Received: 15 March 2023 | Revised: 26 May 2023 | Accepted: 28 May 2023

DOI: 10.1002/hsr2.1341

ORIGINAL RESEARCH

Health Science Reports Open Access WILEY

Research on the socioeconomic factors that influence the development of voluntary, non-remunerated blood donation in China—A correlation and regression analysis based on data from 2012 to 2018

Feng Lin | **Yi Huang** | **Xu He** | **Zhong Liu**

Institute of Blood Transfusion, Chinese Academy of Medical Science and Peking Union Medical College, Chengdu, China

Correspondence
Zhong Liu, Institute of Blood Transfusion, Chinese Academy of Medical Science and Peking Union Medical College, No. 26 Huacai Road, Chengdu 610052, China.
Email: liuz@ibt.pumc.edu.cn

Funding information
Fundamental Research Funds for the Central Universities, Grant/Award Number: 3332019170

Abstract

Background and Aims: Previous research has shown that socioeconomic factors, such as income and education, are associated with blood donation behavior. This study aims to investigate the factors that influence blood donation behavior using a large database of blood donation records and to provide further empirical evidence and insights into the factors that influence blood donation behavior and to identify potential strategies to increase blood donation rates in China.

Methods: This study employed correlation analysis and regression analysis to investigate the impact of socioeconomic factors on blood donation rates in 36 major cities in China. The study also used the K-Means clustering algorithm to identify the socioeconomic factors that influence blood donation behavior at the per capita level. Additionally, the study conducted correlation analysis between the donor's age groups and the socioeconomic factors. The indices of economy and education were determined by correlation analysis.

Results: At the city level, overall economic output and university student populations were positively correlated with total blood donation times. However, at the per capita level, this linear relationship disappeared. By applying the K-Means clustering algorithm, the cities were divided into subgroups where the linear relationship reemerged. The study also found that the number of college students had a linear relationship with blood donors between the ages of 18–34.

Conclusion: City characteristics and demographics significantly impact blood donation rates in different ways. A universal policy may have limited effect given the heterogeneity across cities. Tailored interventions targeting specific city clusters and education levels are needed to promote blood donation behavior, especially among youth and students. Investment in education and targeted education on blood donation appear particularly impactful. A combined approach integrating social and economic policies will likely be most effective.

KEYWORDS
blood donation, correlation analysis, K-Means, regression analysis, socioeconomic factors

1 | INTRODUCTION

Blood transfusion services are a vital component of healthcare systems worldwide. However, ensuring a sustainable supply of blood and blood products remains a challenge for many countries.[1] This is particularly true in China, where demand for blood exceeds supply. To address this issue, it is important to understand the factors that are associated with blood donation. In this context, this study aims to investigate the factors that influence blood donation behavior among the Chinese population.

Previous research has shown that blood donation behavior is associated with various factors, including socioeconomic status and education.[2–12] Specifically, individuals with higher levels of education and income tend to have a higher likelihood of donating blood. However, despite the importance of blood donation, many regions still struggle with low donation rates, which can lead to blood shortages and risks for patients in need. Therefore, identifying effective strategies to increase blood donation rates is crucial for ensuring an adequate supply of blood and blood products.[13–15]

Previous research has shown that blood donation behavior is associated with socioeconomic factors such as income and education. These findings were mainly based on surveys and questionnaires.[8–15] In contrast, the current study aims to investigate the same relationship using correlation and regression analysis based on a large database of blood donation records. By doing so, we aim to provide further empirical evidence and insights into the factors that influence blood donation behavior and to identify potential strategies to increase blood donation rates.

To investigate the relationship between blood donation and socioeconomic data, we obtained our data from two sources: the National Bureau of Statistics and a blood donation database supplied by the Institute of Blood Transfusion.

2 | MATERIALS AND METHODS

2.1 | Materials: Data source

Voluntary, nonremunerated blood donation has been utilized in China for over two decades.[16] In China, whole blood is collected by blood centers. The country has 452 blood centers, including 32 provincial, 321 regional, and 99 county-level blood centers.[17] In addition, Chinese donation databases have kept detailed and interlinked data on donors, donations, blood components and transfusions since 2012. In this study, we choose the data from 2012 to 2018.

This study selects 36 major cities in China as research objects, including directly controlled cities, provincial capital cities and planned units. These cities not only have large populations and influence, but are also key and frontline cities for China's economic and social development, and can more comprehensively reflect the development of cities in China. As the data in Hainan province is not separated by city, the whole of Hainan is regarded as a single city (the area of Hainan is less than that of the city of Chongqing).

To analyze the possible relationships between economic development, education progress and blood donation behavior in the 36 major cities, we obtained annual statistical data and blood donation data from 2012 to 2018 of these 36 cities. Statistical data includes GDP, tertiary industry output value from 2012 to 2018 of these 36 cities to reflect economic development and enrollment numbers of college students from 2012 to 2018 to reflect educational progress in these 36 cities. These data come from the *Annual Statistical Review* released by the National Bureau of Statistics, aiming to build the authority and persuasiveness of the analytical framework.

2.2 | Data processing

The blood donation data were collected from various monitoring stations in each city. To obtain the annual data for each city, the data from all stations within that region were aggregated. In cases where the data for a particular year were missing for the resident population, the value for the previous year's resident population or the registered population for that year was used as a substitute.

2.3 | Method

2.3.1 | Selection of relevant variables for modeling

As the statistical yearbook only provides GDP and the output values of primary, secondary, and tertiary industries for these cities, although GDP is commonly used to measure the total economic activity of a region or country, the output value of the tertiary industry also reflects the level of local economic prosperity. To determine the variables that are more suitable for reflecting the correlation between the economy and blood donation, we used correlation analysis methods. Similarly, as the statistical yearbook only provides the number of college students, we used the number of college students as the local education indicator. Since the number of blood donations can be further divided into total blood donations, individual blood donations, and group blood donations, we calculated the correlation coefficients between economic and education indicators and each of these indicators to determine the most appropriate measure of blood donation behavior.

We used the Spearman correlation coefficient in the non-parametric statistical methods of the scipy library's stats package, rather than the common Pearson correlation coefficient.[18–20] The Spearman correlation coefficient is suitable for data that does not follow a normal distribution and can be used to measure the monotonic relationship between two variables, making it more suitable for handling nonlinear data.[21] By using the Spearman correlation coefficient and *p* value, we evaluated the strength of the correlation between each economic and education indicator and the different measures of blood donation. This analysis helped us determine which economic indicator is best suited for examining the relationship between the economy and blood donation and which measure of blood donation behavior is most appropriate for our study. The variables to be determined include the economy (such as GDP or tertiary industry indicators) and blood donation behavior (such as total blood donation times or other blood donation indicators), which will be screened through correlation coefficients. As economists have stated, changes to economic structure result in long-term employment shifts.[22,23] This implies that if a particular area's tertiary industry output values are high, then it is relatively well developed. We can determine the correlation coefficients for the number of the blood donations and tertiary industry output values, as well as the coefficients for the number of blood donations and GDP. If these two correlation coefficients are similar, we will analyze the regression model to determine which one is more suitable to be the economic index.[24–27]

2.3.2 | Model development and evaluation

We will evaluate the relationship between the economy, education, and blood donation behavior through correlation analysis and select the most relevant indicators to establish a linear model to describe this relationship.

However, the linear relationship between these aggregate indicators does not necessarily indicate a linear relationship between per capita indicators, so we need to observe whether there is a linear relationship between per capita economy, education, and blood donation rates. If a linear relationship exists, we will establish a new linear model to better evaluate this relationship. Specifically, we will use data on per capita economy and education levels and blood donation rates per 1000 population to establish linear regression models in different regions and evaluate the significance of the models.

If there is no linear relationship, we will calculate the ratio of per capita economy (or per capita education) to blood donation rates per 1000 population as a classification indicator and use clustering methods to group regions, and then establish linear models for each group separately.[28,29] We will then use the parameters of these models to evaluate the blood donation behavior in each region and compare the differences between different regions. This way, we can have a more comprehensive understanding of the relationship between per capita economy, education, and blood donation rates and develop public health policies and blood donation promotion plans that are more suitable for different regions.

In summary, we will use correlation analysis, linear regression models, and clustering methods to evaluate the relationship between the economy, education, and blood donation behavior, better understand the relationship between these variables, and provide scientific evidence for public health policies and blood donation promotion plans.

2.3.3 | Analysis of the relationship between age and blood donation behavior

In addition to the main analysis, a simple analysis was conducted on the correlation between the age of blood donors and the number of blood donations, as well as the proportion of the number of blood donations in different age groups to the total number of blood donations. Although this analysis is not the main focus of our study, it provides additional insights into the blood donation behavior of different age groups and complements the overall analysis. By examining the relationship between age and blood donation behavior, we can better understand the factors that influence blood donation and develop targeted strategies to encourage blood donation among different age groups.

2.3.4 | Short summary

Statistical analyses are performed using Python 3.8. All tests are two-tailed with a significance level of 0.05.

Spearman's rank correlation analysis was conducted to assess the relationships between variables. The null hypothesis is that there is no significant correlation. Correlation coefficients are calculated using the scipy.stats library.

Linear regression analyses are performed using the statsmodels.api library to determine the effects of predictors on the outcome variable. The null hypothesis is that there is no linear relationship between predictors and the outcome. *t*-Tests are used to evaluate the significance of individual regression coefficients. A coefficient is considered statistically significant if the absolute value of t is larger than 2 and p is less than 0.05.

The explanatory power of the regression models was assessed using R square and F tests. For R square, the null hypothesis was that the model explains none of the variance. R square represents the proportion of variance explained, with higher values indicating better model fit. For the F test, the null hypothesis is that the model explains no more variance than the intercept-only model. An F test with p less than 0.05 indicates a significant model.

The intercept indicates the predicted value of the outcome when all predictors are 0. Though often nonsignificant, the intercept provides a reference point for interpreting the effects of other predictors.

3 | RESULTS AND ANALYSIS

We will calculate the correlation coefficients between economic and educational indicators from 2012 to 2018 and the number of blood donations following the method proposed in the second section. And we will analyze the regression models in the following subsections.

3.1 | Comparison of different blood donation metrics and their relationship with economic and educational indicators

To analyze blood donation patterns in urban areas, we first identified which economic and blood donation variables to use. Since blood donations can be categorized into total donations, individual donations, and group donations, we had to consider different variables. We then calculated the correlation coefficients between economic and education indicators for each of these categories to determine the most appropriate measure of blood donation behavior. The correlation coefficients are shown in Tables 1–3.

The scatter plots of GDP, tertiary industry output values, and number of college students versus total donation times are presented in Figure 1 in the left column, corresponding to the variables' respective correlation coefficients in Table 1. To facilitate

TABLE 1 Spearman's correlation coefficients between total donation times and the corresponding variables (2012–2018).

	Spearman's correlation coefficients between total donation times and the corresponding variables					
Year	GDP	*p* value	Tertiary industry output values	*p* value	Number of college students	*p* value
2012	0.73	4.36E-07	0.7	2.35E-06	0.79	1.01E-08
2013	0.77	3.36E-08	0.76	9.44E-08	0.72	7.86E-07
2014	0.81	2.56E-09	0.77	2.92E-08	0.81	2.36E-09
2015	0.8	5.5E-09	0.78	1.99E-08	0.81	3.08E-09
2016	0.82	8.96E-10	0.83	4.73E-10	0.8	3.70E-09
2017	0.83	5.96E-10	0.85	8.82E-11	0.82	7.83E-10
2018	0.81	2.00E-09	0.82	6.84E-10	0.84	1.76E-10

TABLE 2 Spearman's correlation coefficients between individual donation times and corresponding variables (2012–2018).

	Spearman's correlation coefficients between individual donation times and the corresponding variables					
Year	GDP	*p* value	Tertiary industry output values	*p* value	Number of college students	*p* value
2012	0.72	8.62E-07	0.7	2.30E-06	0.72	8.85E-07
2013	0.75	1.23E-07	0.73	5.16E-07	0.61	8.09E-05
2014	0.77	4.32E-08	0.74	3.15E-07	0.74	2.22E-07
2015	0.77	4.55E-08	0.76	7.08E-08	0.75	1.09E-07
2016	0.72	6.16E-07	0.74	2.53E-07	0.75	1.55E-07
2017	0.74	1.89E-07	0.78	2.50E-08	0.76	6.32E-08
2018	0.72	6.60E-07	0.75	1.23E-07	0.73	3.74E-07

TABLE 3 Spearman's correlation coefficients between group donation times and the corresponding variables (2012–2018).

	Spearman's correlation coefficients between group donation times and the corresponding variables					
Year	GDP	*p* value	Tertiary industry output values	*p* value	Number of college students	*p* value
2012	0.46	0.0044	0.43	0.0083	0.46	0.0048
2013	0.45	0.0063	0.46	0.0046	0.51	0.0014
2014	0.51	0.0016	0.51	0.0017	0.44	0.0077
2015	0.59	0.00015	0.57	0.00029	0.47	0.0035
2016	0.69	2.67E-06	0.68	5.32E-06	0.55	4.90E-04
2017	0.6	0.00012	0.58	0.00019	0.58	0.00021
2018	0.6	0.00012	0.58	0.00019	0.58	0.0002

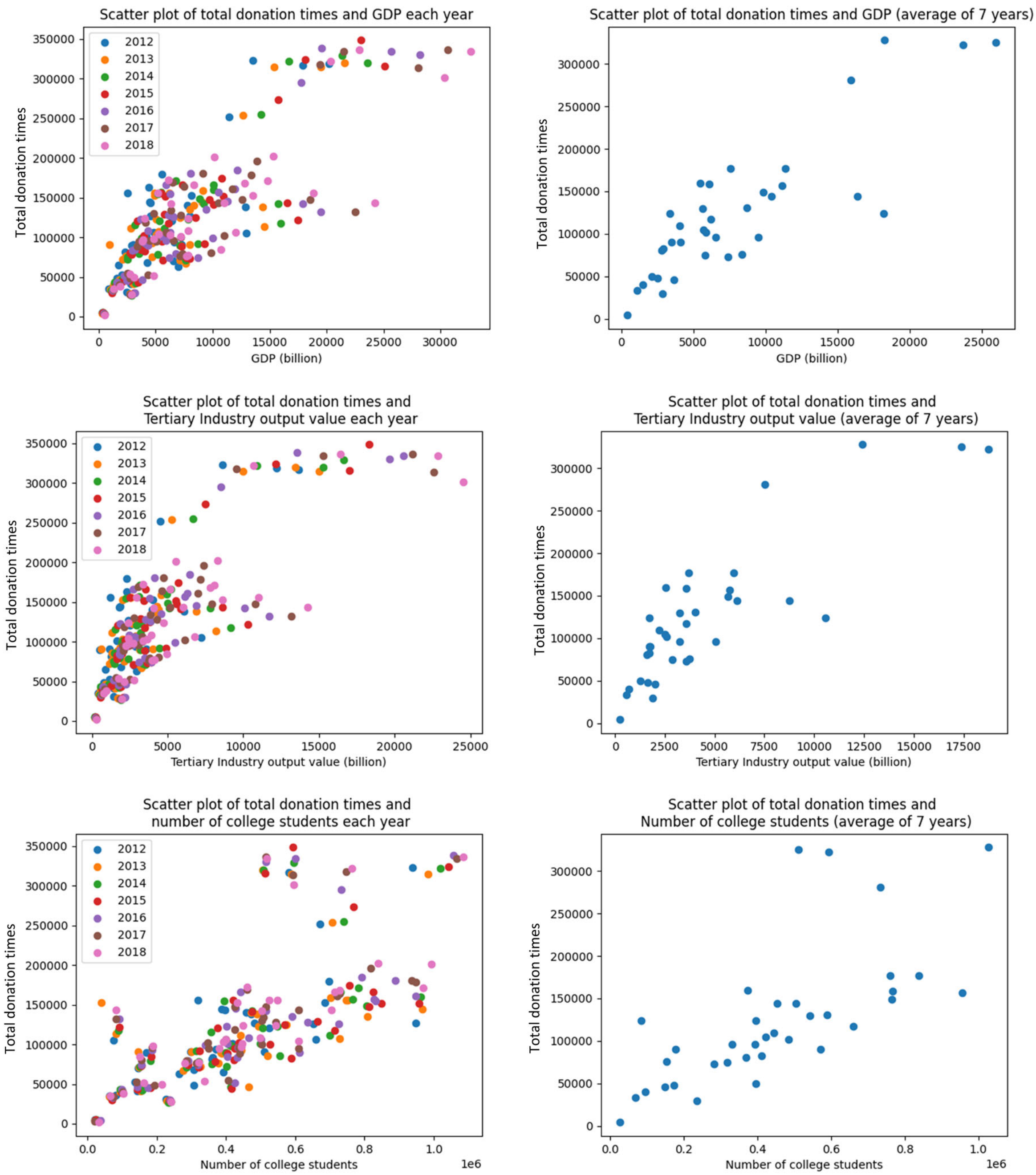

FIGURE 1 Scatter plot of GDP, tertiary industry output value and number of college students versus total donation times.

interpretation, the scatter plots in the right column display the average values of these variables over a 7-year period.

The scatter plots of GDP, tertiary industry output values, and number of college students versus individual donation times are presented in Figure 2 in the left column, corresponding to the variables' respective correlation coefficients in Table 2. To facilitate interpretation, the scatter plots in the right column display the average values of these variables over a 7-year period.

The scatter plots of GDP, tertiary industry output values, and number of college students versus group donation times are presented in Figure 3 in the left column, corresponding to the variables' respective correlation coefficients in Table 3. To facilitate interpretation, the scatter plots in the right column display the average values of these variables over a 7-year period.

In Tables 1–3, we observe that the Spearman's correlation coefficients between total donation times and other variables, such as GDP, output values of tertiary industries, and number of college students, are larger than those for individual donation times and group donation times. These findings are further supported by the data presented in Figure 4 Scatter plot of GDP,

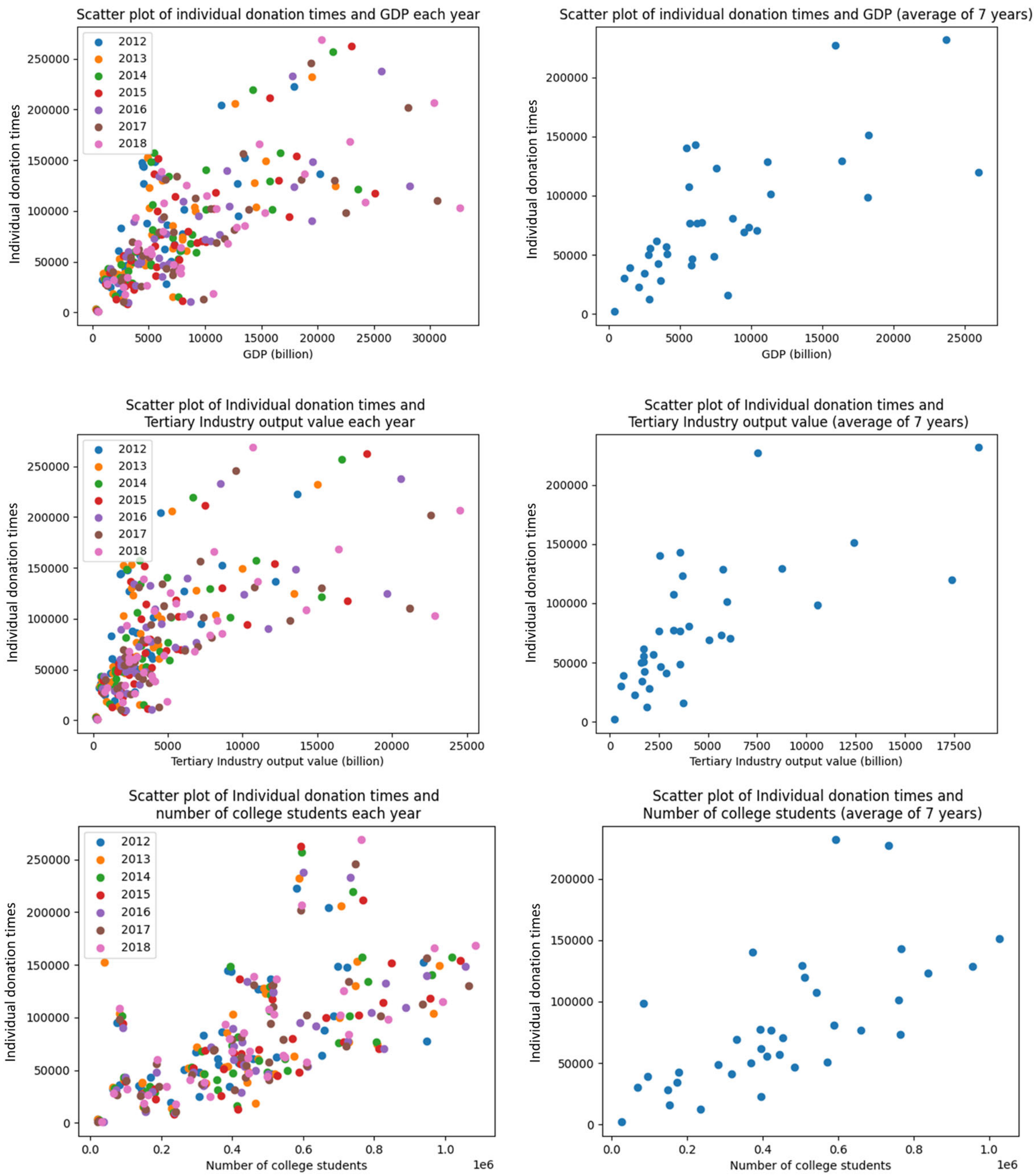

FIGURE 2 Scatter plot of GDP, tertiary industry output value and number of college students versus individual donation times.

tertiary industry output value and number of college students vs group donation times. Consequently, we have chosen to concentrate our analysis on how these variables specifically impact total donation times.

Before proceeding, we determine which variable influences blood donation times by conducting regression analysis. We choose either GDP and number of college students or output values of tertiary industries and number of college students as variables to explain how they influence total donation times. Table 4 shows that the model explains over 79% of the variance in Total Donation times. The effects of GDP and Number of College Students on Total Donation times are statistically significant at $p < 0.05$. Table 5 shows that there is no multicollinearity between GDP and number of college students. The t and p values indicate that the effects of the predictor variables on the outcome are extremely statistically significant.

Table 6 shows that the model explains over 80% of the variance in Total Donation times. The effects of GDP and number of college students on total donation times are statistically significant at $p < 0.05$. Table 7 shows that there is no multicollinearity between output value of tertiary industries and number of college students.

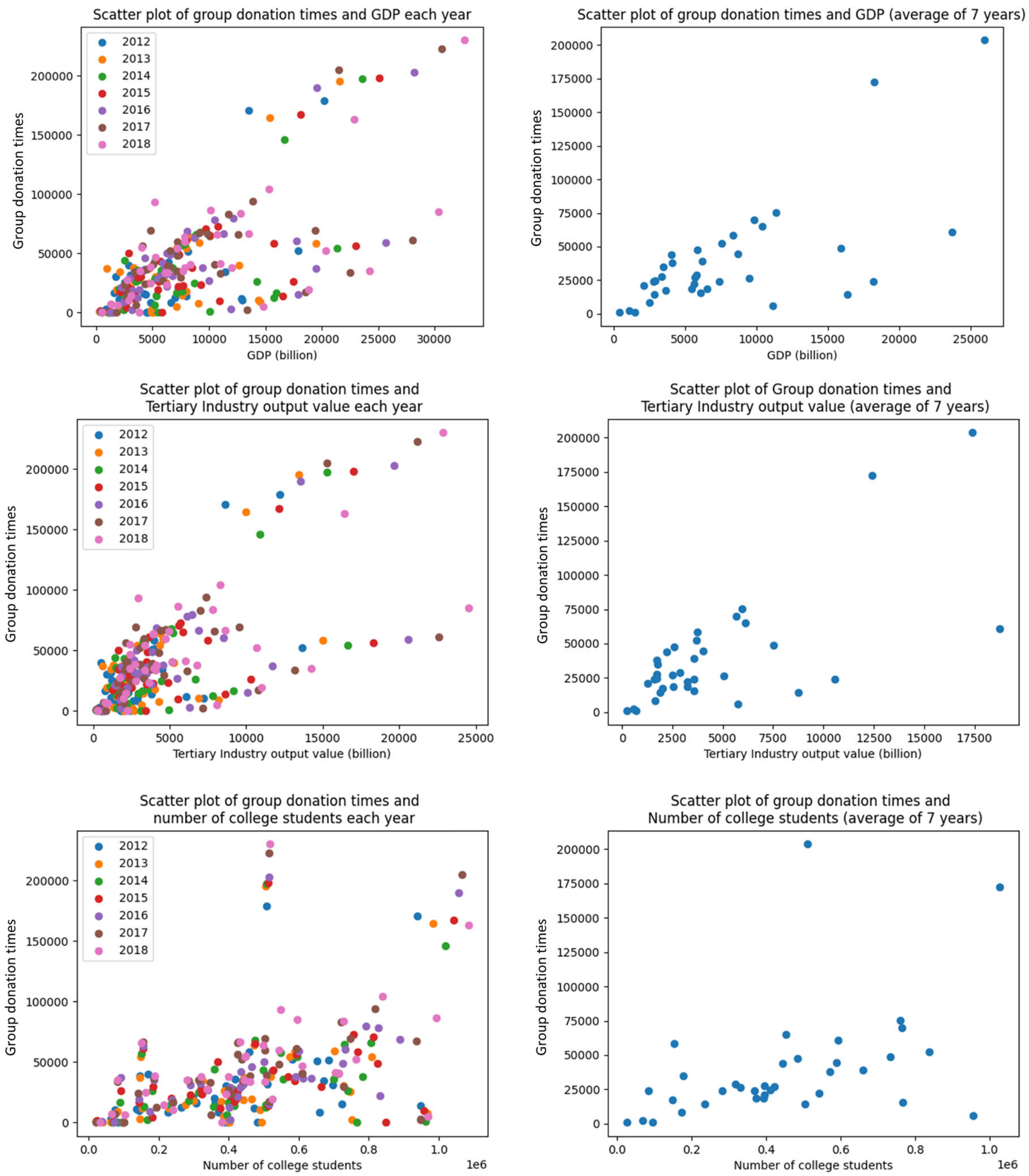

FIGURE 3 Scatter plot of GDP, tertiary industry output value and number of college students versus group donation times.

The t and p values indicate that the effects of the predictor variables on the outcome are extremely statistically significant.

By comparing the R squared and t values of regression models using GDP and tertiary industry output as variables, we observed that the R squared and t values for tertiary industry output were both higher. This can be directly seen in Figures 5 and 6.

According to the regression results, we can conclude that economy and the education are the factors that will influence the total donation times, and output value of tertiary industries is a better indicator which reveals the relation of economy and blood donation times.

3.2 | Correlation and regression analysis of per capita economy and education with blood donation per thousand population

Although the economy, specifically the output values of tertiary industries and the number of college students in a city, can affect the overall number of donations made, it's important to note that resident populations vary from city to city. Therefore, it's crucial to assess whether the output values of tertiary industries and the number of college students per capita still impact the donation rate per capita.

To investigate this, we conducted a correlation analysis and found that there is no linear relationship between the per thousand population blood donation rate and the average GDP, average Tertiary Industry Output Value, and percentage of university students. Table 8 displays the corresponding correlation coefficients.

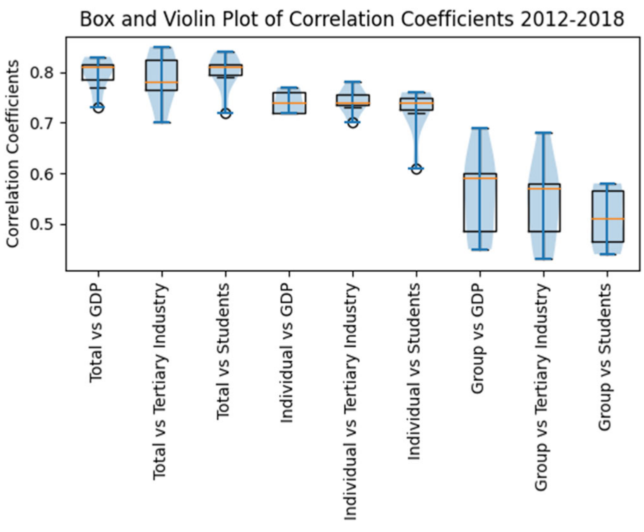

FIGURE 4 Correlation of blood donations, economic indices and education factors: A comparison among total, individual and group levels.

It is noteworthy that the distribution of average tertiary industry output values and percentage of university students across various cities may not align with the distribution of donation times. To better understand the correlation between these factors, it is important to sort the data into subsets.

To identify areas with a linear relationship between the factors, the data was no longer clustered on a yearly basis. This is because clustering the data from different years can result in certain biases between the clustering results. Instead, a clustering analysis was conducted on the arithmetic mean of the data over a 7-year period. The clustering standards were based on two ratios: the ratio of per capita tertiary industry output values to per thousand population blood donation (standard 1), and the ratio of average university student population to per thousand population blood donation (standard 2). Therefore, two clustering results were obtained based on these two standards.

To better represent the economy of each city, the ratios of average GDP and per thousand population blood donation were dropped. Using the K-Means algorithm, the cities were sorted into several categories based on either standard 1 or standard 2. The major cities were divided into 6 categories, which are outlined in Table 9 (for standard 1) and Table 11 (for standard 2). Each group

TABLE 4 Regression analysis of total donation times between GDP and number of college students for the model.

Year	Regression coefficients			Evaluation of the model	
	Constant (C)	GDP (G)	Number of college students (N)	*R* square	*F* value
2012	90	11	0.013	0.79	4.70E-12
2013	710	11	0.009	0.8	2.00E-12
2014	−490	9.8	0.012	0.86	4.60E-15
2015	−580	9.6	0.011	0.86	7.10E-15
2016	−265	8.7	0.012	0.87	1.40E-15
2017	−210	7.6	0.014	0.88	5.00E-16
2018	99	6.9	0.014	0.87	1.60E-15

TABLE 5 Regression analysis of total donation times between GDP and number of college students for each of the regression variables.

Year	Variance inflation factor			*t* values			*p* values		
	C	G	N	C	G	N	C	G	N
2012	4.4	1.3	1.3	0.072	7.20	4.40	9.40E-02	0.027	9.00E-03
2013	4.1	1.2	1.2	0.6	8.30	3.70	5.50E-01	0.012	9.00E-03
2014	4.5	1.2	1.2	−0.46	9.70	5.40	6.50E-01	3.60E-02	5.90E-06
2015	4.5	1.2	1.2	−0.52	9.90	5.10	6.00E-01	2.00E-02	1.20E-02
2016	4.5	1.2	1.2	−0.25	10.40	5.70	8.10E-01	5.40E-02	2.30E-02
2017	4.5	1.2	1.2	−0.2	10.10	6.40	8.40E-01	1.30E-02	3.10E-02
2018	4.4	1.2	1.2	0.092	9.40	6.40	9.30E-01	7.00E-02	2.80E-02

TABLE 6 Regression analysis of total donation times between output value of tertiary industries and number of college students for the model.

	Regression coefficients			Evaluation of the model	
Year	Constant (C)	Output value of tertiary industries (T)	Number of college students (N)	*R* square	*F* value
2012	9.00E+01	11	0.013	0.82	8.00E-13
2013	2.00E+04	16	0.01	0.83	2.10E-13
2014	3.40E+03	15	0.013	0.9	4.20E-17
2015	2.20E+03	14	0.012	0.9	6.50E-17
2016	5.30E+03	12	0.013	0.89	2.50E-16
2017	5.50E+03	11	0.015	0.89	2.20E-16
2018	7.90E+03	9.5	0.015	0.88	6.50E-16

TABLE 7 Regression analysis of total donation times between output value of tertiary industries and number of college students for each of the regression variables.

	Variance inflation factor			*t* values			*p* values		
Year	C	T	N	C	T	N	C	T	N
2012	4.4	1.30	1.30	0.072	7.20	4.40	9.40E-02	0.027	9.00E-03
2013	4.1	1.20	1.20	1.5	9.20	4.40	1.40E-02	0.012	1.10E-02
2014	4.5	1.20	1.20	0.38	12.00	7.00	7.10E-01	3.10E-13	9.00E-08
2015	4.5	1.20	1.20	0.24	12.00	6.60	8.20E-01	1.70E-13	1.90E-07
2016	4.5	1.20	1.20	0.53	11.00	6.70	6.00E-01	9.10E-13	1.40E-07
2017	4.5	1.20	1.20	0.55	10.00	7.20	5.90E-01	5.70E-12	3.30E-08
2018	4.4	1.20	1.20	0.77	9.80	7.10	4.50E-01	2.80E-11	3.90E-08

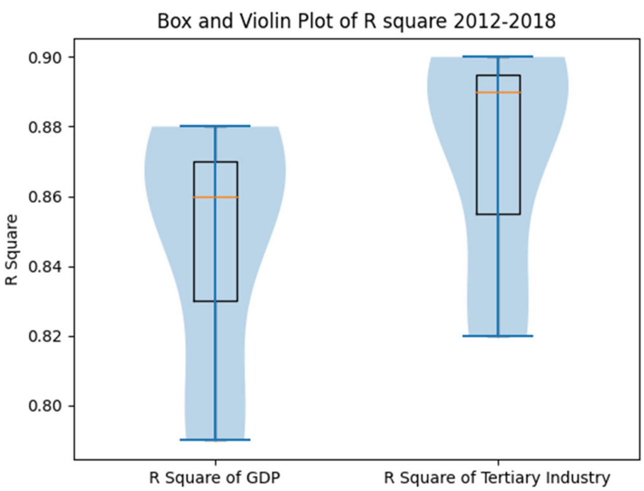

FIGURE 5 *R* square: GDP versus tertiary industry output value.

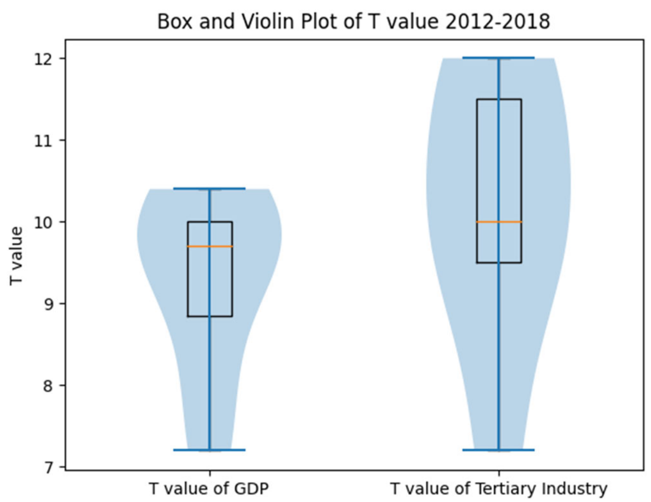

FIGURE 6 *T* value: GDP versus tertiary industry output value.

displayed distinct correlation characteristics, which are further detailed in Table 10 (for standard 1) and Table 12 (for standard 2). Not all categories could be used to generate regression models. Only the groups of cities that passed the correlation analysis, with an absolute correlation coefficient larger than 0.7, were used to generate regression models.

To provide a comprehensive overview of the distribution of cities by groups in China, Figure 7 shows the geographic distribution of the

TABLE 8 Spearman's correlation coefficients between the per thousand population blood donation and the economy indices or the education.

	Spearman's correlation coefficients between per thousand population blood donation and the corresponding variables					
Year	Average GDP	*p* value	Average tertiary industry output value	*p* value	Percentage of college students	*p* value
2012	−0.08	0.631	−0.08	0.631	0.53	0.0008
2013	−0.04	0.798	−0.04	0.798	0.58	0.0002
2014	0.02	0.899	0.02	0.899	0.61	8.20E-05
2015	0.03	0.874	0.03	0.874	0.62	6.50E-05
2016	0.04	0.799	0.04	0.799	0.6	0.00011
2017	0.07	0.678	0.07	0.678	0.66	1.10E-05
2018	−0.01	0.939	−0.01	0.939	0.66	1.40E-05

TABLE 9 Cities clustered by K-Means algorithm based on per capita tertiary industry output-to-blood donation ratio.

Groups	Corresponding cities	Counts
Group 1	Shijiazhuang, Taiyuan, Nanchang, Zhengzhou, Nanning, Hainan, Chengdu, Kunming, Xining, Yinchuan	10
Group 2	Beijing, Tianjin, Hohhot, Lhasa	4
Group 3	Shenyang, Nanjing, Hangzhou, Fuzhou, Wuhan, Guangzhou, Guiyang, Urumqi	8
Group 4	Changchun, Harbin, Hefei, Jinan, Changsha, Chongqing, Xi'an, Lanzhou	8
Group 5	Shenzhen	1
Group 6	Dalian, Shanghai, Ningbo, Xiamen, Qingdao	5

TABLE 10 The correlation and regression analysis of per capita tertiary industry output value and per thousand population blood donation.

Cities	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Correlation						
Coefficient	0.75	1.00	1.00	0.86	N	0.7
p value	0.013	0.0	0.0	0.0065	N	0.19
Regression						
Constant	6.53	−0.206	1.84	4.66	N	N
Per capita tertiary industry output value	3.26E-04	1.70E-04	2.35E-04	2.53E-04	N	N
***t* value**						
Constant	2.53	−0.165	1.57	2.06	N	N
Per capita tertiary industry output value	3.91	8.66	11.71	4.3	N	N
***p* value**						
Constant	0.04	0.88	1.67E-01	0.085	N	N
Average tertiary industry output value	0.0045	0.013	2.34E-05	0.0051	N	N
Evaluation of the model						
R square	0.66	0.97	0.96	0.76	N	N
F value	0.0045	0.013	2.34E-05	0.0051	N	N

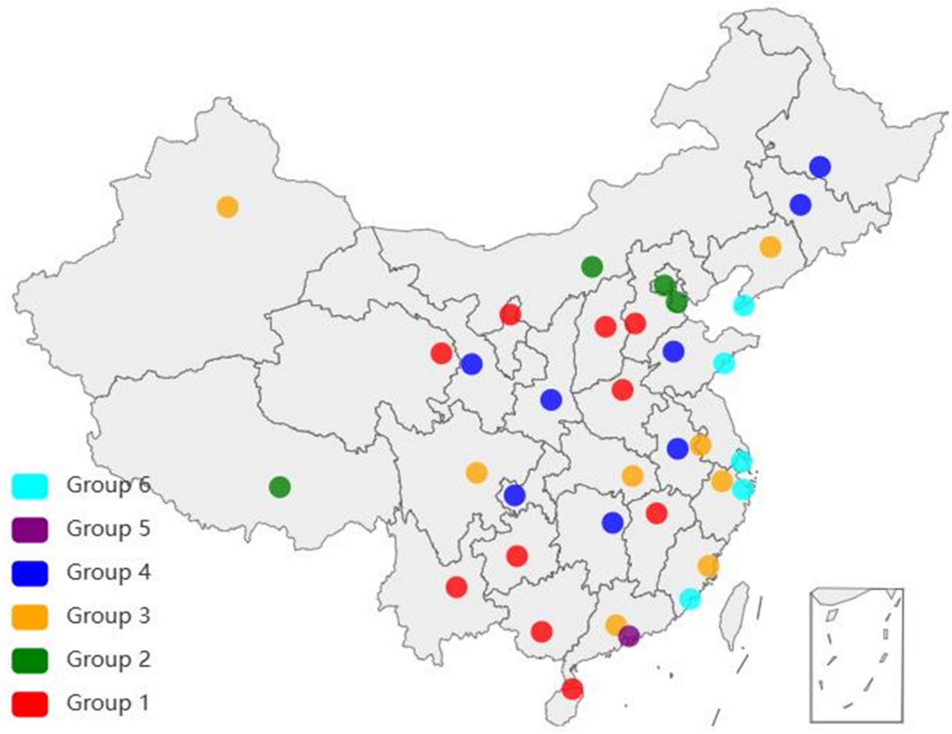

FIGURE 7 Cities clustered by K-Means algorithm based on per capita tertiary industry output-to-blood donation ratio.

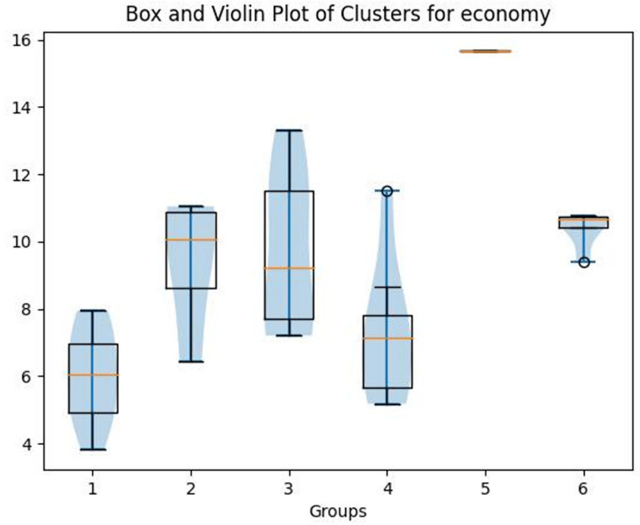

FIGURE 8 The distribution of per capita tertiary sector output values varies across city groups.

TABLE 11 Cities clustered by K-Means algorithm based on average college student population-to-blood donation ratio.

Groups	Corresponding cities	Counts
Group 1	Taiyuan, Shenyang, Changchun, Harbin, Nanjing, Hefei, Fuzhou, Zhengzhou, Changsha, Chengdu, Guiyang, Kunming, Xi'an	13
Group 2	Beijing, Shijiazhuang, Shanghai, Ningbo, Hainan, Chongqing, Xining, Yinchuan	8
Group 3	Nanchang, Jinan, Wuhan, Lhasa	4
Group 4	Hohhot, Lanzhou	2
Group 5	Tianjin, Dalian, Hangzhou, Xiamen, Qingdao, Guangzhou, Nanning, Urumqi	8
Group 6	Shenzhen	1

cities and Figure 8 shows the average tertiary industry output value of each group of cities.

Group 1 cities have the lowest per capita output values, indicating their relatively weak economic strength and underdeveloped economic levels. Geographically, Group 1 cities are separated into northern and southern parts, demonstrating little connection with their locations but potentially with historical reasons or other social factors. According to Figure 5, these cities belong to economically underdeveloped regions. Group 2 cities are mainly concentrated in a particular area except Lhasa. However, their per capita output values vary greatly due to huge gaps in economic development between Lhasa, Hohhot, Beijing and Tianjin. Groups 3 and 4 would connect geographically by excluding a few less developed cities. Group 3 has a wide range of per capita output values resulting from substantial differences in economy between cities. Group 4 sees a concentrated distribution of per capita output values. Group 5 contains only one city named Shenzhen. Unlike other city groups, it shows little or even no linear relationship with per capita blood donation frequency. As the first and most economically dynamic special economic zone in China, Shenzhen's unique status leads to its special socioeconomic characteristics different from other inland cities. Group 6 cities in China's

TABLE 12 The correlation and regression analysis of average Average number of college students and per thousand population blood donation.

Cities	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Correlation						
Coefficient	0.98	0.79	1.0	N	0.95	N
p value	1.61E-09	0.02	0.0	N	2.60E-03	N
Regression						
Constant	3.73	4.09	6.05E-02	N	−2.42	N
Average number of university students	165.90	329.22	1.65E+02	N	350.04	N
***t* value**						
Constant	4.20	1.84	0.036	N	−1.96	N
Average number of university students	13.27	4.24	8.50	N	13.91	N
***p* value**						
Constant	1.48E-03	0.115	0.97	N	9.83E-02	N
Average number of university students	4.10E-08	0.0054	0.01	N	8.59E-06	N
Evaluation of the model						
R square	0.94	0.75	0.97	N	0.97	N
F value	4.10E-08	0.0054	0.014	N	8.59E-06	N

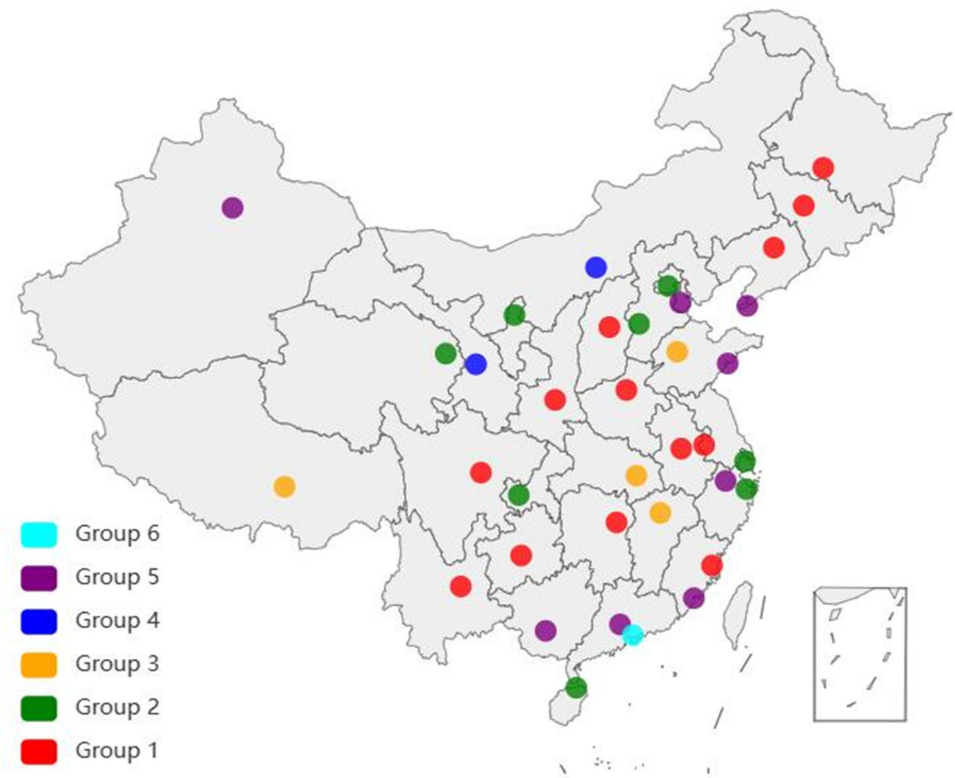

FIGURE 9 Cities clustered by K-Means algorithm based on average college student population-to-blood donation ratio.

coastal region have developed market economies and strong foreign trade, but their per capita blood donation levels are weak. There appears to be a weak linear relationship between the per capita output value of the tertiary industry and the number of blood donations per thousand people in Group 6 cities.

The city clusters identified by K-Means clustering based on the ratio of average university student population to blood donation frequency per capita differs from those identified based on the ratio of per capita tertiary industry output value to blood donation frequency per capita. The results in Tables 9 and 11 are not the same. However, the correlation coefficients between average university student population and blood donation frequency per thousand population are greater than the correlation coefficients between per capita tertiary industry output value and blood donation frequency per thousand population. Readers can compare Table 10 with Table 12. This means that the per capita tertiary

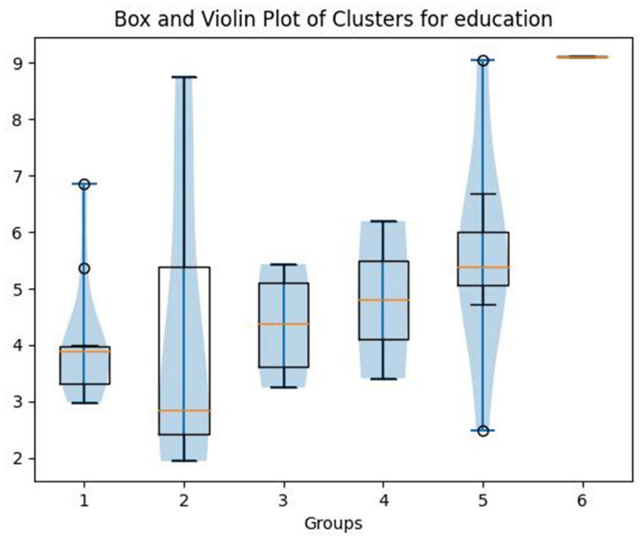

FIGURE 10 Average number of college students Distribution in each groups.

TABLE 13 Proportion of the donors in age groups.

Year	18–24	25–34	Proportion of 18–34
2012	0.416	0.328	0.744
2013	0.409	0.329	0.738
2014	0.409	0.327	0.736
2015	0.408	0.327	0.735
2016	0.414	0.316	0.730
2017	0.420	0.302	0.722
2018	0.433	0.288	0.721

TABLE 14 Spearman's correlation coefficients between the number of college students and the number of donors in each of the age groups.

Year	Donors's age groups	Spearman's correlation coefficients	*p* values
2012	18–24	0.88	1.49E-12
	25–34	0.7	2.50E-06
	35–44	0.68	5.42E-06
	45–54	0.64	2.91E-05
	55–60	0.21	0.22
2013	18–24	0.84	2.20E-10
	25–34	0.61	7.54E-05
	35–44	0.6	9.99E-05
	45–54	0.54	6.07E-04
	55–60	0.31	0.06
2014	18–24	0.9	1.55E-13
	25–34	0.68	5.24E-06
	35–44	0.68	5.48E-06
	45–54	0.62	5.10E-05
	55–60	0.53	1.01E-03
2015	18–24	0.9	7.72E-14
	25–34	0.68	4.56E-06
	35–44	0.68	5.48E-06
	45–54	0.67	6.56E-06
	55–60	0.5	1.79E-03
2016	18–24	0.91	1.99E-14
	25–34	0.69	3.14E-06
	35–44	0.67	7.17E-06
	45–54	0.67	9.22E-06
	55–60	0.38	2.14E-02
2017	18–24	0.93	3.91E-16
	25–34	0.71	1.50E-06
	35–44	0.68	5.48E-06
	45–54	0.67	6.86E-06
	55–60	0.36	3.19E-02
2018	18–24	0.93	3.69E-16
	25–34	0.69	3.25E-06
	35–44	0.66	1.04E-05
	45–54	0.69	3.93E-06
	55–60	0.36	3.22E-02

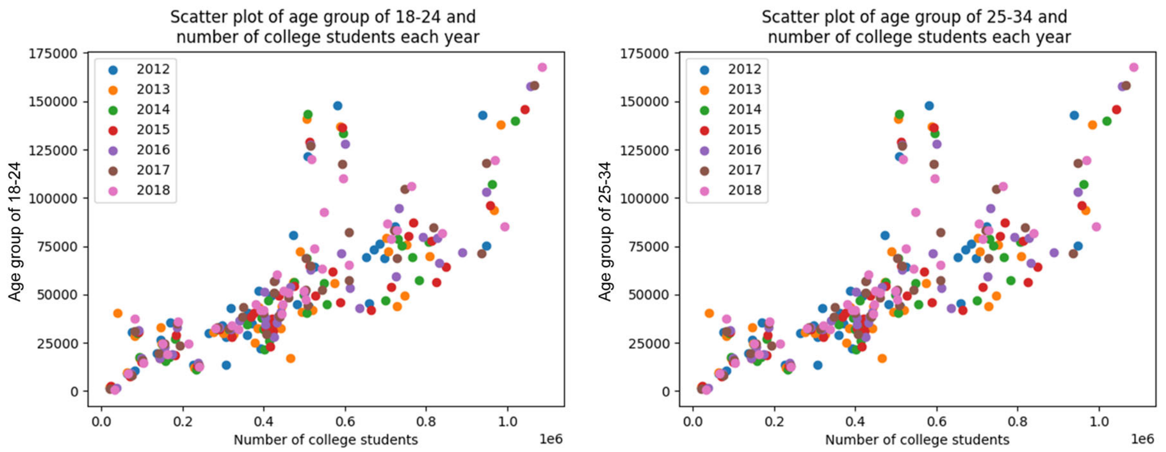

FIGURE 11 Scatter plot of number of college students and number of blood donations by age group (18–24 and 25–34).

industry output value and average college students population reflects 2 different factors that influence the blood donation. The variable of "average number of number students" has a more significant and important impact on the outcome variable of "per thousand population blood donation." This suggests that education level may be a key factor in determining blood donation rates among urban residents. Increasing public education can more directly and effectively improve their awareness and willingness to donate blood.

Figure 9 indicates that most of the selected cities are categorized into Group 1. Groups 1 and 3 have similar regression constants but different intercepts, together accounting for almost half of the selected cities. Similar to Figure 7, most of the cities in Group 5 are located in coastal areas. According to Figure 10, Groups 2 and 5 have similar distribution of average number of college students. These two groups of cities are distributed along the coastal regions (Group 5) and the boundary of the central plains (Group 2), forming a circular shape when combined. Group 6 only consists of Shenzhen, which shares a unique status similar to per capita tertiary industry output values and has distinct socioeconomic characteristics compared to other inland cities. Shenzhen is also unique in terms of the average number of college students.

Table 11 suggests that improving education can have a greater impact on blood donation. Among all education levels, college students contribute the most to blood donation. Therefore, it would be worth examining the age distribution of blood donors and its relation to the number of college students. In the next subsection, we will analyze the correlation between the age of blood donors and the number of blood donations, as well as the proportion of blood donations in different age groups to the total number of blood donations. Although this analysis is not the primary focus of our study, it provides additional insights into the blood donation behavior of different age groups and complements our overall analysis.

3.3 | Age distribution of blood donors and correlation analysis with education indicators

In China, donors are required to be 18–60 years old. According to Chinese donation databases, the greatest donors are between the ages of 18 and 34. Indeed, Table 13 shows that 70% of China's donors are between the age of 18 and 34. This cohort represents the majority of donors. Most of the college students are between the age of 18–24, and most of the postgraduate are between the age of 22–35. The two age groups contributes to the most of the donors.

The correlation coefficients for the numbers of college students and the numbers of donors in each of the age groups are given in Table 14. Table 14 shows that the correlation coefficients for the donors in each age group with the number of college students demonstrate a linear relationship.

Consulting Tables 13 and 14, our findings suggest that the age group of 18–34 is the most active group of blood donors in China. This group overlaps significantly with the age range of college students and postgraduates, who are also found to be the primary donors. This highlights the importance of targeting educational institutions for blood donation campaigns and promoting blood donation among students. To emphasize the key points, we present a scatter plot showing the number of blood donations by age group (18–24 and 25–34) against the number of college students which are shown in Figure 11.

Furthermore, our analysis shows that there is a strong linear relationship between the number of college students and blood donation (Table 1). This implies that increasing the number of college students in a region could lead to an increase in the number of blood donors. Therefore, investing in education and expanding access to higher education could be an effective way to increase the blood donation rate in China.

4 | CONCLUSION

Through conducting a multiregression analysis of 36 cities, our study have investigated the impact of socioeconomic factors on blood donation rate from the perspectives of city level, per capita level and population structure.

At the city level, overall economic output and university student populations were positively correlated with total blood donor figures, indicating a positive relationship between education, economy and blood donation rate. However, at the per capita level, this linear relationship disappeared, suggesting that a city's blood donation times, tertiary industry output and student population were closely related to its demographic structure.

When we applied the K-Means clustering algorithm, cities could be divided into subgroups where the linear relationship reemerged. This indicates that cities differ in characteristics, and a one-size-fits-all policy may have limited effect. Tailored policy interventions are needed to suit different city types.

The results show that while development in economy and education was associated with higher blood donation rate at the city level, the impact of these factors on per capita participation varied in different city clusters. Factors influencing blood donation are complex, depending on a city's demographic composition and characteristics.

Anyway, the impact of economic factors is less significant compared with education factors. This indicates that education level, especially university education, has a more substantial effect on blood donation behavior, particularly among the youth. Policy makers should consider increasing investment in education and conducting targeted education on blood donation knowledge and social responsibility. Policy measures should focus on increasing education investment and conducting targeted education. Economic means should be combined with other social policies as its effect may be limited if adopted alone.

AUTHOR CONTRIBUTIONS

Feng Lin: conceptualization; formal analysis; funding acquisition; methodology; writing—original draft; writing—review & editing. **Yi Huang**: conceptualization; data curation; funding acquisition; project administration; resources; writing—review & editing. **Xu He**: data curation; resources. **Zhong Liu**: funding acquisition; resources. All authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We would like to express our appreciation to the anonymous reviewers for their thoughtful comments and suggestions on this manuscript. This work was supported by Fundamental Research Funds for the Central Universities (3332019170).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The economic education data used in this study can be accessed from the Annual Statistical Review (http://www.stats.gov.cn/sj/ndsj/), which is publicly available. Part of the blood donation data used in this study can be accessed from the National Annual Blood Safety Report, which is available in print format. Due to the unavailability of an online version of the report, interested researchers can contact the corresponding author for a copy of the relevant sections. Yi Huang and Zhong Liu had full access to all of the data in this study and takes complete responsibility for the integrity of the data and the accuracy of the data analysis.

ETHICS STATEMENT

This study was conducted in accordance with the ethical guidelines of Institute of blood transfusion Chinese academy of medical science and Peking union medical college and complied with all relevant laws and regulations. The study did not involve human or animal participants, and the data used in this study were obtained in compliance with data protection laws and regulations. All data were treated with strict confidentiality and privacy protection measures.

TRANSPARENCY STATEMENT

The lead author Zhong Liu affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

ORCID

Feng Lin http://orcid.org/0000-0002-9976-8030
Yi Huang http://orcid.org/0000-0001-7821-9692
Xu He http://orcid.org/0000-0001-9305-7468
Zhong Liu http://orcid.org/0000-0001-8173-9656

REFERENCES

1. WHO Expert Group. Expert Consensus Statement on achieving self-sufficiency in safe blood and blood products, based on voluntary non-remunerated blood donation (VNRBD)*: self-sufficiency in blood and blood products based on VNRBD. *Vox Sang*. 2012;103(4):337-342.
2. https://www.who.int/news-room/fact-sheets/detail/blood-safety-and-availability
3. Cruz JR, Pérez-Rosales MD. Availability, safety, and quality of blood for transfusion in the americas. *Revista Panamericana de Salud Pública*. 2003;13:103-110.
4. Wang ZF, Li JB, Zhang SQ, Wang ZD. Red cross blood center Henan. Discussion on the correlation between voluntary blood donation and economic development in zhengzhou area. *Occup Health*. 2014;19:2014.
5. Slonim R, Wang C, Garbarino E. The market for blood. *J Econ Perspect*. 28(2). Spring; 2014:177-196.
6. Zhu YM. Ensure blood supply, while demands are suppressed. *ISBT Science Series*. 2007;2(2):147-149.
7. World Health Organization. The 2016 global status report on blood safety and availability. 2017: http://apps.who.int/iris/bitstream/10665/254987/1/9789241565431-eng.pdf
8. Mirutse G, Fisseha G, Abebe L, Birhanu Z, Alemayehu M. Intention to Donate Blood among the Eligible Population in Mekelle City.

Northern Ethiopia: Using the Theory of Planned Behavior; *Am J Health Res*, 2014:2(4). 158-163. doi:10.11648/j.ajhr.20140204.19

9. Jemberu YA, Esmael A, Ahmed KY. Knowledge, attitude and practice towards blood donation and associated factors among adults in Debre Markos town, Northwest Ethiopia. *BMC Hematol*. 2016;16:23. doi:10.1186/s12878-016-0062-8
10. Melku M, Terefe B, Asrie F, et al. Knowledge, Attitude, and Practice of Adult Population towards Blood Donation in Gondar Town, Northwest Ethiopia: a Community Based Cross-Sectional Study. *J Blood Transfus*. 2016, 2016:7949862. doi:10.1155/2016/7949862
11. Jibril AD, Mohammed A, Hashi A, et al. Knowledge, attitude, practice and associated factors of adult population towards blood donation in Jigjiga Town, Somali Region, Ethiopia. *J Drug Delivery Ther*. 2020;10(6):149-165.
12. Tarlabelen Karaytuğ E, Boyraz S. Knowledge and attitudes of apheresis donors regarding apheresis blood donation. *Transfus Med*. 2022;33:39-48. doi:10.1111/tme.12885
13. Gao L, Wang Q. Survey on knowledge, attitude and practice about blood donation among continuing medical education (cme) students in sichuan province, China. *Transfus Apher Sci*. 2017;56(3):454-458.
14. Charles KS, Friday M, Lall D, et al. A university–led initiative to promote voluntary non-remunerated blood donation in a developing country. *Transfus Apher Sci*. 2019;58(5):674-679.
15. Suen LKP, Siu JY, Lee YM, Chan EA. Knowledge level and motivation of Hong Kong young adults towards blood donation: a cross-sectional survey. *BMJ Open*. 2020;10(1):e031865.
16. Yang BC, Shao CP, Zhang YY, Kong L, Xu YP. Two decades of voluntary nonremunerated blood donation in Shenzhen, China. *Transfusion*. 2015;55(5):1108-1114.
17. Shi L, Wang JX, Stevens L, Ness P, Shan H. Blood safety and availability: continuing challenges in China's blood banking system. *Transfusion*. 2014;54(2):471-482.
18. Pearson K. VII. Note on regression and inheritance in the case of two parents. *Proc R Soc London*. 1895;58:240-242.
19. Cohen J, Cohen P, West SG, Aiken LS. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd ed. Erlbaum; 2003.
20. Field A. *Discovering statistics using IBM SPSS statistics*. 5th ed. Sage Publications; 2018.
21. Spearman C. The proof and measurement of association between two things. *Am J Psychol*. 1904;15(1):72-101.
22. Quatraro F. *The economics of structural change in knowledge*. Routledge; 2012. https://www.jstor.org/stable/1885848?seq=1
23. Ngai LR, Pissarides CA. Structural change in a multisector model of growth. *American Economic Review*. 2007;97(1):429-443.
24. Neter J, Kutner MH, Nachtsheim CJ, Wasserman W. *Applied Linear Statistical Models*. 4th ed. McGraw-Hill Education; 2004.
25. Draper NR, Smith H. *Applied regression analysis[M]*. John Wiley & Sons; 1998.
26. Fox J. *Applied Regression Analysis and Generalized Linear Models*. Sage Publications; 2015. ISBN: 9781452205663. doi:10.4135/9781452205663
27. Hair JF, Black WC, Babin BJ, Anderson RE. *Multivariate Data Analysis*. 8th ed. Cengage; 2019:976. ISBN: 978-1-337-55833-2.
28. MacQueen J. Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. 1967;1(14):281-297.
29. Hastie T, Tibshirani R, Friedman J *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media; 2009. 745 p. ISBN: 978-3-642-05760-5.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Lin F, Huang Y, He X, Liu Z. Research on the socioeconomic factors that influence the development of voluntary, non-remunerated blood donation in China—A correlation and regression analysis based on data from 2012 to 2018. *Health Sci Rep*. 2023;6:e1341. doi:10.1002/hsr2.1341