



ARTICLE



<https://doi.org/10.1057/s41599-025-04859-z>

OPEN

# Self-rated health differences among different workers in China with an agricultural hukou: the chain mediation effect of income satisfaction and medical service quality

Li He<sup>1,2</sup>, Jiangyin Wang<sup>1</sup>, Yang Yang<sup>3</sup>, Yue Yu<sup>1</sup>, Zihan Feng<sup>1</sup> & Zhixiong Yang<sup>2,4</sup>✉

Workers with agricultural hukou (WAH) in China, due to the prevalence of their group and their vulnerability in terms of health, have long been the focal point of research. This study utilizes data from the 2016, 2018, and 2020 *China Family Panel Studies* (CFPS) to analyze the self-rated health status and influencing factors of different subgroups of WAH, including subsistence farmers, agricultural laborers, and non-agricultural workers. This study employed the two-way fixed-effects model, treatment effects model, double machine learning (DML), and structural equation modeling (SEM). The results show that, within the WAH population, subsistence farmers have the lowest self-rated health levels. In contrast, no statistically significant health differences were found between agricultural laborers and non-agricultural workers. Drawing upon public service accessibility theory, this study further confirms the chain mediation effect of income satisfaction and medical service quality on these health disparities. The findings provide important insights for the government to formulate more targeted and practical health promotion policies.

<sup>1</sup> Zhongnan University of Economics and Law, Wuhan, China. <sup>2</sup> The United Research Center of The Chinese Government Development, Nankai University, Tianjin, China. <sup>3</sup> Central China Normal University, Wuhan, China. <sup>4</sup> Ocean University of China, Qingdao, China. ✉email: [yzxouc@163.com](mailto:yzxouc@163.com)

## Introduction

The term “nongmin” is commonly translated into English as “farmer,” “peasant,” or “peasantry,” but these translations fail to fully capture its multidimensional meaning in the Chinese context. In China, “nongmin” not only refers to individuals engaged in primary industries but, more importantly, to those who hold an agricultural hukou. The household registration system plays a critical role in Chinese society, distinguishing between agricultural and non-agricultural hukou, determining legal status, and affecting the distribution of social welfare and resources (Gong et al., 2012). This study focuses on the working-age population holding an agricultural hukou, particularly those involved in material or service production activities, aiming to explore health disparities within this group and their causes.

However, workers with agricultural hukou (WAH) represents a large social group with significant and non-negligible internal differences, necessitating their classification into types. Drawing on other works (Schneider, 2015; Zhang and Donaldson, 2010), and considering the realities of China, this study categorizes WAH according to the type of employer and the nature of work into subsistence farmers, agricultural laborers, and non-agricultural workers.

Subsistence farmers are a widely prevalent group in China,<sup>2</sup> typically referring to laborers who cultivate their own land and rely primarily on agricultural income for their livelihood. The lifestyle of subsistence farmers is relatively independent, and their economic activities depend less on external labor markets. However, with the development of the socioeconomic environment, the rigid hukou system in China has also led to a migration pattern from rural to urban areas that differs from those in other developing countries (Treiman et al., 2001), forming the unique Chinese migrant worker group. A defining characteristic of this group is that they engage in non-agricultural work in cities, but their hukou remains in rural areas.

Besides these two types, there exists a group known as agricultural laborers. Although they are also engaged in agricultural production, they do not work on their own land but for agricultural enterprises or cooperatives, with wages as their main source of livelihood. The characteristics of this group are as follows: Labor forms are more modernized, and production relations are increasingly corporate and collectivized. It should be noted that, in more economically developed countries, agricultural laborers are often referred to as farm workers and are included in the worker category. However, in China, because they still retain an agricultural hukou, they are classified as WAH rather than general workers.

According to the *China Statistical Yearbook 2023*, the number of workers engaged in agricultural production (including subsistence farmers and agricultural laborers) reached 177 million in 2022, accounting for 24.1% of the total employed population in China.<sup>3</sup> Additionally, in 2023, the total number of migrant workers in China reached 298 million.<sup>4</sup> In comparison, globally, the agricultural laborer group is also substantial. According to the International Labor Organization, in 2019, 1.239 billion people worldwide were engaged in primary industries, such as agriculture, forestry, animal husbandry, and fisheries, while the number of workers living in rural areas reached 1.508 billion.<sup>5</sup>

At the same time, with the advancement of urbanization, more rural residents are moving to cities for work. This has further highlighted the heterogeneity within agricultural laborer and rural resident groups. Focusing on the health of these groups and their internal subgroups is not only pertinent to China's national context but also has global significance. The WAH subgroups face unique health risks, and their health status urgently requires improvement. They are more susceptible to multiple factors, such as occupational health risks, insufficient access to medical

services, and social psychological factors. These factors are intertwined, posing significant challenges to the health of WAH and potentially exacerbating health inequalities. In-depth research into health disparities within WAH and their causes is crucial for developing effective health intervention strategies and promoting social equity and harmonious development.

Based on the context presented above, this study poses two core research questions: 1) Are there differences in self-rated health among the three subgroups within WAH—subsistence farmers, agricultural laborers, and non-agricultural workers? 2) What are the pathways that influence the health disparities among these three groups?

The contributions of this study are reflected in the following two aspects: 1) On a theoretical level, by integrating the public service accessibility theory, this study deepens the understanding of self-rated health differences among various WAH. The application of this methodology not only enriches research on WAH's health but provides a new perspective on how WAH assess their health levels based on the relationship between their income status and the accessibility of medical public services. 2) On a practical level, by exploring how the accessibility of medical services influences self-rated health differences among WAH subgroups, the findings of this research help policymakers better understand the specific health challenges faced by these groups. This provides an empirical foundation for formulating more precise and targeted health promotion and disease prevention strategies. Furthermore, the research emphasizes the importance of improving public service accessibility, particularly enhancing access to medical services, in improving the health of WAH and other vulnerable groups.

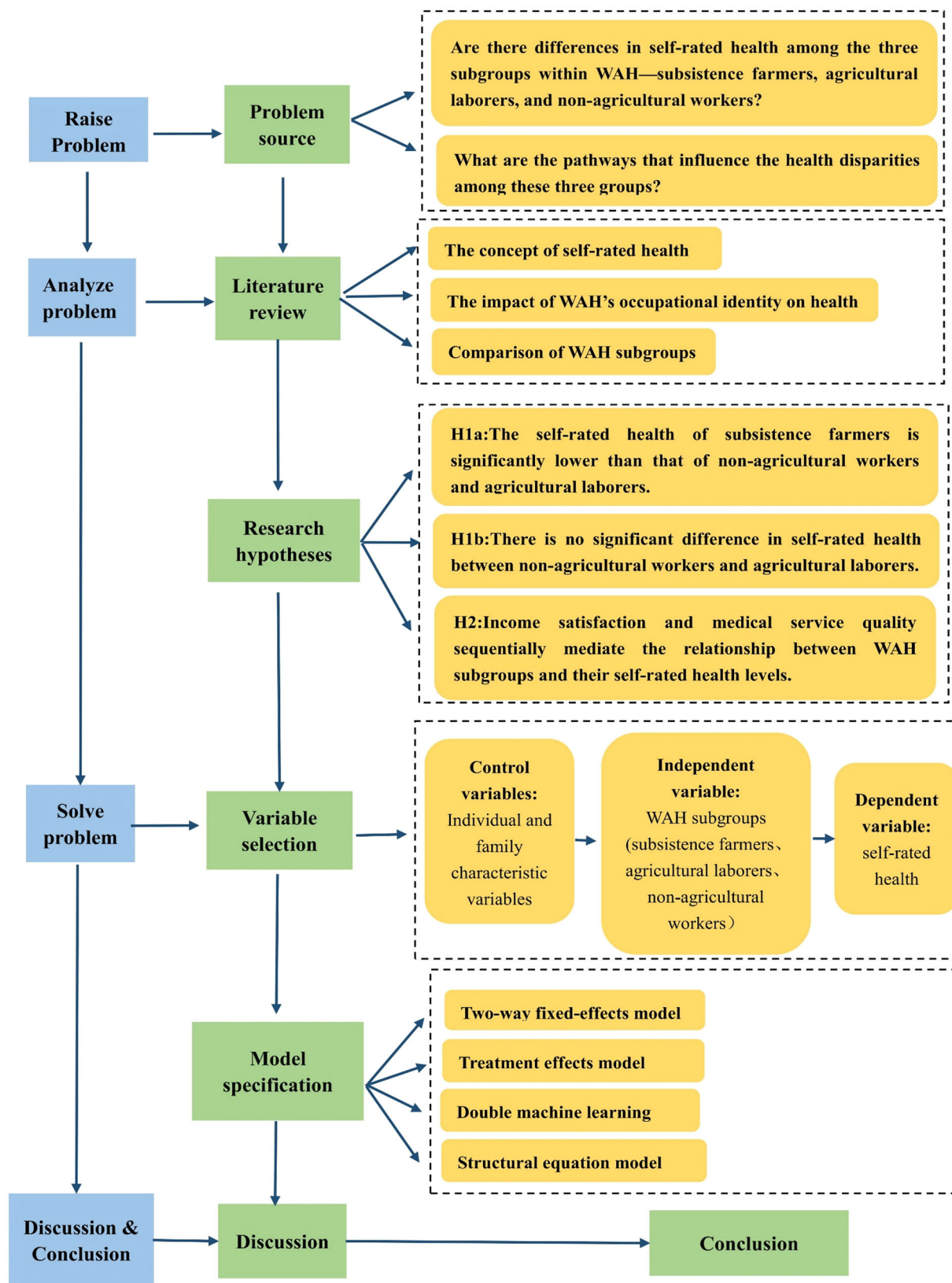
Figure 1 summarizes the structure of this paper's remaining sections, Fig. 2 shows the theoretical framework of the study.

## Literature review

**The concept of self-rated health.** Self-rated health, also known as self-perceived health, self-assessed health, or subjective health, refers to an individual's subjective evaluation of their own health status. Although it is a subjective assessment, it is not only an important measure of health, reflecting both objective and subjective aspects of health (Jylhä, 2009; Kaplan and Baron-Epel, 2003), but also reflects other relevant factors that influence health (Meng et al., 2014). Therefore, the concept of self-rated health is widely applied in public health research and has become an effective tool for exploring the health status of WAH in this study.

**The impact of WAH's occupational identity on health.** There is some debate in the academic community about WAH's self-rated health status. Agriculture is characterized by demanding and unpredictable occupational requirements. Farmers typically work 50 to 80 h per week, with work hours varying by season, month, and day and often experience lower profit margins (Lilley et al., 2012; Sanne et al., 2004). These challenges expose agricultural laborers to a broad range of mental health risks (Younker and Radunovich, 2021), including anxiety, depression, and most concerning, suicide (Jones-Bitton et al., 2020; Kim et al., 2019; Klingelschmidt et al., 2018). Numerous studies have found that due to prolonged engagement in heavy physical labor and high-risk tasks, WAH reports higher rates of chronic diseases, functional disabilities, and occupational comorbidities (Tonelli et al., 2014), resulting in lower self-rated health compared to other socioeconomic groups (van Oostrom et al., 2021; Rask et al., 2017).

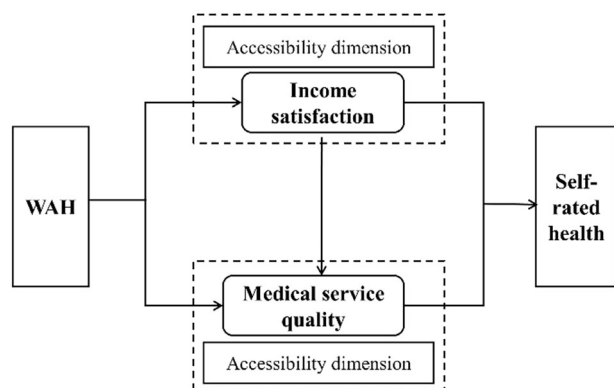
Vandenbosch et al. (2016) further observed that low access to medical services and lower health literacy exacerbate health



**Fig. 1** Research and analysis framework.

problems among farmers, leading to poorer overall health compared to other occupational groups. Additionally, Zhang et al. (2018) found that the mental constraints shaped by traditional family ethics and negative social culture make farmers

reluctant to seek help when ill, further increasing their health risks. However, there are differing opinions in the research about WAH's self-rated health levels. Some researchers, while acknowledging the prevalent poor self-rated health status among



**Fig. 2** Theoretical framework diagram.

agricultural workers, argue that such lower levels of self-rated health are common across all occupations with low-income and education levels (Mueller, 2021). Others believe that, compared to other labor groups, WAH have better self-rated health levels (Haseli-Mashhadi et al., 2009). The reasoning is that WAH engage in more physical activity in their daily work, making them less prone to chronic diseases such as diabetes and hypertension. In other words, compared to other occupational groups, WAH is more active, which could lead to better overall health status and, consequently, better reported health levels (Yu et al., 2019).

**Comparison of WAH subgroups.** In addition to the overall discussion of self-rated health levels among WAH, researchers have also focused on the health status of different types of WAH. Existing studies primarily reveal health disparities among various subgroups based on factors such as gender, age, income level, and education.

First, gender plays an important role in the health of agricultural laborers. Specifically, female WAH tend to rate their health lower than their male counterparts (Tian et al., 2023) and are more likely to suffer from conditions such as hypertension and musculoskeletal pain (Lee et al., 2019), and face higher levels of psychological and emotional distress (Brew et al., 2016). These disparities are closely related to social gender roles, the division of labor, and the dual pressures that women face in agricultural labor (both agricultural work and domestic responsibilities) (Adejoh et al., 2022). Second, regarding age, studies have found that older WAH experience poorer physical and mental health compared to younger WAH (Martins et al., 2023; Okyay et al., 2018; Zhang et al., 2011) and face higher risks of anxiety and depression (Akpınar et al., 2023). Furthermore, Rinsky-Halivni et al. (2022) observed that the work environment and social support play significant roles in the mental health of agricultural laborers across different age groups. A safe working environment and a strong social network can help mitigate agricultural stress and promote health. Finally, income level and education are also key factors influencing the health status of WAH's subgroups. Many studies have found that, compared to low-income and low-education farming groups, high-income and highly educated farmers are more likely to understand the health risks associated with pesticide exposure (O'Brien et al., 2012), are more likely to adopt healthier lifestyles and practices (Wu and Zeng, 2024), and thus enjoy better overall health (Kongtip et al., 2018; Nguyen et al., 2020).

Overall, current academic discussions on the relationship between agricultural occupational identity and self-rated health, physical health, and mental health are extensive. However, comparative studies on the health of WAH subgroups are relatively rare, and few studies categorize subgroups based on

labor characteristics. Additionally, there is limited research analyzing the mechanisms influencing health disparities between subgroups. In fact, the mechanisms contributing to the health issues of agricultural subgroups vary. Clarifying the roles of economic factors such as income and public service factors such as medical service is a critical prerequisite for improving the health levels of different WAH groups in a targeted manner.

## Research hypotheses

**Comparison of self-rated health among WAH subgroups.** This study categorizes WAH into three types—subsistence farmers, non-agricultural workers, and agricultural laborers—based on the nature of their work and working environment. It aims to explore the differences in self-rated health levels among these three groups.

First, for the subsistence farmer group, which primarily depends on their own land for livelihood, they mostly live in underdeveloped areas and engage in low-level agricultural production activities. Not only are they highly vulnerable to the negative impacts of sudden changes in the work environment (Li et al., 2018a; Wang et al., 2014), but their income is also difficult to secure effectively. Therefore, they are likely to have lower self-rated health levels (Nguyen et al., 2019).

Second, for the increasingly large group of non-agricultural workers, a study by Shao et al. (2016) found that, due to their lower income levels, they have poorer self-rated health. However, they can improve their self-rated health through means such as registering for health insurance (Wang, 2018) and increasing awareness of the value of their work (Yang et al., 2021).

Third, for agricultural laborers working in agricultural enterprises or cooperatives, the corporate production model provides them with a safer work environment and modern agricultural equipment, reducing their risk of occupational diseases and work intensity (Wossen et al., 2017). However, agricultural laborers' health still faces negative impacts from pesticide exposure (Li et al., 2018b) and injuries related to specific tasks (Perkio-Makela, 2000). Thus, the following hypotheses are proposed:

**Hypothesis 1a:** The self-rated health of subsistence farmers is significantly lower than that of non-agricultural workers and agricultural laborers.

**Hypothesis 1b:** There is no significant difference in self-rated health between non-agricultural workers and agricultural laborers.

**The chain mediation role of income satisfaction and medical service quality.** In response to the self-rated health status observed in WAH subgroups, this study further analyzed its potential mechanisms under the framework of public service accessibility theory. The accessibility of public services means the ability of individuals or groups to obtain the public services they need. This ability consists of five dimensions: accessibility, acceptability, availability, affordability, and geographical accessibility (Li and Wang, 2024; Penchansky and Thomas, 1981). Among them, accessibility is the extent to which public services are sufficient and able to meet the needs of the target group, emphasizing the quantity and variety of services provided. Acceptability is the extent to which public services are consistent with the cultural, social, and value preferences of the target group, meaning services must take the audience's background and needs into account. Availability is the extent to which services are provided when the target group can access them and whether they meet the specific needs of the group, emphasizing the timeliness and adaptability of services. Affordability is the extent to which the target group has the financial ability to pay for the



services required. Finally, geographical accessibility is the extent to which individuals can physically reach the locations where public services are provided. These dimensions interact and collectively determine the actual level and quality of public service access available to different groups, thereby affecting their quality of life and health levels.

Income is a key factor in predicting differences in medical utilization across various income groups (Chen et al., 2020), reflecting the affordability dimension of public service accessibility theory. Income enhances the likelihood of individuals accessing necessary medical services and increases the freedom to choose higher-quality medical resources (Li et al., 2014), thereby predicting personal health levels to some extent (Onarheim et al., 2018). This explains why individuals with higher annual incomes are more likely to report a good self-rated health status (Strauss et al., 2010; Sun et al., 2011).

As an individual's cognitive attitude toward the state of medical service, medical service quality reflects the accessibility, acceptability, and geographical accessibility dimensions of public service accessibility theory (Li and Wang, 2024). High satisfaction with medical services indicates an advantage in the quality and accessibility of medical services, not only enhancing individuals' evaluations of health services but positively impacting their self-rated health status (Arraras et al., 2013; Iskandarsyah et al. 2013; Wong and Fielding, 2008). Therefore, individuals who are more satisfied with the quality of medical services tend to have better self-rated health levels (Gupta et al., 2015; Pita-Fernandez et al., 2013).

Furthermore, there is a close relationship between income satisfaction and the evaluation of medical services, reflecting the interaction between economic resources and service quality in the theory of accessibility to public services. Generally, individuals with higher incomes and satisfaction are more likely to opt for higher standard medical service and healthcare (Wang et al. 2013). Those who receive high-quality medical services often achieve higher satisfaction with medical and life quality, thereby obtaining higher self-rated health levels (Perlman and Bobak, 2008). Moreover, previous studies have demonstrated a significant positive correlation between income satisfaction and satisfaction with medical services (Sha et al., 2019; Zheng et al., 2023).

Finally, income satisfaction is a relative concept, emphasizing the importance of relative income in the theory of social comparison (Crosby and Hamilton, 2017; Yu and Chen, 2016). An individual's satisfaction with their income depends not only on their absolute income level but also their socioeconomic environment (D'Ambrosio and Frick, 2007) and the influence of their reference group (Prus, 2011). Based on the above analysis, there are significant differences in the work environment and income satisfaction among subsistence farmers, non-agricultural workers, and agricultural laborers within the WAH. These differences further affect their evaluations of medical service quality and judgments about self-health status. Hence, the following hypothesis is proposed:

**Hypothesis 2:** Income satisfaction and medical service quality sequentially mediate the relationship between WAH subgroups and their self-rated health levels.

## Methods

**Data.** The data used in this study were derived from the *China Family Panel Studies* (CFPS) and the *China City Statistical Yearbook*. The CFPS is a national, large-scale, multidisciplinary social tracking survey organized by Peking University that reflects the changes in Chinese society, economy, population, education, and health. This survey has been conducted every two years since

2010, with the latest data available up to 2020. The CFPS focuses on the economic and non-economic welfare of Chinese residents, with surveys from 25 provinces/municipalities/autonomous regions, and a target sample size of 16,000 households. The survey included all family members in the sampled households, ensuring the representation, reliability, and validity of the sample data. This study selected the CFPS data from 2016, 2018, and 2020 to construct a panel dataset. The *China City Statistical Yearbook*, published by the National Bureau of Statistics of China, is an important statistical resource that comprehensively reflects the socioeconomic development of Chinese cities.

The statistical analysis for this study was conducted using Stata 17. Initially, CFPS and the *China City Statistical Yearbook* data were merged, and samples with missing variables were excluded. Then, a balanced panel dataset was constructed, excluding samples with observations in only one or two periods. The final effective sample size for the study was 16,439.

## Variables

**Dependent variable.** The dependent variable in this study was self-rated health. Drawing from the research by Dorélien and Xu (2020), this variable was obtained from the CFPS2018 question, "What do you think of your health status?" The respondents' answers ranged from "unhealthy," "average," "relatively healthy," to "very healthy," and "extremely healthy." Values from 1 to 5 were assigned, with the score positively correlated with the level of self-rated health.

**Independent variable.** The independent variable in this study was the type of WAH. Based on the research and classification of WAH in China by other researchers (Schneider, 2015; Zhang and Donaldson, 2010), this study identified the target group—WAH—based on the "current hukou status" question in the CFPS. There were significant differences within the WAH, and these groups could be categorized according to employer type and job nature, which may lead to differences in self-rated health. Therefore, it was crucial to explore these health disparities.

Thus, based on the two questions—whether "the main job is working for oneself/one's own family or employed by others/families/organizations/units/companies" and whether "the main job is agricultural or non-agricultural work"—WAH are divided into three categories: 1) subsistence farmers (engaged in their own agricultural production and management), 2) agricultural laborers (working in agriculture for others), and 3) non-agricultural workers (engaged in non-agricultural work). This classification facilitated subsequent pairwise comparative analysis of self-rated health levels across these groups.

**Control variables.** Based on the research questions, this study considered various factors when selecting the control variables. Specifically, the control variables were in two categories: 1) individual characteristics and 2) family characteristics.

Individual characteristics included age, gender, marital status, education level, belief, medical history, medical insurance, life satisfaction, social status evaluation, exercise, smoking, and drinking. These variables aim to control for the individual's basic social characteristics, psychological state, and health behaviors.

Family characteristic variables included household size, household net income, and household total expenditure; they were primarily used to control for the potential influence of family structure and economic status on the research framework.

**Instrumental variable.** The primary analysis in this study focused on the differences in self-rated health levels among three WAH

groups: subsistence farmers, agricultural laborers, and non-agricultural workers. However, endogeneity may be present in the study, particularly in the process of occupational choice, in which self-selection bias might exist among WAH. To address this issue, this study employed a treatment-effect model. The required instrumental variable for this model is the urbanization rate, which reflects the degree of urbanization in a region. The urbanization rate is calculated as: Urban population/(Urban population + Rural population).

The validity of the instrumental variable requires meeting two criteria: relevance and exogeneity. In terms of relevance, as urbanization levels increase, economic conditions in the region improve, and the industrial structure gradually diversifies, providing more employment opportunities (including agricultural jobs) (Dorosh and Thurlow, 2021). Therefore, WAH may transition into agricultural laborers or non-agricultural workers, leading to the formation of different groups. As a result, there is a strong correlation between the urbanization rate and the distinctions among subsistence farmers, agricultural laborers, and non-agricultural workers.

Regarding exogeneity, although urbanization may indirectly affect individual health through factors such as economic conditions and living environment (Wang et al., 2023), its direct impact on individual health status is relatively limited. Therefore, the urbanization rate meets the exogeneity requirement and can serve as a valid instrumental variable in comparing self-rated health differences among subsistence farmers, agricultural laborers, and non-agricultural workers.

**Mediating variables.** Based on the studies by Sha et al. (2019) and Zheng et al. (2023), both wage satisfaction and satisfaction with medical services have a certain impact on individuals' self-rated health. Therefore, this study analyzed the mediation mechanisms of income satisfaction and medical service quality as mediators in the self-rated health differences between subsistence farmers and agricultural laborers and between subsistence farmers and non-agricultural workers. In the selection of specific variables, the income satisfaction variable was directly obtained from the CFPS question, "How satisfied are you with the income from this job?" with values assigned from 1 to 5, low to high.

Since medical service quality as a latent variable is difficult to observe directly, this study selected three observable variables to measure it: rank of healthcare institutions, satisfaction with healthcare institutions, and assessment of the level of healthcare institutions. These observable variables are obtained from the CFPS responses to "If you need to see a doctor, where do you usually go?", "How satisfied are you with the overall medical conditions at the place you receive treatment?", and "How do you rate the medical level at the place you receive treatment?" The responses to these three observable variables were assigned values from 1 to 5, low to high.

## Model selection

**Two-way fixed-effects model.** To explore the differences in self-rated health levels among the three subgroups of WAH (subsistence farmers, agricultural laborers, and non-agricultural workers), this study employed a two-way fixed-effects model for preliminary estimates. The model is specified as follows:

$$SRH_{it} = \alpha_1 Peasant_{it} + Individual_{it}\beta + Family_{it}\gamma + \lambda_c + \delta_t + \varepsilon_{it} \quad (1)$$

Where  $SRH_{it}$  represents the self-rated health level of the  $i$ -th respondent at time  $t$ , and  $Peasant_{it}$  represents the identity of the  $i$ -th respondent at time  $t$ , indicating whether they are a subsistence farmer, agricultural laborer, or non-agricultural worker.  $Individual_{it}\beta$  represents a set of variables that measure individual characteristics, and  $Family_{it}\gamma$  represents a set of variables that

measure family characteristics.  $\lambda_c$  denotes the regional fixed effects,  $\delta_t$  denotes the year fixed effects, and  $\varepsilon_{it}$  represents the random error term.  $\alpha_1$  is the coefficient to be estimated in this study, reflecting the magnitude and direction of the effect of different WAH identities on self-rated health levels.

**Treatment effects model.** To address the estimation bias caused by self-selection bias, this study adopted the treatment effects model proposed by Maddala (1983). This model aims to mitigate the impact of unobservable factors on individual decisions and is particularly suitable for situations with endogeneity issues. The core idea is to introduce a selection equation to control for self-selection bias, thereby improving the precision of treatment effect estimation. The treatment effects model resolves the self-selection bias problem through two-stage regression.

In the first stage, a probit model is used to estimate the selection equation, which describes the probability of sample selection. Specifically, the selection equation is expressed as follows:

$$P(D_i = 1|Z_i) = \Phi(Z_i'\delta) \quad (2)$$

Where  $D_i$  is a binary variable indicating whether individual  $i$  receives the treatment (0 or 1);  $Z_i$  is the set of explanatory variables related to the selection, including control variables and exogenous instrumental variables;  $\delta$  is the parameter vector to be estimated; and  $\Phi$  denotes the cumulative distribution function of the standard normal distribution. By estimating this equation, the estimated value  $\hat{\delta}$  is obtained, and the inverse Mills ratio (IMR), denoted as  $\hat{\lambda}_i$ , is calculated.

In the second stage, ordinary least squares regression is performed with the following model:

$$Y_i = X_i\beta + D_i\gamma + \hat{\lambda}_i\rho + \varepsilon_i \quad (3)$$

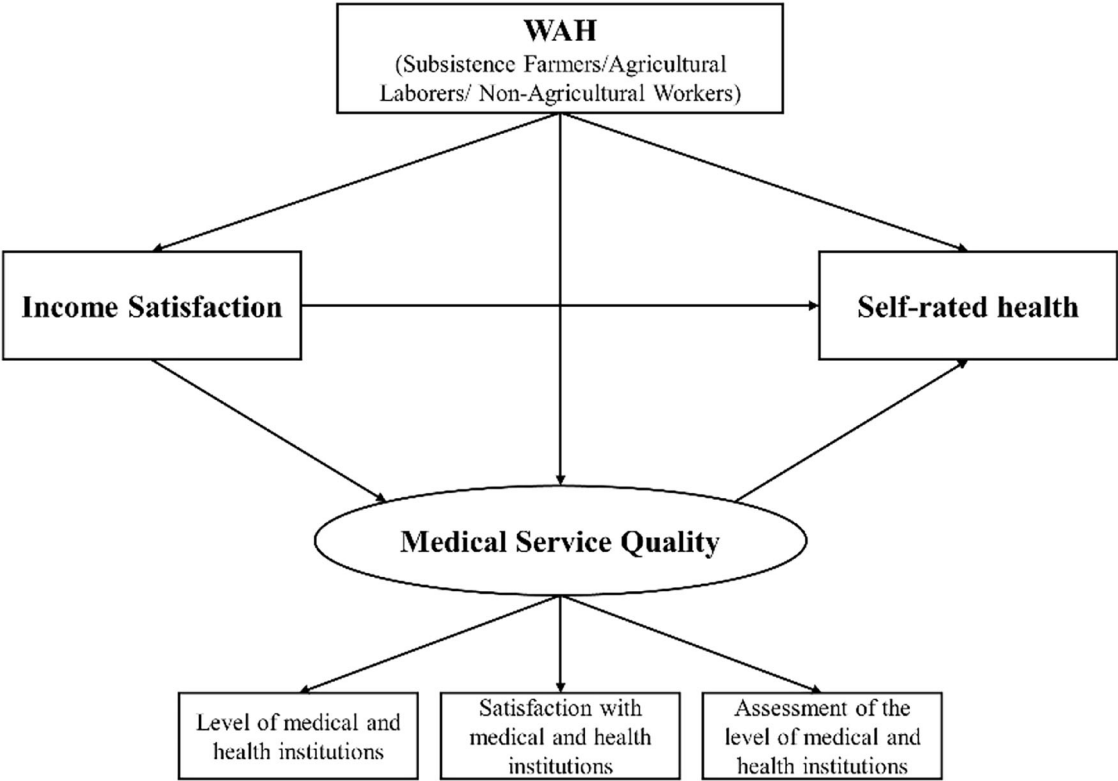
Where  $Y_i$  is the dependent variable,  $X_i$  represents the control variables,  $D_i$  is the core explanatory variable (the binary variable indicating whether the treatment was received), and  $\hat{\lambda}_i$  is the IMR obtained in the first stage, included as an additional control variable in the regression. The regression provides estimates of  $\hat{\beta}$ ,  $\hat{\gamma}$ ,  $\hat{\rho}$ .

In this model, the inclusion of  $\hat{\lambda}_i$  as a control variable helps eliminate potential endogeneity problems caused by self-selection bias. If the estimated coefficient  $\hat{\rho}$  is significant, it indicates that self-selection bias has a significant impact on the results; the treatment effect estimation should be interpreted after controlling for this bias. Conversely, if  $\hat{\rho}$  is not significant, it suggests that self-selection bias has a minimal impact on the regression results and that the baseline regression results are reliable.

**Double machine learning (DML).** To further address the endogeneity problem, this study employs the DML method. DML effectively reduces potential endogeneity biases by combining traditional regression models with machine-learning algorithms, thus improving the accuracy of causal effect estimation. This study performs the following steps:

First, supervised machine learning was employed to estimate conditional expectations. The partial linear model is used to estimate the conditional expectations of the outcome variable  $Y$  and the treatment variable  $D$  under the control variables  $X$ , specifically  $E[Y|X]$  and  $E[D|X]$ . The partial linear model is specified as follows, where  $g_0(X)$  and  $m_0(X)$  represent the nonparametric parts related to the control variables. This setup allows the estimation of the true value of  $\theta_0$  under the control of  $X$ .

$$Y = D\theta_0 + g_0(X) + \zeta, E(\zeta|D, X) = 0 \quad (4)$$



**Fig. 3** The chain mediation model of self-rated health differences among WAH.

$$D = m_0(X) + V, E(V|X) = 0 \tag{5}$$

Next, cross-validation and fitting were conducted. In this step, cross-fitting involves repeatedly randomizing the sample data, ensuring that each sample point is fully utilized for both training and validation. This enhances the model’s stability and robustness. Cross-fitting not only reduces endogeneity issues but also mitigates selection bias, strengthening the model’s predictive capability for causal effects.

Finally, causal effects are estimated. Using the previous steps, more precise estimations of causal relationships are obtained, providing more reliable evidence for policy analysis and practical decision-making.

*Structural equation model (SEM).* Based on the study by Werneck et al. (2018), this study employs the SEM to conduct a chain mediation analysis, exploring the self-rated health differences between subsistence farmers and agricultural laborers, as well as between subsistence farmers and non-agricultural workers. Specifically, this study analyzes the chain mediating effects of income satisfaction and medical service quality on health disparities. This research was conducted using Stata 17, with the results of the SEM represented by standardized regression coefficients (*r*). To estimate the overall fit of the model, the study employs the following parameters: chi-square value and its significance, root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residual (SRMR). Acceptable model fit criteria were RMSEA < 0.08, CFI > 0.90, and SRMR < 0.08; whereas RMSEA < 0.06, CFI > 0.95, and SRMR < 0.08 were considered good fit values (Lei and Wu, 2007).

Figure 3 shows the chain mediation model of self-rated health differences among WAH. Among them, WAH, income satisfaction, and medical service quality as predictors, with the level of self-rated health as the theoretical outcome of the model. The

Table 1 Sample size and proportion of each group.		
Type	Sample Size	Proportion of WAH
WAH	16,439	—
Subsistence Farmers	8853	53.85%
Agricultural Laborers	497	3.02%
Non-Agricultural Workers	7089	43.12%

model uses a 95% significance level to determine significant correlations.

**Results**

**Descriptive statistics.** This study presents the specific sample sizes and proportions of the three subgroups of WAH. As shown in Table 1, subsistence farmers comprise the largest proportion, accounting for 53.85% (8853 individuals), followed by non-agricultural workers, who comprise 43.12% (7089 individuals). Agricultural laborers make up a smaller proportion, only 3.02% (497 individuals). These data indicate that subsistence farmers and non-agricultural workers are the primary components of the WAH. Although the proportion of agricultural laborers is relatively low in the sample, considering China’s large population base, this group still represents a significant absolute number nationwide.

Table 2 shows the descriptive statistics for the main variables in the study. The results show that the average self-rated health level for the overall sample was 3.011, suggesting that the health status of the respondents was generally at a moderate level. Regarding individual objective characteristics, the average age of the sample was 46.9 years, with males accounting for 52.9%, indicating a relatively balanced gender ratio. In terms of marital status, 89.4% of the sample was married, reflecting a stable marital situation among the respondents. The average education level was 8.2,

**Table 2 Variable definition and descriptive statistics.**

Variable	Variable definition	Sample size	Mean	Variance	Minimum	Maximum
Self-rated health	This value is proportional to the level of health and is assigned from 1 to 5	16,439	3.000	1.219	1	5
Age	The age of the individual	16,439	47.459	13.050	16	87
Gender	Male = 1, Female = 0	16,439	0.522	0.500	0	1
Marital status	Married = 1, unmarried = 0	16,439	0.903	0.297	0	1
Education Level	No schooling = 0, Primary/Private school = 6, Junior High = 9, High School = 12, College = 15, Bachelor's = 16, Master's/PhD = 19	16,439	7.153	4.086	0	19
Belief	Has religious belief = 1, No religious belief = 0	16,439	0.294	0.456	0	1
Medical history	Chronic illness in the past 6 months: yes = 1, not = 0	16,439	0.163	0.369	0	1
Medical insurance	Medical insurance status: yes = 1, not = 0	16,439	0.942	0.233	0	1
Life satisfaction	This value is directly proportional to life satisfaction and is assigned from 1 to 5	16,439	3.870	1.021	1	5
Social status evaluation	This value is directly proportional to life satisfaction and is assigned from 1 to 5	16,439	3.065	1.057	1	5
Exercise	Number of times exercised in the past week	16,439	1.688	2.653	0	7
Drinking	Drinking: yes = 1, not = 0	16,439	0.162	0.368	0	1
Smoking	Smoking: yes = 1, not = 0	16,439	0.324	0.468	0	1
Household Size	Number of people in the household	16,439	4.353	1.887	1	13
Household Net Income	Household net income, ln-transformed	16,439	10.838	0.977	0	15.613
Household Total Expenditure	Household total expenditure, ln-transformed	16,439	10.733	0.890	0	15.458
Income satisfaction	Satisfied = 1, dissatisfied = 0	16,439	3.241	1.065	1	5
Rank of medical and health institutions	Ranking of medical institutions in China: Clinics =1; Community health service station/village clinic =2; Community health service center/township health =3; Specialized hospital =4; General hospital =5	16,439	2.827	1.491	1	5
Satisfaction with medical and health institutions	This value is proportional to satisfaction and is assigned from 1 to 5	16,439	3.293	0.960	1	5
Assessment of the level of medical and health institutions	This value is proportional to the level evaluation and is assigned from 1 to 5	16,439	3.232	0.965	1	5

indicating that most respondents had completed junior high school. Additionally, 29.4% of the sample practiced religion, 15.9% of respondents reported having a chronic illness in the past six months, and 92.7% had health insurance.

In terms of individual subjective characteristics, the average life satisfaction score was 3.863, suggesting that most respondents were relatively satisfied with their lives. The average self-assessment of social status was 3.018, indicating that respondents generally perceived their social status to be at a moderate level. Regarding health behaviors, the frequency of exercise among respondents was low, and the proportions of those who drink alcohol and smoke were 16.2% and 32.1%, respectively, indicating certain health risk behaviors. In terms of family characteristics, the average family size was 4.227, indicating a moderate number of family members. The average household net income (ln-transformed) was 11.012, and the average total household expenditure (ln-transformed) was 10.883, reflecting economic diversity.

In the analysis of mediation variables, the average income satisfaction score was 3.253, indicating that the income satisfaction level of the sample was moderate. In terms of healthcare, the average rating of healthcare institutions chosen by respondents was 3.046, suggesting that most respondents selected medium-tier healthcare institutions. The satisfaction and service evaluation scores for these institutions were 3.297 and 3.235, respectively, indicating generally high levels of satisfaction.

**Baseline regression.** To analyze the health differences among various subgroups of WAH, this study conducted a regression analysis comparing subsistence farmers, agricultural laborers, and non-agricultural workers. Table 3 presents the results of the

comparative analysis. Specifically, columns (1), (3), and (5) show the regression results with individual characteristics controlled, while columns (2), (4), and (6) further control for family characteristics. All regression models controlled for regional and year fixed effects to ensure the accuracy of the analysis.

Columns (1) and (2) focus on the self-rated health difference between subsistence farmers and agricultural laborers (where subsistence farmers are coded as 1 and agricultural laborers as 0). The results show that the self-rated health level of subsistence farmers is significantly lower than that of agricultural laborers, and this difference holds at the 5% significance level. Columns (3) and (4) compare the self-rated health levels between agricultural laborers and non-agricultural workers (where agricultural laborers are coded as 1 and non-agricultural workers as 0). The analysis revealed that there was no significant difference in self-rated health between the two groups. This suggests that within the current statistical framework, the health status of agricultural laborers and non-agricultural workers shows no noticeable statistical difference. Finally, columns (5) and (6) show the self-rated health differences between subsistence farmers and non-agricultural workers (with subsistence farmers coded as 1 and non-agricultural workers as 0). The results show that the self-rated health level of subsistence farmers is significantly lower than that of non-agricultural workers, with this difference holding at the 1% significance level. This further emphasizes the relative health disadvantage of subsistence farmers.

Overall, the regression results revealed significant differences in self-rated health across different WAH subgroups. The self-rated health level of subsistence farmers was the lowest, while no significant health differences were found between agricultural laborers and non-agricultural workers. These findings empirically validate *Hypotheses 1a and 1b*.



**Table 3 Comparative analysis of self-rated health among different WAH subgroups.**

	(1) SRH	(2) SRH	(3) SRH	(4) SRH	(5) SRH	(6) SRH
Subsistence Farmers vs Agricultural Laborers	−0.112** (0.055)	−0.111** (0.055)				
Agricultural Laborers vs. Non-Agricultural Workers			0.034 (0.054)	0.034 (0.054)		
Subsistence Farmers vs. Non-Agricultural Workers					−0.091*** (0.022)	−0.090*** (0.022)
Individual Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables						
Family Characteristics		Yes		Yes		Yes
Control Variables						
Year Fixed	Yes	Yes	Yes	Yes	Yes	Yes
Regional Fixed	Yes	Yes	Yes	Yes	Yes	Yes
_cons	2.718*** (0.115)	2.697*** (0.217)	2.846*** (0.092)	2.851*** (0.239)	2.725*** (0.069)	2.697*** (0.159)
N	9349	9349	7584	7584	15940	15940
r <sup>2</sup>	0.236	0.238	0.199	0.200	0.227	0.229

(1) \*\* and \*\*\* denote statistical significance at the 5% and 1% significance levels, respectively. (2) Individual characteristic control variables: age, gender, marital status, education level, belief, medical history, medical insurance, life satisfaction, social status evaluation, exercise, smoking, and drinking. Family characteristic variables: household size, household net income, and household total expenditure.

**Table 4 Results of robustness checks.**

	Unbalanced Panel			Control Variables Trimming 1%			Change of Dependent Variable Assignment Method		
	SRH	SRH	SRH	SRH	SRH	SRH	SRH	SRH	SRH
Subsistence Farmers vs. Agricultural Laborers	−0.105** (0.042)			−0.110** (0.055)			−0.036* (0.020)		
Agricultural Laborers vs. Non-Agricultural Workers		0.022 (0.041)			0.035 (0.054)			−0.007 (0.019)	
Subsistence Farmers vs. Non-Agricultural Workers			−0.094*** (0.017)						
Individual Characteristics						−0.089*** (0.022)			−0.046*** (0.008)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Regional Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
_cons	2.365*** (0.161)	2.796*** (0.175)	2.496*** (0.117)	2.695*** (0.222)	2.768*** (0.247)	2.667*** (0.164)	0.494*** (0.086)	0.454*** (0.092)	0.487*** (0.063)
N	16192	12946	27475	9349	7584	15940	9349	7584	15940
r <sup>2</sup>	0.221	0.198	0.233	0.239	0.200	0.229	0.196	0.168	0.202

(1) \*, \*\*, and \*\*\* denote statistical significance at the 10%, 5%, and 1% significance levels, respectively. (2) The control variables are consistent with those used in Table 3.

**Robustness checks.** To enhance the credibility of the conclusions, this study conducted robustness checks to further verify the differences in self-rated health levels between different WAH. Specifically, three robustness checks were employed. First, an unbalanced panel dataset was used, retaining samples with two or three periods of data. Second, 1% trimming was applied to control variables to eliminate the impact of potential outliers on the regression results. Third, the dependent variable assignment method was changed, where samples with self-rated health levels greater than or equal to 3 were assigned a value of 1 and those with values less than 3 were assigned a value of 0. The regression analysis was conducted on this binary dependent variable.

Table 4 presents the regression results from the robustness checks. The results show that, whether using unbalanced panel data, applying 1% trimming to control variables, or changing the dependent variable assignment method, the regression results are consistent with the baseline results in most cases. This further validates the main conclusions of this study: The health status of subsistence farmers is significantly lower than that of agricultural laborers and non-agricultural workers, while no significant health

difference exists between agricultural laborers and non-agricultural workers. These results demonstrate consistency across different data-handling methods, strengthening the robustness and credibility of the conclusions.

### Endogeneity treatment

**Treatment effects model.** When individuals choose to become different types of WAH (such as subsistence farmers, agricultural laborers, or non-agricultural workers), their decisions may be influenced by unobservable factors, leading to sample self-selection bias. To address this issue, this study employed the treatment effects model to analyze the self-rated health differences between subsistence farmers and agricultural laborers as well as between subsistence farmers and non-agricultural workers.

Table 5 shows the regression results of the treatment effects model. First, in the selection equation, the urbanization rate serves as a valid instrumental variable with good identification. Theoretically, regional urbanization levels influence employment

opportunities and land transfer markets, which significantly correlate with the career choices of WAH (the estimated coefficient is significantly negative at the 1% level), but do not directly affect individual health levels, satisfying the exogeneity condition for the instrumental variable. Second, in the outcome equation, the regression coefficient of the hazard function (the IMR, Lambda) is positive and statistically significant at the 1% level. This suggests that sample selection bias did exist, and unobservable heterogeneous factors significantly affected the self-rated health levels of WAH. Therefore, it was necessary to use the treatment effects model to correct for selection bias.

After controlling for self-selection bias, the study found that significant health-level differences exist between different WAH groups. Specifically, subsistence farmers are at a greater health disadvantage compared to agricultural laborers, with their self-rated health score on average being 1.985 units lower (regression coefficient =  $-1.985$ ,  $p < 0.01$ ). When the comparison group is switched to non-agricultural workers, the health disadvantage of subsistence farmers remains significant, but the effect size is notably reduced (regression coefficient =  $-0.370$ ,  $p < 0.01$ ).

**Double machine learning (DML).** To address potential endogeneity issues, this study employed the DML method to conduct robustness checks on the self-rated health differences between subsistence farmers and agricultural laborers as well as between

subsistence farmers and non-agricultural workers. The model used partial linear regression to estimate conditional expectations and cross-validation to ensure robustness. The number of folds for the subsistence farmers and agricultural laborers was 3733, while the number of folds for the subsistence farmers and non-agricultural workers group was 5703, ensuring sufficient training and validation to reduce the risk of overfitting. Additionally, the resampling frequency was set to 1, meaning that each training instance was validated using a single sample.

Table 6 presents the main estimation results. Column (1) shows the self-rated health differences between subsistence farmers and agricultural laborers. The results indicate that, compared to agricultural laborers, subsistence farmers have a lower self-rated health level, and this difference is statistically significant at the 1% level. Column (2) presents the comparison between subsistence farmers and non-agricultural workers, revealing a similar significant self-rated health difference. The results indicate that compared to subsistence farmers, non-agricultural workers have significantly higher self-rated health levels.

Through robustness checks using the DML, the study further validated the self-rated health differences between subsistence farmers and agricultural laborers, as well as between subsistence farmers and non-agricultural workers. The findings not only demonstrate that subsistence farmers have relatively poorer health status but also show that the DML method successfully addresses endogeneity issues, thereby ensuring the reliability of causal effect estimates.

**Mechanism analysis.** The empirical results above show that there are certain differences in self-rated health levels between subsistence farmers, agricultural laborers, and non-agricultural workers. Specifically, compared to agricultural laborers and non-agricultural workers, subsistence farmers have lower self-rated health levels. However, no significant health difference exists between agricultural laborers and non-agricultural workers; thus, no further analysis of the influencing mechanism is necessary. A more in-depth study will continue to explore the mechanisms through which different WAH affect their self-rated health levels.

This study used the SEM to examine the potential multiple mediation effects. Multiple mediation models can be divided into single-step multiple mediation models and multiple-step multiple mediation models, where the latter refers to the existence of mutual effects between multiple mediation variables, and these variables exhibit sequential characteristics, forming a mediation chain (e.g.,  $X \rightarrow M1 \rightarrow M2 \rightarrow Y$  pathway) (Hayes, 2009).

**Table 5 Treatment effects model results.**

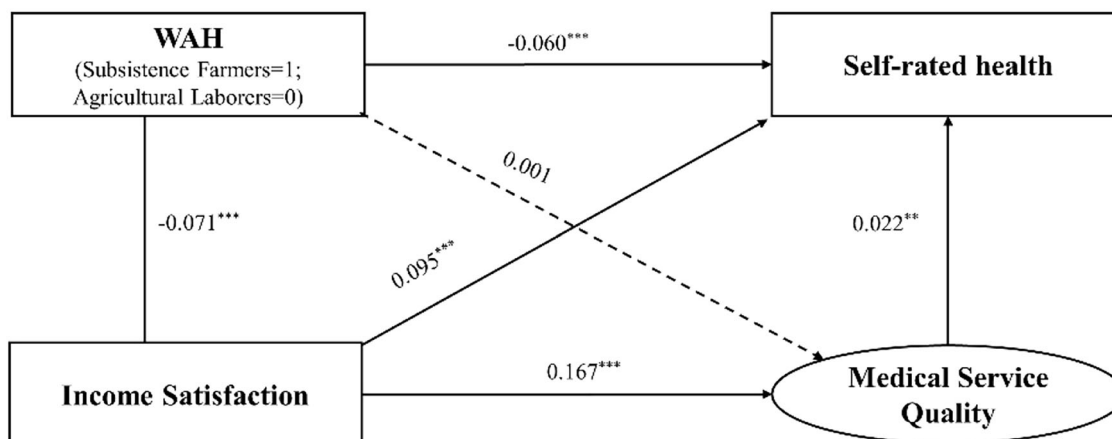
	(1) SRH	(2) SRH
Subsistence Farmers vs Agricultural Laborers	$-1.985^{***}$ (0.379)	
Subsistence Farmers vs Non-Agricultural Workers		$-0.370^{***}$ (0.079)
Individual Characteristics	Yes	Yes
Control Variables		
Family Characteristics Control Variables	Yes	Yes
_cons	$4.718^{***}$ (0.428)	$3.211^{***}$ (0.161)
Subsistence Farmers Urbanization Rate	$-1.806^{***}$ (0.160)	$-2.233^{***}$ (0.075)
_cons hazard:	$2.597^{***}$ (0.092)	$1.360^{***}$ (0.042)
lambda	$0.867^{***}$ (0.178)	$0.177^{***}$ (0.049)
N	9350	15942
r <sup>2</sup>	0.032	0.042

(1) \*\*\* denotes statistical significance at the 1% level. (2) The control variables are consistent with those used in Table 3.

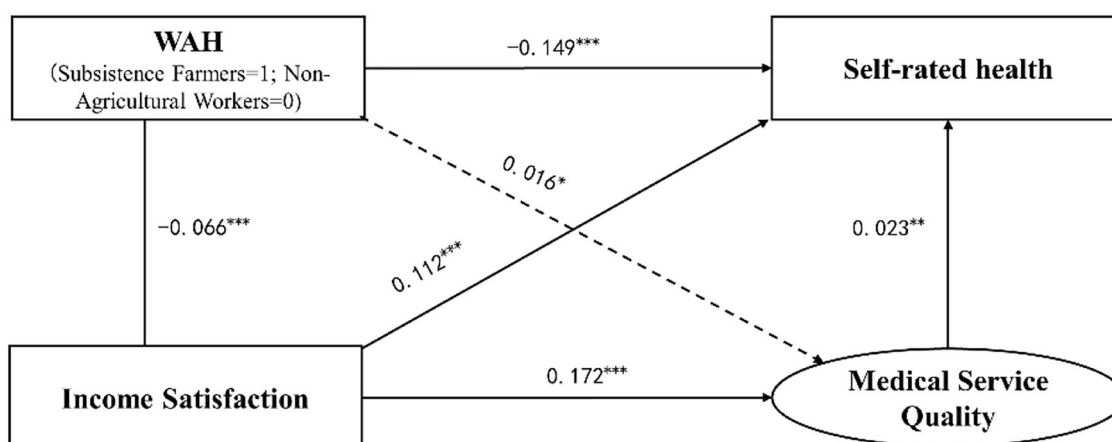
**Table 6 Double machine learning results.**

	(1) SRH	(2) SRH
Subsistence Farmers vs Agricultural Laborers	$-0.113^{**}$ (0.054)	
Subsistence Farmers vs Non-Agricultural Workers		$-0.090^{***}$ (0.022)
Individual Characteristics Control Variables		Yes
Family Characteristics Control Variables	Yes	Yes
Subsistence Farmers vs Agricultural Laborers	Yes	Yes
Year Fixed	Yes	Yes
Regional Fixed	Yes	Yes
_cons	$-0.003$ (0.012)	$-0.000$ (0.009)
N	9350	15,942

(1) \*\*, \*\*\* denotes statistical significance at the 5% and 1% level, respectively. (2) The control variables are consistent with those used in Table 3.



**Fig. 4** The chain mediation of differences in self-rated health levels between subsistence farmers and agricultural laborers. Note: \*\* and \*\*\* are significant at 5% and 1% levels, respectively.



**Fig. 5** The chain mediation of differences in self-rated health levels between subsistence farmers and non-agricultural workers. Note: \*\* and \*\*\* are significant at 5% and 1% levels, respectively.

Figures 4 and 5 present the path coefficients of the SEM models for the subsistence farmers vs. agricultural laborers and subsistence farmers vs non-agricultural workers groups, respectively. As shown in the figures, compared to the agricultural laborer and non-agricultural worker groups, subsistence farmers have lower income satisfaction, which negatively affects their medical service quality, subsequently reducing their self-rated health levels. In contrast, non-agricultural workers exhibit relatively better income satisfaction, which enhances their medical service evaluation and, in turn, improves their self-rated health levels.

Furthermore, to ensure that the model can effectively explain the data, model fit tests were conducted, the results are presented in Table 7. RMSEA, CFI, SRMA, chi-square, and their significance are indicators of the overall fit of the SEM. The results show that the model fits well in all aspects. Finally, bootstrap analysis was used to verify the mediation effects, with 500 repeated samples and a 95% confidence interval. Table 8 shows that the path results of the model indicate that bootstrap 95% confidence intervals for all three paths do not include 0, indicating significant mediation paths. Based on this analysis, Hypothesis 2 is validated, showing that income satisfaction and medical service quality play a chain mediation role in self-rated health differences among different WAH groups.

## Discussion

This study aimed to explore the differences in self-rated health levels among WAH subgroups and the chain mediation effect of

Table 7 Overall fit of the models.		
	Subsistence Farmers vs Agricultural Laborers	Subsistence Farmers vs Non-Agricultural Workers
RMSEA	0.048	0.046
CFI	0.979	0.981
SRMA	0.025	0.022
chi2_bs	6036.864	10415.173
P>chi2	0.000	0.000

income satisfaction and medical service quality on health outcomes. The findings in Table 9 are consistent with and extend previous research in several key areas. For example, building on the other works on WAH in China (Schneider, 2015; Zhang and Donaldson, 2010), we examined the differences in self-rated health levels among WAH subgroups, supporting and expanding upon the research by Mueller (2021) and van Oostrom et al. (2021) on occupational identity, work environment, and self-health assessments. Our investigation of income satisfaction and medical service quality as mediating factors was based on the work of Sha et al. (2019) and Zheng et al. (2023) on public health and well-being. Integrating public service accessibility theory further revealed the chain mediation effect of income satisfaction and medical service quality. This highlights how access to economic resources affects the efficiency of medical resource

**Table 8 Bootstrap analysis of the significance test for mediating effects.**

	The mediating effect affects the path	Standardized Coefficient	Standard error	95% Confidence interval	
				Lower limit	Upper limit
Subsistence Farmers vs Agricultural Laborers	Direct effects: subsistence farmers → self-rated health	−0.343***	0.055	−0.451	−0.235
	Subsistence farmers → income satisfaction → medical service quality → self-rated health	−0.378***	0.056	−0.487	−0.269
Subsistence Farmers vs Non-Agricultural Workers	Direct effects: Non-agricultural workers → self-rated health	−0.364***	0.020	−0.403	−0.325
	Non-agricultural workers → income satisfaction → medical service quality → self-rated health	−0.383***	0.020	−0.422	−0.343

\*\*\* denotes statistical significance at the 1% level.

**Table 9 Comparison of related research findings.**

Study	Focus	Key Findings	Relation to Our Study
Schneider (2015) Zhang and Donaldson (2010)	Conceptual changes and type differentiation of WAH in China under the background of agricultural industrialization	<b>Concept:</b> The meaning and usage of WAH in China have become more complex. <b>Classification:</b> The transformation of rural land in China has altered agricultural production forms and production relations, leading to the differentiation of subsistence peasants into six categories: Commercial Farmer, Entrepreneurial Farmer, and Contract Farmer etc.	This study, based on China's reality, classifies WAH into three categories—subsistence farmers, agricultural laborers, and non-agricultural workers—according to employer type and work nature.
Mueller (2021) van Oostrom et al. (2021) Brennan et al. (2022) Yunker and Radunovich (2021)	Agricultural occupational identity and health welfare	There is ongoing debate in the academic community regarding the health status of WAH.	We investigated the self-rated health levels of WAH subgroups and explained the reasons behind these differences.
Penchansky and Thomas (1981) Li and Wang (2024)	Definition and dimensions of public service accessibility theory	Public Service Accessibility Theory emphasizes that the accessibility of public services refers to the ability of individuals or groups to obtain the public services they need. This ability consists of five dimensions: accessibility, acceptability, availability, affordability, and geographical accessibility.	In combination with Public Service Accessibility Theory, this study analyzes the role of economic factors such as income and public service factors such as medical service quality in the manifestation of self-rated health disparities among WAH subgroups.
Tian et al. (2023) Adejoh et al. (2022) Martins et al. (2023) Wu and Zeng (2024) OBrien et al. (2012)	Health disparities among WAH subgroups under gender, age, income, and education differences	Female, elderly, low-income, and low-educated WAH tend to exhibit poorer health outcomes.	This study investigates the self-rated health disparities among the three WAH subgroups—subsistence farmers, agricultural laborers, and non-agricultural workers—classified according to employer type and work nature.
Sha et al. (2019) Zheng et al. (2023)	The mediating role of income satisfaction and medical service quality	<b>Income satisfaction</b> mediates the relationship between commuting time and subjective well-being <b>Perceived attitudes toward medical services</b> significantly mediate the relationship between overall public satisfaction with medical service and self-rated health.	This study further reveals the chain mediation effect of income satisfaction and medical service quality, providing a more systematic explanation of the mechanisms behind health disparities.

utilization, thus providing a more systematic explanation of the mechanisms behind health disparities.

The research findings indicate that, compared to agricultural laborers and non-agricultural workers, subsistence farmers have the lowest self-rated health levels. There was no significant difference in self-rated health between agricultural laborers and non-agricultural

workers. The self-rated health levels of these subgroups are closely related to their work nature and work environment. Subsistence farmers typically engage in heavy agricultural labor in rural areas, with low incomes (Brennan et al., 2022), and face risks associated with changes in the natural environment (Wang et al., 2014). These factors often lead to poorer self-rated health (Ryder et al., 2011).



In contrast, agricultural laborers employed by agricultural enterprises or cooperatives do not bear the risks associated with agricultural production (Abate et al., 2014) and generally have better work and healthcare conditions (e.g., agricultural companies provide modern farming equipment and healthcare) (Ayantoyinbo et al., 2024; Wossen et al., 2017), resulting in better self-rated health. Non-agricultural workers in cities tend to have better self-rated health due to more job opportunities and higher wages (Yang et al., 2021). When comparing agricultural laborers and non-agricultural workers, agricultural laborers are still affected by the seasonality of agricultural production and face issues such as job instability and income fluctuations (Wei et al., 2019). Meanwhile, migrant workers in cities also encounter difficulties with social integration (Zhou et al., 2023) and high levels of life stress (James, 2021). As a result, there was no significant health disparity between the two groups.

Additionally, based on public service accessibility theory, this study found that the relationship between subsistence farmers and their self-rated health levels may be mediated by income satisfaction and medical service quality. Subsistence farmers in China have a higher prevalence of chronic diseases (Yu et al., 2022), making it common for them to seek medical attention for minor illnesses or chronic conditions. However, their lower socioeconomic status and income level often force subsistence farmers to seek medical care in rural health centers, village clinics, and other primary medical institutions with scarce medical resources and poor medical levels (Kim and Kim, 2016; Song et al., 2019; Zhang et al., 2015). Yet, primary medical institutions often fail to fulfill the health system's functions and might be perceived to offer low-quality services. These factors affect the accessibility, availability, and adaptability of individuals to public services, leading to their distrust (Zhang et al., 2017); hence, they might report lower satisfaction with medical services, further lowering their evaluation of their health status.

## Conclusions

This study assessed the self-rated health outcomes of WAH subgroups in China and explored the continuous mediating effects of income satisfaction and medical service quality on the self-rated health levels of these subgroups. The results of the study indicate the following: 1) Within the WAH, the subsistence farmer subgroup had the lowest self-rated health levels; in contrast, no statistically significant health differences were found between agricultural laborers and non-agricultural workers. 2) Compared to agricultural laborers and non-agricultural workers, the subsistence farmer group performed worse in terms of income satisfaction, which in turn lowered their evaluation of medical service quality and self-rated health levels.

The limitations of this study and plans for further research are as follows: First, considering that the data used in this study primarily comes from the 2016–2020 CFPS, the sample's geographic and temporal scope is limited, which may affect the generalizability of the research findings. Therefore, future research plans to incorporate additional databases, especially those with more representative samples from different regions of the world, to enhance the reliability and external validity of the results. Second, the next phase could involve a mixed-methods research design, combining qualitative research (such as in-depth interviews or case studies) with quantitative analysis for a more comprehensive perspective. Additionally, the study's measurement of WAH's health status mainly relies on self-rated health indicators. However, self-rated health may be influenced by personal subjective perceptions and socio-cultural differences, potentially not fully reflecting actual health status. Hence, future research could integrate multiple data sources and more precise

measurement tools (such as physiological health data, psychological health assessments, etc.) to conduct a more objective and comprehensive health evaluation that addresses the subjectivity issues inherent in self-rated health measurements.

## Data availability

Due to the restrictions outlined in the CFPS Data Use Agreement (<https://www.issp.pku.edu.cn/cfps/cjwt/fbxg/1379024.htm>), the study is not permitted to share the raw or processed datasets on external websites. However, the raw CFPS data is publicly available on the official CFPS website (<https://www.issp.pku.edu.cn/cfps/>).

Received: 31 October 2024; Accepted: 2 April 2025;

Published online: 15 April 2025

## Notes

- 1 Schneider elaborated on the meaning and complex usage of the term “farmer” in the context of China's agricultural industrialization. Zhang, Q., and Donaldson argued that rural land transformations in China altered agricultural production forms and relations, leading to the differentiation of subsistence peasants into six categories: Commercial Farmer, Entrepreneurial Farmer, Contract Farmer, Semiproletarian Farm Worker with Chinese Characteristics, Semiproletarian Farm Worker, and Proletarian Farm Worker.
- 2 The third national agricultural census in China indicates that subsistence farmers are the main form of agricultural production and operation in the country. As of 2016, they still accounted for more than 98% of the total number of contracted agricultural households. <https://www.stats.gov.cn/sj/pcsj/nypc/202302/U02030223531273769774.pdf>. Accessed 22 Feb 2025.
- 3 National Bureau of Statistics (2023) China Statistical Yearbook 2023. <https://www.stats.gov.cn/sj/ndsj/2023/indexch.htm>. Accessed 20 Feb 2025.
- 4 National Bureau of Statistics (2023) Monitoring Survey Report on Migrant Workers. [https://www.stats.gov.cn/sj/zxfb/202404/t20240430\\_1948783.html](https://www.stats.gov.cn/sj/zxfb/202404/t20240430_1948783.html). Accessed 20 Feb 2025.
- 5 International Labor Organization (2025). [https://rshiny.ilo.org/dataexplorer35/?lang=en&segment=indicator&id=EMP\\_2IFL\\_SEX\\_RT\\_A](https://rshiny.ilo.org/dataexplorer35/?lang=en&segment=indicator&id=EMP_2IFL_SEX_RT_A). Accessed 20 Feb 2025.

## References

- Abate GT, Francesconi GN, Getnet K (2014) Impact of agricultural cooperatives on smallholders' technical efficiency: empirical evidence from Ethiopia. *Ann Public Coop Econ* 85(2):257–286. <https://doi.org/10.1111/apce.12035>
- Adejoh SO, Attoh F, Akinbode GA et al. (2022) Socio-demographic, community characteristics and fear of crime as correlates of self-reported health status in Lagos, Nigeria. *Ibom J Soc Issues* 11(2). <https://doi.org/10.36108/ijsi/2202.11.0270>
- Akpınar VC, Mandıracıoğlu A, Özvurmaz S et al. (2023) Mental well-being as a predictor of quality of life in elderly agricultural workers. *Acıbadem Univ Sağlık Bilimleri Dergisi* 14(2). <https://doi.org/10.31067/acusaglik.1233982>
- Arraras JL, Illarramendi JJ, Viudez A et al. (2013) Determinants of patient satisfaction with care in a Spanish oncology Day Hospital and its relationship with quality of life. *Psychooncology* 22(11):2454–2461. <https://doi.org/10.1002/pon.3307>
- Ayantoyinbo AA, Alufolia GO, Babalola DO et al. (2024) Effect of agricultural cooperative businesses on the wellbeing of members in Ogun State, Nigeria. *Afr J Agric Food Sci* 7(3):22–38. <https://doi.org/10.52589/ajafs-2bl7vvue>
- Brennan M, Hennessy T, Meredith D et al. (2022) Weather, workload and money: determining and evaluating sources of stress for farmers in Ireland. *J Agromed* 27(2):132–142. <https://doi.org/10.1080/1059924X.2021.1988020>
- Brew B, Inder K, Allen J et al. (2016) The health and wellbeing of Australian farmers: a longitudinal cohort study. *BMC Public Health* 16:988. <https://doi.org/10.1186/s12889-016-3664-y>
- Chen J, Qiang L, Qian Z et al. (2020) Cluster analysis of differences in medical economic burden among residents of different economic levels in Guangdong Province, China. *BMC Health Serv Res* 20(1):988. <https://doi.org/10.1186/s12913-020-05817-y>
- Crosby F, Hamilton V (2017) Social comparison theory. In: *The Wiley-Blackwell encyclopedia of social theory*, Wiley, New York, p 1–2
- D'Ambrosio C, Frick JR (2007) Income satisfaction and relative deprivation: an empirical link. *Soc Indic Res* 81:497–519. <https://doi.org/10.1007/s11205-006-0020-0>
- Dorélien A, Xu HW (2020) Estimating rural–urban disparities in self-rated health in China: impact of choice of urban definition. *Demogr. Res* 43:1429–1460. <https://doi.org/10.4054/DemRes.2020.43.49>

- Dorosh PA, Thurlow J (2021) Agricultural growth, urbanization, and poverty reduction. In: *Agricultural Development: new perspectives in a changing world*. IFPRI, Washington DC, p 285–320
- Gong P, Liang S, Carlton EJ et al. (2012) Urbanisation and health in China. *Lancet* 379(9818):843–852. [https://doi.org/10.1016/S0140-6736\(11\)61878-3](https://doi.org/10.1016/S0140-6736(11)61878-3)
- Gupta D, Patel K, Lis CG (2015) Self-rated health supersedes patient satisfaction with service quality as a predictor of survival in prostate cancer. *Health Qual Life Outcomes* 13:137. <https://doi.org/10.1186/s12955-015-0334-1>
- Haseli-Mashhadi N, Pan A, Ye X et al. (2009) Self-rated health in middle-aged and elderly Chinese: distribution, determinants and associations with cardiometabolic risk factors. *BMC Public Health* 9:368. <https://doi.org/10.1186/1471-2458-9-368>
- Iskandarsyah A, de KC, Suardi DR et al. (2013) Satisfaction with information and its association with illness perception and quality of life in Indonesian breast cancer patients. *Support Care Cancer* 21:2999–3007. <https://doi.org/10.1007/s00520-013-1877-5>
- James T (2021) The Covid-19 pandemic and structural barriers for migrant agricultural workers in Ontario. *West J Leg Stud* 12(1). <https://doi.org/10.5206/uwojls.v12i1.13636>
- Jones-Bitton A, Best C, MacTavish J et al. (2020) Stress, anxiety, depression, and resilience in Canadian farmers. *Soc Psychiatry Psychiatr Epidemiol* 55(2):229–236. <https://doi.org/10.1007/s00127-019-01738-2>
- Jylhä M (2009) What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Soc Sci Med* 69(3):307–316. <https://doi.org/10.1016/j.socscimed.2009.05.013>
- Kaplan G, Baron-Epel O (2003) What lies behind the subjective evaluation of health status? *Soc Sci Med* 56(8):1669–1676. [https://doi.org/10.1016/s0277-9536\(02\)00179-x](https://doi.org/10.1016/s0277-9536(02)00179-x)
- Kim I, Koo MJ, Lee HE et al. (2019) Overwork-related disorders and recent improvement of national policy in South Korea. *J Occup Health* 61(4):288–296. <https://doi.org/10.1002/1348-9585.12060>
- Kim JI, Kim G (2016) Country-level socioeconomic indicators associated with healthy life expectancy: income, urbanization, schooling, and internet users: 2000–2012. *Soc Indic Res* 129:391–402. <https://doi.org/10.1007/s11205-015-1107-2>
- Klingelschmidt J, Milner A, Khireddine-Medouni I et al. (2018) Suicide among agricultural, forestry, and fishery workers: a systematic literature review and meta-analysis. *Scand. J Work Environ Health* 44(1):3–15. <https://doi.org/10.5271/sjweh.3682>
- Kongtip P, Nankongnab N, Tipayamongkhogul M et al. (2018) A cross-sectional investigation of cardiovascular and metabolic biomarkers among conventional and organic farmers in Thailand. *Int J Environ Res Public Health* 15(11):2590. <https://doi.org/10.3390/ijerph15112590>
- Lee H, Cho SY, Kim JS et al. (2019) Difference in health status of Korean farmers according to gender. *Ann Occup Environ Med* 31:7. <https://doi.org/10.1186/s40557-019-0287-7>
- Lei P-W, Wu Q (2007) Introduction to structural equation modeling: issues and practical considerations. *Educ Meas Issues Pract* 26(3):33–43. <https://doi.org/10.1111/j.1745-3992.2007.00099.x>
- Li C, Wang MY, Song Y (2018a) Vulnerability and livelihood restoration of landless households after land acquisition: evidence from peri-urban China. *Habitat Int* 79:109–115. <https://doi.org/10.1016/j.habitatint.2018.08.003>
- Li J, Wang Y (2024) How public service accessibility affects health of migrants: evidence from China. *Risk Manag Health Policy* 17:3065–3084. <https://doi.org/10.2147/rmhp.s475634>
- Li JP, Dong LJ, Tian DN et al. (2018b) Association between pesticide exposure intensity and self-rated health among greenhouse vegetable farmers in Ningxia, China. *PLOS One* 13(12):e0209566. <https://doi.org/10.1371/journal.pone.0209566>
- Li Y, Wu Q, Liu G et al. (2014) Catastrophic health expenditure and rural household impoverishment in China: what role does the new cooperative health insurance scheme play? *PLoS One* 9(4):e93253. <https://doi.org/10.1371/journal.pone.0093253>
- Lilley R, Day L, Koehncke N et al. (2012) The relationship between fatigue-related factors and work-related injuries in the Saskatchewan Farm Injury Cohort Study. *Am J Ind Med* 55(4):367–375. <https://doi.org/10.1002/ajim.22003>
- Maddala GS (1983) *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press, Cambridge
- Martins CA, do Prado CB, Ferreira JRS et al. (2023) Self-rated health status and associated factors in rural workers. *BMC Public Health* 23(1):680. <https://doi.org/10.1186/s12889-023-15548-4>
- Meng Q, Xie Z, Zhang T (2014) A single-item self-rated health measure correlates with objective health status in the elderly: a survey in suburban Beijing. *Front Public Health* 2:27. <https://doi.org/10.3389/fpubh.2014.00027>
- Mueller JT (2021) Decomposing differences in poor self-rated health between those in agriculture and natural resource occupations and the rest of the labor force. *J Agromed* 26(2):109–119. <https://doi.org/10.1080/1059924X.2020.1713275>
- National Bureau of Statistics (2023) *Monitoring Survey Report on Migrant Workers*. [https://www.stats.gov.cn/sj/zxfb/202404/t20240430\\_1948783.html](https://www.stats.gov.cn/sj/zxfb/202404/t20240430_1948783.html). Accessed 20 Feb 2025
- Nguyen DN, Nguyen LH, Nguyen CT et al. (2019) Health status and health service utilization among Vietnamese farmers in a mountainous province. *Int J Environ Res Public Health* 16(23):4768. <https://doi.org/10.3390/ijerph16234768>
- Nguyen TM, Vu TT, Nguyen NA et al. (2020) Quality of farmer's life in rural areas in Vietnam. *E3S Web Conf* 175:10006. <https://doi.org/10.1051/e3sconf/202017510006>
- OBrien LV, Berry HL, Hogan A (2012) The structure of psychological life satisfaction: Insights from farmers and a general community sample in Australia. *BMC Public Health* 12:976. <https://doi.org/10.1186/1471-2458-12-976>
- Okuy RA, Tamir F, Ağaoğlu PM (2018) Occupational health and safety characteristics of agricultural workers in Adana, Turkey: a cross-sectional study. *PeerJ* 6:e4952. <https://doi.org/10.7717/peerj.4952>
- Onarheim K, Sisay M, Gizaw M et al. (2018) Selling my sheep to pay for medicines—household priorities and coping strategies in a setting without universal health coverage. *BMC Health Serv Res* 18(1):153. <https://doi.org/10.1186/s12913-018-2943-y>
- Penchansky R, Thomas JW (1981) The concept of access: definition and relationship to consumer satisfaction. *Med Care* 19(2):127–140. <https://doi.org/10.1097/00005650-198102000-00001>
- Perkio-Makela MM (2000) Finnish farmers' self-reported morbidity, work ability, and functional capacity. *Ann Agric Environ Med* 7(1):11–16
- Perlman F, Bobak M (2008) Determinants of self-rated health and mortality in Russia—are they the same? *Int J Equity Health* 7:19. <https://doi.org/10.1186/1475-9276-7-19>
- Pita-Fernandez S, Pertega-Diaz S, Lopez-Calvino B et al. (2013) Diagnostic and treatment delay, quality of life and satisfaction with care in colorectal cancer patients: a study protocol. *Health Qual Life Outcomes* 11:117. <https://doi.org/10.1186/1477-7525-11-117>
- Prus SG (2011) Comparing social determinants of self-rated health across the United States and Canada. *Soc Sci Med* 73(1):50–59. <https://doi.org/10.1016/j.socscimed.2011.04.010>
- Rask MT, Ørnboel E, Rosendal M et al. (2017) Long-term outcome of bodily distress syndrome in primary care: a follow-up study on health care costs, work disability, and self-rated health. *Psychosom Med* 79(3):345–357. <https://doi.org/10.1097/PSY.0000000000000405>
- Rinsky-Halivni L, Brammli-Greenberg S, Christiani DC (2022) Ageing workers' mental health during COVID-19: a multilevel observational study on the association with the work environment, perceived workplace safety and individual factors. *BMJ Open* 12(12):e064590. <https://doi.org/10.1136/bmjopen-2022-064590>
- Ryder AB, Wilkinson AV, McHugh MK et al. (2011) The advantage of imputation of missing income data to evaluate the association between income and self-reported health status (SRH) in a Mexican American cohort study. *J Immigr Minor Health* 13(6):1099–1109. <https://doi.org/10.1007/s10903-010-9415-8>
- Sanne B, Mykletun A, Moen BE et al. (2004) Farmers are at risk for anxiety and depression: the Hordaland Health Study. *Occup Med* 54(2):92–100. <https://doi.org/10.1093/occmed/kqh007>
- Schneider M (2015) What, then, is a Chinese peasant? Nongmin discourses and agroindustrialization in contemporary China. *Agric Hum Values* 32:331–346. <https://doi.org/10.1007/s10460-014-9559-6>
- Sha F, Li B, Law YW et al. (2019) Beyond the resource drain theory: salary satisfaction as a mediator between commuting time and subjective well-being. *J Transp Health* 15(2):100631. <https://doi.org/10.1016/j.jth.2019.100631>
- Shao CY, Meng XH, Cui SC et al. (2016) Income-related health inequality of migrant workers in China and its decomposition: an analysis based on the 2012 China labor-force dynamics survey data. *J Chin Med Assoc* 79(10):531–537. <https://doi.org/10.1016/j.jcma.2016.02.009>
- Song S, Yuan B, Zhang L et al. (2019) Increased inequalities in health resource and access to health care in rural China. *Int J Environ Res Public Health* 16(1):49. <https://doi.org/10.3390/ijerph16010049>
- Strauss J, Lei X, Park A et al. (2010) Health outcomes and socio-economic status among the elderly in China: evidence from the CHARLS Pilot. *J Popul Ageing* 3(3–4):111–142. <https://doi.org/10.1007/s12062-011-9033-9>
- Sun S, Chen J, Johannesson M et al. (2011) Population health status in China: EQ-5D results, by age, sex and socio-economic status, from the National Health Services Survey 2008. *Qual Life Res* 20(3):309–320. <https://doi.org/10.1007/s11136-010-9762-x>
- Tian Y, Zhan Y, Wu M (2023) Gender differences in migrant workers health in China. *Int J Public Health* 68:1605018. <https://doi.org/10.3389/ijph.2023.1605018>
- Tonelli S, Culp K, Donham K (2014) Work-related musculoskeletal disorders in senior farmers: Safety and health considerations. *Workplace Health Saf* 62(8):333–341. <https://doi.org/10.1177/216507991406200804>
- Treiman DJ, Mason WM, Lavelly W (2001) Determinants and consequences of the migration of rural Chinese from their village of origin. Paper presented at the International Forum on Rural Labor Mobility in China

- van Oostrom SH, Nachat A, Loef B et al. (2021) The mediating role of unhealthy behaviors and body mass index in the relationship between high job strain and self-rated poor health among lower educated workers. *Int Arch Occup Environ Health* 94(1):95–105. <https://doi.org/10.1007/s00420-020-01565-y>
- Vandenbosch J, Van den Broucke S, Vancorenland S et al. (2016) Health literacy and the use of healthcare services in Belgium. *J Epidemiol Community Health* 70(10):1032–1038. <https://doi.org/10.1136/jech-2015-206910>
- Wang F, Liu S, Chen T et al. (2023) How urbanization affects residents' health risks: evidence from China. *Environ Sci Pollut Res* 30(13):35554–35571. <https://doi.org/10.1007/s11356-022-24767-y>
- Wang J, Huang J, Yang J (2014) Overview of impacts of climate change and adaptation in China's agriculture. *J Integr Agric* 13(1):1–17. [https://doi.org/10.1016/S2095-3119\(13\)60588-2](https://doi.org/10.1016/S2095-3119(13)60588-2)
- Wang SH (2018) Inching up and socio-economic differentiation: exploring self-rated health of China's rural-to-urban migrants from 2005 to 2015. *Am J Health Behav* 42(5):117–125. <https://doi.org/10.5993/AJHB.42.5.10>
- Wang TF, Shi L, Nie X et al. (2013) Race/ethnicity, insurance, income and access to care: the influence of health status. *Int J Equity Health* 12:29–36. <https://doi.org/10.1186/1475-9276-12-29>
- Wei X, Önel G, Guan Z et al. (2019) Substitution between immigrant and native farmworkers in the United States: does legal status matter? *IZA J Dev Migr* 10(1):3. <https://doi.org/10.2478/izajodm-2019-0007>
- Werneck AO, Silva DR, Agostinete RR et al. (2018) Relationship of parental and adolescents' screen time to self-rated health: a structural equation modeling. *Health Educ Behav* 45(5):764–771. <https://doi.org/10.1177/1090198118757825>
- Wong WS, Fielding R (2008) The association between patient satisfaction and quality of life in Chinese lung and liver cancer patients. *Med Care* 46:293–302. <http://www.jstor.org/stable/40221657>
- Wossen T, Abdoulaye T, Alene AD et al. (2017) Impacts of extension access and cooperative membership on technology adoption and household welfare. *J Rural Stud* 54:223–233. <https://doi.org/10.1016/j.jrurstud.2017.06.022>
- Wu M, Zeng S (2024) Exploring factors influencing farmers' health self-assessment in China based on the LASSO method. *BMC Public Health* 24(1):333. <https://doi.org/10.1186/s12889-024-17809-2>
- Yang F, Jiang Y, Paudel KP (2021) Impact of work value awareness on self-rated physical health of rural-to-urban migrant workers in China. *Healthcare* 9(5):505. <https://doi.org/10.3390/healthcare9050505>
- Younker T, Radunovich HL (2021) Farmer mental health interventions: a systematic review. *Int J Environ Res Public Health* 19(1):244. <https://doi.org/10.3390/ijerph19010244>
- Yu C, Zhang X, Gao J (2022) Multi-dimensional comparison of the impact mechanism of the self-rated health status of urban and rural residents in Chinese social environments. *Int J Environ Res Public Health* 19(17):10625. <https://doi.org/10.3390/ijerph191710625>
- Yu T, Jiang Y, Gamber M et al. (2019) Socioeconomic status and self-rated health in China: findings from a cross-sectional study. *Medicine* 98(12):e14904. <https://doi.org/10.1097/MD.00000000000014904>
- Yu Z, Chen L (2016) Income and well-being: relative income and absolute income weaken negative emotion, but only relative income improves positive emotion. *Front Psychol* 7:2012. <https://doi.org/10.3389/fpsyg.2016.02012>
- Zhang QF, Donaldson JA (2010) From peasants to farmers: peasant differentiation, labor regimes, and land-rights institutions in China's agrarian transition. *Polit Soc* 38(4):458–489. <https://doi.org/10.1177/0032329210381236>
- Zhang R, Eschler J, Reddy M (2018) Online support groups for depression in China: culturally shaped interactions and motivations. *Comput Support Coop Work* 27:327–354. <https://doi.org/10.1007/s10606-018-9322-4>
- Zhang SC, Wei CN, Fukumoto K et al. (2011) A comparative study of health-promoting lifestyles in agricultural and non-agricultural workers in Japan. *Environ Health Prev Med* 16(2):80–89. <https://doi.org/10.1007/s12199-010-0167-9>
- Zhang T, Xu Y, Ren J et al. (2017) Inequality in the distribution of health resources and health services in China: hospitals versus primary care institutions. *Int J Equity Health* 16(1):42. <https://doi.org/10.1186/s12939-017-0543-9>
- Zhang X, Zhao L, Cui Z et al. (2015) Study on equity and efficiency of health resources and services based on key indicators in China. *PLoS One* 10(12):e0144809. <https://doi.org/10.1371/journal.pone.0144809>
- Zheng W, Liang Y, Lee WS et al. (2023) The mediation effect of perceived attitudes toward medical service on the association between public satisfaction with the

overall medical service and self-rated health among the general population in China: a cross-sectional study. *Int J Environ Res Public Health* 20(4):3369. <https://doi.org/10.3390/ijerph20043369>

Zhou J, Zhang J, Chi M et al. (2023) Does migrant workers' subjective social status affect their mental health? Job satisfaction as a longitudinal mediator. *Int J Soc Psychiatry* 69(6):1420–1431. <https://doi.org/10.1177/00207640231164014>

## Acknowledgements

Data used were from China Family Panel Studies (CFPS) in this research, originally collected by the Institute of Social Science Survey (ISSS) of Peking University, China. This research was funded by the Fundamental Research Funds for the Central Universities, Zhongnan University of Economics and Law (Program No. 202410211), and the Shandong Provincial Science and Technology Department Youth Face Special Project (ZR2024QG092).

## Author contributions

Li He: Conceptualization, Formal analysis, Methodology, Writing – Original Draft, Writing – Review & Editing. Jiangyin Wang: Conceptualization, Formal analysis, Methodology, Writing – Original Draft, Writing – Review & Editing. Yang Yang: Conceptualization, Formal analysis, Writing – Original Draft, Writing – Review & Editing. Yue Yu: Conceptualization, Writing – Review & Editing. Zihan Feng: Writing – Review & Editing. Zhixiong Yang: Writing – Review & Editing, Resources, Funding acquisition.

## Competing interests

The authors declare no competing interests.

## Ethical approval

The data employed in this study were obtained from the China Family Panel Studies (CFPS). The CFPS dataset is a secondary dataset, so ethical approval was obtained by the CFPS project team, not by the authors of this paper. The CFPS project and its data collection procedures were granted ethical approval by the Institutional Review Board of Peking University (IRB PU). The ethical approval number is IRB00001052-14010.

## Informed consent

The informed consent of the CPFS project can be found here: <http://www.issp.pku.edu.cn/cfps/docs/20200615141215123435.pdf>.

## Additional information

**Correspondence** and requests for materials should be addressed to Zhixiong Yang.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025