



OPEN ACCESS

EDITED BY

Mohsen Khosravi,
Zahedan University of Medical Sciences, Iran

REVIEWED BY

Pasquale Scognamiglio,
ASL Napoli 3 Sud, Italy
Massimo Tusconi,
University of Cagliari, Italy
Isa Multazam Noor,
YARSI University, Indonesia

*CORRESPONDENCE

Fang Yang
✉ 20191087@zcmu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 04 February 2025

ACCEPTED 14 April 2025

PUBLISHED 08 May 2025

CITATION

Jiang MH, Lei SY and Yang F (2025) Analysis of influencing factors and equity in health education for mobile populations based on random forest model: evidence from the China migrants dynamic survey. *Front. Public Health* 13:1570633. doi: 10.3389/fpubh.2025.1570633

COPYRIGHT

© 2025 Jiang, Lei and Yang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Analysis of influencing factors and equity in health education for mobile populations based on random forest model: evidence from the China migrants dynamic survey

Meng Han Jiang^{1†}, Shan Yan Lei² and Fang Yang^{1*†}

¹School of Humanities and Management, Zhejiang Chinese Medical University, Hangzhou, China,

²Department of Rehabilitation Medicine, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

Background: Strengthening health education is an important measure to improve the health of mobile populations and a key objective of China's basic public health services. Existing studies demonstrate that health education affects the health of mobile populations, but insufficient attention is paid to the importance of factors that influence health education. Moreover, few studies examine how these factors contribute to health education equity among mobile populations in China. Therefore, this study aims to reveal the importance of factors affecting health education based on a comprehensive understanding of mobile populations' overall health education status. Furthermore, the contribution of these important factors to health education equity is analyzed to inform differentiated intervention strategies, thereby providing a reference for enhancing mobile populations' health level and achieving equal access to basic public health services.

Methods: This study utilized data from the 2018 China Migrants Dynamic Survey (CMDS), with a final sample of 103,910 participants after data cleaning. Chi-square tests were first conducted to examine differences in health education across various characteristics of the mobile population. The relative importance of influencing factors was then assessed using a random forest model, followed by key factor identification through LASSO regression. Subsequently, binary logistic regression was performed to quantify the effects of these key factors. Finally, concentration indices were calculated to identify these factors' contributions to health education equity.

Results: The self-assessed health status of China's mobile population was good, with 81.89% reporting receipt of health education, while 18.11% had not received any health education. Seven key factors were identified as most influential in determining health education access among the mobile population: income, education, age, health record, scope of mobility, reason for mobility and gender. The health education concentration index of the mobile population was 0.0121, indicating a significant inequality in the utilization of health education services. Each important factor had a significant negative amplifying effect on health education equity among the mobile population.

Conclusion: Health education among the mobile population requires further enhancement. Special attention should be directed toward vulnerable groups,

including low-income individuals, the older adult, those with lower educational levels, and those with wider migration ranges. Moreover, the impact of key factors on health education equity among the mobile population should be carefully considered to improve health education equity.

KEYWORDS

mobile population, health education, influencing factors, equity, random forest model

Introduction

Mobile population is a concept that emerged from the household registration system in China in the past 40 years (1). It refers to people aged 15 and above who have lived in the inflow area for more than 1 month and are not household members of their own district (city or county). According to the seventh national census data released by the Chinese government in 2021, the number of the mobile population reached nearly 380 million in 2020, representing an increase of 150 million from 2010 (2). The mobile population has the characteristic of “mobility,” which may cause social integration, policy neglect and other problems, and bring certain difficulties to social governance. Compared with other groups of people, the mobile population is a group that cannot be ignored, and the research results of scholars in different countries on the mobile population show that most of the mobile population have a low education level, mostly engaged in labor-intensive work, poor working environment and living conditions, which makes them face higher health risks (3, 4). At the same time, their physical and mental health problems are often neglected, resulting in inadequate social support and protection. Their access to public health services is lower than that of local residents (5, 6). Therefore, the health protection of the mobile population should receive greater attention.

Health education is an educational activity that promotes the acquisition of health knowledge, fosters health awareness, and encourages voluntary adoption of healthy behaviors and lifestyles among individuals or groups through information dissemination and behavioral interventions (7). Health education is not only one of the main contents of China's basic public health service program (8), but also a strategic health care measure recognized worldwide (9). Effective health education is conducive to improving the health knowledge of individuals or groups, helping them to establish good health beliefs, and improving poor health-related behaviors (10). To enhance the health status of the mobile population, China has consistently prioritized their health education, positioning it as a crucial initiative to achieve equal access to basic public health services and promote health equity. The State Council of the Central Committee of the Communist Party of China (CPC) issued the “Healthy China 2033” Planning Outline in 2016, which emphasizes strengthening health education and actively promoting equal access to basic public health services for mobile populations. This aims to further reduce the disparities in health levels among populations and promote social equity (11).

With regard to health education for the mobile population, relevant studies can be summarized in the following three areas. First, in terms of the current situation, a study shows that the acceptance rate of health education for China's mobile population is 73.1%. The primary content focuses on smoking control, maternal and child health care and eugenics, reproductive health and contraception, while mental health, tuberculosis prevention and treatment and

prevention of occupational diseases are relatively lacking in the type of health education. The main delivery methods include publicity materials (paper, television programs, film and television), bulletin boards, public health counseling (12, 13). Other studies reveal that the level of health education received by urban and rural mobile populations in China is uneven, with urban mobile populations receiving higher levels of health education (14). Second, in terms of influencing factors, studies have pointed out that gender, age, education level, marital status, monthly household income, annual per capita GDP, occupation type, inflow time, inflow area, whether they suffer from chronic diseases, and whether they have had a health record can influence the level of health education of the mobile population (15–17). Third, in terms of fairness, current studies have paid more attention to the health status of the mobile population and the fairness of the utilization of public health services such as family doctor contracting and health records, but have paid less attention to the equity of health education (18, 19).

Current literature demonstrates extensive scholarly attention to health education among China's mobile population, but there is still some room for enhancement. Firstly, current research generally employs descriptive statistics and logistic regression methods to analyze influencing factors, with little comparison of the relative importance of these factors. Secondly, the current research primarily focuses on the fairness of family doctor contracts and health records among China's floating population, with limited studies addressing the fairness of health education service utilization.

To address the above deficiencies, this study takes health education for the mobile population as the research object. First, the current situation of health education for the mobile population is analyzed. Second, the Random Forest model is employed to rank the relative importance of factors and identify key variables. Third, regression analysis is conducted to explore the influencing factors of health education for the mobile population. Finally, the Concentration Index is applied to identify the key factors influencing the equity of health education. The findings are expected to inform the development of targeted health education interventions and promote equal access to basic public health services for mobile populations.

Compared with existing studies, this study offers several contributions: First, the study uses large-scale micro-survey data specifically targeting the migrant population in China, and the current situation of health education for the mobile population is comprehensively analyzed from multiple dimensions and levels. This helps to understand the health of the mobile population from the overall level. Second, current research on health education for the migrant population typically explores the independent effects of different factors or the interactive influences among them, but lacks a comparison of the relative importance of these variables. This study uses a random forest model to measure the importance of various influencing factors and identifies the most critical ones, which helps to provide a more accurate evaluation of these factors. Thirdly, this

study focuses on the equity of health education in the utilization of public health services by the migrant population. This not only enriches existing analyses of equity among migrant populations but also provides data support and theoretical foundations for developing more targeted health education policies.

Research methods

Theoretical framework

The theoretical framework can further clarify the determinants of health education and the inequalities among the migrant population. Current theoretical frameworks related to healthcare utilization in academic literature include the Social Determinants of Health Action Framework (20), the Health Equity Model (21), and Anderson's Behavioral Model of Health Service Utilization (22), among others. Among these, Anderson's Behavioral Model of Health Service Utilization is a significant theoretical model for studying healthcare utilization behavior and has been widely applied in the healthcare research field (23, 24). Additionally, this model has been used in relevant studies to analyze the utilization of public health services by the migrant population (25). Therefore, this study adopts Anderson's Behavioral Model of Health Service Utilization as its theoretical framework. Based on the characteristics of the migrant population and relevant literature (26, 27), the study constructs an analytical framework for the factors influencing health education service utilization among China's migrant population, with a specific classification framework shown in Figure 1.

Data source

The data for this study were obtained from the 2018 China Migrants Dynamic Survey (CMDS), which is an annual large-scale

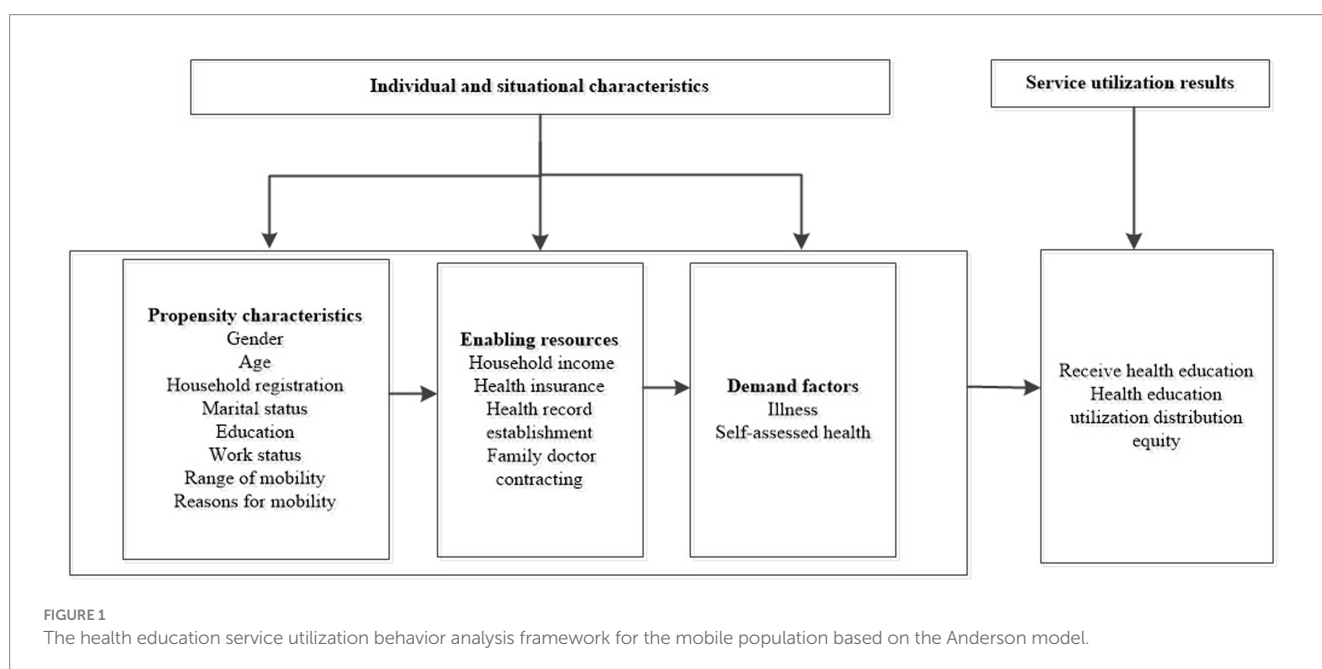
national mobile population sample survey conducted by the National Health Commission of China since 2009. The CMDS is considered to be a good representative sample, and has a small sampling error (28). The survey adopted a hierarchical, multi-stage, and proportionate-to-size PPS method for sampling, i.e., in the first stage, the townships (towns and streets) were selected according to the PPS method. In the second stage, village (neighborhood) committees were sampled within the selected townships (towns and streets) according to the PPS method. In the third stage, individual respondents were drawn within the selected village (neighborhood) committees. The survey also adopted strict methods to ensure data quality, including scientific design of questionnaires, training of surveyors, setting up survey supervisors to verify questionnaires, and quality checking by means of telephone callbacks.

The survey selects as its sample mobile population in 32 provincial administrative regions of China who are at least 15 years of age, have lived in their current location for more than 1 month, and are not household members of their own district (county or city). The survey covers basic demographic information on the mobile population and household members, the scope and tendency of mobility, employment and social security, income and expenditure, residence, basic public health services, marriage and family planning services, children's mobility and education, and psycho-culture, which makes the survey a rich set of variables. A total of 152,000 samples of the mobile population were collected in this survey, and the final sample comprised 103,910 after processing the missing values and outliers of the data.

Variables selection

Dependent variable

The core explanatory variable in this paper is health education. A 0–1 dummy variable was constructed based on the questionnaire: "Have you received any of the following types of health education in



your current community/workplace during the past year?” The answers to this question include occupational disease prevention and control, infectious disease prevention and control, reproductive health and maternal and child health, chronic disease prevention and control, mental health, self-help in public emergencies, and others. Referring to related studies (13, 29), this study assigns a value of 1 to the mobile population who received one or more types of education, and 0 to those who did not.

Independent variables

Based on the analytical framework for the utilization of health education services for mobile populations established in the theoretical foundation, the independent variables in this study were classified into three dimensions: (1) propensity characteristics: including gender, age, education, marital status, work status, scope of mobility, reasons for mobility and household registration; (2) enabling resources: including household income (in this study, data about household income were transformed into rankings within each province: < 20th percentile, 20th–39th percentile, 40th–59th percentile, 60th–79th percentile and $\geq$ 80th percentile for data analysis), health insurance, health record establishment and family doctor contracting; (3) demand factors: including self-assessed health and illness, with variable assignments detailed in Table 1.

Data processing

Random forest model

The random forest model not only allows for the comparison of the importance of various variables, which identifies relatively important ones, but also offers high accuracy, efficiency in testing, and stability, making the results more reliable (30). Compared to more complex models such as XGBoost, the advantages of the random

forest model are as follows: (1) It provides a more intuitive analysis of feature importance and requires less adjustment. (2) It reduces the risk of overfitting, as it is more stable due to built-in regularization techniques such as bootstrapping and bagging, making it less susceptible to parameter selection. (3) While some machine learning algorithms (such as decision trees) are effective in understanding feature importance, they typically do not capture complex interactions between features as well as random forests do (31, 32). Moreover, the random forest algorithm strikes a good balance between model complexity and performance. Given these advantages, it has been widely applied in fields such as medicine, management, and public health (33, 34). Recent studies have also used this model to analyze issues related to the utilization of public health services among China's migrant population (35). Therefore, building on similar studies, this research also uses the random forest model to measure the importance of various factors influencing health education for China's migrant population.

Random forest algorithm consists of multiple decision trees, with the help of bagging algorithm to generate multiple training sets through random sampling of the sample. Decision trees serve as the base classifiers for each training set, with the final prediction determined by majority voting across trees. This approach is suitable for classification, regression, and prediction tasks. Importance analysis, on the other hand, takes the best variable selected in the decision tree as the classification node, so as to rank the importance of variables (36). The specific steps are as follows: (1) For each decision tree, the corresponding out of bag data (OOB) are selected to calculate the out of bag data error, which is recorded as err_{OOB1} ; (2) Randomly add noise interference to the feature X of all the samples of OOB (which can randomly change the value of the samples at the feature X), and then calculate the out-of-bag data error again, which is recorded as err_{OOB2} ; (3) Assuming that there are N trees in the forest, the importance of the feature X can be calculated by the following formula.

TABLE 1 Variable selection and definitions.

Variables		Definition
Implicit variable	Health education	No = 0, Yes = 1
Independent variable	Gender	Female = 0, Male = 1
	Age	15–30 = 1, 31–45 = 2, 46–60 = 3, $\geq$ 61 = 4
	Educational attainment	Illiterate = 1, Elementary school = 2, Middle school = 3, High school = 4, University and above = 5
	Marital status	Not in marriage = 0, In marriage = 1
	Working conditions	No occupation = 0, Employed = 1
	Household registration	Rural = 1, Urban = 2
	Range of mobility	Cross-county within city = 1, Cross-city within province = 2, Cross-province = 3
	Reasons for mobility	Family = 1, Work = 2, Other = 3
	Household income ranking	Lowest (<20%) = 1, Lower (20–39%) = 2, Medium (40–59%) = 3, Higher (60–79%) = 4, Highest ($\geq$ 80%) = 5
	Insurance location	Place of domicile = 1, Place of inflow = 2
	Whether one is sick or not	No = 0, Yes = 1
	Health level	Unhealthy = 1, Basically healthy = 2, Healthy = 3
	Availability of health records	No = 0, Yes = 1
	Whether the family doctor signed up	No = 0, Yes = 1

$$\text{OOB_store} = \frac{\sum_{i=1}^N (\text{erroob2i} - \text{erroob1i})}{N}$$

This value can indicate feature importance because if the out-of-bag data accuracy decreases substantially (i.e., err OOB2 rises) after the addition of random noise, it indicates that this feature has a significant impact on the prediction results of the sample, which in turn indicates that the importance level is relatively high.

Binary logistic regression model

In this study, health education is a binary variable. Therefore, a binary logit model is used to analyze the impact of other factors on health education. The basic model is as follows:

$$\text{logit} \frac{P_i}{1-P_i} = \beta_0 + \beta_1 X_i + \varepsilon_i$$

In which, P_i denotes the probability that the mobile population will receive health education, β_0 is the intercept term, X_i denotes other variables that will have an impact on health education, β_1 denotes the coefficient of influence of the relevant independent variable on health education, ε_i is the error term.

Analysis of the concentration index

In research related to public health and health, methods for measuring inequality typically include the Concentration Index (CI), Gini coefficient, Oaxaca-Blinder decomposition, Lorenz curve, and range method, among others (37, 38). The Concentration Index, compared to other methods, has the following advantages (39–41): (1) The Concentration Index quantifies the tendency of health service utilization or health inequality, while the Gini coefficient only reflects the overall level of disparity. (2) The Concentration Index directly establishes a relationship between the relevant dependent variable (such as health education) and socioeconomic status, revealing structural inequalities, whereas the range method focuses solely on the distribution of the dependent variable itself. (3) The Concentration Index allows for decomposition techniques (such as Wagstaff decomposition) to attribute inequality to specific influencing factors. (4) Results obtained using the Concentration Index are not affected by the absolute level of the research indicators (e.g., through standardized Concentration Index), allowing for comparisons across populations, time periods, or countries. Currently, the Concentration Index is widely used not only to measure income-related inequality in the health sector but also to assess the equity of public health service utilization, medical service utilization, and health among China's migrant population (42, 43). Based on the above analysis, this study also employs the Concentration Index to analyze the equity of health education for China's migrant population and measure the contribution of related influencing factors to the equity of health education. Its calculation formula is as follows:

$$\text{CI} = 2 \text{cov}(y_i, R_i) / \mu$$

In which y_i represents the outcome variable of public health service utilization, i denotes the average level of the variable in the population, and R_i denotes the fractional rank of the sample i in the income

distribution. The value of CI ranges from $(-1, 1)$, $\text{CI} > 0$ indicates that there is pro-rich inequality in the outcome variable and $\text{CI} < 0$ indicates that there is pro-poor inequality in the outcome variable. The larger the absolute value of CI, the more sensitive the distribution of the outcome variable is to the level of income, indicating a greater degree of inequality.

In this study, the concentration index decomposition method proposed by Wagstaff (44) was used to decompose the factors that may affect the equity of public health service utilization. By examining each factor's contribution to inequality after decomposition, the main sources of inequity was identified, which helps target specific factors for control or elimination.

The decomposition formula is as follows:

$$C = \sum_j (\beta_j \bar{X}_j / \mu) C_j + GC_\varepsilon / \mu$$

C is the unstandardized concentration index, $\beta_j, \bar{X}_j, C_j$ are the regression coefficient (replaced by the marginal effect), mean and concentration index of the influence factor j , respectively, $\beta_j \bar{X}_j / \mu$ denotes the magnitude of the contribution of the influence factor j to the inequality in health, GC_ε is the concentration index of the residual term, and μ is the mean value of the outcome of the utilization of the health services (i.e., the dependent variable).

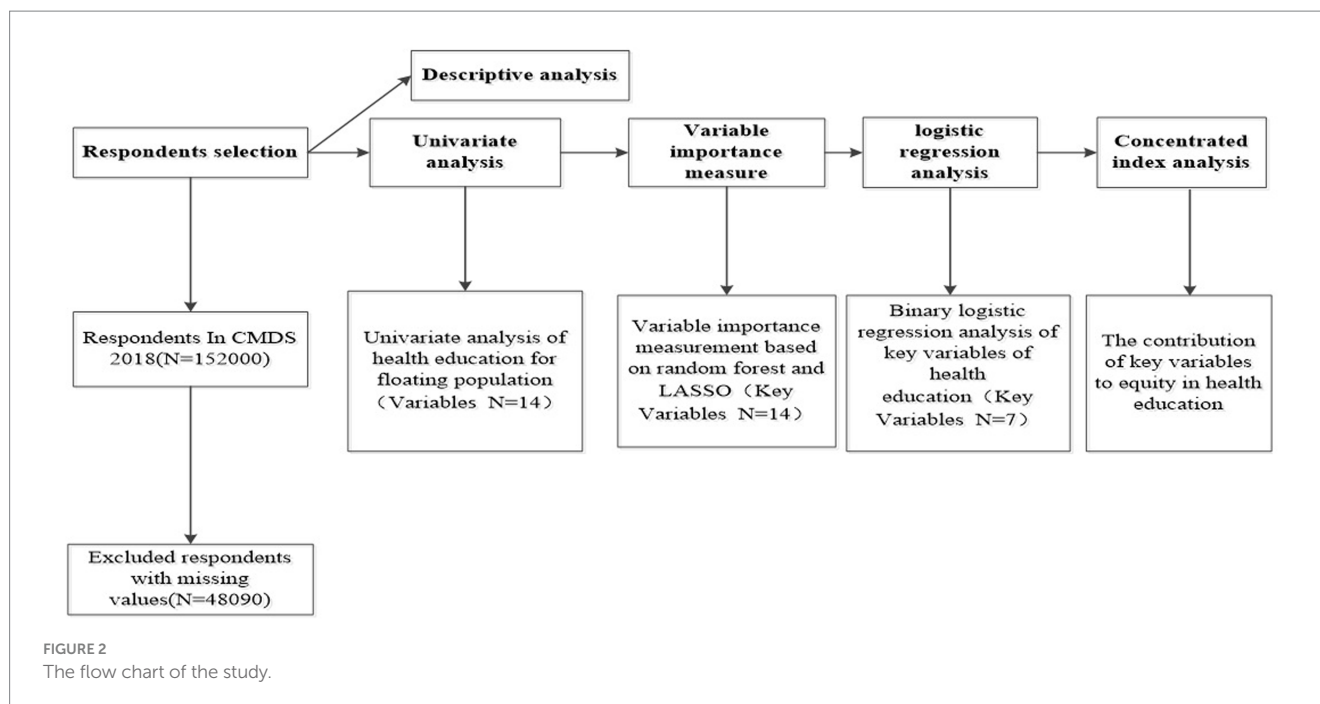
Statistical methods

Data were organized and analyzed using Stata22.0 and RStudio software. Count data were expressed as the number of cases and percentage (%), and χ^2 test was used for one-way analysis. Random forest model analysis was performed using RStudio software, and variables with statistically significant differences in the unifactorial analysis were included in the random forest model to derive and rank the variable importance scores. LASSO analysis was used for variable selection, and the screened variables were used for multifactorial analysis using binary logistic regression analysis. Finally, the decomposition of concentration index was used to identify the contribution of important variables to equity in health education. The flow chart is shown in Figure 2.

Results

Basic characteristics of the respondents

Table 2 shows the descriptive statistics of the main variables. Among 103,910 respondents, 86.08% report being healthy and 11.63% report having a basic health status. Regarding health education, the percentage of mobile population receiving health education is higher at 81.89%. In terms of health records, only 32.42% of the mobile population establish health records locally, and 67.58% still do not have health records. For family doctor services, only 15.09% sign up with local family doctors, while the percentage of those who do not sign up is 84.91%. Educational attainment among the mobile population is generally low, with only 19.91% holding university degrees or above, and 80.09% having high school education or below. Regarding household registration, 67.60% are registered as rural residents and 32.40% as urban residents. In terms of mobility, the proportion of inter-provincial mobility is the highest, followed by intra-provincial inter-city and intra-city inter-county



mobility. In terms of income, the majority of the mobile population have below-medium income levels (62.85%). Regarding health insurance, 69.86% are covered by health insurance in their place of domicile, while 30.14% are covered by health insurance in their place of inflow.

Univariate analysis of health education for the mobile population

The results of the univariate analysis with whether the mobile population receives health education as the dependent variable are shown in [Table 3](#). Statistically significant differences in health education are observed across various demographic characteristics including genders, ages, education levels, marriages, jobs, household registration status, mobility ranges, reasons for mobility, incomes, places of insurance coverage, illnesses, availability of health records, whether the family doctor signed up and health levels.

Importance measurement of factors influencing health education for mobile populations

Random forest data selection

This study randomly divided the samples into a training set (80%) and a validation set (20%). In the model training, we employed a Random Forest classifier for the classification task. To optimize model performance, we utilized Randomized Search for hyperparameter tuning. Randomized Search CV performs 5-fold cross-validation by randomly selecting parameter combinations multiple times, and the parameter set that yields the highest accuracy score is selected as the final result. The range of parameter values and their definitions are shown in [Table 4](#).

The process of Randomized Search is as follows: Due to the large number of parameters, we present the search process with the maximum depth on the x-axis and the minimum samples per leaf on the y-axis. The color of the scatter points represents the accuracy, with darker colors indicating higher accuracy. It can be observed that the point with a maximum depth of 18 and a minimum samples per leaf of 7 achieves the highest accuracy. This combination of parameters represents the optimal hyperparameters we were seeking (see [Figure 3](#)).

Ranking of importance of influencing factors

Using health education as the dependent variable, variables with statistically significant differences in the one-way analysis were included in the random forest model. Using the Random Forest program package in RStudio to output the results, %Inc. MSE is the mean decrease in accuracy, and the larger the %Inc. MSE, the higher the importance of the variable in the influencing factors ([45](#)). [Figure 4](#) shows the ranking of the importance of the factors influencing health education among the mobile population. In descending order of importance, these factors are: income, education, age, health record, range of mobility, reason for mobility, gender, health level, household registration status, contracted by the family doctor, marriage, place of insurance participation, illness, and work.

Characteristic variable screening

LASSO regression method achieves the feature selection and dimensionality reduction by setting unimportant feature coefficients to zero through L1 regularization. This method can screen the variables in order to avoid the function of covariance and overfitting. Compared with the traditional stepwise regression, LASSO can process all independent variables simultaneously, enhancing model stability. In this study, due to the inclusion of numerous independent variables which are prone to multicollinearity problems ([46](#), [47](#)), the

TABLE 2 Distribution of respondents' characteristics.

Variable		Quantity	Proportions (%)
Health status	Unhealthy	2,384	2.29
	Basic health	12,080	11.63
	Health	89,446	86.08
Health education	No	18,821	18.11
	Yes	85,089	81.89
Availability of health records	No	70,226	67.58
	Yes	33,684	32.42
Whether the family doctor signed up	No	88,235	84.91
	Yes	15,675	15.09
Genders	Female	50,899	48.98
	Male	53,011	51.02
Age	15–30	29,641	28.53
	31–45	48,358	46.54
	46–60	21,396	20.59
	≥61	4,515	4.35
Educational attainment	Illiterate	2,542	2.45
	Elementary school	14,179	13.65
	Middle school	43,614	41.97
	High school	22,890	22.03
	University and above	20,685	19.91
Marital status	Not in marriage	16,618	15.99
	In marriage	87,292	84.01
Working conditions	No occupation	17,723	17.06
	Employed	86,187	82.94
Household registration	Rural	70,248	67.60
	Urban	33,662	32.40
Range of mobility	Cross-county within city	18,533	17.84
	Cross-city within province	35,347	34.02
	Cross-province	50,030	48.15
Reasons for mobility	Family	14,985	14.41
	Work	87,036	83.76
	Other	1,889	1.82
Household income ranking	Lowest (<20%)	22,866	22.02
	Lower (20–39%)	21,207	20.41
	Medium (40–59%)	21,216	20.42
	Higher (60–79%)	20,038	19.28
	Highest (≥80%)	18,563	17.89
Insurance location	Place of domicile	72,588	69.86
	Place of inflow	31,322	30.14
Whether one is sick or not	No	91,544	88.10
	Yes	12,366	11.98

LASSO regression model is therefore used for variable selection, and the important variables are further included in the logistic regression model for analysis. The LASSO regression method in this study is implemented by using glmnet package in R, with 100 iterations and

the optimal λ value determined through 10-fold cross-validation. By calculation, the λ corresponding to the smallest mean square error is 0.000677 (see Figure 5). Figure 6 shows the LASSO variable screening diagram, based on the results of this diagram, seven variables were

TABLE 3 Differences in health education for mobile populations across baseline characteristics.

Variable		Health education				χ^2	p
		No		Yes			
		n	%	n	%		
Genders	Female	9,001	47.82	41,989	49.24	12.365	<0.001
	Male	9,820	52.18	43,191	50.76		
Age	15–30	4,561	24.23	25,080	29.48	747.923	<0.001
	31–45	8,172	43.42	40,186	47.23		
	46–60	4,869	25.87	16,257	19.42		
	≥61	1,219	6.48	3,296	3.87		
Educational attainment	Illiterate	757	4.02	1785	2.10	1.1e+03	<0.001
	Elementary school	3,528	18.75	10,651	12.52		
	Middle school	8,288	44.04	35,326	41.52		
	High school	3,343	17.76	19,547	22.97		
	University and above	2,905	15.43	17,780	20.90		
Marital status	Not in marriage	2,671	14.19	13,947	16.39	55.496	<0.001
	In marriage	16,150	85.81	71,142	83.61		
Working conditions	No occupation	3,956	21.02	13,767	16.18	255.154	<0.001
	Employed	14,865	78.98	71,322	83.82		
Household registration	Rural	13,491	71.68	56,757	66.70	174.348	<0.001
	Urban	5,330	28.32	28,332	33.30		
Range of mobility	Cross-county within city	2,823	15.00	15,710	18.46	671.508	<0.001
	Cross-city within province	5,329	28.31	30,018	35.28		
	Cross-province	10,669	56.69	39,361	46.26		
Reasons for mobility	Family	3,021	16.05	11,964	14.06	62.338	<0.001
	Work	15,405	81.85	71,631	84.18		
	Other	395	2.10	1,494	1.76		
Household income ranking	Lowest (<20%)	4,942	26.26	17,944	21.09	290.686	<0.001
	Lower (20–39%)	3,878	20.60	17,329	20.37		
	Medium (40–59%)	3,721	19.77	17,495	20.56		
	Higher (60–79%)	3,140	16.68	16,898	19.86		
	Highest (≥80%)	3,140	16.68	15,423	18.13		
Insurance location	Place of domicile	14,412	76.57	58,176	68.37	492.533	<0.001
	Place of inflow	4,409	23.43	26,913	31.63		
Whether one is sick or not health level	No	15,841	84.17	75,703	88.97	399.049	<0.001
	Yes	2,980	15.83	9,386	11.03		
Whether the family doctor signed up	No	18,164	96.51	70,071	82.35	2.4e+03	<0.001
	Yes	657	3.49	15,018	17.65		
Availability of health records	No	16,428	87.57	53,744	63.16	4.2e+03	0.000
	Yes	2,339	12.43	31,345	36.84		
Health level	Unhealthy	746	3.96	1,638	1.93	622.733	<0.001
	Basically healthy	2,880	15.30	9,200	10.81		
	Healthy	15,195	80.73	74,251	87.36		

TABLE 4 Parameter ranges and descriptions for the random forest model.

Parameters	Definition	Range of values	Optimal value
n_estimators(ntree)	Specifies the number of decision trees to be built in the random forest	1–500	187
max_depth	Specifies the maximum depth of each decision tree to control tree growth	1–20	17
min_samples_split(mtry)	Specifies the minimum number of samples needed to split an internal node, which influences the model complexity	2–20	18
min_samples_leaf	The minimum number of samples required to be at a leaf node, controlling the minimum size of leaf nodes	1–10	8

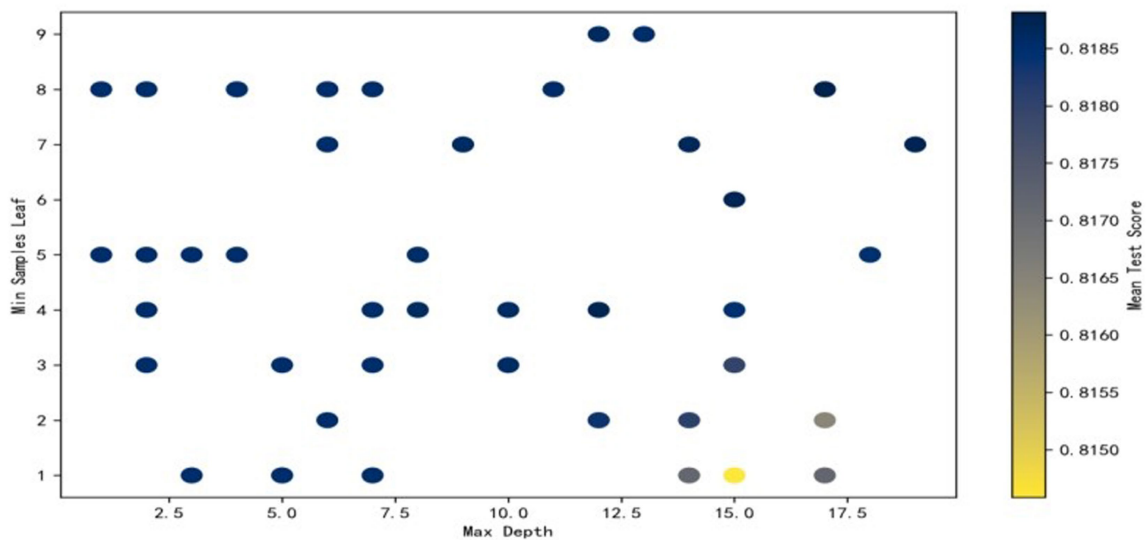


FIGURE 3
Randomized search process figure.

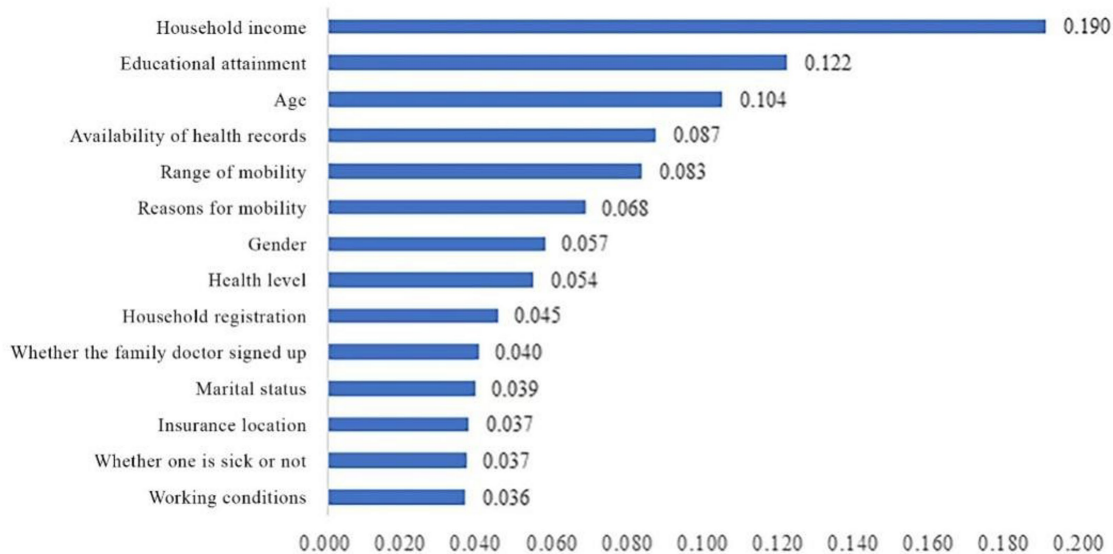


FIGURE 4
Ranking of importance eigenvalues of factors influencing health education for the mobile population.

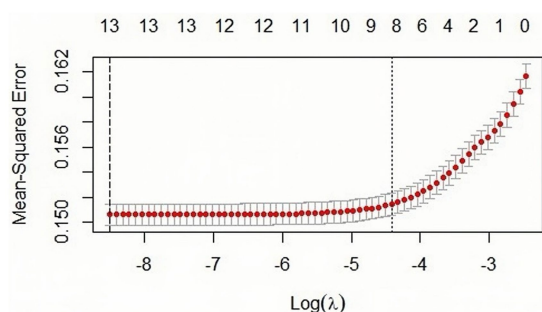


FIGURE 5
Characteristic variable screening based on LASSO analysis. The figure is based on the minimum criterion to filter the number of feature variables by plotting dashed lines at $\lambda_{\min}$ (left dashed line) and λ_{1se} (right dashed line), respectively.

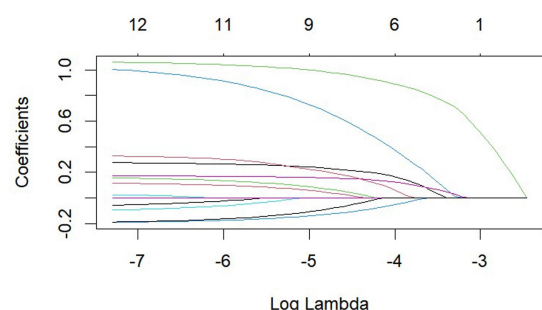


FIGURE 6
A total of 14 variables were included in this study and the figure shows 14 lines of different shades of color. Each curve represents the trajectory of the coefficient of each independent variable, with the vertical coordinate being the value of the coefficient and the lower horizontal coordinate being $\log(\lambda)$.

selected, namely income, education, age, health record, mobility range, reason for mobility and gender.

Regression analysis of factors affecting health education of mobile population

A binary logistic regression analysis was conducted using health education as the dependent variable and the top 7 variables in terms of importance screened by the random forest model as the independent variables. The results are shown in Table 5. The findings indicate that higher income levels are associated with increased access to health education compared to the lowest income group. As age increases, the propensity of mobile population to access health education decreases. In terms of educational level, an increase in educational level promotes the mobile population's access to health education compared to those who did not receive any health education. The mobile population with health records is more likely to receive health education compared to those who do not have health records ($OR = 1.413$, $p < 0.001$). Compared with intra-city cross-county mobility, intra-provincial cross-city and inter-provincial mobility was more likely to receive health education, indicating that the propensity of the mobile population to obtain health education

increases as the scope of mobility expands. Compared with those who moved for family reasons, those who moved for other reasons (e.g., study, cross-country retirement, marriage, etc.) were more likely to receive health education ($OR = 1.327$, $p < 0.001$). The male mobile population was more likely to receive health education compared to the female mobile population ($OR = 1.087$, $p < 0.001$).

The influence of key factors on the equity of health education for mobile populations

Equity analysis of health education

The concentration index of health education for the mobile population is $CI = 0.0121$. The concentration index is greater than 0, and the concentration curve is below the absolute fairness line, which indicates that the distribution of health education service utilization is inclined to the population with higher income level, (see Figure 7).

Analysis of the degree of contribution of key factors to health education

The important variables of the seven health education influencing factors selected by LASSO were used as independent variables, and the health education was used as the dependent variable. The contribution of key factors to health education for the migrant population is further explored. The results are shown in Table 6. From the results of the decomposition of the concentration index, the effects of income, age, education level of junior high school and below, health records, range of mobility, reason for mobility and gender on the health education are all positive, indicating that the above factors increase the inequality of health education among the mobile population. In contrast, education levels of high school and above show negative concentration index for health education, suggesting that higher education levels help reduce health education inequality.

Discussion

The purpose of this study was to analyze the key influencing factors of health education for mobile populations and their contribution to the equity of health education. This evidence from China not only enriches the theoretical understanding of the utilization of public health services among the floating population, but also provides practical support for improving health education policies for this group. Ultimately, it contributes to the dual objectives of achieving equality in the utilization of public health services and improving health outcomes for the floating population. The findings reveal that the proportion of mobile populations with a healthy status was 97.71%, indicating high overall health levels. However, since health forms the foundation for better utilization of human capital advantages (48), continued improvement in mobile populations health remains necessary. Health education is an important means of maintaining and improving health (49), and the results of existing studies also indicate that health education in the form of personal behavior change, policy and regulations, economic support, and organizational action can promote individuals' acquisition of health knowledge and skills, which in turn can lead to changes in health behaviors. The results of existing studies have also shown that health education in the form of personal

TABLE 5 Binary logistic regression analysis of influencing the acceptance of health education by the migrant population.

Variables		β	p	OR	95% CI
Household income ranking	Lowest (<20%) (reference group)				
	Lower (20–39%)	0.140	<0.001	1.150	1.096–1.208
	Medium (40–59%)	0.151	<0.001	1.163	1.107–1.221
	Higher (60–79%)	0.229	<0.001	1.257	1.192–1.324
	Highest (≥80%)	0.088	0.002	1.092	1.034–1.153
Age	15–30 (reference group)				
	31–45	−0.038	0.069	0.962	0.923–1.003
	46–60	−0.280	<0.001	0.756	0.719–0.794
	61–	−0.557	<0.001	0.573	0.529–0.629
Educational attainment	Illiterate (reference group)				
	Elementary school	0.143	0.004	1.155	1.048–1.272
	Middle school	0.369	<0.001	1.447	1.317–1.588
	High school	0.629	<0.001	1.876	1.699–2.072
	University and above	0.628	<0.001	1.873	1.690–2.076
Availability of health records	NO (reference group)				
	Yes	1.413	<0.001	4.108	3.924–4.301
Range of mobility	Cross-county within city (reference group)				
	Cross-city within province	0.428	<0.001	1.535	1.466–1.606
	Cross-province	0.432	<0.001	1.541	1.486–1.597
Reasons for mobility	Family (reference group)				
	Work	0.051	0.403	1.052	0.934–1.184
	Other	0.283	<0.001	1.327	1.185–1.485
Genders	Female (reference group)				
	Male	0.083	<0.001	1.087	1.052–1.123

behavior change, policies and regulations, economic support, and organizational action can promote individuals' acquisition of health knowledge and skills, which in turn can change health behaviors. While this study found that 81.89% of the mobile population received health education, 18.81% still lacked access. This falls significantly short of China's target of 95% health education coverage for mobile populations by 2020, suggesting generally low health education coverage and levels among mobile population. These findings indicate a need for strengthened health education.

Regarding the influencing factors of health education for the mobile population, the results of this study show that health education is influenced by a variety of factors, which is similar to the results of previous studies (12, 50). However, this study advances the understanding by employing a random forest model to quantify the importance of various influencing factors. The analysis reveals varying degrees of importance among different characteristic factors affecting health education in mobile populations. The main factors that have significant influence on health education are income, age, education level, health record, scope of mobility, reason for mobility and gender. Certainly, the identification of these key factors is based on the evaluation of the 2018 CMDS cross-sectional survey data. In the future, with the availability of long-term longitudinal data, further analysis could be conducted to explore the trends of various influencing factors on health education for the floating population

over time, thereby effectively identifying the most critical factors influencing health education within a long-term timeframe.

The study reveals that higher income levels facilitate better access to health education among mobile populations. This correlation may be attributed to higher-income mobile populations having both greater financial capacity for disease prevention and treatment, and enhanced health awareness. These groups can acquire more health information through health education, thereby maintaining or improving their health status (51). However, the analysis shows that most of the mobile population's income is below the middle level. Therefore, health education initiatives should prioritize low-income mobile populations, such as further expanding the coverage of medical insurance for low-income people to reduce the cost burden of health education for such people (52), establishing personalized health education files for low-income people, and continuing to follow up on the changes in their health.

The impact of age on health education shows that acceptance rates decline as age increases. This trend may be attributed to older adult individuals having limited access to health education channels, while younger and middle-aged populations can access health education through new media platforms such as the internet and mobile phones (53, 54). Therefore, health education programs for mobile populations should develop age-specific content that addresses the unique health needs of older adult individuals. Examples include increasing

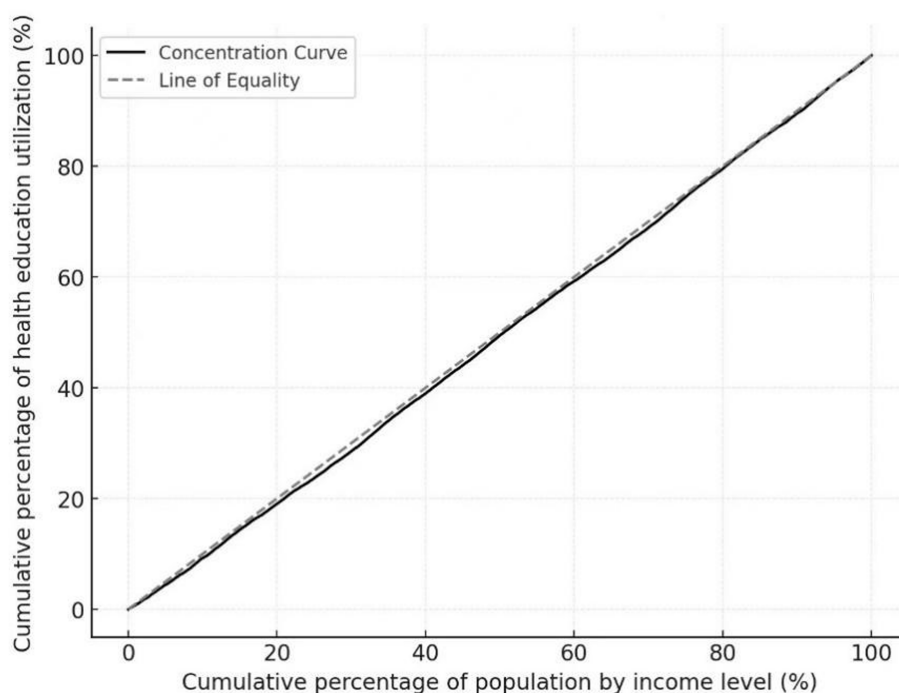


FIGURE 7
Concentration curve of health education service utilization.

awareness of prevention and treatment strategies for common chronic diseases among the older adult, which can enhance their interest in understanding and acquiring health-related knowledge. Furthermore, government agencies and community organizations can offer free health knowledge lectures and educational campaigns, and recruit healthcare professionals as volunteers to provide face-to-face health services to the older adult mobile populations.

Education level also has a significant effect on the acquisition of health education, which is also consistent with the results of existing studies (14). Education level objectively reflects the ability of individuals to obtain health information, learn health knowledge and master health skills (55). Higher education levels not only increase mobile populations' health awareness but also provide access to more diverse channels for receiving health education. However, this study indicates that the education level of the mobile population is generally low, with the proportion of mobile people with high school or less education about four times higher than the proportion with university and above. The low level of education may lead to the inability of the mobile population to accurately and objectively evaluate their own health status, which may cause them to neglect the importance of health education. Therefore, future health education content should be simple, easily understandable, and appealing to target groups. Additionally, health education promotion methods should be optimized by leveraging the convenience and immediacy of mass media platforms, such as WeChat and websites, to stimulate mobile populations' engagement in health education and enable faster, better access to health knowledge.

The establishment of health records also helps the mobile population to obtain health education. This may be because health records contain comprehensive health information including disease status and treatment history, making it easier for individuals to understand their health conditions and obtain individualized, targeted

health education based on their needs. Moreover, mobile populations with established health records may demonstrate stronger health awareness and better compliance with basic public health service policies and organizational initiatives. However, this study found that only 32.42% of mobile populations have established health records, indicating insufficient utilization of basic public health services. Therefore, improving the health education of the mobile population, we should actively promote the establishment of health records, continuously improve the management of health records, strengthen the management and dynamic monitoring of the health data for the mobile population, and provide regular health services to them. This will maximize the role of health records in promoting health education.

In terms of the scope of mobility, this study indicates that as the scope of mobility expands, the mobile populations are more likely to have access to health education. This finding differs from those of Wang et al. on the utilization of public health services for access to the older adult mobile population. The results of this study indicated that older adult mobile populations moving within city limits had higher rates of health education acceptance and health record establishment compared to those moving across provinces (56). The reason for this disparity may be that the smaller the migration span, the stronger the stability, and the greater the reliance on medical service resources in the destination area. The differing results in this study may be related to the differences in the research subjects included.

Regarding reasons for mobility, populations moving for work, education, or retirement are more likely to receive health education compared to those moving for family reasons. This may be because populations migrating for reasons such as work, study, or relocating for older adult care are often in formal and structured environments, which provide more opportunities for health education. In contrast, those migrating for family reasons tend to focus more on family matters and may have less exposure to public health resources and educational

TABLE 6 Concentration index decomposition of the equity in health education service utilization among the mobile population.

Variable		Coefficient of elasticity	Concentration index	Contribution margin	Contribution rate
Household income ranking	Lowest (<20%) (reference group)				
	Lower (20–39%)	0.514	0.205	0.105	8.684
	Medium (40–59%)	0.554	0.199	0.111	9.156
	Higher (60–79%)	0.841	0.177	0.149	12.302
	Highest (≥80%)	0.323	0.192	0.062	5.136
Age	15–30 (reference group)				
	31–45	−0.096	0.004	−0.001	−0.033
	46–60	−0.708	0.016	−0.011	−0.926
	61–	−1.410	0.019	−0.028	−2.318
Educational attainment	Illiterate (reference group)				
	Elementary school	0.622	0.012	0.008	0.617
	Middle school	1.604	0.005	0.008	0.663
	High school	2.735	<0.001	−0.002	−0.023
	University and above	2.730	−0.012	−0.032	−2.663
Availability of health records	NO (reference group)				
	Yes	0.546	0.004	0.002	0.180
Range of mobility	Cross-county within city (reference group)				
	Cross-city within province	1.239	0.015	0.019	1.557
	Cross-province	1.251	0.011	0.013	1.075
Reasons for mobility	Family (reference group)				
	Work	0.119	0.012	0.001	0.122
	Other	0.665	0.029	0.019	1.599
Genders	Female (reference group)				
	Male	0.053	0.013	0.001	0.056

activities. Furthermore, this research result also indicates that beyond individual social characteristics, factors related to mobility patterns also warrant attention. As China's mobile population continues to grow in size and mobility frequency increases, the impact of mobility-related factors on health education will become increasingly significant. Therefore, strengthening health education for the mobile population should include increasing health knowledge training for family members, such as offering family health education courses for parents and caregivers in migrant families to help them promote the health of family members in the home environment. Additionally, the construction of a health information platform for the mobile population should be accelerated, and cross-regional resource and information sharing should be realized as soon as possible in order to improve the continuity of the utilization of public health services for the mobile population.

Gender is also a significant factor influencing health education. This is both similar to and somewhat different from the results of previous studies. For example, the study by Liping et al. concluded that women had better health education (13), while the results of Wang Xiaohui et al. showed that a higher percentage of men received health education compared to women (56). This study concluded that the reason for the difference in health education service utilization between genders is that the health status of women is generally higher than that of men, and health differences affect the health awareness

and public health service utilization of mobile populations of different genders (57).

This study found that the seven key variables selected through LASSO contribute to increasing health education inequality among the mobile population, which is consistent with the findings of related studies (18). Further analysis of the contribution of important factors to the equity of health education reveals significant disparities in health education service utilization among mobile populations, with a bias toward high-income groups. This is also consistent with previous studies on the equity of public health service utilization among the mobile population in China. For example, relevant research shows that the concentration index for inpatient service utilization among China's mobile population is 0.154, indicating a significant pro-rich inequity in the utilization of inpatient services (43). Furthermore, the study by Deng Yuhua et al. suggests that the utilization of family doctor contract services among China's mobile population is more likely to benefit high-income groups (58). Additionally, the research by Zheng Cihui et al. reveals an inequity in the establishment of health records for China's mobile population (59). Therefore, there is still a need to strengthen public health education for low-income mobile populations. Future efforts should focus not only on increasing overall health education acceptance rates among mobile populations, but also on addressing inequities in health education utilization to achieve the goal of equalizing public

health service access. Certainly, this study's assessment of the equity of health education service utilization among the mobile population is based on cross-sectional data, which allows for an analysis of the static situation of health education distribution at a specific point in time. However, with longitudinal data spanning a longer period, it would be possible to analyze the dynamic changes in this distribution over time, thereby providing a comprehensive understanding of the distribution changes in health education from both static and dynamic perspectives.

Conclusion

Based on the 2018 China Migrants Dynamic Survey (CMDS), this study analyzed factors affecting health education and equity among mobile populations. The main conclusions are as follows. Firstly, the acceptance rate of health education for the mobile population still needs further improvement. Secondly, the health education of mobile population is affected by multiple factors with varying degrees of impact. Income, age, education level, health record establishment, mobility range, reasons for mobility, and gender were identified as the seven most significant influencing factors. All these factors have significant negative amplifying effects on health education of the mobile population. Thirdly, significant inequalities exist in health education utilization, showing a gap between current reality and the goal of equalized public health services.

Based on these findings, future efforts to strengthen health education for mobile populations should implement differentiated strategies according to the varying impacts and significance of different factors. Government and health departments should increase resource allocation for health education targeting mobile populations, with particular attention to vulnerable groups such as those with low income, older adult individuals, those with lower education levels, those without established health records, and those with wide mobility ranges. The ultimate goals are to enhance health literacy and improve the overall health conditions of mobile populations through continuously enriching health education content, innovating education methods, and expanding communication channels. However, when implementing health education interventions, it is essential to consider the specific characteristics of the mobile population and address potential ethical issues or other factors from four aspects: the implementing parties, the intervention process, the target audience, and the evaluation of intervention effects. For instance, the implementers of health interventions should possess relevant health-related professional knowledge; participants in the intervention should be informed of the objectives, methods, and potential benefits of the health education intervention; and post-intervention, it is crucial to focus on the follow-up and support of the intervention's effects. These considerations will enhance the mobile population's willingness to participate in health education interventions, promote the smooth implementation of intervention programs, and ensure that the interventions achieve the desired outcomes.

Limitations

This study had several important limitations. First, the research relied on secondary data from a public dataset (with the latest version released in 2018), which presented certain time lag issues. As newer data become available, future research can analyze new characteristics

of health education among mobile populations using more current information. In addition, the data used in this study are collected from a large-scale cross-sectional survey, which explored the relationship between health education and key factors among the mobile population in China. However, we were unable to draw causal inferences from the existing cross-sectional data. In the future, with the availability of more comprehensive data, more recent datasets (such as the Chinese General Social Survey or the China Labor-force Dynamic Survey) or longitudinal research data can be used to further explore the relationship between various influencing factors and health education for the mobile population. Additionally, it will be possible to examine the temporal changes in the key factors affecting health education and the dynamic changes in the equity of health education service utilization. Second, Although the China Migrants Dynamic Survey (CMDS) is highly representative, it also has certain limitations, such as respondents' subjective biases when answering the questionnaire and uneven coverage of survey areas. Third, factors affecting health are complex. When selecting the influencing factors on the health of the mobile population, this study may overlook other potential factors (e.g., type of disease, severity of disease, health service utilization, etc.). Future research needs to consider incorporating additional potential influencing factors into the analysis. Fourth, although this study combined random forest models and logistic regression methods to analyze factors affecting health education among mobile populations, which helped overcome the limitations of using a single method, it did not utilize other machine learning techniques such as decision trees or XGBoost. In the future, these methods could be considered for in-depth research on issues such as health education for the mobile population, thereby providing stronger evidence to support improvements in mobile populations' health status.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: data supporting the results of this study came from the National Mobile Population Dynamics Monitoring Survey (CMDS). The datasets generated and/or analyzed during the current study are available in the official website (<https://www.ncmi.cn/project/project-showProjectList.html?type=data&id=d3>).

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

MJ: Conceptualization, Data curation, Formal analysis, Methodology, Software, Writing – original draft, Writing – review & editing, Supervision. SL: Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. FY: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was financially supported by the Natural Science Foundation of Zhejiang Province (LTGY23H270005); the National Administration of Traditional Chinese Medicine of China-Zhejiang and the Provincial Administration of Traditional Chinese Medicine Co-Construction Project (GZY-ZJ-KJ-23062); the National Natural Science Foundation of China Faceted Project (71874157); Zheijiang Chinese Medical University 2024 Provincial Key Cultivation of Philosophical and Social Sciences Research Base Self submitted Subjects (2024JDYB09).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Fei C, Zhu Y, Jiang L, Zhou H, Yu H. Social integration, physical and mental health and subjective well-being in the floating population-a moderated mediation analysis. *Front Public Health*. (2023) 11:1167537. doi: 10.3389/fpubh.2023.1167537
2. National Bureau of Statistics of China. (2021). Communiqué of the seventh National Population Census (no.7) urban and rural population and floating population. Available at: http://www.stats.gov.cn/english/PressRelease/202105/t20210510_1817192.html (Accessed January 6, 2025).
3. Zheng L, Hu R, Dong Z, Hao Y. Comparing the needs and utilization of health services between urban residents and rural-to-urban migrants in China from 2012 to 2016. *BMC Health Serv Res*. (2018) 18:717. doi: 10.1186/s12913-018-3522-y
4. Zheng Y, Ji Y, Dong H, Chang C. The prevalence of smoking, second-hand smoke exposure, and knowledge of the health hazards of smoking among internal migrants in 12 provinces in China: a cross-sectional analysis. *BMC Public Health*. (2018) 18:655. doi: 10.1186/s12889-018-5549-8
5. Boardman J. Social exclusion and mental health - how people with mental health problems are disadvantaged: an overview. *Ment Health Soc Incl*. (2011) 15:112–21. doi: 10.1108/20428301111165690
6. Zhang J, Ye YL, Wang JC. Social exclusion and psychopathologies of Chinese floating populations. *J Affect Disord Rep*. (2021) 6:100263. doi: 10.1016/j.jadr.2021.100263
7. Hua Y, Wentao L, Xin P, Ping Z, Libin A. Research status of community hypertension health education evaluation in China. *Chin Gen Med*. (2013) 16:4190–3. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlbheLHWEgNiDyZx5HecXQeGnYp8YlTrgQdXxjbTclqIT_C6nWX2r6nfGZJUboUIptDbvUf144GIn41R013TxyDXyNlJr5neEalmln8fb4kuKu8zOfTavBrftPQnhUvBH9ZpcqYoWwt4s1Vp3fv3q9pdSZHtqZv5hrOHTBc61Bj9omYYU6g==&uniplatform=NZKPT&language=CHS
8. Zhu ZK, Leng CX, Chen SL, Zheng YX, Chen DQ. Does public health education improve migrant workers' health status in China?—evidence from China migrants dynamic survey. *Health Educ Res*. (2022) 37:292–313. doi: 10.1093/her/cyao020
9. Shao S, Zhang H, Chen X, Xu X, Zhao Y, Wang M, et al. Health education services utilization and its determinants among migrants: a cross-sectional study in urban-rural fringe areas of Beijing, China. *BMC Fam Pract*. (2021) 22:23. doi: 10.1186/s12875-021-01368-1
10. Levy CR, Phillips LM, Murray CJ, Tallon LA, Caron RM. Addressing gaps in public health education to advance environmental justice: time for action. *Am J Public Health*. (2022) 112:69–74. doi: 10.2105/AJPH.2021.306560
11. Zhou SF. Make joint efforts to realize the great goal of the "healthy China 2030" planning outline. *Chin J Rehabil Med*. (2017) 32:3. doi: 10.3969/j.issn.1001-1242.2017.01.002
12. Luyao W, Qin Y, Kai Z. China's floating population health education present situation and influence factors analysis. *China Health Educ*. (2021) 37:193–7. doi: 10.11847/zgggws1125696
13. Liping X, Hui F, Jing G. Study on the status quo of health education of floating population and its influencing factors. *Health Educ China*. (2017) 33:771–4. Available at: <https://link.cnki.net/doi/10.16168/j.cnki.issn.1002-9982.2017.09.001>

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpubh.2025.1570633/full#supplementary-material>

14. Yu Y, Li NL. Analysis on the status quo of health education and its influencing factors among urban and rural floating population in China. *Modern Prev Med*. (2021) 48:3769–73. Available at: <https://link.cnki.net/doi/10.20043/j.cnki.mpm.2021.20.024>

15. Jie D, Weng Z, Xiaohua C, Na W, Min Z, Junchang L. Analysis on the status quo and influencing factors of health education on prevention and treatment of infectious diseases among floating population in China. *Chin J Prev Med*. (2023) 24:299–304. doi: 10.16506/j.1009-6639.2023.04.004

16. Ya L, Yingzhe Z, Xingru L, Shuying C, Xiaolei W, Qi J, et al. Research on the utilization of mental health education and its influencing factors of the elderly migrant population based on Anderson model. *Modern Prev Med*. (2024) 51:2517:2522. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlbheLHWEiMlE963dNwsQEmOv-w0n5RBgYj76PHjT_aGeLmFtJzhFUF1bWCg4IYA1iH9LDKRsc_oppP1f6Y3P18Zh0TFIiHSsI5v1YzeG1r1gRvIxaoiDS31Q26OveGVF7NzRQuiipLC7f7ZlXy11EXzy_slcrNgeLqUq881hKIWlutey==&uniplatform=NZKPT&language=CHS

17. Xiaohua Y, Jingyi X, Yue L, Chunfang M, Rong T. An analysis of the importance of influencing factors of chronic disease prevention and control education for middle-aged and elderly migrant population in China: based on the 2018 CMDs data. *Modern Prev Med*. (2023) 50:1410–6. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlbheLHWEgWqUzDvW6dU2GJWeQn8iKyyC2_VRFD2Tmh7_9Fn33fj1C9xFA_KYQ0n8l34-P6m6Y0aRqrsxBACvUKTKRxeYldbnSjk6yqt-aHTceO-36KosNcc1C1ywi4iQYXMyUOLxdd-essdDcIPYVt5TsOQqbZ5gvye-PYTAW2xMEtWgJlVg==&uniplatform=NZKPT&language=CHS

18. Shuhui M, Yi H, Xiaofen C, Ruiqi G, Bei L. Research on public health service utilization and health equity of the Mobile elderly in China. *Chin Gen Pract*. (2023) 26:1938–45. doi: 10.12114/j.issn.1007-9572.2022.0561

19. Fang Z, Jing C, Ying J. Research on equity of health status of floating population and its influencing factors. *China Rural Health Serv Manag*. (2023) 43:235–41. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlbheLHWEj699YLYUP0mB_yAGmVPUPnNIqoo6SggDH6i0AfwMG2m65BlqAHOEym0RBKXdaOSmJi0MEeOY7beVgCfC0KpBvVnAojZfQtposXr-Khliz5Y_8NEOuRwOjuESC2ZU70Qpb51F8lgQZQAftsng_ebmiME16M_8d7v7rW_RFX6c5w==&uniplatform=NZKPT&language=CHS

20. Yan G, Zheng X. Closing the gap in a generation: the theory of social determinants of health and international experiences. *Pek Univ J Med Sci*. (2009) 41:125–8. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlbheLHWEjgXDU1XKpMUisidXJ_TAWGRr4ZZpEfNrxDkRdEf8dYj6UHidzfV9dJ9iMeup-EyvkCYbW1pP_lruVdujpNghiOAAh8wOWm_rLEHVdma9-WMVuv7hmAE23pNxy5HnsRq0qm7h0B9rIS2vCIRo-K5fSU2LijTV9YbKKHsLac-nk-A==&uniplatform=NZKPT&language=CHS

21. Diderichsen F, Evans T, Margaret W. The social basis of disparities in health In: T Evans, editor. *Challenging inequities in health: from ethics to action*. New York: Oxford Academic (2001)

22. Yuee L, Shan L. The development, application, and implications of the Andersen model in the field of healthcare. *Chin J Health Policy*. (2017) 10:77–82. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlbheLHWEikQLWLau77M2ypIGZy0rQkzTBh3LdkDhAZZ2d4dg-Kpqh6ASbkZ9cswlKviyMA3kH2Qr6lBaScirAmnKe5xR6jjDEqT4KLAX4rd9HZDKiY9kQXfBo8qZujskdhY54D2hO9PVkXiRQw5izeclgiwB5_1QljiGiTaYnETHA0z3Q==&uniplatform=NZKPT&language=CHS

23. Shao S, Wang M, Jin G, Zhao Y, Lu X, Du J. Analysis of health service utilization of migrants in Beijing using Anderson health service utilization model. *BMC Health Serv Res.* (2018) 18:462. doi: 10.1186/s12913-018-3271-y
24. Yang Y, Yu M. Disparities and determinants of maternal health services utilization among women in poverty-stricken rural areas of China: a cross-sectional study. *BMC Pregnancy Childbirth.* (2023) 23:115. doi: 10.1186/s12884-023-05434-7
25. Qiang Y, Zhiheng Z, Zhaohui C, Hanxuan L. Health inequality among migrant populations in China from the perspective of health insurance: theoretical framework, measurement, and mechanism analysis. *Chin J Health Policy Res.* (2024) 17:26–34. doi: 10.3969/j.issn.1674-2982.2024.11.004
26. Xufeng A, Li Q, Wanchen W, Jia S, Chengxin F, Zhiqiang F, et al. Analysis of the current status and influencing factors of inpatient service utilization among the migrant elderly population based on the Andersen model. *Modern Prev Med.* (2022) 49:3170–5. doi: 10.20043/j.cnki.MPM.202201081
27. Haoran C, Hui F, Jinru Y, Shanju H. Study on the status and influencing factors of family doctor contracting among the migrant elderly population based on the Andersen model. *Modern Prev Med.* (2023) 50:2578–82. Available at: <https://link.cnki.net/doi/10.20043/j.cnki.MPM.202303320>
28. Su B, Wu Y, Yihao Z, Chen C, Panliang Z, Zheng X. The effect of equalization of public health services on the health China's migrant population: evidence from 2018 China migrants dynamic survey. *Front Public Health.* (2023) 10:1043072. doi: 10.3389/fpubh.2022.1043072
29. Zhang J, Cai JL, He ZC, Huang YY, Tang GZ. Analysis on the status quo and influencing factors of health education of floating population in China. *Health Educ China.* (2021) 37:291–6. Available at: <https://link.cnki.net/doi/10.16168/j.cnki.issn.1002-9982.2021.04.001>
30. Hapfelmeier A, Ulm K. A new variable selection approach using random forests. *Comput Stat Data Anal.* (2013) 60:50–69. doi: 10.1016/j.csda.2012.09.020
31. Kong Y, Yu T. A deep neural network model using random Forest to extract feature representation for gene expression data classification. *Sci Rep.* (2018) 8:16477. doi: 10.1038/s41598-018-34833-6
32. Cutler AD, Richard C, John RS. Random forests In: C Zhang and YQ Ma, editors. *Ensemble machine learning*. New York: Springer (2012)
33. Hui T, Chunmei L, Yangjun T, Zhou T, Mengni Z, Qingxia S. Study on the current status and influencing factors of stroke Caregivers' self-care contribution based on the random forest model. *Mil Nurs.* (2025) 42:1–5. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlBheLHWEibMxB0gVd_sMkHUpjwGzN2atFNRTWKik57WAPu9kOevHAN0VK0wJMyuTlKaXbtCPHH30rSBuOCLTwB9JaSWOFADV6OkY1YaF28Z7d9Vz4D9HXHTiz6mr9bqdcQq0-CUKu6zyq6Gu_WkF4Ot_7VzneglZluyCXZKyE1E73z_EEA==&uniplatform=NZKPT&language=CHS
34. Xing M, Zhao Y, Li Z, Zhang L, Yu Q, Zhou W, et al. Development and validation of a stacking ensemble model for death prediction in the Chinese longitudinal healthy longevity survey (CLHLS). *Maturitas.* (2024) 182:107919. doi: 10.1016/j.maturitas.2024.107919
35. Xiumin Z, Wenjing Z, Hong S, Xihe Y, Hanfang Z, Shixun W, et al. The acceptance and influencing factors of self-rescue health education for emergencies among migrant populations in China. *Med Soc.* (2023) 36:59–85. doi: 10.13723/j.yxysh.2023.12.010
36. Kim KM, Kim JH, Rhee HS, Youn BY. Development of a prediction model for the depression level of the elderly in low-income households: using decision trees, logistic regression, neural networks, and random forest. *Sci Rep.* (2023) 13:11473. doi: 10.1038/s41598-023-38742-1
37. Epo BN, Baye FM. Decomposing poverty-inequality linkages of sources of deprivation by men-headed and women-headed households in Cameroon. *J Econ Dev.* (2016) 41:57–79. doi: 10.35866/caujed.2016.41.1.004
38. Chirwa GC, Suhrcke M, Moreno-Serra R. Socioeconomic inequality in premiums for a community-based health insurance scheme in Rwanda. *Health Policy Plan.* (2021) 36:14–25. doi: 10.1093/heapol/czaa135
39. Chen M, Fang G, Wang L, Wang Z, Zhao Y, Si L. Who benefits from government healthcare subsidies? An assessment of the equity of healthcare benefits distribution in China. *PLoS One.* (2015) 10:e0119840. doi: 10.1371/journal.pone.0119840
40. Marfels C. Absolute and relative measures of concentration reconsidered. *Kyklos.* (1971) 24:753–66. doi: 10.1111/j.1467-6435.1971.tb00631.x
41. Kjellsson G, Gerdtham UG. On correcting the concentration index for binary variables. *J Health Econ.* (2013) 32:659–70. doi: 10.1016/j.jhealeco.2012.10.012
42. Nanfu H, Yajiao P, Zhongqi L, Hong S, Hanfang Z, Yuyu L, et al. Influencing factors and equity of health status among the migrant elderly population in China. *Chin J Gerontol.* (2021) 41:4398–401. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlBheLHWEiEGEG4sP7t0rSspGFb1uNxr83t-ODlyMblJAUWx6yvWgD4GpZQQiNv8FLEx3KmwZ2uqqjy3mdAW0VkBjM0_N13io8PrHy5CSKsXm9fZyj1fzpepBQ11f6krsEQE2r-Org2uQQQYZ_6M5wQ5x9zRe8OVtXetjOeR8TbVfpvzsg0g==&uniplatform=NZKPT&language=CHS
43. Xiangrong L, Xiumin Z, Zheng R, Xinwen F, Xia G, Hong S, et al. Evaluation of the equity in inpatient service utilization among migrant populations in eastern China. *Med Soc.* (2021) 34:12–5. doi: 10.13723/j.yxysh.2021.09.003
44. Wagstaff A, Paci P, Van Doorslaer E. On the measurement of inequalities in health. *Soc Sci Med.* (1991) 33:545–57. doi: 10.1016/0277-9536(91)90212-U
45. Grömping U. Variable importance assessment in regression: linear regression versus random forest. *Am Stat.* (2009) 63:308–19. doi: 10.1198/tast.2009.08199
46. Nannan Q. General causality test based on lasso improvement. *Stat Inform Forum.* (2008) 23:5. doi: 10.3969/j.issn.1007-3116.2008.02.004
47. Jingmin Y, Ying C, Jie J, Lin H, Ruijun L, Zhuozhi S, et al. Influencing factors of outcomes of anti-tuberculosis treatment in TB/HIV patients based on LASSO-logistic regression. *Modern Prev Med.* (2023) 50:4039–45. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlBheLHWEhHvMuFpANGvVbSdUkacpKs-2uLVbhlw2zvF_w1sMfZMHX-ZJMX69iioX2B_PzXbBj31TQhl8caJ00MaoKRErLcMPckRmxUfdZtzhEYaYxyfPr7w-t881PKoVoYgJW9wj7Bi4hZ3sfh62vvegto7GTT70FhO0XyZiwrXPM8i5-g==&uniplatform=NZKPT&language=CHS
48. Zivin JG, Neidell M. Environment, health, and human capital. *J Econ Lit.* (2013) 51:689–730. doi: 10.1257/jel.51.3.689
49. Belcastro PA, Ramsaroop-Hansen H. Addressing the antinomy between health education and health literacy in advancing personal health and public health outcomes. *J Sch Health.* (2017) 87:968–74. doi: 10.1111/josh.12570
50. Liao ZL, Wang DW. Analysis of the status quo of mental health education and its influencing factors for the elderly floating population. *Chin Public Health.* (2022) 38:922–7. doi: 10.11847/zgggws1135395
51. Vu HM, Vu LG, Tran TH, Boyer L, Auquier P, Fond G, et al. Economic burden and financial vulnerability of injuries among the elderly in Vietnam. *Sci Rep.* (2023) 13:19254. doi: 10.1038/s41598-023-46662-3
52. Qiong Y, Lian T. Analysis on utilization of basic public health services and its influencing factors among young floating population. *Chin J Public Health.* (2019) 35:680–4. doi: 10.11847/zgggws1119401
53. Navabi N, Ghaffari F, Jannat-Alipoor Z. Older adults' attitudes and barriers toward the use of mobile phones. *Clin Interv Aging.* (2016) 11:1371–8. doi: 10.2147/CIA.S112893
54. Mulinde P. Elderly people's perceptions on the use of mobile phones to support the self-management of long-term illnesses at Kiruddu National Referral Hospital. *Afr Health Sci.* (2024) 24:269–78. doi: 10.4314/ahs.v24i1.32
55. Guo J, Guo YM, Zhu L, Fu LL, Dai Y, Chen SL. Analysis on health literacy of infectious diseases among floating population in China and its influencing factors. *Chin J Public Health.* (2021) 37:209–13. doi: 10.11847/zgggws1131104
56. Xiao Hui W, Li JB, Yang Y. Status quo of health education and establishment of health records among elderly migrant population in China and its influencing factors. *Chin J Public Health.* (2021) 37:203–8. doi: 10.11847/zgggws1128961
57. Kincaid R. Status, masculinity, and femininity at the intersection of age and gender. *Soc Sci Res.* (2022) 105:102695. doi: 10.1016/j.ssresearch.2022.102695
58. Yuhua D, Chao W, Yong G, Zuxun L. Analysis on the current situation and influencing factors of family doctor contract service utilization in China: based on a survey of 31 provinces and cities in China. *Chin Health Policy Res.* (2020) 13:47–54. Available at: https://kns.cnki.net/kcms2/article/abstract?v=RNlBheLHWEiivw4fKHeSn c O c f R 8 A D j s N g _ m l q v z 6 h p K k A Q 3 P R L i _ a - 6 T W Q 0 Z p T B H HdQGZ6owo7Itqt4Sj3-ykn0srE_YRkQyO8uM44RxLM6OYcilqwxrgS5en4ix2ECuEOK3jnJlB7eXvAP8UMsj-hW9ukqQp4bEYME-bz-zJAm0lyDyidyqZw==&uniplatform=NZKPT&language=CHS
59. Cihui Z, Siyu Z, Shuting L, Leying H, Qian Y, Xinxin Y, et al. Analysis on influencing factors and equity of basic public health service utilization among Chinese floating population. *Chin J Soc Med.* (2023) 40:605–10. doi: 10.3969/j.issn.1673-5625.2023.05.023