
Xiao Zhang, Yuwen Tang, Beibei Zhang, Yongjing Zhang, Jifeng Dai, Junhui Zhang

ABSTRACT

Objective The present study investigated the province-level distribution and drivers of infant mortality rate (IMR) in mainland China.

Design Ecological analysis based on publicly available data for all 31 provinces in mainland China.

Data sources Data on province-level IMRs in 2020 were obtained from the official websites of the healthcare commissions within each province and from the China Health Statistics Yearbook 2021. Data on potential IMR drivers were retrieved from the China Statistical Yearbook 2021.

Data analysis GeoDa V.1.12.1 and ArcMap V.10.2 software were used to examine province-level distribution of IMR. Global and local spatial autocorrelations were performed, and Getis-ord G* hotspots and coldspots were identified. Geodetector was used to analyse the individual and joint influence of drivers on IMR.

Results IMRs in 2020 varied from 1.91 to 7.60 per 1000 live births across provinces. The following statistically significant drivers with q values >0.5 were identified: health literacy of the population (0.6673), male illiteracy rate (0.6433), proportion of the population older than >65 years (0.6369), per capita government health expenditure (0.6216), forest coverage rate (0.5820), per capita disposable income (0.5785), per capita number of hospitals (0.5592), per capita gross regional product (0.5410) and sulfur dioxide concentration in the atmosphere (0.5158). The following three interactions among these drivers emerged as strongest influences on province-level IMR: proportion of population >65 years ∩ per capita gross regional product (q=0.9653), forest coverage rate ∩ per capita gross regional product (0.9610) and per capita government health expenditure ∩ sulfur dioxide (0.9295).

Conclusion IMR in mainland China varies substantially across the country, being generally high-west and low-east. Several factors, on their own and interacting together, contribute to IMR. Policies and programmes to reduce IMR should be formulated according to local conditions and should focus on western provinces of the country.

INTRODUCTION

The infant mortality rate (IMR), usually defined as the number of infants who die before the age of 1 year per 1000 live births (%) in a given year, is often taken as an important economic and developmental indicator of a country or region. One of the United Nation’s Sustainable Development Goals is for IMR to fall below 12‰ by 2030. Many countries classified by the World Bank as having high or upper-middle national incomes have already achieved this target, but numerous developing countries in South Asia and Africa have not. To reduce its IMR, China announced the ‘Health China 2030’ initiative to reduce IMR below 5.0‰ by 2030, and the government has called for reducing geographical disparities in IMR. So far, China has succeeded in reducing its IMR to 5.4‰, and it now ranks at the top of middle-income and high-income countries globally. However, strong geographical variations in IMR remain, reflecting disparities in socioeconomic development and healthcare infrastructure, highlighting the need to

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study used Geodetector to analyse drivers of infant mortality rate (IMR), and their interactions with each other, across all 31 provinces in mainland China.
⇒ Our analysis provides concrete insights relevant to China as well as other countries struggling with maternal and child health disparities.
⇒ Our method may prove useful more generally for understanding health indicators at greater geographic resolution.
⇒ The analysis was restricted to a single year (2020), so the dynamics of IMR and its drivers over time were not analysed.
⇒ This study was ecological, so the findings cannot necessarily be extrapolated down to the level of smaller geopolitical or administrative units such as cities or counties.
identify where IMR reduction efforts and policies should focus.

Numerous studies have examined the factors influencing IMR in China. These studies have identified key factors such as the economics, health resources, education level of the population, health and living standards, and the natural environment. However, most of these studies used classical linear regression or spatial regression, which overlooked potential complex interactions among these factors and assumed no multicollinearity among independent variables. In contrast, Geodetector in spatial econometric models can effectively handle this complexity, which can analyse the independent effects of each factor and their interactions, without the need to consider multicollinearity issues. We are unaware of studies using Geodetector to analyse IMR in China.

The present study used Geodetector and publicly available data from 2020 to analyse the province-level distribution of IMR across mainland China and its drivers.

METHODS
Study design and data sources
This ecological study divided all 31 provinces of mainland China, which include municipalities and autonomous regions, into eastern, central and western parts according to the regional classification criteria of the Chinese National Bureau of Statistics, which are based on geographical location and level of economic development. IMR data for all 31 provinces for 2020 were obtained from the official websites of the healthcare commissions of each province and from the China Health Statistics Yearbook 2021 (www.stats.gov.cn/tjsj/ndsj/2021/indexeh.htm). Provincial-level data on 27 factors that might influence IMR were also taken from the China Statistical Yearbook 2021. These factors were based on previous work and on the data available to us (table 1).

Data analysis
Data for all factors/indicators were normalised using SPSS V.26.0 (IBM, Chicago, IL, USA) and rasterised using Arcmap V.10.2 (ESRI, Redlands, CA, USA). The spatial distribution of IMR was mapped using Arcmap V.10.2. Spatial autocorrelation analysis using Geoda V.1.12.1 (Luc Ancelin at UChicago) was performed to determine whether province-level IMR was dispersed, clustered or randomly distributed based on respective values of Global Moran’s I. The statistical significance of index values was checked using the Z-test. For hot spots, confidence levels were defined as follows: 99% confidence indicates a z-score greater than 2.58, p<0.01; 95% confidence indicates a z-score between 1.96 and 2.58, p<0.05; and 90% confidence indicates a z-score between 1.65 and 1.95, p<0.10. For cold spots, confidence levels were defined as follows: 99% confidence indicates a z-score less than −2.58, p<0.01; 95% confidence indicates a z-score between −1.96 and −1.65, p<0.05; and 90% confidence indicates a z-score between −1.95 and −1.65, p<0.10.

We used the ‘factor detector’ module within Geodetector, in order to analyse the extent to which certain factors (table 1) explained the spatial heterogeneity of IMR. We quantified the influence q of each factor on IMR using the equation:

\[ q = 1 - \frac{\sum_{h=1}^{L} N_h \bar{Y}_h^2}{\sigma^2} = 1 - \frac{SSW}{SST} \] (1)

where

\[ SSW = \sum_{h=1}^{L} N_h \bar{Y}_h^2 \]  \[ SST = \sigma^2 \] (2)

and h=1, 2, ..., L denotes the strata of IMR or its drivers, N_h is the number of units in stratum h, N is the number of units in the whole region, SSW is the sum of variance within the stratum and SST is the total variance in the whole region. A larger q value for a given driver indicates stronger power to explain the observed spatial distribution of IMR. The statistical significance of q was determined by calculating the non-central F distribution:

\[ F = \frac{N-1-q}{1-q} \sim F(L - 1, N - L; \lambda) \] (3)

where λ is the non-central parameter defined as:

\[ \lambda = \frac{1}{\sigma^2} \left[ \sum_{h=1}^{L} \frac{\bar{Y}_h^2}{N} - \frac{1}{N} \sum_{h=1}^{L} \bar{Y}_h \right]^2 \] (4)

and \( \bar{Y}_h \) is the mean of layer h.

We also used the ‘interaction detector’ within the software to analyse the extent to which two-way interactions among the factors (table 2) explained the spatial heterogeneity of IMR.

RESULTS
In 2020, provincial IMRs varied from 1.91‰ to 7.60‰, with a median of 3.40‰ (figure 1A). In general, IMRs were higher in western provinces than in eastern ones. The three provinces with the highest IMRs were Tibet (7.60‰), Qinghai (7.01‰) and Xinjiang (6.75‰); the three with the smallest IMRs were Guangdong (2.13‰), Beijing (1.98‰) and Zhejiang (1.91‰).

Among the 27 potential IMR drivers that we considered, male illiteracy rate (figure 1B) and per capita government health expenditures (figure 1C) were higher in western provinces and lower in eastern ones, similar to the spatial

Clusters of provinces with high IMR (‘hotspots’) or low IMR (‘coldspots’) were identified using Getis-ord G* index in Arcmap V.10.2. The significance of index values was checked using the Z-test. For hot spots, confidence levels were defined as follows: 99% confidence indicates a z-score greater than 2.58, p<0.01; 95% confidence indicates a z-score between 1.96 and 2.58, p<0.05; and 90% confidence indicates a z-score between 1.65 and 1.95, p<0.10. For cold spots, confidence levels were defined as follows: 99% confidence indicates a z-score less than −2.58, p<0.01; 95% confidence indicates a z-score between −1.96 and −1.65, p<0.05; and 90% confidence indicates a z-score between −1.95 and −1.65, p<0.10.
distribution of IMR. Conversely, health literacy of the population (figure 1D) and per capita gross regional product (figure 1E) were lower in western provinces and higher in eastern ones. The remaining potential drivers showed spatial distributions different from that of IMR. Figure 1 displays the spatial distribution maps of the nine drivers with q values >0.5 from factor detector, while maps of the remaining drivers can be found in online supplemental figure 1A–R.

Spatial autocorrelation analysis indicated that IMR differed significantly across the 31 provinces (Global Moran’s I=0.5090, p=0.001), Local Moran’s I spatial autocorrelation analysis showed that there was spatial clustering of IMR in 8 of the 31 provinces in 2020 (figure 2). So-called ‘high-high’ areas, where a given province as well

| Table 1 | Factors analysed in this study as potential drivers of infant mortality in mainland China |
| --- | --- | --- | --- |
| Dimension | Factor/indicator | Unit | Variable ID |
| Economy | Per capita gross regional product (Per.GRP) | RMB | X1 |
| | Per capita disposable income (Per.DI) | RMB | X2 |
| | Proportion of urban population at year-end (POUP) | % | X3 |
| | Expenditure for healthcare by government (EHCG) | Billion RMB | X4 |
| | Per capita government health expenditure (Per.GHE) | RMB | X5 |
| | Percentage of health expenditure (Pct.HE)* | % | X6 |
| Health resource | Per capita number of healthcare institutions (Per.HCI) | No. per 1000 people | X7 |
| | Per capita number of hospitals (Per.Hosp) | No. per 1000 people | X8 |
| | Per capita number of beds in healthcare institutions (Per.BHCI) | No. per 1000 people | X9 |
| | Per capita number of beds in hospitals (Per.BHosp) | No. per 1000 people | X10 |
| | Per capita number of health technical personnel in healthcare institutions (Per.HTP) | No. per 1000 people | X11 |
| Education | Average years of schooling in the population (AYS) | years | X12 |
| | Percentage of total population at least 15 years old that is illiterate (Pct.IPT) | % | X13 |
| | Male illiteracy rate (MIR)† | % | X14 |
| | Female illiteracy rate (FIR)‡ | % | X15 |
| Social welfare | Urban Engel’s coefficient (UEC)§ | NA | X16 |
| | Rural Engel’s coefficient (REC) | NA | X17 |
| | Per capita healthcare and medical services expenditure (Per.HCMSE) | RMB | X18 |
| | Percentage of gross domestic product spent on healthcare (Pct.GDPsh) | % | X19 |
| | Participation rate of basic medical insurance (Par.BasMI)¶ | % | X20 |
| | Health literacy of the population (HLP) | % | X21 |
| Natural environment | Forest coverage rate (FCR) | % | X22 |
| | Sulfur dioxide (SO2) 10 000 tons | X23 |
| | Nitrogen oxides (NOx) 10 000 tons | X24 |
| | Particulate matter (PM) 10 000 tons | X25 |
| Population | Sex ratio, where all female=100% | % | X26 |
| | Percentage of the population that is older than >65 years (Pct.EPOP) | % | X27 |

*The ratio of healthcare expenditures to the size of public expenditures.†The rate of illiteracy among males at least 15 years old.‡The rate of illiteracy among females at least 15 years old.§The proportion of per capita income that is spent on food.¶The percentage of total per capita consumption expenditure that is spent on healthcare.

| Table 2 | Two-way interactions between potential drivers of infant mortality rate |
| --- | --- | --- | --- |
| Interaction | Strength | Type of modelling |
| q(X_i∩X_j)<Min(q(X_i), q(X_j)) | Weak | Non-linear |
| Min(q(X_i), q(X_j))<q(X_i∩X_j)<Max(q(X_i), q(X_j)) | Weak | Univariate |
| q(X_i∩X_j)>Max(q(X_i), q(X_j)) | Strong | Bivariate |
| q(X_i∩X_j)=q(X_i)+q(X_j) | Independent |
| q(X_i∩X_j)>q(X_i)+q(X_j) | Strong | Non-linear |

X_i and X_j refers to the variables in table 1.
Figure 1  Provincial distributions of infant mortality rates (IMRs) and their potential drivers in mainland China in 2020. (A) IMR. (B) Male illiteracy rate (MIR). (C) Per capita government health expenditure (Per.GHE). (D) Health literacy of the population (HLP). (E) Per capita gross regional product (Per.GRP). (F) Percentage of the population that is older than >65 years (Pct.EPOP). (G) Per capita disposable income (Per.DI). (H) Per capita number of hospitals (Per. Hosp). (I) Forest coverage rate (FCR). (J) Sulfur dioxide (SO2).
as its neighbouring provinces had high IMR, were distributed in the western areas of Xinjiang, Qinghai, Tibet, Sichuan and Yunnan. Shanghai was a so-called ‘low-low’ area, where the province and its neighbours had low IMR. Jiangxi was a ‘high-low’ outlier area, where the province had a high IMR and its neighbours had low IMR. Gansu was a ‘low-high’ outlier area where the province had a low IMR and its neighbours had high IMR. IMR hotspots (99% confidence) were located in the west (Xinjiang, Qinghai and Tibet), while coldspots were located in the center-east region. The coldspots (99% confidence) included Hubei, Anhui and Fujian; the coldspots (95% confidence) included Zhejiang, Jiangsu, Jiangxi, Tianjin, Henan and Shandong (figure 3).

The most significant drivers (q>0.5) of IMR in 2020 were health literacy of the population (q=0.6673), male illiteracy rate (0.6433), proportion of the population older than >65 years (0.6369), per capita government health expenditure (0.6216), forest coverage rate (0.5820), per capita disposable income (0.5785), per capita number of hospitals (0.5592), per capita gross regional product (0.5410) and sulfur dioxide concentration in the atmosphere (0.5158) (online supplemental table 1). Most of these drivers mutually reinforced one another, with the strongest interactions between proportion of the population older than >65 years and per capita gross regional product (q=0.9653), forest coverage rate and per capita gross regional product (0.9610) and per capita government health expenditure and SO2 (0.9295).

**DISCUSSION**

This study based on data from 2020, perhaps the most detailed so far of province-level IMR and its potential drivers in mainland China, indicates that IMR varies from 1.91‰ to 7.60‰ across all 31 provinces. Our finding of higher IMR in the west than in the east of the country is consistent with a previous study based on data from 2010.24 Moreover, we found high-high clustering in Qinghai, Tibet, Yunnan and Sichuan, while we found low-low clustering in Shanghai. IMR hotspots were located in the west, while coldspots were located in the center-east. These geographic variations may reflect the spatial distribution of socioeconomic characteristics such as economic development, health resource allocation, educational level of the population and forest coverage. Notably,
some of these factors showed similar or opposite spatial patterns to that of IMR, whereas the remaining potential drivers showed spatial distributions different from those observed in IMR.

Our study appears to be the first to investigate not only individual but also joint effects of drivers on IMR in China. Factor detector revealed that all 27 drivers had statistical significance \( (p<0.01) \). Among them, we will specifically discuss the following nine drivers with \( q \) values greater than 0.5: health literacy of the population, male illiteracy rate, percentage of the population that is older than >65 years, per capita government health expenditure, forest coverage rate, per capita disposable income, per capita number of hospitals and sulfur dioxide concentration in the atmosphere. These geographic variations may reflect the spatial distribution of socioeconomic characteristics including economic development, health resource allocation, educational level of the population and forest coverage.

We found that the most influential driver of IMR may be health literacy of the population, which has been shown to correlate negatively with IMR in China.\(^{23}\) This could be due to the fact that individuals with low health literacy lack basic hygiene knowledge and health skills, which hinders their ability to effectively access and use the health information and services provided by the government.\(^{23}\) The second most influential driver of IMR in our study was the male illiteracy rate, which varied inversely with IMR. This may reflect that more educated men may be more likely than less educated men to provide appropriate care and monitoring to pregnant women and newborns, as well as to behave in ways that protect or increase maternal or fetal health.\(^{18}\) The third most influential driver in our study was percentage of the population that is older than >65 years, which also varied inversely with IMR. We attribute the influence of this driver to its positive association with overall levels of health resources and economic development.\(^{26}\) Indeed, general health among elderly in China is lower in the west than in the east,\(^{27}\) similar to IMR.

The observed negative associations of IMR with per capita disposable income and per capita gross regional product suggest that risk of infant mortality varies inversely with the level of socioeconomic development. In fact, previous work has suggested that economic development may exert a stronger influence on IMR in poorer regions than in wealthier ones.\(^{28}\) Wealthier, more developed provinces may be more likely to provide adequate maternal and child health services,\(^{10}\) and their populations may have higher levels of education\(^{29}\) and literacy,\(^{30}\) helping them foster supportive, safe environments for mothers and newborns.\(^{16,14}\)

Per capita government health expenditure theoretically can provide more healthcare services and improve healthcare infrastructure, thereby reducing IMR. However, strikingly, this driver in our study was higher in the west and lower in the east, yet IMR showed the similar trend. This suggests that per capita government health expenditure plays a relatively minor role in reducing IMR compared with other drivers, possibly because health expenditures in western regions are focused on factors other than maternal and child health. Additionally, the western regions may be hindered by factors such as economic underdevelopment, adverse natural environment, inconvenient transportation, poor accessibility to medical...
resources, shortage of healthcare professionals and lower educational level compared with the eastern regions. The per capita number of hospitals showed spatial distributions different from that of IMR. Notably, Tibet has the highest per capita number of hospitals but also high IMR, it may be that most hospitals are primary care centres and township health centres, with few hospitals specialised in women and children. Therefore, it is evident that relying solely on per capita government health expenditure and per capita number of hospitals may not completely solve the IMR issue.  

We found that forest coverage rate was negatively associated with provincial IMR, which we attribute to the association between environmental quality and personal health. This may also explain our observation of a positive association between SO2 levels and IMR, consistent with a study in Wuhan, China. In fact, a study of 175 Chinese cities found that during the period 1991–2000, reduction in industrial emissions of SO2 was associated with a 20% reduction in IMR. Consistent with these observations, we found that forest coverage rate and per capita gross regional product strongly interacted to jointly influence IMR in our study, as did per capita government health expenditure and SO2 levels.

Among the drivers of IMR and their interactions, the strongest interaction was between percentage of the population that is older than >65 years and per capita gross regional product. Percentage of the population that is older than >65 years indirectly reflects economic level: larger percentage of the population that is older than >65 years means greater government expenditure, including in healthcare. At the same time, regions with higher per capita gross regional product may also tend to have higher government health expenditures. Thus, the two indicators interact to affect IMR jointly. The second strongest interaction was between forest coverage rate and per capita gross regional product. Per capita gross regional product is an important indicator of regional economic level, while forest coverage rate is an important indicator of both regional ecological environment and regional economic level. Thus, the interaction of these two factors essentially reflects the economic level and air quality of a region. As per capita gross regional product increases, so does forest coverage rate, contributing to better air quality and potentially lower IMR. The third strongest interaction was between per capita government health expenditure and SO2. Per capita government health expenditure reflects the level of economic development, while SO2 reflects air quality and, indirectly, the level of economic development. Therefore, the interaction of these two factors also indirectly reflects the impact of economic level on IMR. The joint influence of percentage of the population that is older than >65 years and per capita gross regional product on IMR in our study probably reflects that both variables are positively associated with government expenditure in general and, by extension, in healthcare.

While our study provides numerous insights for researchers and policymakers about IMR and how it is influenced individually and jointly by factors for which public data are readily available, our findings should be interpreted with caution in light of several limitations. This study examined only the provincial level, so our results should be verified and deepened through analyses on municipal or county scales. Our data came from only 1 year, so future work should examine changes in IMR and its drivers over time in order to ensure that future research and policy align with ongoing changes in society. In addition, this study was ecological, so the findings cannot necessarily be extrapolated down to the level of smaller geopolitical or administrative units such as cities or counties.

CONCLUSION
Our analysis across all provinces of mainland China suggests that several factors related to the economy, education and healthcare coverage have acted individually and jointly to lead to low IMR in eastern regions of the country and high IMR in western regions. In order to narrow this east-west gap and continue to reduce IMR, the country should formulate policies according to local conditions, with a focus on western regions. Policies should support, in parallel, economic development, education and health literacy of the population, per capita government expenditure on health, access to health resources and environmental quality.

Acknowledgements We thank the provincial health commissions and the National Bureau of Statistics for providing access to the data used in this study.

Contributors Study conceptualisation: XZ, YT and JZ. Data collection and analysis: XZ, YT, JD and YZ. Manuscript writing: XZ, YT, BZ and JZ. Manuscript review and editing: XZ, YT, BZ and JZ. Supervision and funding acquisition: JZ. All authors contributed to revisions of the paper. JZ, as the guarantor, accepts full responsibility for the finished work and the conduct of the study, had access to the data, and controlled the decision to publish.

Funding This study was supported by grants from the National Statistical Science Research Programme of China (2021LZ31), the Scientific Research Programme of Southwest Medical University (2021KMS002) and the Innovation and Entrepreneurship Training Programme for University Students in Southwest Medical University (2021-308, 2022-366).

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Permission for data access was obtained from the provincial health commissions and the National Bureau of Statistics through an online request to the China Statistical Yearbook 2021 (www.stats.gov.cn/tjsj/ndsj/2021/indexen.htm). The data in this study are publicly available without any personal identifying information.
REFERENCES


7. Analysis and prediction of urban and rural maternal and child health levels based on GM (1, 1) model. *AAM* 2021;10:2511–8.


