of respondents were investigated by questions: 1) Have you participated in any local activity since 2016; 2) Have you ever served as community volunteer since 2016; 3) Your views on local en-

vironment and residents; 4) healthcare insurance you have; 5) Have you applied for a social security card; 6) Have you applied for a temporary residence permit.

### Common diseases and influencing factors

The investigation was mainly focused on common diseases including hypertension, type 2 diabetes, symptoms of diarrhea, fever, rash, icterus, conjunctival redness, cold, and other illnesses or injuries. Regarding the potential influencing factors, there were 13 questions on aspects of living environment, sanitary facility, whether have a pet, dietary habits, hygiene of drinking water, life style, and vaccination (**Table 3**).

## DATA ACCESS AND ANALYSIS

The dataset is freely available online to public from two data repositories, where the data are identical with each other. One is the National Population Health Data Center (NPHDC)/Population Health Data Archive (PHDA)(<https://www.ncmi.cn>)<sup>[2]</sup>, with China Science and Technology Resource (CSTR):A0006.11.A000T.201906.000225; the other is the China Migrant Population Data Platform (<https://www.chinaldrk.org.cn/wjw/#/home>).

**Table 3.** Examples of influencing factors of common diseases in the questionnaire

Influencing factors	Questions in the questionnaire
Living environment	How big is your area of residence? Are there cockroaches, flies, or mice in your house?
Sanitary facility	Is there an air purifier in your house? What is the type of toilet in your house?
Whether have a pet	What kind of pets do you keep? Is your pet vaccinated?
Dietary habits	What is your usual dining schedule? Do you separate raw food from cooked food when cooking?
Hygiene of drinking water	What is your main source of drinking water? Do you usually boil water before drinking?
Life style	Do you share the washcloth with others? Do you share water glasses with others?
Vaccination	What kind of vaccination did you have?

The dataset at PHDA complies with CC0 license. Researchers who need to access the dataset through PHDA should send application to PHDA committee along with their research protocol, and follow the data usage license rigorously. A coding handbook is accompanied with the dataset on the platform to facilitate users. Data analysis and mining could be carried out through the PHDA platform at <https://dataservice.imi-cams.ac.cn/vpn/index.html>, where analysis software such as SPSS, Stata, R, and Python have been installed for use. The dataset is inadmissible to be downloaded and saved to local computers.

Researchers who need to access dataset through China Migrant Population Data Platform should register an account firstly, submit a research plan and an application form with authorization by the researchers' institution; once approved, sign an agreement on copyright and reuse of the dataset with the Migrant Population Service Center of the National Health Commission of China. The data platform will send the dataset to user on the platform, and then user can download the data from the platform.

### Limitations

The questionnaires were designed by the former NHFPC, and the validity of some questions was not detailed enough. Because the lack of employment and expenditure information, the correlation between employment and health condition would be difficult to verify. Due to the impreciseness of the terminology in the questionnaire, and the lack of medical common sense of the respondents, the respondents' self-reported illness conditions may be ambiguous or lack of medical accuracy, which should be noticed when researchers handling this part of the data.

### Ethical statement

*The 2017 China Migrants Dynamic Survey was approved by the China National Bureau of Statistics [No. NBS (2015)72], and written informed consents were obtained from all respondents at the time of data collection. All procedures in this study were in accordance with the 1975 Helsinki Declaration and its later amendments or comparable ethical standards. The Migrant Population Service Center of the National Health Commission of China approved this study.*

### Conflict of interest

*The authors declared no conflict of interest.*

### Author contributions

*Wang ZL, Chen J, and Zhou F were designers of the survey project, and they were responsible for the data collection and data audit. Zhang SF drafted the manuscript. Luo W, Zhou W, Ma YH, and Zhang XY involved in data processing. Chen J and Zhou W are responsible for all aspects of the manuscript and questions related to accuracy or integrity of the dataset. All the authors approved the final version of the manuscript.*

### REFERENCES

1. Yu X, Liang J, Zhang Y. Air pollution and settlement intention: evidence from the China migrants dynamic survey. *Int J Environ Res Public Health* 2022; 19(8):4924. doi: 10.3390/ijerph19084924
2. Migrant Population Service Center NHCPRC. Special survey on health and family planning dynamic monitoring of migrants in china-survey data on epidemic factors of key diseases in eight cities (2017) [dataset]. National Population Health Data Center (NPHDC) and Population Health Data Archive (PHDA) 2019. Available from: <https://www.ncmi.cn/phda/dataDetails.do?id=CSTR:A0006.11.A000T.201906.000570-V1.0>. Accessed: September 15, 2022. doi: 10.12213/11.A000T.201906.000570.
3. Zhu Z, Guo M, Petrovsky DV, et al. Age and regional disparity in HIV education among migrants in China: migrants population dynamic monitoring survey, 2014-2015. *Int J Equity Health* 2019; 18(1):104. doi: 10.1186/s12939-019-0999-x.
4. Zhu Z, Guo M, Dong T, et al. Do migrants receive tuberculosis education in China? Evidence from the China migrants dynamic survey. *Health Educ Behav* 2021; 10901981211000309. doi: 10.1177/10901981211000309.
5. Migrant Population Service Center NHCPRC. Health data Dynamic Monitoring on Health and Family planning of Chinese Migrants from 2009 to 2017 [dataset]. National Population Health Data Center PHDA 2019. Available from: <https://www.ncmi.cn/phda/dataDetails.do?id=CSTR:A0006.11.A000T.201906.000225>. Accessed: September 15, 2022. doi: 10.12213/11.A000T.201906.000225.
6. Yang Z, Jiang CH, Hu J. Moderating effects of regional disparities on the relationship between individual determinants and public health service utilization among internal migrants: evidence from the China migrant dynamic survey in 2017. *BMC Public Health* 2022; 22(1):564. doi: 10.1186/s12889-022-12870-1.
7. Huang W, He L, Lan H. The impact of self-employment on the health of migrant workers: evidence from China migrants dynamic survey. *Int J Environ Res Public Health* 2022; 19(10):5856. doi: 10.3390/ijerph19105868.

# 中国流动人口常见疾病及其影响因素：2017 年横断面调查数据集

张胜发<sup>1</sup>，罗葳<sup>2</sup>，马玉环<sup>2</sup>，王志理<sup>3</sup>，陈晶<sup>4\*</sup>，周芳<sup>4</sup>，张晓宇<sup>2</sup>，周伟<sup>2\*</sup>

<sup>1</sup>北京师范大学法学院，北京 100875，中国

<sup>2</sup>中国医学科学院国家人口健康科学数据中心，北京 100730，中国

<sup>3</sup>中国人口与发展研究中心，北京 100081，中国

<sup>4</sup>国家卫生健康委流动人口服务中心，北京 100088，中国

## 摘要

本文描述了 2017 年中国流动人口常见疾病及其影响因素的全国性横断面调查数据集。该调查是 2017 年中国流动人口卫生计生动态监测调查（一项针对全国流动人口的年度大规模全国性调查）的一个重要组成部分。调查以 15 岁以上的流动人口及本地户籍人口为调查对象，根据 2016 年中国流动人口年报数据为抽样框，采用分层、多阶段、与规模成比例的概率比例规模抽样法（probability-proportional-to-size sampling, PPS）。数据集中获得的抽样人群包括来自国内 8 个城市的 13,998 名流动人口和 14,000 名本地户籍居民。该调查搜集调查对象及其家庭成员的人口学信息、流动或居住意向、健康状况、医疗卫生服务、社会融合状况、包括高血压、2 型糖尿病、腹泻、发热、皮疹、黄疸、结膜发红、感冒等身体损害在内的常见疾病患病情况以及其相关影响因素的信息。该数据集是中国调查流动人口常见疾病及其影响因素的规模最大和最具代表性的数据集，可用于研究中国流动人口常见疾病的患病情况及其影响因素，并为政府改善流动人口的医疗服务可及性和公共卫生服务公平性提供数据支持。

**关键词：**卫生服务；流动人口；人口调查；常见疾病；影响因素；中国

**\* 通讯作者：**陈晶 okokok\_cj@163.com；周伟 rkjkpt@pumc.edu.cn