

Mismatch between Urban Eco-sanitation Improvement and Health Service Supply Capacity: Evidence from Chinese City Panel Data and Interpretable Machine Learning

Zhen Peng

pengzhen@qut.edu.cn

Qingdao University of Technology

Lin Wang

Qingdao University of Technology

Research Article

Keywords: health service supply capacity, healthcare resource allocation, urban eco-sanitation environment, health service governance, Chinese cities, interpretable machine learning

Posted Date: July 2nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10075851/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Health Services Research on August 25th, 2026. See the published version at <https://doi.org/10.1186/s12913-026-15444-8>.

Mismatch between Urban Eco-sanitation Improvement and Health Service Supply Capacity: Evidence from Chinese City Panel Data and Interpretable Machine Learning

Zhen Peng^{*a} and Lin Wang^a

^a School of Architecture and Urban Planning, Qingdao University of Technology, Qingdao, 266033, China; pengzhen@qut.edu.cn; wanglinar@qut.edu.cn

* Correspondence: pengzhen@qut.edu.cn

Abstract

Background: Health service supply capacity depends not only on the stock of medical resources, but also on urban eco-sanitation conditions, public infrastructure, and resource-allocation mechanisms. Existing studies have mainly examined the effects of environmental exposure on health outcomes, whereas limited attention has been paid, from a health services research perspective, to whether improvements in the eco-sanitation environment can be converted into urban health service supply capacity.

Methods: This study constructed a panel dataset of Chinese prefecture-level and above cities for 2002-2024. The urban health service supply capacity index was used as the dependent variable, and the eco-sanitation environment index was used as the core explanatory variable. Two-way fixed effects models, dimension-specific analysis, mechanism tests, moderation analysis, heterogeneity analysis, robustness checks, and interpretable machine learning were combined to identify mismatch mechanisms and boundary conditions in the conversion from eco-sanitation improvement to health service supply capacity.

Results: The results show that the eco-sanitation environment, green infrastructure, and municipal sanitation facilities in Chinese cities improved continuously, whereas health service supply capacity increased more slowly. The baseline model does not provide evidence that eco-sanitation improvement can be stably converted into stronger health service supply capacity. Dimension-specific results indicate differentiated roles of green infrastructure, municipal sanitation facilities, and pollution pressure; pollution pressure mainly reflects a resource-allocation response rather than an environmental benefit. Mechanism tests show that urban construction investment and public fiscal capacity have not formed a stable transmission chain. Heterogeneity and interpretable machine learning results further suggest that this relationship is jointly shaped by the stage of urban development, population size, fiscal capacity, the energy-carbon context, and nonlinear boundaries.

Conclusions: Eco-sanitation improvement should not be regarded as a sufficient condition for enhancing health service supply capacity. Healthy city development should integrate eco-sanitation governance, municipal infrastructure investment, primary care network planning, health workforce allocation, and public fiscal arrangements into a unified health service governance framework, so that environmental improvement can be translated into real health service supply capacity.

Keywords: health service supply capacity; healthcare resource allocation; urban eco-sanitation environment; health service governance; Chinese cities; interpretable machine learning

1 Background

Urbanization is reshaping the distribution of population, resources, environmental risks, and public services. Urban agglomeration improves the efficiency of medical care, employment, and infrastructure

provision, but it also intensifies air pollution, energy consumption, heat exposure, solid waste, wastewater treatment pressures, and public health service demand. Previous studies have shown that urbanization changes not only health exposure and chronic disease risks, but also healthcare resource allocation, service accessibility, and the structure of public health demand [1-3]. Urban health, therefore, is no longer only an issue within medical institutions; it is the joint outcome of the built environment, ecological infrastructure, municipal sanitation conditions, pollution exposure, fiscal capacity, and the health service system.

Research on high-quality health systems emphasizes that health improvement depends not only on the quantity of medical resources, but also on whether these resources can be effectively organized, made accessible, and converted into real service capacity [4]. Donabedian's structure-process-outcome framework and the ecology of medical care further indicate that health service capacity should be understood in relation to resource structure, service processes, and the broader social environment [5,6]. Accordingly, this study defines urban health service supply capacity as the ability of a city, under given population size, eco-sanitation conditions, and public infrastructure constraints, to allocate, support, and continuously provide basic health service resources. The central issue is not the mere stock of medical resources, but whether resources can be effectively allocated and transformed within the urban governance system.

Existing health services research has explained health service capacity from the perspectives of medical utilization, accessibility, and resource allocation [7-9]. At the macro level, the density of the health workforce, primary care systems, and medical service networks is closely related to health system performance [10,11]. Since China's health system reform, health resources, insurance coverage, and public health systems have improved continuously, yet resource-allocation disparities remain across regions, urban and rural areas, and institutional tiers [12-14]. Studies on the Healthcare Access and Quality Index and effective coverage index also show that mortality or disease burden alone is insufficient to identify weaknesses in health systems; comprehensive assessment should incorporate supply capacity, service coverage, and population demand [15-17].

At the same time, a growing body of evidence suggests that urban ecological environments and municipal sanitation conditions are important upstream factors shaping health service demand and service capacity. Green space may affect residents' health by promoting physical activity, relieving psychological stress, reducing heat exposure, mitigating air pollution, and enhancing social interaction [18,19]. Systematic reviews and meta-analyses have reported significant associations between green-space exposure and all-cause mortality, cardiovascular disease, mental health, pregnancy outcomes, and physiological stress indicators [20-23]. Research on urban green infrastructure further emphasizes that green space is not merely a landscape element, but a joint carrier of ecosystem services, climate adaptation, public health, and social equity [24,25]. Maas et al. found a stable relationship between green space and residents' health, while Mitchell and Popham showed that exposure to the natural environment may reduce health inequalities [26,27]. These studies imply that green infrastructure may not only directly reduce health risks, but also indirectly influence healthcare resource allocation and health service capacity by improving urban livability, governance capacity, and the attractiveness of public services.

Environmental pollution is another key pathway linking the eco-sanitation environment with health service supply capacity. Strong evidence has linked long-term air pollution exposure to mortality risk, cardiopulmonary disease, lung cancer, and loss of life expectancy [28-31]. Global burden of disease studies and the Lancet Commission on pollution and health have shown that air pollution and broader pollution risks cause substantial mortality and disease burden [32-34]. Source-attribution research further indicates that industrial, energy, transport, and residential combustion emissions increase premature mortality through different pathways, while transboundary pollution transport changes the spatial attribution of health damages [35,36]. In China, evidence from the Huai River policy, national air pollution mapping, and estimates of pollution-related health damages all demonstrate that long-term pollution exposure is closely related to life expectancy losses and health burden [37-40]. Higher pollution pressure may increase latent health service demand and may also crowd out public fiscal and infrastructure resources through environmental governance and disease-burden pressures, thereby weakening the capacity of urban health service systems.

In addition to green space and pollution pressure, municipal sanitation infrastructure, such as water supply, wastewater treatment, environmental sanitation, and solid-waste management, provides the basic foundation of urban public health systems. Studies on water, sanitation, and hygiene interventions have shown that improved water and sanitation conditions can substantially reduce diarrhoeal disease risks [41,42]. Mara et al. and Bartram and Cairncross emphasized that sanitation facilities are not peripheral to traditional medical systems, but constitute a foundational component of public health [43,44]. Global burden of disease studies have further estimated the health losses attributable to inadequate water, sanitation, and hygiene, while cost-benefit analyses have argued that health benefits should be incorporated into evaluations of water-supply and sanitation infrastructure investments [45-47]. Thus, sewage treatment rates, centralized sewage treatment rates, harmless domestic waste treatment rates, water supply coverage, and gas coverage can be regarded as important components of the urban eco-sanitation environment.

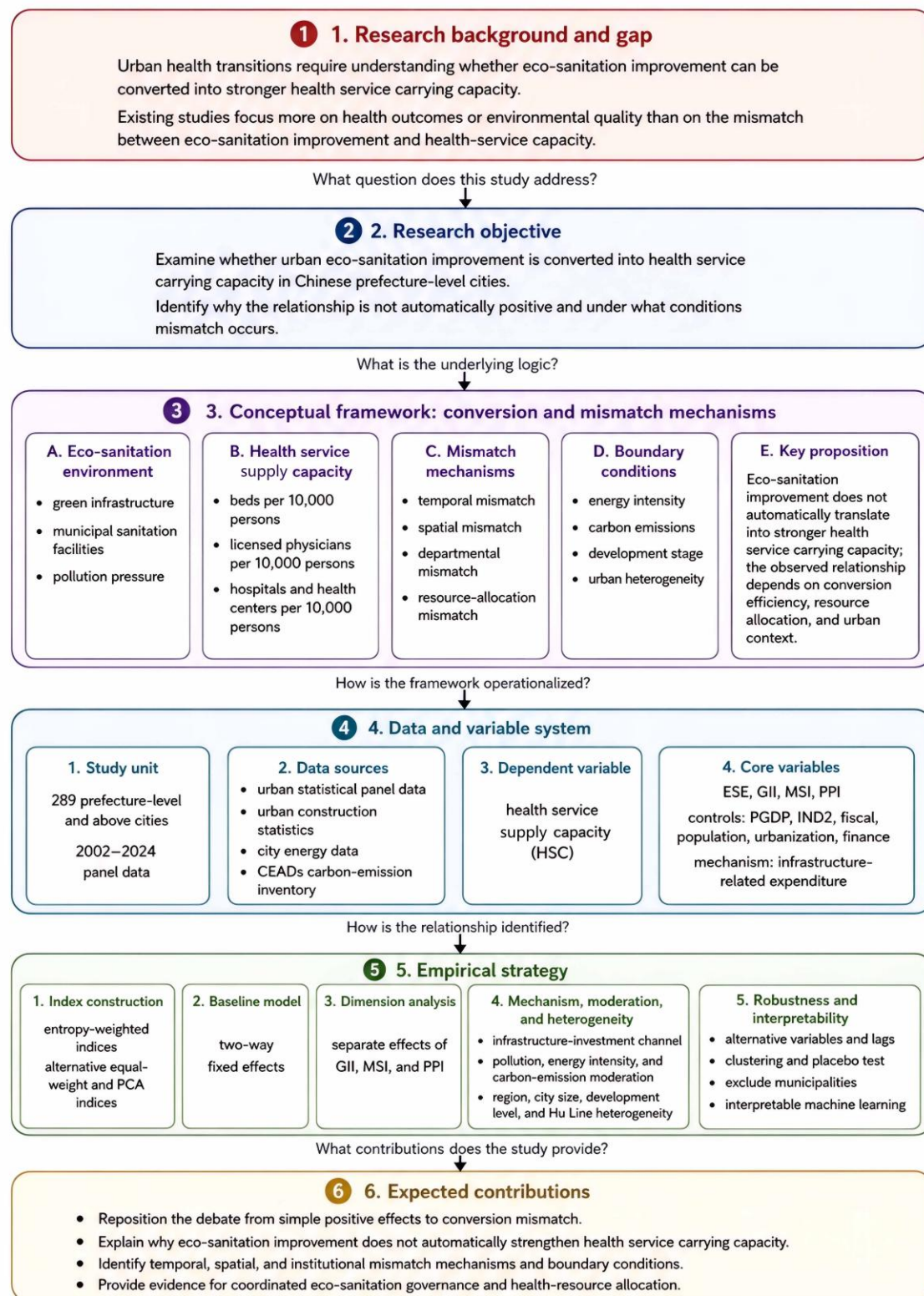
Health service supply capacity is also affected by spatial accessibility and the structure of urban resource allocation. Classic accessibility research argues that access depends not only on whether services exist, but also on the spatial, economic, and organizational match between resources and demand [48]. Luo and Wang systematically examined measures of spatial accessibility to healthcare in a GIS environment, and Wang and Luo further integrated spatial and non-spatial factors to identify health professional shortage areas [49,50]. Methods such as the enhanced two-step floating catchment area approach have advanced healthcare accessibility research from static supply indicators to spatial supply-demand matching measures [51]. However, fine-grained geographic accessibility and residents' healthcare-seeking behavior data are not always available in long-term cross-city panel research. Therefore, city-level indicators such as hospital and health-center beds per 10,000 persons, licensed physicians per 10,000 persons, and hospitals and health centers per 10,000 persons cannot fully represent micro-level accessibility, but they still provide long-term, comparable proxies for supply-side health service capacity.

Energy consumption and carbon emissions provide a broader environmental pressure context for understanding the relationship between the eco-sanitation environment and health service supply capacity. City-level carbon-emission accounting in China has developed into a relatively mature framework and can support research on urban low-carbon governance, environmental risks, and public health [52,53]. Energy consumption, industrial structure, and carbon emissions are often intertwined with pollution exposure, infrastructure efficiency, and urban governance capacity [54,55]. Therefore, high energy intensity and high carbon emissions may reflect not only resource and environmental pressures associated with urban development patterns, but also changes in the marginal relationship between eco-sanitation improvement and health service supply capacity. In other words, whether eco-sanitation construction can be converted effectively into stronger health service supply capacity may depend on whether a city can simultaneously move beyond development constraints characterized by high pollution, high energy consumption, and high carbon emissions.

In recent years, the combination of panel econometric models and interpretable machine learning has provided a new methodological pathway for urban environmental health research. Random forests and gradient boosting models can capture complex nonlinear relationships and variable importance [56,57], while interpretability tools such as SHAP and LIME further improve the explainability of machine learning results [58,59]. However, machine learning mainly supports prediction, identification of variable contributions, and discovery of nonlinear patterns; it cannot replace fixed effects models, lag structures, and robustness checks in identifying average associations and directional relationships. Therefore, in city panel studies, a more appropriate strategy is to use panel econometric models to identify average relationships between the eco-sanitation environment and health service supply capacity, while using interpretable machine learning to supplement the analysis of variable importance, threshold patterns, and potential heterogeneity.

In summary, three gaps remain in the literature. First, most environmental health studies focus on mortality, disease burden, or individual health outcomes, while relatively few studies examine how the urban eco-sanitation environment affects health service supply capacity and resource allocation from a health services research perspective. Second, green infrastructure, water and gas supply, wastewater treatment, solid-waste management, and pollution pressure are often studied separately, and there is a lack of an integrated analytical framework that connects urban eco-sanitation conditions with the supply capacity of

health service systems. Third, few studies of Chinese cities have simultaneously examined whether eco-sanitation improvement can be converted into stronger health service supply capacity and why this conversion may be constrained by energy-carbon conditions, public finance, healthcare resource allocation, and urban heterogeneity. Against this background, this study constructs a city panel dataset for China covering 2002-2024, uses urban health service supply capacity as the dependent variable and the urban eco-sanitation environment as the core explanatory variable, and combines two-way fixed effects models, mechanism tests, heterogeneity analysis, robustness checks, and interpretable machine learning to identify mismatch mechanisms and boundary conditions in the conversion from eco-sanitation improvement to health service supply capacity. Figure 1 presents the research framework and methodological route.



From isolated eco-sanitation improvement to coordinated health-service resource allocation

Figure 1 Research framework and methodological route

Accordingly, this study focuses on the conversion relationship between eco-sanitation improvement and health service supply capacity and proposes the following research questions: RQ1, can improvements in the urban eco-sanitation environment be stably converted into stronger health service supply capacity?

RQ2, do green infrastructure, municipal sanitation facilities, and pollution pressure play differentiated roles in this relationship? RQ3, do urban construction investment and public fiscal capacity constitute effective channels through which the eco-sanitation environment is converted into health service supply capacity? RQ4, do pollution pressure, energy intensity, and carbon emissions modify the marginal relationship between the eco-sanitation environment and health service supply capacity? RQ5, does this relationship exhibit city-type heterogeneity and nonlinear boundaries?

The contributions of this study are threefold. First, it incorporates urban health service supply capacity into the eco-sanitation environment research framework, but does not presuppose an automatic positive conversion between the two; instead, it examines from a health services research perspective whether eco-sanitation improvement can be translated into comparable service supply capacity. Second, it constructs an urban eco-sanitation analytical framework that includes green infrastructure, municipal sanitation facilities, pollution pressure, energy intensity, and carbon emissions, thereby identifying structural mismatches between the eco-sanitation environment and health service supply capacity. Third, by combining two-way fixed effects models with interpretable machine learning, it reveals how such mismatches manifest across time, space, administrative sectors, and stages of urban development, providing empirical evidence for coordinated healthcare resource allocation, primary service network planning, and healthy city governance.

2 Methods

2.1 Data sources and sample construction

This study adopts a long-term city-level panel design, taking Chinese prefecture-level and above cities as observational units and city-year observations as the basic analytical unit. It examines the conversion relationship and mismatch mechanisms between eco-sanitation improvement and urban health service supply capacity. The object of analysis is not individual health outcomes or the performance of a single medical institution, but the supply-side capacity of urban health service systems. Therefore, the urban health service supply capacity index (HSC) is used as the dependent variable, and the ecological sanitation environment index (ESE) is used as the core explanatory variable. Considering that eco-sanitation infrastructure construction, green-space improvement, and enhancement of municipal sanitation conditions may affect healthcare resource allocation with a time lag, the baseline model uses the one-year lagged eco-sanitation environment index ($ESE_{i,t-1}$).

The data come from four types of city-level statistical sources[60]. The first is a comprehensive urban panel dataset, which provides GDP, population, fiscal revenues and expenditures, industrial structure, health resources, industrial pollution emissions, and $PM_{2.5}$, and is used to construct health service supply capacity, control variables, and pollution pressure indicators. The second is the panel dataset from the China Urban Construction Statistical Yearbook, which provides water supply, gas supply, wastewater treatment, waste treatment, green space, and built-up area indicators, and is the main source for constructing the eco-sanitation environment, green infrastructure, and municipal sanitation facility indices. The third is prefecture-level city energy consumption data, which provide total energy consumption, energy intensity per unit GDP, and per capita energy consumption for energy-constraint and extended analyses. The fourth is the CEADs city-level sectoral carbon-emission inventory, which is used to incorporate the carbon-emission dimension. The data sources and uses are shown in Table 1.

Table 1 Data sources and uses

Data source	Year range	Main variables	Use in this study
Comprehensive urban panel data	2000-2024	GDP, population, industrial structure, fiscal revenues and expenditures, health resources, industrial pollution emissions, $PM_{2.5}$	Construction of health service supply capacity, control variables, and pollution pressure indicators
China Urban Construction Statistical Yearbook panel	2002-2024	Water supply, gas supply, sewage treatment, waste treatment, green space, and built-up area indicators	Construction of eco-sanitation environment, green infrastructure, and municipal

			sanitation facility indices
Prefecture-level city energy consumption data	2006-2024	Total energy consumption, energy intensity per unit GDP, and per capita energy consumption	Energy constraints, moderation analysis, and extended analysis
CEADs city-level sectoral carbon-emission inventory	1997-2019	City-level sectoral CO ₂ emissions	Carbon-emission moderation analysis and extended analysis; only successfully matched city-code samples are used

The sample was constructed by merging data using city administrative division codes and year. For CEADs data without unified city codes, province and city names were first standardized and then matched to administrative division codes through city-name aliases. Observations with multiple candidates, unidentifiable administrative boundary changes, or unstable city-name matches were excluded from carbon-emission extended models. For cities that experienced administrative boundary adjustments, name changes, or code changes during the study period, the analysis used continuously traceable prefecture-level city administrative codes as the baseline. City-year observations that could not be stably traced to the same administrative unit were excluded from the baseline model to ensure consistency in spatial units.

The main sample covers 2002-2024 and includes 289 cities and 6,647 city-year observations. The baseline model requires non-missing values for the health service supply capacity index, the one-year lagged eco-sanitation environment index, and the main control variables, resulting in 5,070 usable city-year observations. The energy-extended sample covers 2006-2024 and includes 279 cities and 5,301 observations, of which 4,178 satisfy the extended model conditions. The CEADs extended sample matches 257 cities. The main_ceads_2002_2019 sample contains 5,202 observations, of which 3,147 are usable; the energy_ceads_2006_2019 sample contains 3,906 observations, of which 2,748 are usable. The sample structure is shown in Table 2.

Table 2 Sample structure

Sample	Data sheet	Year range	Cities	Observations	Model-usable observations	Main use
Main sample	main_2002_2024	2002-2024	289	6647	5070	Baseline two-way fixed effects model
Energy-extended sample	energy_2006_2024	2006-2024	279	5301	4178	Models including energy consumption, energy intensity, and per capita energy consumption
CEADs extended sample	main_ceads_2002_2019	2002-2019	257	5202	3147	Models including city-level sectoral carbon emissions
Energy-CEADs extended sample	energy_ceads_2006_2019	2006-2019	257	3906	2748	Models including both energy variables and carbon emissions

Data preprocessing included standardizing city names, city codes, and year formats; removing duplicate city-year records; harmonizing variable units and statistical definitions; and winsorizing continuous indicators at the 1st and 99th percentiles. Derived variables were further constructed, including health resources per 10,000 persons, pollution emissions per unit GDP, fiscal expenditure as a share of GDP, fixed-asset investment as a share of GDP, loan balance as a share of GDP, and the logarithm of per capita GDP. The baseline estimation uses complete-case samples and does not impute core variables. For monetary variables such as GDP, fiscal expenditure, fixed-asset investment, loan balance, and urban construction investment, this study preferentially uses ratios to GDP or logarithmic forms; where cross-year absolute monetary comparisons are involved, values are converted to comparable prices using the corresponding price indices. In the mechanism tests, urban maintenance and construction expenditure, landscaping investment, wastewater treatment investment, and sanitation and environmental appearance investment are all converted to shares of GDP, while year fixed effects further absorb national price

changes and common macroeconomic trends.

2.2 Variable definitions

The dependent variable is the urban health service supply capacity index (HSC_{it}), which measures a city's ability to allocate and provide basic health service resources for a given population size. The index consists of three positive indicators: hospital and health-center beds per 10,000 persons ($health_beds_per_10k$), licensed assistant and practicing physicians per 10,000 persons ($licensed_doctors_per_10k$), and hospitals and health centers per 10,000 persons ($hospitals_health_centers_per_10k$). These indicators reflect medical facility capacity, health workforce supply, and service network density, respectively.

It should be noted that the urban health service supply capacity referred to in this study is operationalized as a supply-side proxy index. Because all three component indicators are standardized per 10,000 persons, they incorporate population demand to some extent; however, they primarily reflect the stock of health resources and the density of service outlets. They cannot directly represent quality of care, actual service utilization, or residents' health outcomes. To avoid overextending the concept, the index is interpreted below as a comparable city-level proxy for health service supply capacity rather than as a complete measure of health system performance.

The core explanatory variable is the one-year lagged eco-sanitation environment index ($ESE_{i,t-1}$). ESE_{it} is composed of park green area per capita, built-up area green coverage rate, built-up area green-space rate, water supply coverage, gas coverage, sewage treatment rate, centralized sewage treatment rate, and harmless domestic waste treatment rate. It comprehensively reflects urban green space, municipal sanitation facilities, and basic eco-sanitation conditions. The baseline model uses lagged ESE to reduce the risk of reverse association between contemporaneous expansion of health services and eco-sanitation construction.

To identify the internal structure of the eco-sanitation environment, this study further constructs a green infrastructure index (GII), a municipal sanitation facility index (MSI), and a pollution pressure index (PPI). GII_{it} is composed of park green area per capita, built-up area green coverage rate, and built-up area green-space rate. MSI_{it} is composed of water supply coverage, gas coverage, sewage treatment rate, centralized sewage treatment rate, and harmless domestic waste treatment rate. PPI_{it} is composed of industrial wastewater, sulfur dioxide, smoke and dust, and nitrogen oxide emissions per unit GDP, together with annual mean $PM_{2.5}$ concentration. A higher PPI indicates greater pollution exposure and governance pressure.

The control vector X_{it} is included to reduce omitted-variable bias arising from differences in urban economic development, industrial structure, fiscal capacity, population size, urbanization, investment, and financial conditions. The baseline model includes the logarithm of per capita GDP ($\ln_pgdp_yuan$), the share of secondary industry in GDP ($secondary_industry_share$), and local fiscal expenditure as a share of GDP ($fiscal_expenditure_gdp$). Extended models further include the tertiary industry share, population size, urbanization rate, population density, fixed-asset investment, loan balance, education expenditure, science and technology expenditure, and R&D intensity. Energy-extended analyses use total energy consumption, energy intensity per unit GDP, and per capita energy consumption. Carbon-emission extended analyses use CEADs city-level sectoral CO_2 emissions. Mechanism tests use variables such as urban maintenance and construction expenditure, landscaping investment, sanitation and environmental appearance investment, wastewater treatment investment, and public fiscal capacity. Heterogeneity analyses use grouping variables for region, the Hu Huanyong line, city size, economic development level, and fiscal capacity. The main variable definitions are shown in Table 3.

Table 3 Variable definitions and model roles

Category	Symbol	Variable name	Measurement	Model role
Dependent variable	HSC_{it}	health_capacity_index	Entropy-weighted index based on beds, licensed physicians, and hospitals/health centers per 10,000 persons	Measures urban health service supply capacity
Core	$ESE_{i,t-1}$	eco_sanitation_index	One-year lagged eco-sanitation environment	Core variable in the

explanatory variable		_11	index including green space, water supply, gas supply, sewage treatment, and waste treatment indicators	baseline model
Green infrastructure	$GII_{i,t-1}$	green_infrastructure_index_11	Park green area per capita, built-up area green coverage rate, and built-up area green-space rate	Dimension-specific test
Municipal sanitation facilities	$MSI_{i,t-1}$	municipal_sanitation_index_11	Water supply coverage, gas coverage, sewage treatment rate, centralized sewage treatment rate, and harmless domestic waste treatment rate	Dimension-specific test
Pollution pressure	$PPI_{i,t-1}$	pollution_pressure_index_11	Industrial wastewater, SO ₂ , smoke/dust, and NO _x emissions per unit GDP, plus PM _{2.5} ; higher values indicate greater pollution pressure	Control variable, dimension-specific variable, or moderator
Energy variables	$Energy_{i,t-1}$	energy_intensity_10k_tce_per_100m_yuan_11, etc.	Total energy consumption, energy intensity per unit GDP, and per capita energy consumption	Energy-extended model
Carbon-emission variable	CO_{2it}	ceads_sector_co2_million_ton	CEADs city-level sectoral CO ₂ emissions, measured in million tons	Carbon-emission extended model
Control variables	X_{it}	ln_pgdp_yuan, secondary_industry_share, fiscal_expenditure_gdp, etc.	Economic development, industrial structure, fiscal capacity, population size, urbanization, investment, and financial conditions	Reduce omitted-variable bias
Mechanism variables	M_{it}	Variables related to construction, green space, sanitation, sewage, and fiscal capacity	Urban maintenance and construction expenditure, landscaping investment, sanitation and environmental appearance investment, wastewater treatment investment, and public fiscal capacity	Mechanism test
Grouping variables	G_i	region, city_size, huhuan_yong, pgdp_group, fiscal_group, etc.	Region, city size, Hu Huanyong line, economic development level, and fiscal capacity	Heterogeneity analysis

2.3 Index construction method

This study uses the entropy weighting method to construct composite indices. The basic logic is that if an indicator varies more across city-year observations, it contains more information and receives a larger weight in the composite index; if an indicator varies less, it receives a lower weight. To ensure that the indices are comparable across years, all composite indices are weighted using the pooled city-year sample rather than being weighted separately by year.

Let x_{ijt} denote the j th raw indicator for city i in year t . For positive indicators, values are first winsorized at the 1st and 99th percentiles and then standardized using min-max normalization:

Formula (1): Min-max normalization

$$z_{ijt} = \frac{x_{ijt} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (1)$$

where z_{ijt} is the standardized indicator value, x_{ijt} is the raw indicator value, and $\min(x_j)$ and $\max(x_j)$ are the minimum and maximum values of indicator j in the full city-year sample, respectively. For pollution-pressure indicators, this study focuses on the level of pollution pressure itself; therefore, pollution emissions per unit GDP and PM_{2.5} concentration are also treated as positive pressure indicators, meaning that higher values indicate greater pollution pressure.

After standardization, the share of indicator j in observation i is calculated as follows:

Formula (2): Indicator share

$$p_{ijt} = \frac{z_{ijt} + \varepsilon}{\sum_i \sum_t (z_{ijt} + \varepsilon)} \quad (2)$$

where ε is a very small positive number used to avoid undefined logarithms when values are zero. The information entropy of each indicator is calculated as follows:

Formula (3): Information entropy

$$e_j = -\frac{1}{\ln(n_j)} \sum_i \sum_t p_{ijt} \ln(p_{ijt}) \quad (3)$$

where n_j denotes the number of non-missing observations for indicator j . The information utility and indicator weight are then calculated as follows:

Formula (4): Information utility

$$d_j = 1 - e_j \quad (4)$$

Formula (5): Entropy weight

$$w_j = \frac{d_j}{\sum_j d_j} \quad (5)$$

where d_j denotes the information utility value of indicator j , and w_j denotes its entropy weight. The final composite index is calculated as follows:

Formula (6): Composite index

$$\text{Index}_{it} = \frac{\sum_j w_j z_{ijt}}{\sum_j w_j I(z_{ijt} \text{ is nonmissing})} \quad (6)$$

where $I(\cdot)$ is an indicator function. When a city-year observation has some missing component indicators, the index is recalculated by normalizing over the weights of the indicators actually available for that observation. To ensure stable index interpretation, the health service supply capacity index requires at least two valid components, the green infrastructure index requires at least two valid components, the municipal sanitation facility index requires at least three valid components, the eco-sanitation environment index requires at least four valid components, and the pollution pressure index requires at least two valid components. Observations below the minimum number of valid components are set to missing. The components and entropy weights of each composite index are shown in Table 4.

Table 4 Components and entropy weights of composite indices

Index system	Original indicator	Variable name	Entropy weight
Health service supply capacity index (HSC)	Hospital and health-center beds per 10,000 persons	health_beds_per_10k	0.2626
Health service supply capacity index (HSC)	Licensed physicians per 10,000 persons	licensed_doctors_per_10k	0.3209
Health service supply capacity index (HSC)	Hospitals and health centers per 10,000 persons	hospitals_health_centers_per_10k	0.4165
Green infrastructure index (GII)	Park green area per capita	park_green_area_pc	0.5754
Green infrastructure index (GII)	Built-up area green coverage rate	builtup_green_coverage_rate	0.2106
Green infrastructure index (GII)	Built-up area green-space rate	builtup_green_space_rate	0.2140
Municipal sanitation facility index (MSI)	Water supply coverage rate	water_supply_coverage_rate	0.1136
Municipal sanitation facility index (MSI)	Gas coverage rate	gas_coverage_rate	0.2038

Municipal sanitation facility index (MSI)	Sewage treatment rate	sewage_treatment_rate	0.2457
Municipal sanitation facility index (MSI)	Centralized sewage treatment rate	centralized_sewage_treatment_rate	0.3085
Municipal sanitation facility index (MSI)	Harmless domestic waste treatment rate	harmless_garbage_treatment_rate	0.1284
Eco-sanitation environment index (ESE)	Park green area per capita	park_green_area_pc	0.2757
Eco-sanitation environment index (ESE)	Built-up area green coverage rate	builtup_green_coverage_rate	0.1009
Eco-sanitation environment index (ESE)	Built-up area green-space rate	builtup_green_space_rate	0.1025
Eco-sanitation environment index (ESE)	Water supply coverage rate	water_supply_coverage_rate	0.0592
Eco-sanitation environment index (ESE)	Gas coverage rate	gas_coverage_rate	0.1062
Eco-sanitation environment index (ESE)	Sewage treatment rate	sewage_treatment_rate	0.1280
Eco-sanitation environment index (ESE)	Centralized sewage treatment rate	centralized_sewage_treatment_rate	0.1607
Eco-sanitation environment index (ESE)	Harmless domestic waste treatment rate	harmless_garbage_treatment_rate	0.0669
Pollution pressure index (PPI)	Industrial wastewater emissions per unit GDP	industrial_wastewater_per_gdp	0.1892
Pollution pressure index (PPI)	Industrial sulfur dioxide emissions per unit GDP	industrial_so2_per_gdp	0.2970
Pollution pressure index (PPI)	Industrial smoke/dust emissions per unit GDP	industrial_smoke_dust_per_gdp	0.2625
Pollution pressure index (PPI)	Industrial nitrogen oxide emissions per unit GDP	industrial_nox_per_gdp	0.1913
Pollution pressure index (PPI)	Annual mean PM _{2.5} concentration	pm25_ug_m3	0.0600

2.4 Econometric model specification

This study first estimates a city and year two-way fixed effects model. The baseline model is specified as follows:

Formula (7): Baseline two-way fixed effects model

$$HSC_{it} = \alpha + \beta ESE_{i,t-1} + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (7)$$

where HSC_{it} denotes the health service supply capacity of city i in year t , $ESE_{i,t-1}$ denotes the one-year lagged eco-sanitation environment index, X_{it} is the vector of control variables, μ_i denotes city fixed effects, λ_t denotes year fixed effects, and ε_{it} denotes the random error term. β is the coefficient of primary interest, indicating whether health service supply capacity changes within the same city after eco-sanitation improvement, that is, whether eco-sanitation improvement can be converted into stronger health service supply capacity. Standard errors in all baseline regressions are clustered at the city level to allow serial correlation and heteroskedasticity within cities.

To distinguish the roles of different dimensions of the eco-sanitation environment, this study further decomposes the eco-sanitation environment into green infrastructure, municipal sanitation facilities, and pollution pressure, and estimates the following dimension-specific model:

Formula (8): Dimension-specific model

$$HSC_{it} = \alpha + \beta_1 GII_{i,t-1} + \beta_2 MSI_{i,t-1} + \beta_3 PPI_{i,t-1} + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (8)$$

where $GII_{i,t-1}$, $MSI_{i,t-1}$, and $PPI_{i,t-1}$ denote the one-year lagged green infrastructure index, municipal

sanitation facility index, and pollution pressure index, respectively. This model examines whether different dimensions of the eco-sanitation environment play different roles in changes in health service supply capacity and identifies whether green space, municipal sanitation infrastructure, and pollution pressure constitute different sources of conversion mismatch.

Pollution pressure, energy intensity, and carbon emissions may modify the marginal effect of eco-sanitation improvement. To test this moderation effect, the following model is specified:

Formula (9): Moderation model

$$HSC_{it} = \alpha + \beta ESE_{i,t-1} + \delta Z_{i,t-1} + \theta(ESE_{i,t-1} \times Z_{i,t-1}) + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (9)$$

where $Z_{i,t-1}$ sequentially denotes the pollution pressure index, energy intensity per unit GDP, per capita energy consumption, or CEADs city-level sectoral carbon emissions. If the interaction term θ is significant, pollution, energy, or carbon-emission conditions modify the marginal relationship between the eco-sanitation environment and health service supply capacity; the sign of θ is used to determine whether the effect represents constraint, compensation, or a resource-allocation response. To reduce multicollinearity in interaction terms, continuous interacting variables are mean-centered before entering the model.

The mechanism tests focus on the urban public infrastructure investment pathway: eco-sanitation improvement may enhance material conditions for health service systems by increasing urban maintenance and construction expenditure, landscaping investment, wastewater treatment investment, and sanitation and environmental appearance investment. Pollution pressure, energy intensity, and carbon emissions are no longer treated as mechanism variables; instead, their boundary roles are examined in the moderation model described above. The first-stage model for mechanism variable M_{it} is specified as follows:

Formula (10): First-stage mechanism model

$$M_{it} = \alpha + \rho ESE_{i,t-1} + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (10)$$

In Formula (10), M_{it} denotes the mechanism variable of city i in year t , and ρ captures the association between lagged eco-sanitation environment and the mechanism variable; the other terms are defined as in Formula (7). To avoid excessive dispersion of mechanism pathways, this study tests only channels directly related to the components of the eco-sanitation environment and urban public infrastructure provision. These include urban maintenance and construction expenditure as a share of GDP, landscaping investment as a share of GDP, wastewater treatment investment as a share of GDP, sanitation and environmental appearance investment as a share of GDP, and local fiscal expenditure as a share of GDP. Investment and expenditure variables are uniformly expressed as shares of GDP to reduce the influence of city size and price-level differences on the mechanism tests. In the second stage, the mechanism variable is added to the baseline model:

Formula (11): Model with mechanism variable

$$HSC_{it} = \alpha + \beta ESE_{i,t-1} + \phi M_{it} + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (11)$$

where M_{it} denotes the mechanism variable, including urban maintenance and construction expenditure, landscaping investment, sanitation and environmental appearance investment, wastewater treatment investment, and fiscal expenditure. If ϕ is significantly positive, the coefficient β changes after M_{it} is added, and ρ is significant, this may indicate that the corresponding variable constitutes a possible transmission channel through which the eco-sanitation environment affects health service supply capacity. Because this study uses city-level observational data, the mechanism tests are used to identify possible pathways and are not interpreted as strict causal mediation effects.

2.5 Heterogeneity analysis

To examine whether the conversion relationship between the eco-sanitation environment and health service supply capacity varies across urban conditions, this study conducts four types of heterogeneity analysis. First, cities are grouped into eastern, central, western, and northeastern regions to compare differences across regions with different stages of development and fiscal capacity. Second, cities are grouped by their location on either side of the Hu Huanyong line to examine how population density, resource endowments, and urban-system differences affect the estimates. Third, cities are grouped by city size or population size to test whether large and small-to-medium cities differ in the agglomeration of health service resources and the conversion efficiency of eco-sanitation infrastructure. Fourth, cities are divided into high and low groups according to the median of per capita GDP or fiscal expenditure as a share of GDP to examine whether economic development and fiscal capacity affect the conversion of eco-sanitation improvement into health service supply capacity.

Heterogeneity is assessed using both subgroup regressions and interaction-term regressions. Subgroup regressions maintain the same fixed effects and control-variable settings as the baseline model, while interaction regressions add the interaction between the eco-sanitation environment index and city-type variables to the full sample. The interaction model is specified as follows:

Formula (12): Heterogeneity interaction model

$$HSC_{it} = \alpha + \beta ESE_{i,t-1} + \eta(ESE_{i,t-1} \times G_i) + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (12)$$

where G_i denotes the grouping variable, including region, Hu Huanyong line grouping, city size, economic development level, or fiscal capacity. η reflects whether the conversion relationship between the eco-sanitation environment and health service supply capacity varies by city type.

2.6 Robustness checks and endogeneity treatment

This study evaluates robustness in several ways. First, it replaces core variable definitions. For the dependent variable, the three components of health service supply capacity - hospital and health-center beds per 10,000 persons, licensed physicians per 10,000 persons, and hospitals and health centers per 10,000 persons - are used as alternative dependent variables. For the explanatory variables, green infrastructure, municipal sanitation facilities, and pollution pressure are used as alternative explanatory variables. Second, lag structures are adjusted by using contemporaneous ESE, two-year lagged ESE, and lagged controls to test sensitivity to time specification. Third, sample ranges are adjusted by estimating the models on the main sample, energy-extended sample, and CEADs extended sample, and by excluding municipalities, extreme years, or cities with small samples. Fourth, control-variable sets are adjusted by gradually adding population size, urbanization rate, fixed-asset investment, financial development, education expenditure, science and technology expenditure, and R&D intensity. Fifth, city-clustered standard errors and two-way clustered standard errors by city and year are used to test sensitivity to standard error specification. In addition, the entropy-weighted index is replaced with equal-weighted and principal component analysis (PCA) indices to examine whether the estimation results depend on the index-construction method; consistent results under alternative indices would indicate that the findings are not driven by a single weighting scheme.

With respect to endogeneity, this study faces three main risks: reverse causality, omitted variables, and common trends. To mitigate reverse causality, the baseline model uses the one-year lagged eco-sanitation environment index. To reduce omitted-variable bias, the models include city fixed effects, year fixed effects, and multidimensional controls. To address common trends, year fixed effects absorb national shocks, and heterogeneity and sample-adjustment analyses test whether the results are driven by specific regions or years. In addition, this study includes future ESE as a placebo variable and conducts a joint test using one-year, two-year, and three-year leads of ESE to identify possible systematic pre-trends. If the multi-period lead terms are jointly insignificant, current changes in health service supply capacity are unlikely to be driven entirely by future eco-sanitation changes or pre-existing trends; if weak significance

is present, the baseline results should be interpreted as directional evidence rather than strict causal effects. The placebo model is specified as follows. It should be emphasized that this study is not based on an exogenous policy shock or a natural experimental design. Fixed effects, lagged variables, and placebo tests can only partially mitigate endogeneity concerns. Therefore, the baseline results are interpreted mainly as conditional within-city associations and directional evidence between changes in the eco-sanitation environment and changes in health service supply capacity, rather than as strict causal effects.

Formula (13): Future-period placebo test model

$$HSC_{it} = \alpha + \beta_f ESE_{i,t+1} + \gamma' X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (13)$$

where $ESE_{i,t+1}$ denotes the future one-year eco-sanitation environment index, and β_f is the placebo coefficient. If β_f is insignificant, the baseline result is unlikely to be fully driven by reverse causality or pre-trends.

2.7 Interpretable machine learning as a supplementary analysis

In addition to panel econometric models, this study uses interpretable machine learning as a supplementary analysis. The machine learning models take the urban health service supply capacity index as the prediction target and use the eco-sanitation environment index, green infrastructure index, municipal sanitation facility index, pollution pressure index, energy intensity, carbon emissions, and control variables as features. The models include random forest and gradient boosting, and prediction performance is reported using R^2 , RMSE, and MAE. To avoid temporal leakage, observations from 2018 and earlier are used as the training set, and observations from 2019 onward are used as the test set, rather than using a completely random split. Hyperparameters are selected through time-blocked cross-validation and fixed in the final reproducible configuration: for the random forest, $n_estimators = 500$, $max_depth = None$, $max_features = sqrt$, and $min_samples_leaf = 2$; for the gradient boosting model, $n_estimators = 300$, $max_depth = 3$, $learning_rate = 0.05$, $subsample = 0.80$, and $max_features = sqrt$. Time-blocked cross-validation uses rolling year-based validation within the training period to avoid information leakage from adjacent years of the same city.

The machine learning analysis has three purposes. First, it compares the relative contributions of eco-sanitation variables with economic, fiscal, and population variables. Second, it identifies nonlinear relationships and potential thresholds, such as whether the marginal contribution of the eco-sanitation environment index, sewage treatment rate, built-up area green coverage rate, or harmless domestic waste treatment rate strengthens or weakens after reaching a certain level. Third, SHAP values are used to explain the direction and magnitude of variable contributions to predictions, helping assess whether average effects in panel models conceal city heterogeneity. It should be emphasized that machine learning results are used only for pattern recognition and interpretive supplementation; because they cannot automatically control for unobserved city heterogeneity or reverse causality, they are not used as the main basis for causal inference.

The predictive performance of the machine learning models is evaluated using the following metrics:

Formula (14): Coefficient of determination R^2

$$R^2 = 1 - \frac{\sum_{i,t} (HSC_{it} - \widehat{HSC}_{it})^2}{\sum_{i,t} (HSC_{it} - \overline{HSC})^2} \quad (14)$$

Formula (15): Root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i,t} (HSC_{it} - \widehat{HSC}_{it})^2} \quad (15)$$

Formula (16): Mean absolute error (MAE)

$$MAE = \frac{1}{N} \sum_{i,t} |HSC_{it} - \widehat{HSC}_{it}| \quad (16)$$

where HSC_{it} denotes the observed health service supply capacity, $\hat{HSC}_{it}$ denotes the value predicted by the machine learning model, $\bar{HSC}$ denotes the sample mean of HSC , and N denotes the number of observations in the test sample.

3 Results

3.1 Descriptive statistics and index evolution

Table 5 reports descriptive statistics for the main variables in the model sample. The mean value of the urban health service supply capacity index is 0.2201, with a standard deviation of 0.1236, indicating substantial differences in health resource supply capacity across cities. The mean eco-sanitation environment index is 0.6419, which is much higher than the health service supply capacity index, suggesting that urban eco-sanitation infrastructure improved relatively rapidly but was not fully synchronized with healthcare resource allocation. The mean green infrastructure index is 0.4860, lower than the municipal sanitation facility index of 0.8048, indicating that water supply, gas coverage, sewage treatment, and waste treatment infrastructure had relatively high coverage, whereas green-space provision still showed larger inter-city variation. The mean pollution pressure index is 0.1442, and its maximum reaches 1, indicating that some cities still faced strong environmental pressure.

To examine the external validity of the HSC indicator, this study further tests the correlation between HSC and the number of all health institutions per 10,000 persons, which is not included in the index construction. The Pearson correlation coefficient is 0.379 and the Spearman rank correlation coefficient is 0.495, indicating a moderate positive correlation between HSC and urban health service network density. However, this test provides only supply-side external validity evidence and cannot substitute for indicators of quality of care, healthcare accessibility, or service utilization.

Table 5 Descriptive statistics of main variables

Variable	N	Mean	SD	Min	Median	Max
Health service supply capacity	5070	0.2201	0.1236	0.0084	0.1987	0.7892
Eco-sanitation environment	5068	0.6419	0.1644	0.0071	0.6853	0.9663
Green infrastructure	5062	0.4860	0.1543	0.0000	0.4968	0.9636
Municipal sanitation facilities	4879	0.8048	0.1860	0.0400	0.8710	1.0000
Pollution pressure	4941	0.1442	0.1442	0.0000	0.0966	1.0000
Hospital and health-center beds per 10,000 persons	5069	39.5731	18.1379	8.6000	37.0235	136.9930
Licensed physicians per 10,000 persons	5069	21.0893	11.4158	3.7554	18.7942	98.3317
Hospitals and health centers per 10,000 persons	5070	0.5205	0.6512	0.0543	0.4068	11.4770
Log per capita GDP	5070	10.2887	0.8425	7.6009	10.3590	12.4565
Secondary industry share	5070	46.6420	11.3195	10.6000	46.9600	90.9700
Fiscal expenditure/GDP	5070	0.1701	0.0998	0.0313	0.1455	1.1817

Figure 2 shows the annual evolution of the core indices during 2002-2024. The eco-sanitation environment, green infrastructure, and municipal sanitation facility indices generally increased, while the pollution pressure index generally declined, indicating evident improvement in urban eco-sanitation infrastructure and pollution governance. The health service supply capacity index also increased overall, but breaks appear due to missing health service statistical data in 2020, 2023, and 2024. Therefore, subsequent regional comparisons use 2022, the latest year with available health service supply capacity data.

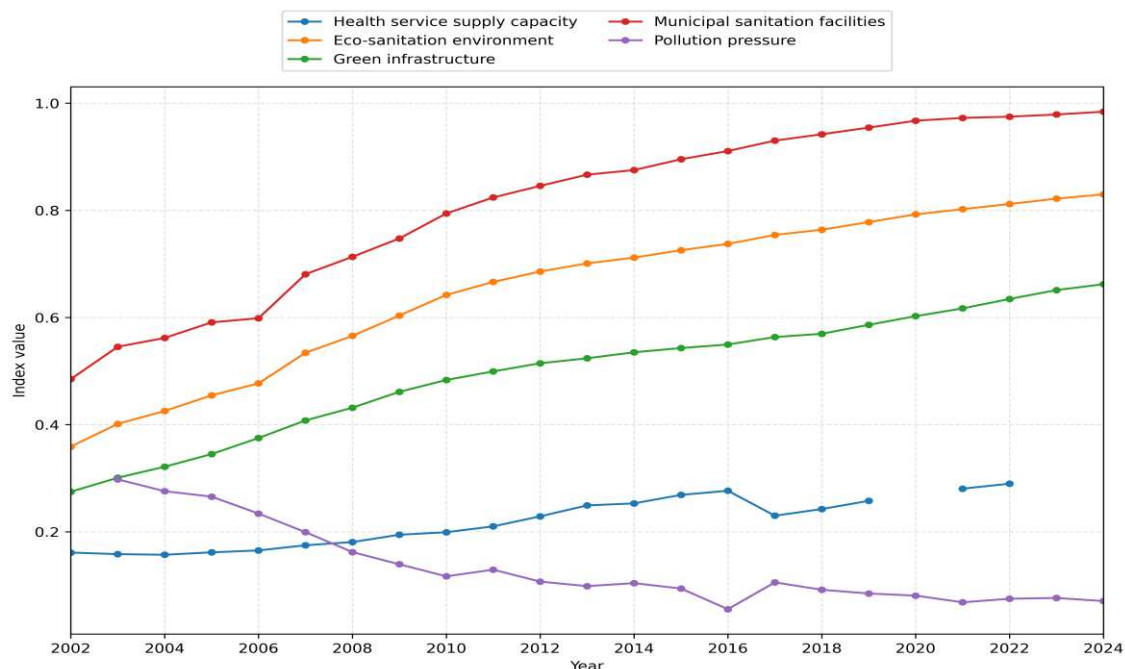


Figure 2 Temporal evolution of core indices

In terms of the components of health service supply capacity, Figure 3 shows that hospital beds per 10,000 persons and licensed physicians per 10,000 persons increased continuously, indicating overall expansion of physical medical resources and health workforce resources. By contrast, the number of hospitals and health centers per 10,000 persons changed only modestly, suggesting that the improvement in health service supply capacity mainly came from the expansion of beds and physicians rather than from a simultaneous increase in the number of medical institutions.

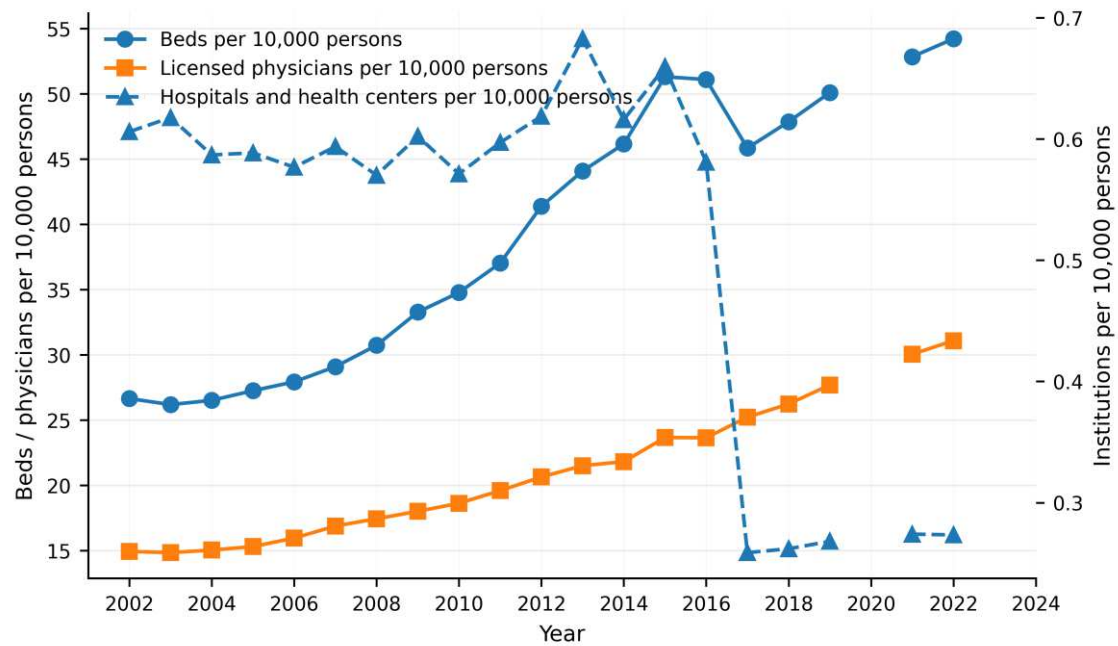


Figure 3 Temporal evolution of health service supply capacity components

Figure 4 further decomposes the component indicators of the eco-sanitation environment. Park green area per capita, built-up area green coverage rate, and built-up area green-space rate increased steadily, reflecting continuous improvement in green infrastructure. Water supply coverage and gas coverage remained at relatively high levels overall. Sewage treatment rate, centralized sewage treatment rate, and harmless domestic waste treatment rate increased more rapidly, indicating that municipal sanitation facilities were the main drivers of the increase in the eco-sanitation environment index.

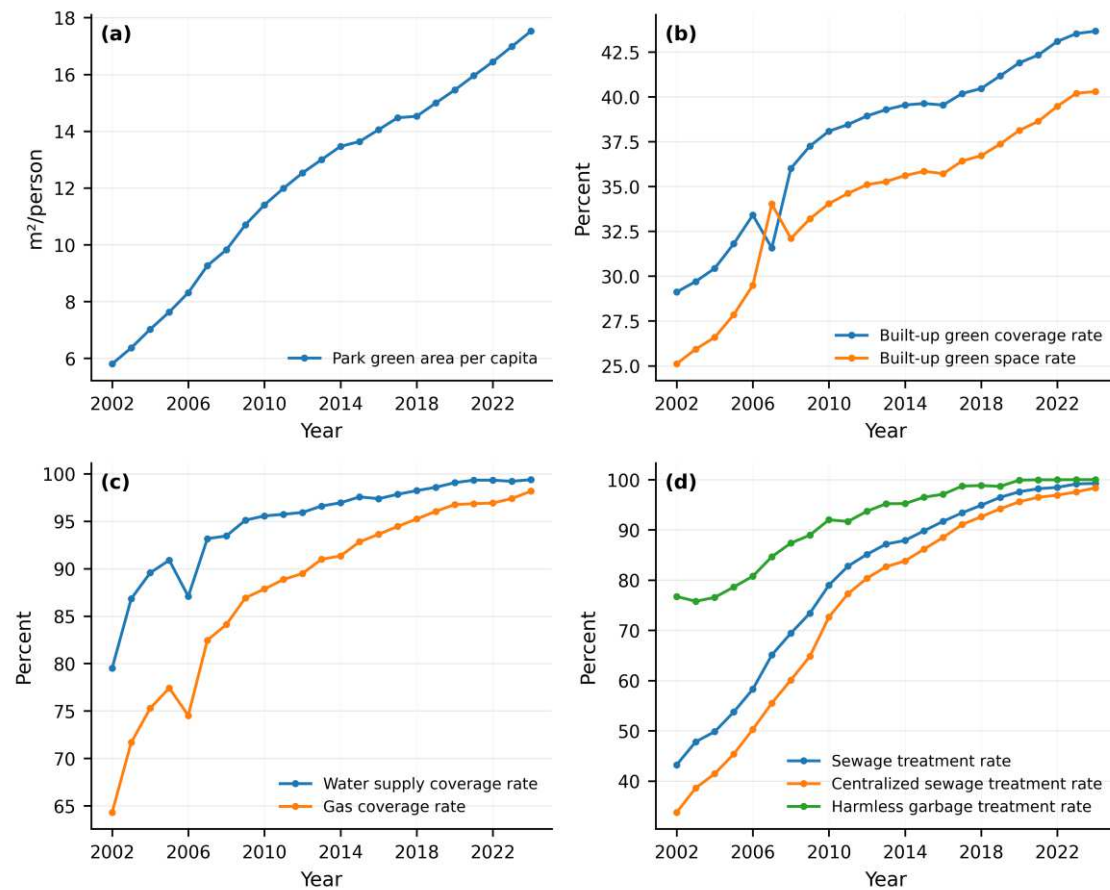


Figure 4 Temporal evolution of eco-sanitation environment components

Figure 5 presents the distribution of the main composite indices. The eco-sanitation environment and municipal sanitation facility indices have relatively high medians, and the latter is more concentrated. The health service supply capacity index and the green infrastructure index are more dispersed, indicating strong inter-city differences in health service resources and green infrastructure. The pollution pressure index is generally low but has a long right tail, suggesting that pollution pressure in a small number of cities was substantially higher than the sample average.

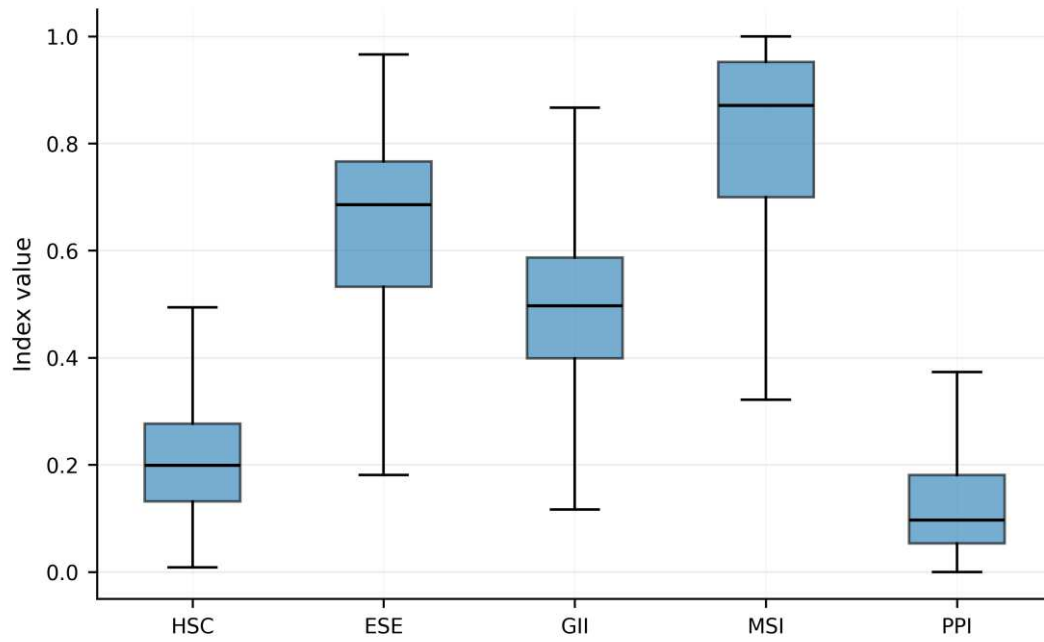


Figure 5 Distribution characteristics of main composite indices

The correlation matrix shows that health service supply capacity is more strongly correlated with economic development, population size, and fiscal capacity, whereas its simple correlations with eco-sanitation variables are unstable (Figure 6). This means that cross-sectional correlations alone cannot determine whether eco-sanitation improvement increases health service supply capacity; it is necessary to further control for unobserved city characteristics and common year shocks through city and year fixed effects.

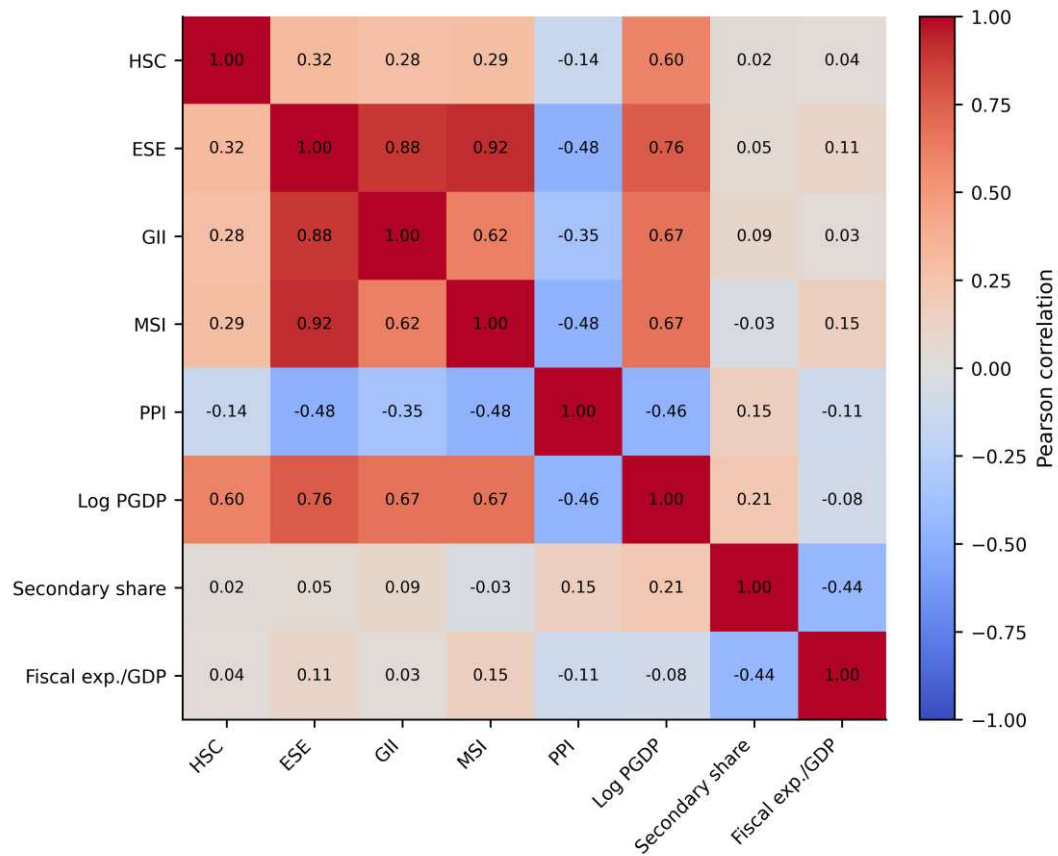


Figure 6 Correlation matrix of core variables

Figure 7 compares the core indices across regions in 2022. Eastern cities had relatively higher health service supply capacity and higher levels of eco-sanitation environment and municipal sanitation facilities. Although the eco-sanitation environment differed less between central and western cities, their health service supply capacity remained lower than that of eastern cities. This result suggests that eco-sanitation improvement is not necessarily equivalent to synchronous improvement in health service supply capacity; the relationship between the two may be jointly shaped by economic development, population size, fiscal capacity, and the functional hierarchy of urban services.

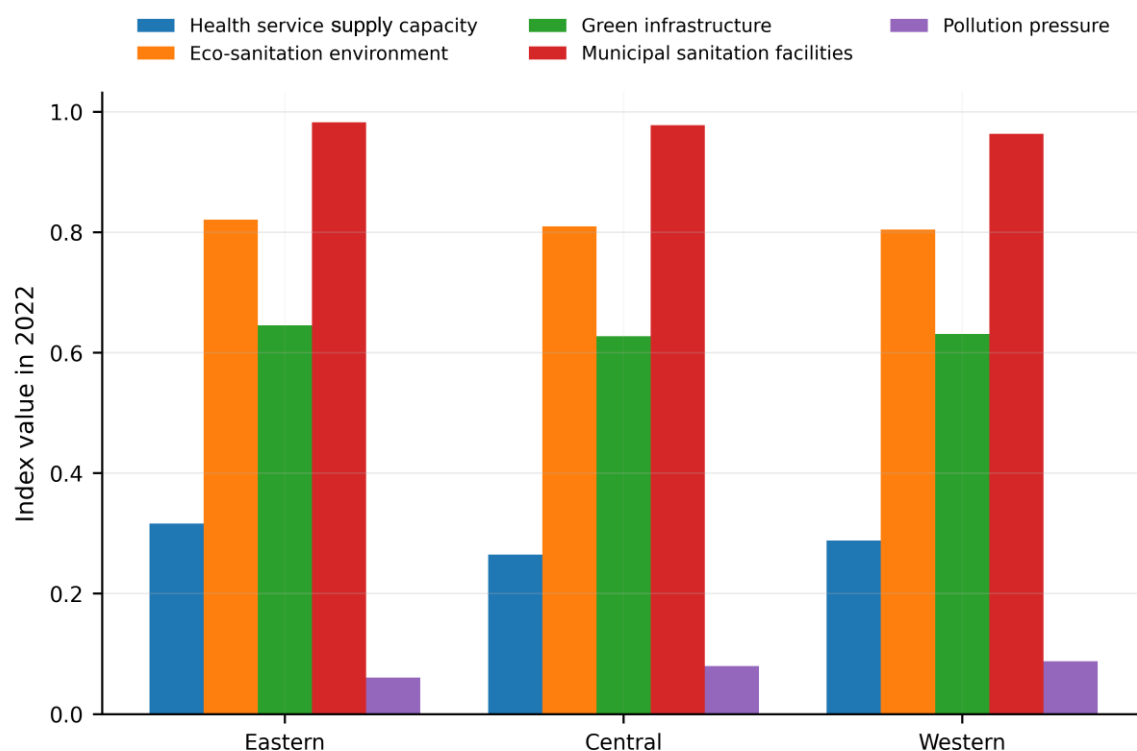


Figure 7 Regional comparison of core indices in 2022

Figure 8 shows the scatter relationship and binned mean curve between the eco-sanitation environment and health service supply capacity. The two do not show a simple monotonic positive relationship. In the high range of the eco-sanitation environment index, the mean health service supply capacity levels off or even declines. This pattern is consistent with the specification of nonlinear and stage-dependent effects in Section 2 and provides preliminary evidence for the subsequent two-way fixed effects and interpretable machine learning analyses.

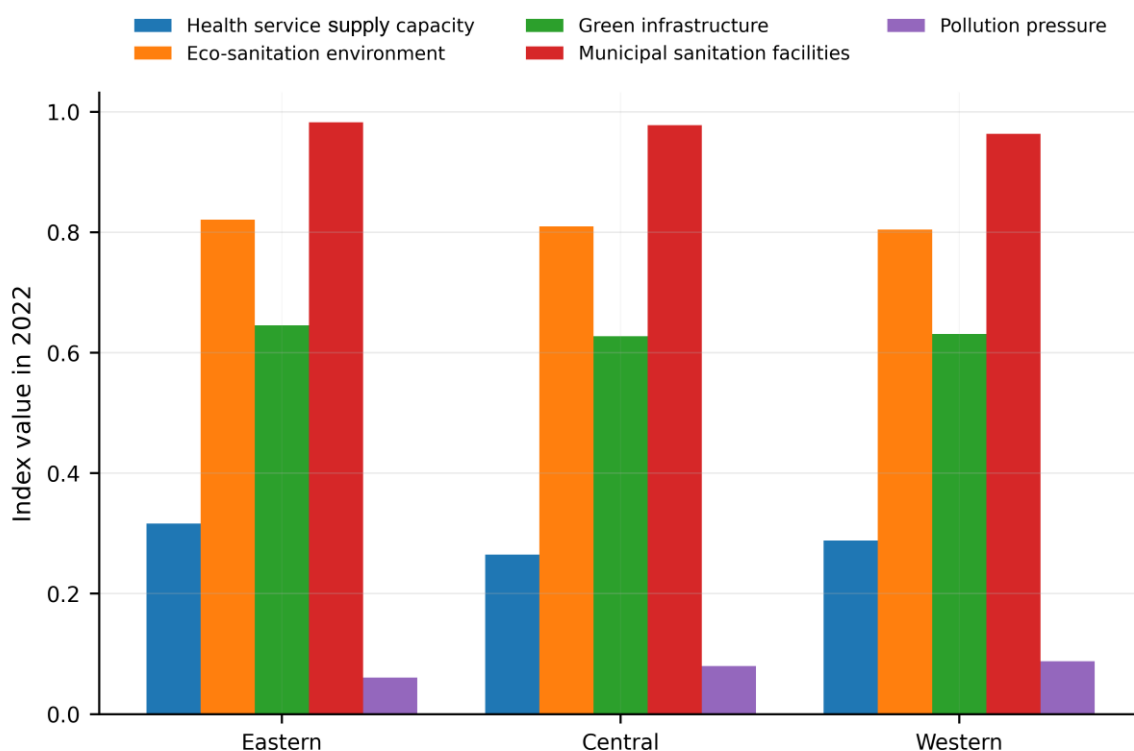


Figure 8 Binned mean relationship between the eco-sanitation environment and health service supply capacity

3.2 Baseline relationship: does eco-sanitation improvement translate into health service supply capacity?

Table 6 reports the baseline two-way fixed effects estimates of the effect of lagged eco-sanitation environment on urban health service supply capacity. All models control for city fixed effects and year fixed effects, and standard errors are clustered at the city level. For ease of interpretation, the eco-sanitation environment index is scaled by 0.1 units.

Table 6 Baseline two-way fixed effects regression results

Variable	(1)	(2)	(3)	(4)
ESE (lagged, per 0.1 increase)	-0.0069*** (0.0020)	-0.0057*** (0.0020)	-0.0053*** (0.0020)	-0.0006 (0.0020)
Log per capita GDP	—	-0.0149 (0.0111)	-0.0123 (0.0130)	0.0015 (0.0163)
Secondary industry share	—	-0.0007** (0.0003)	-0.0004 (0.0005)	-0.0009 (0.0006)
Tertiary industry share	—	—	0.0003 (0.0006)	-0.0003 (0.0006)
Fiscal expenditure/GDP	—	-0.0936** (0.0425)	-0.1390*** (0.0464)	-0.0427 (0.0475)
Education expenditure/GDP	—	—	0.3025 (0.2523)	-0.1231 (0.2014)
Science expenditure/GDP	—	—	2.6187*** (0.8750)	0.7774 (0.5786)
Log population	—	—	-0.0266 (0.0382)	-0.0819** (0.0346)
Loan balance/GDP	—	—	0.0174** (0.0075)	0.0037 (0.0061)
Fixed asset	—	—	—	0.0153

investment/GDP				(0.0109)
City FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	5070	5070	5069	3919
Cities	288	288	288	285
Adjusted R ²	0.011	0.025	0.042	0.011
Note	Robust standard errors clustered at city level are shown in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.10.			

The results show that in model (1), which controls only for city and year fixed effects, the coefficient of the lagged eco-sanitation environment index is -0.0069 and significant at the 1% level. After adding baseline controls such as economic development, industrial structure, and fiscal capacity, the coefficient in model (2) is -0.0057 and remains significant at the 1% level. After further adding extended controls such as the tertiary industry share, education expenditure, science and technology expenditure, population size, and financial development, the coefficient in model (3) is -0.0053 and remains significant. When fixed-asset investment as a share of GDP is included, the sample size falls to 3,919 observations and the coefficient becomes insignificant. Figure 9 visualizes this change in coefficients.

To address the decline in sample size from 5,070 to 3,919 observations in model (4), this study conducted additional tests of missing-sample characteristics. Compared with excluded observations, the sample retained in model (4) has a lower mean health service supply capacity (0.206 vs. 0.268), lower lagged eco-sanitation environment (0.571 vs. 0.778), lower log per capita GDP (10.079 vs. 11.004), and lower loan balance/GDP (0.821 vs. 1.277), but a higher secondary industry share (48.482 vs. 40.374). Most of these differences are statistically significant, indicating that the missingness in model (4) is not completely random. Therefore, the coefficient becoming insignificant in model (4) should not be interpreted as a direct rejection of the baseline result, but as a supplementary test on a non-randomly reduced sample after adding the fixed-asset investment variable.

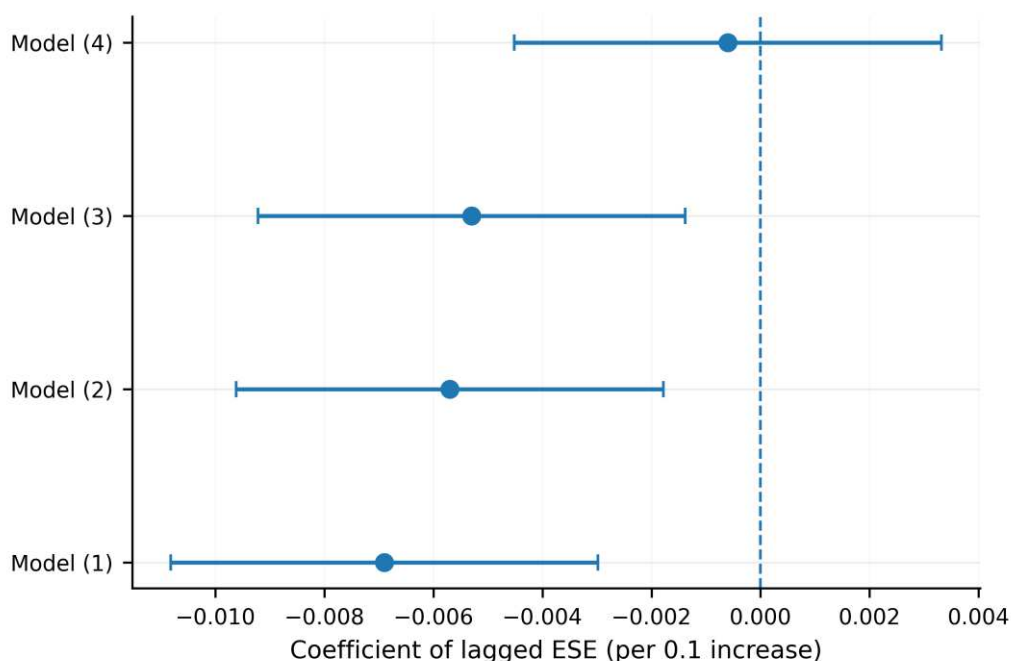


Figure 9 Coefficient estimates of the lagged eco-sanitation environment index in baseline models

These results indicate that after controlling for unobserved time-invariant city characteristics and common national year shocks, eco-sanitation improvement was not stably converted into stronger health service supply capacity. This finding should not be interpreted as evidence that eco-sanitation improvement weakens health service capacity. Rather, it more likely reflects two forms of conversion mismatch: first, a

time lag between eco-sanitation infrastructure construction and healthcare resource allocation; and second, marginally diminishing room for eco-sanitation improvement in some cities with already high health service supply capacity. Thus, RQ1 does not show a stable positive conversion relationship.

3.3 Dimension-specific differences: different roles of green infrastructure, municipal sanitation facilities, and pollution pressure

To identify the roles of different dimensions within the eco-sanitation environment, this study decomposes it into green infrastructure, municipal sanitation facilities, and pollution pressure. Figure 10 shows that when the three variables enter the model simultaneously, the coefficient of green infrastructure is significantly negative, the coefficient of municipal sanitation facilities is no longer significant, and the coefficient of pollution pressure is significantly positive.

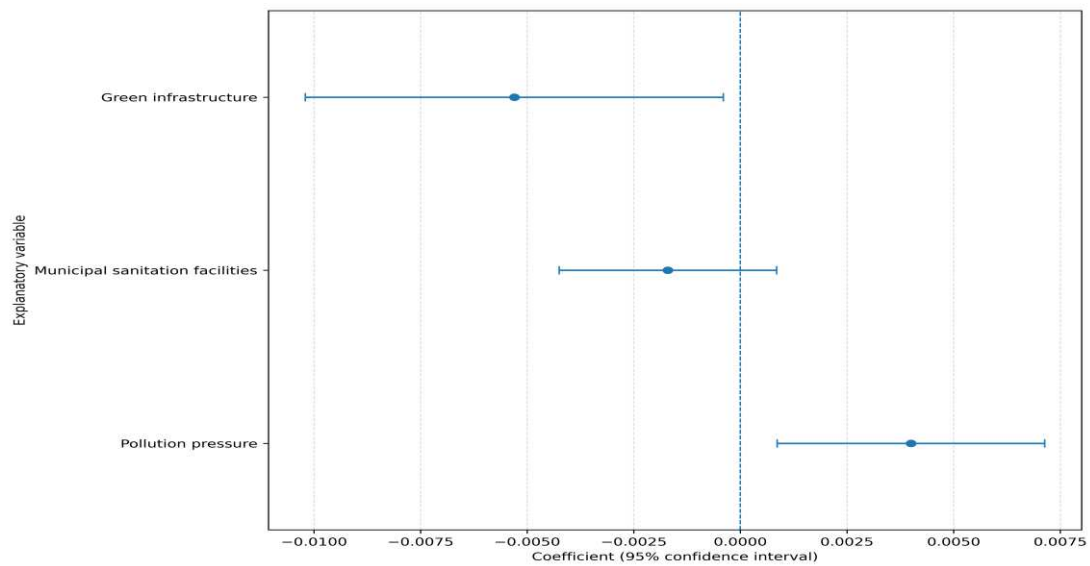


Figure 10 Dimension-specific coefficients of green infrastructure, municipal sanitation facilities, and pollution pressure

This result suggests that the relationship between urban health service supply capacity and the eco-sanitation environment is not a simple linear positive association. Cities with faster improvements in green infrastructure and municipal sanitation facilities do not necessarily have higher or faster-growing health service resources. Conversely, highly industrialized cities with greater pollution pressure often have larger populations, stronger fiscal capacity, and more concentrated healthcare demand, and thus may allocate more health service resources. Therefore, the positive coefficient of pollution pressure is more reasonably interpreted as a pressure-resource allocation response, rather than as a positive health effect of pollution itself.

To examine the risk of multicollinearity when the three dimension-specific indicators enter the model simultaneously, this study calculated variance inflation factors (VIFs). The VIFs of the green infrastructure, municipal sanitation facility, and pollution pressure indices are 1.96, 2.39, and 1.58, respectively. After including the main control variables, the VIFs of all core eco-sanitation variables remain below 3, and the maximum VIF among all explanatory variables is 5.05, below the commonly used empirical threshold of 10. Thus, the dimension-specific results are unlikely to be mainly driven by severe multicollinearity.

3.4 Conversion channel test: does urban construction investment form an effective transmission pathway?

Section 2 narrowed the mechanism pathway to channels directly related to the components of the eco-sanitation environment and urban public infrastructure provision, including urban maintenance and construction expenditure as a share of GDP, landscaping investment as a share of GDP, wastewater treatment investment as a share of GDP, sanitation and environmental appearance investment as a share of GDP, and fiscal expenditure as a share of GDP. Figure 11 shows the first-stage effect of the eco-sanitation environment on the main mechanism variables. The eco-sanitation environment has a weak positive association with landscaping investment, a significant negative association with wastewater treatment investment, and insignificant effects on urban maintenance and construction expenditure and sanitation and environmental appearance investment.

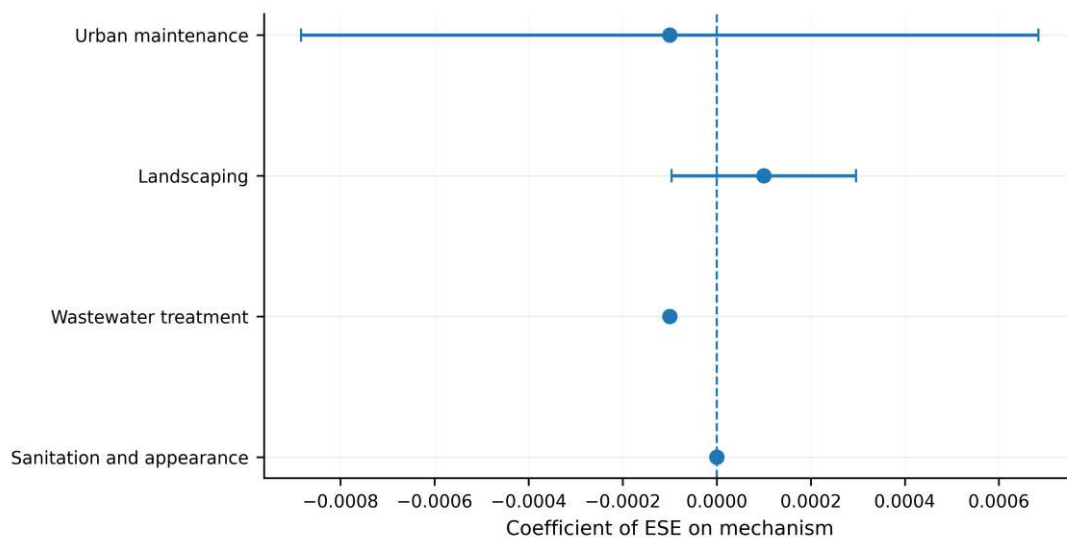


Figure 11 First-stage effects of the eco-sanitation environment on mechanism variables

Figure 12 further presents the coefficients of the mechanism variables after they are added to the health service supply capacity model. The results show that the direct effects of the relevant investment and expenditure variables on health service supply capacity do not reach conventional significance levels. After the mechanism variables are added, the eco-sanitation environment coefficient remains mostly negative. This suggests that the current data do not provide sufficient evidence to support a stable transmission chain of eco-sanitation environment $\rightarrow$ urban construction investment $\rightarrow$ health service supply capacity. A possible explanation is that urban construction investment and health service resource allocation belong to different fiscal and governance systems, and eco-sanitation infrastructure investment may not be converted into medical resource expansion in the short term. In addition, missing values and large annual fluctuations in investment variables may reduce the statistical power of the mechanism tests. Therefore, the direct effects of the mechanism variables on health service supply capacity do not reach conventional significance levels, and the transmission chain is not supported under the current data conditions.

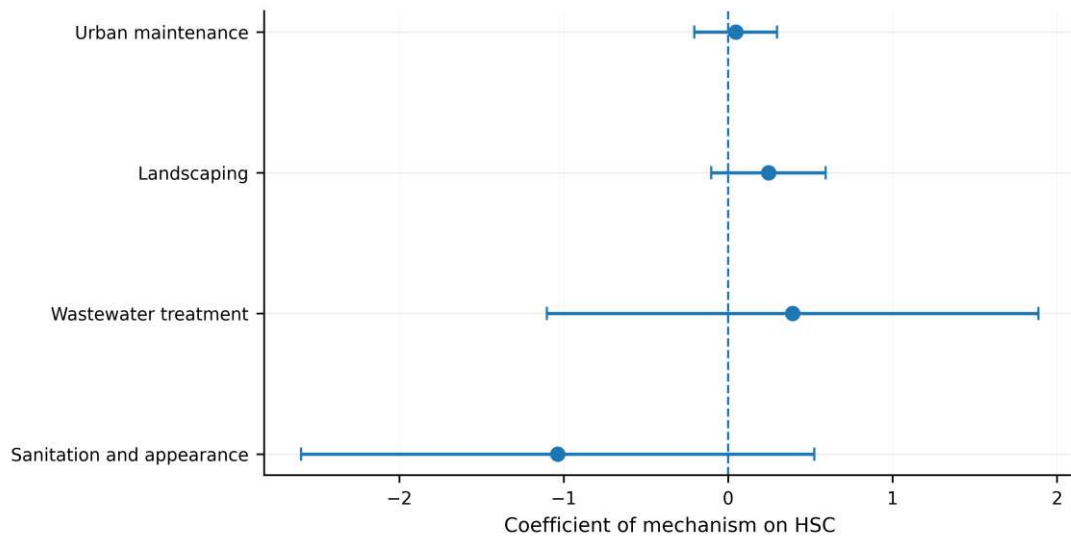


Figure 12 Direct effects of mechanism variables on health service supply capacity

3.5 Moderating effects of energy and carbon emissions

Figure 13 reports the moderating effects of pollution pressure, energy intensity per unit GDP, per capita energy consumption, and CEADs city-level sectoral carbon emissions. The interaction terms for pollution pressure, energy intensity, and per capita energy consumption are insignificant, indicating that the current model specification does not confirm that these variables significantly modify the marginal effect of the eco-sanitation environment on health service supply capacity. The interaction term for CEADs carbon emissions is positive and significant at the 5% level, indicating that in cities with higher carbon emissions, the relationship between the eco-sanitation environment and health service supply capacity is strengthened.

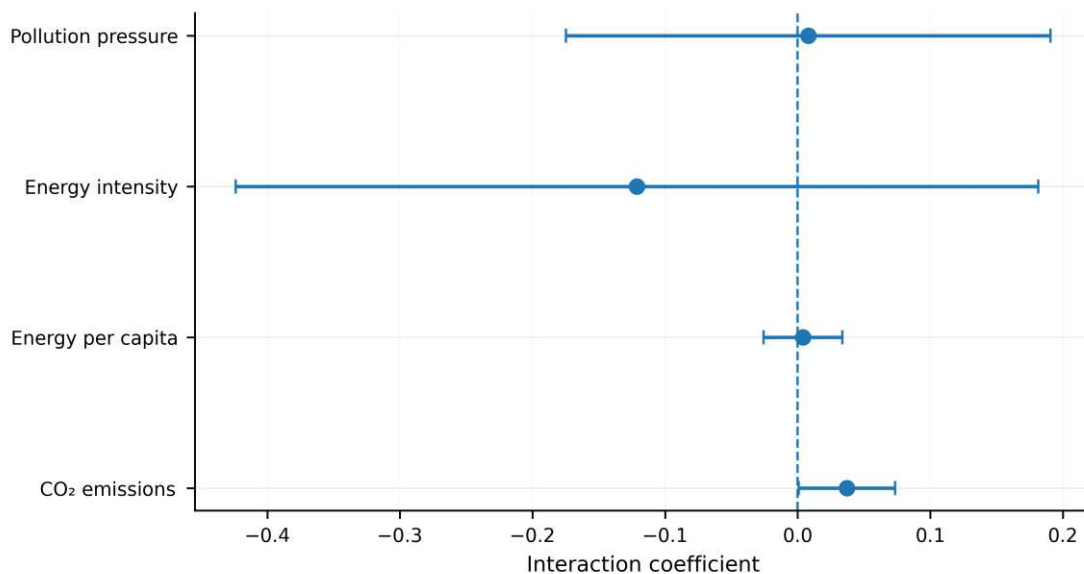


Figure 13 Moderating effects of pollution pressure, energy intensity, and carbon emissions

This result indicates that energy intensity and carbon emissions may modify the marginal relationship between the eco-sanitation environment and health service supply capacity, but their direction is not a simple negative constraint. A more plausible explanation is that high-carbon cities usually have stronger industrial bases, fiscal capacity, population agglomeration, and public service demand; therefore,

eco-sanitation improvement in these cities is more likely to occur together with the expansion of public service systems. The positive significance of the carbon-emission interaction term thus reflects the combined effect of urban development stage and resource-allocation capacity, rather than any positive role of carbon emissions themselves.

3.6 Heterogeneity analysis

Figure 14 reports heterogeneous effects of the eco-sanitation environment on health service supply capacity. Regional grouping shows that the coefficients are significantly negative in eastern and central cities, but insignificant in western cities. When grouped by economic development, the coefficient is significantly negative in high-GDP cities and insignificant in low-GDP cities. When grouped by city size, the coefficient is significantly negative in large-population cities and insignificant in small-population cities. Grouping by the Hu Huanyong line shows that the coefficient is significantly negative on the southeastern side and insignificant on the northwestern side.

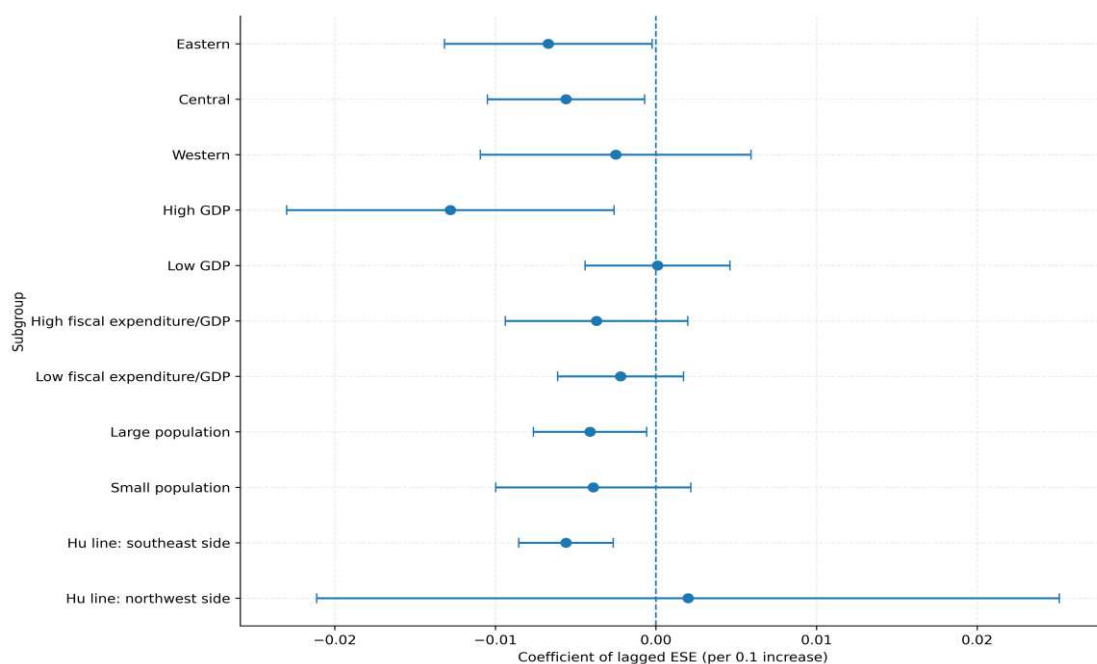


Figure 14 Heterogeneous coefficients of the eco-sanitation environment across city types

These results indicate that the relationship between the eco-sanitation environment and health service supply capacity mainly appears in regions with higher economic development, greater population concentration, and more mature urban systems. In these cities, both eco-sanitation infrastructure and health service systems have reached relatively high levels, and their relationship is more likely to reflect structural adjustment rather than simple synchronous expansion. In other words, eco-sanitation improvement does not automatically translate into stronger health service supply capacity; its role depends on urban development stage, resource-allocation mechanisms, and the degree of coordination in public service planning. Therefore, RQ5 regarding city-type heterogeneity and nonlinear boundaries is supported.

3.7 Robustness checks

Figure 15 shows the main robustness-check results. When the lagged eco-sanitation environment is replaced with contemporaneous eco-sanitation environment, the coefficient remains significantly negative. When the dependent variable is replaced with the three component indicators of health service supply capacity, the eco-sanitation environment has a weakly significant negative effect on the number of licensed physicians per 10,000 persons, but insignificant effects on beds and institutional density. After excluding

municipalities, the core coefficient remains significant. In the energy-extended and CEADs extended samples, the eco-sanitation environment coefficient is also significantly negative.

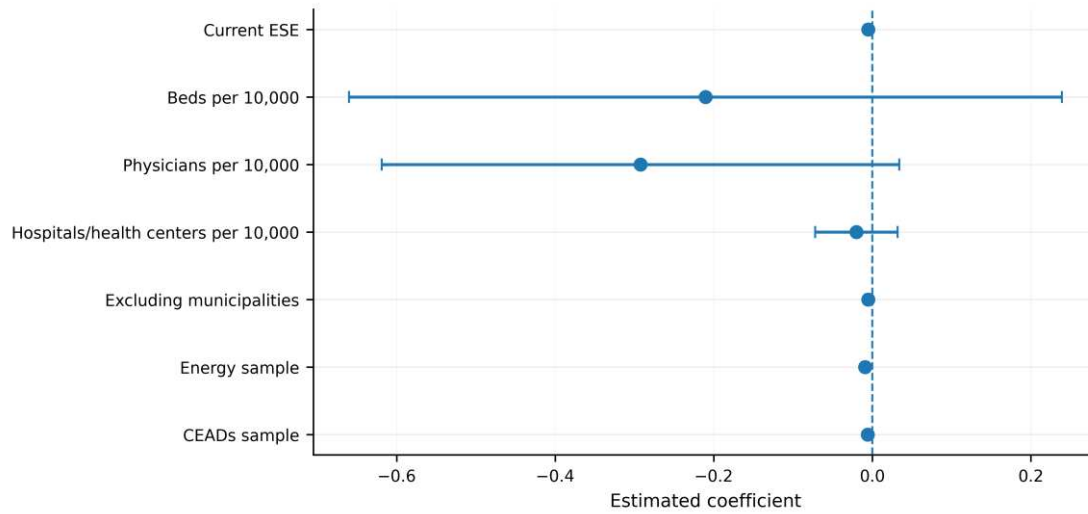


Figure 15 Core coefficient estimates in robustness checks

The robustness checks indicate that the baseline results are not driven by a specific sample, municipalities, or the lag specification. However, results using alternative dependent variables also suggest that the association between the eco-sanitation environment and the composite health service supply capacity index mainly arises from changes in health workforce and the composite index rather than from a single indicator such as beds or number of institutions. This further indicates that urban health service supply capacity is a multidimensional supply-side capacity and cannot be explained by a single health resource indicator. It should be emphasized that because this study uses city panel fixed effects models, the results mainly represent directional evidence under within-city changes rather than strict natural-experiment causal effects.

Further robustness checks show that when the entropy-weighted eco-sanitation environment index is replaced with an equal-weighted index, the core coefficient is -0.0050 (SE = 0.0018) and significant at the 1% level. When it is replaced with a PCA index, the core coefficient is -0.0035 (SE = 0.0017) and significant at the 5% level, indicating that the results do not depend on a single weighting method. With two-way clustered standard errors by city and year, the core coefficient remains -0.0053 (SE = 0.0026) and significant at the 5% level, indicating that the standard error specification does not change the main conclusion. For placebo tests, the coefficient of future one-year ESE is -0.0042 (SE = 0.0021), weakly significant at the 5% level. After including one-year, two-year, and three-year leads of ESE, the Wald statistic for the joint test of multi-period leads is 7.53, with a p-value of 0.057. The joint test does not pass the 5% significance level, but is close to the 10% level. This result suggests that the risk of pre-trends is weaker than indicated by the single lead test, but unobserved dynamic factors still cannot be completely excluded. Therefore, the baseline estimates are interpreted as conditional within-city associations and directional evidence rather than strict causal effects.

3.8 Interpretable machine learning results

Table 7 reports the predictive performance of the machine learning models. To avoid temporal leakage, observations from 2018 and earlier are used as the training set, and observations from 2019 onward are used as the test set. The random forest model achieves an R^2 of 0.5936, an RMSE of 0.0763, and an MAE of 0.0589 on the test set, outperforming the gradient boosting model. Model hyperparameters are selected using the time-blocked cross-validation procedure described in Section 2 and then fixed to ensure reproducibility of variable importance and SHAP explanations. Figure 16 further shows that the random forest model has better overall performance in explaining nonlinear differences in urban health service

supply capacity.

Table 7 Predictive performance of machine learning models

Model	R ²	RMSE	MAE	Train N	Test N
Random forest	0.5936	0.0763	0.0589	3665	793
Gradient boosting	0.5637	0.0791	0.0630	3665	793

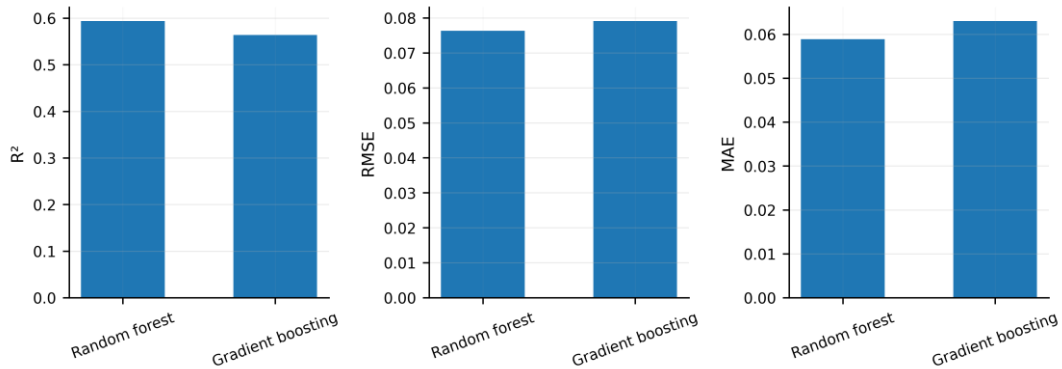


Figure 16 Comparison of predictive performance of machine learning models

Figure 17 reports SHAP-based variable importance for the random forest model. The results show that per capita GDP, population size, the tertiary industry share, and loan balance as a share of GDP are the most important variables explaining differences in urban health service supply capacity. The importance of eco-sanitation-related variables is lower than that of economic and population variables, but pollution pressure and green infrastructure still rank among the explanatory features, indicating that eco-sanitation variables contribute to explaining health service supply capacity, although their strength is weaker than that of economic development, population size, and industrial-structure factors.

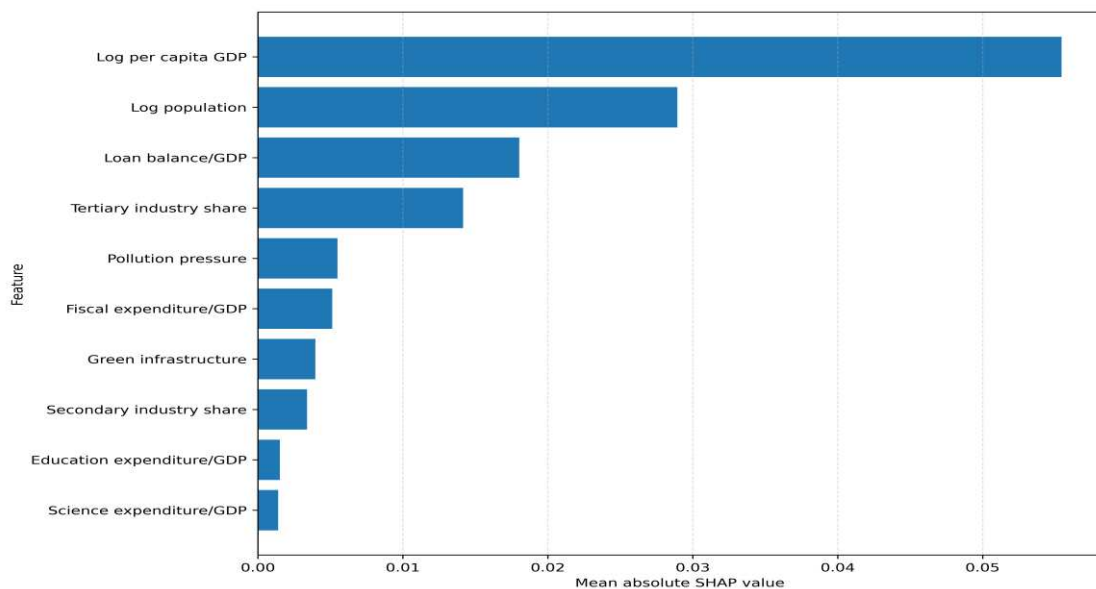


Figure 17 SHAP variable importance ranking

Figure 18 further shows the quantile dependence relationships between the eco-sanitation environment, municipal sanitation facilities, pollution pressure, and health service supply capacity. The marginal relationships of the eco-sanitation environment and municipal sanitation facilities tend to flatten at high values, while pollution pressure shows a clear nonlinear increasing relationship with health service supply

capacity. This is consistent with the earlier interpretation of pollution pressure as a resource-allocation response and also suggests that average effects in conventional linear models may conceal differences across cities at different stages of development.

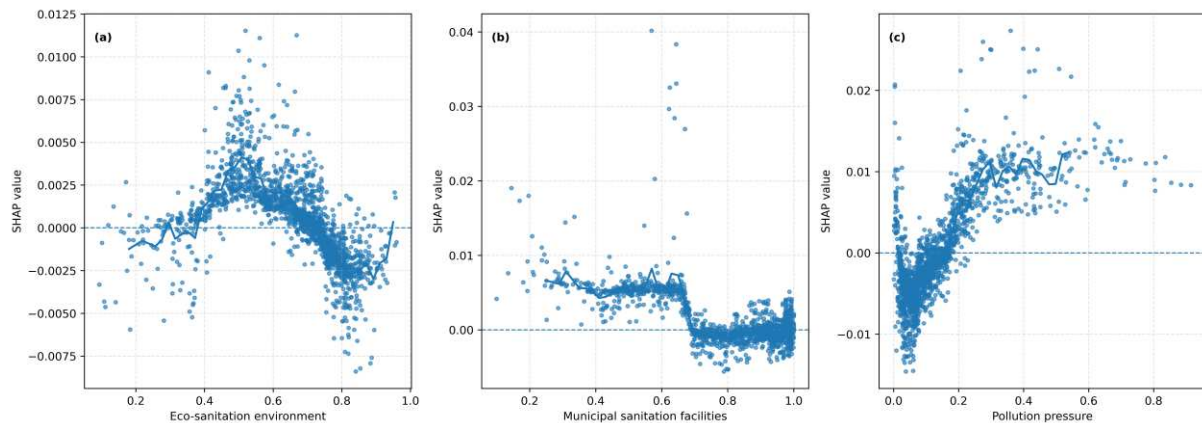


Figure 18 Nonlinear dependence between key variables and health service supply capacity

Finally, Figure 19 summarizes the empirical evidence for the five research questions. The results show that eco-sanitation improvement was not stably converted into stronger health service supply capacity; its internal dimensions played differentiated roles; the transmission chain through urban construction investment was weak; pollution pressure, energy intensity, and carbon-emission conditions modified the marginal relationship; and city-type heterogeneity and nonlinear boundaries were evident. Overall, the core finding of this study is not that the eco-sanitation environment has a unidirectional promoting effect, but that structural mismatches exist between eco-sanitation improvement and the enhancement of health service supply capacity.

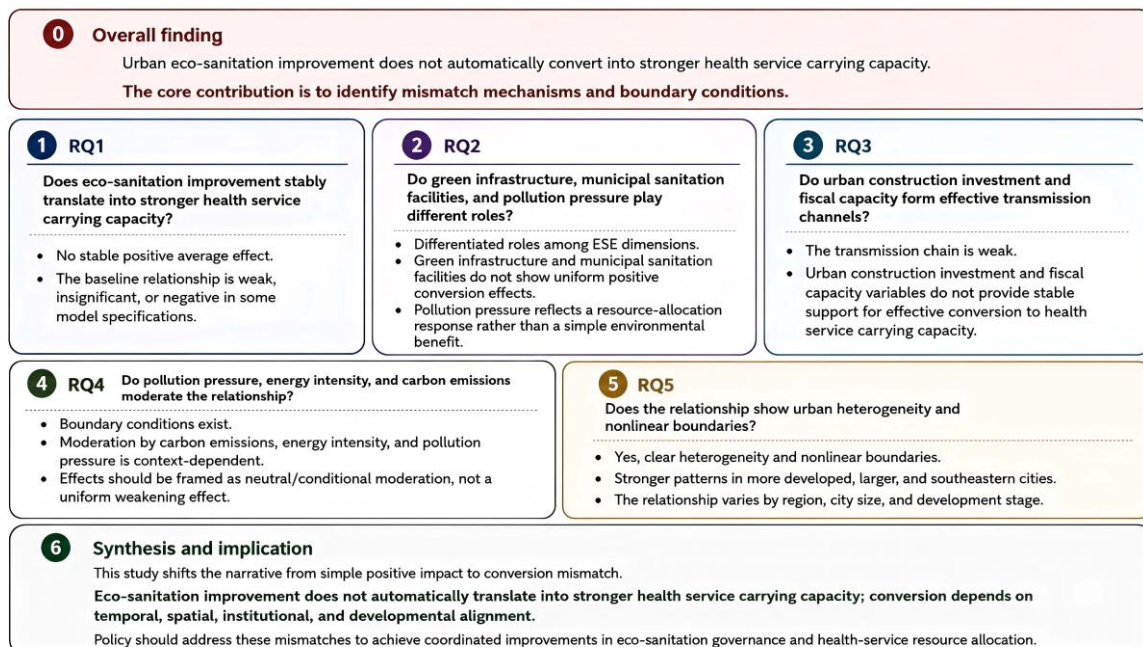


Figure 19 Summary of empirical evidence for the research questions

4 Discussion

4.1 From positive-effect assumptions to mismatch mechanism identification

The core finding of this study is not that eco-sanitation improvement has a stable positive effect on health service supply capacity, but that there is a clear conversion mismatch between the two. During 2002-2024, the eco-sanitation environment, green infrastructure, and municipal sanitation facilities of Chinese cities improved continuously, and pollution pressure generally declined; however, health service supply capacity improved more slowly and relied mainly on the expansion of beds and licensed physicians rather than on a synchronous increase in the number of medical institutions. The two-way fixed effects model further shows that the lagged eco-sanitation environment does not exhibit a stable positive average effect. This finding suggests that although eco-sanitation construction can improve the foundation for public health, its conversion into health service supply capacity is jointly constrained by healthcare resource allocation, interdepartmental coordination, population demand, primary service networks, and the stage of urban development.

The theoretical implication of this study is therefore to move eco-sanitation research beyond the single question of whether it produces positive health effects and toward a health service governance perspective that asks whether environmental improvement can be converted into health service supply capacity. Eco-sanitation improvement can reduce health risks, improve urban livability, and strengthen the foundations of public health, but it does not automatically expand beds, physicians, primary service outlets, or medical institutions. Health service supply capacity is more directly affected by population size, fiscal capacity, service network planning, health workforce allocation, and the service functions of cities.

4.2 Temporal, spatial, and institutional mismatches

The baseline regressions, dimension-specific models, and mechanism tests jointly indicate temporal, spatial, and institutional mismatches between eco-sanitation improvement and health service supply capacity. First, the two systems adjust over different time horizons. Green space, wastewater treatment, waste treatment, and water and gas facilities are often promoted by urban construction and environmental governance systems, whereas the expansion of health resources involves hospital siting, physician training, and fiscal budgeting, which require longer cycles. Second, their spatial distributions do not fully coincide. Cities with rapid eco-sanitation improvement are not necessarily cities where medical resources are concentrated; regional centers, population-inflow cities, and cities with higher administrative status are more likely to attract health resources. Third, governance is segmented across departments. The eco-sanitation environment is mainly promoted by departments responsible for housing and urban-rural development, urban management, and ecology and environment, whereas health service supply capacity is shaped by health, finance, and personnel systems. Differences in departmental objectives and budget channels weaken policy coordination.

The mechanism tests do not support a stable transmission chain of eco-sanitation environment -> urban construction investment -> health service supply capacity. The eco-sanitation environment has a weak positive relationship with landscaping investment, a negative relationship with wastewater treatment investment, and insignificant effects on urban maintenance and construction expenditure and sanitation and environmental appearance investment. After these investment and expenditure variables enter the health service supply capacity model, they still do not show stable effects. Thus, the direct effects of the mechanism variables on health service supply capacity do not reach conventional significance levels, and the transmission chain is not supported under current data conditions. This suggests that if eco-sanitation infrastructure construction is not jointly planned with medical service layouts, primary care networks, and fiscal resource allocation, it is unlikely to naturally generate stronger health service supply capacity.

4.3 Pollution pressure, energy-carbon conditions, and resource-allocation responses

The dimension-specific results show a positive coefficient for pollution pressure, but this should not be interpreted as evidence that pollution has health benefits. A more reasonable interpretation is a pressure-resource allocation response: cities with higher pollution pressure usually have stronger industrial bases, larger populations, higher fiscal capacity, and more concentrated healthcare demand, and therefore may allocate more beds, physicians, and medical institutions. Higher pollution pressure also increases latent health service demand and may prompt governments and markets to strengthen the supply of health service resources.

The moderating effects of energy and carbon emissions also reflect differences in stages of urban development. The interaction terms for pollution pressure, energy intensity per unit GDP, and per capita energy consumption are unstable, whereas the CEADs city-level carbon-emission interaction term is positive and significant. This indicates that energy intensity and carbon emissions may modify the marginal relationship between the eco-sanitation environment and health service supply capacity, but the direction is not a simple negative constraint. One possible explanation is that high-emission cities often simultaneously have stronger industrial foundations, population agglomeration, fiscal capacity, and public service demand, making eco-sanitation improvement more likely to occur alongside public service expansion. Therefore, the positive moderation effect of carbon emissions reflects the resource-allocation capacity behind high-emission cities rather than any beneficial effect of carbon emissions themselves.

4.4 Urban heterogeneity and nonlinear boundaries

Heterogeneity analysis shows that the relationship between the eco-sanitation environment and health service supply capacity mainly appears in eastern and central cities, high-GDP cities, large-population cities, and cities on the southeastern side of the Hu Huanyong line, while it is insignificant in low-GDP cities, small-population cities, and cities on the northwestern side. This suggests that the role of the eco-sanitation environment depends on urban development stage and resource-allocation conditions. In mature cities, both eco-sanitation infrastructure and health service systems have reached relatively high levels, and their relationship is more likely to take the form of structural adjustment, diminishing marginal returns, or resource substitution. In less developed cities, eco-sanitation improvement alone is insufficient to promote medical resource expansion, and health service supply capacity remains constrained by fiscal capacity, population size, and regional service functions.

Interpretable machine learning further supports this interpretation. The random forest model shows good predictive performance. SHAP results indicate that per capita GDP, population size, the tertiary industry share, and financial development are the main variables explaining differences in health service supply capacity. Eco-sanitation variables have relatively lower importance, but green infrastructure and pollution pressure still contribute to model explanations. SHAP dependence relationships show that the marginal effects of the eco-sanitation environment and municipal sanitation facilities flatten at high values, while pollution pressure has a nonlinear relationship with health service supply capacity. Thus, the evidence supports RQ5 regarding urban heterogeneity and nonlinear boundaries.

4.5 Policy implications and limitations

The results indicate that healthy city development cannot rely solely on eco-sanitation improvement; instead, ecological governance, municipal infrastructure, and health service resource allocation should be incorporated into the same health service governance framework. Green space, water supply, wastewater treatment, waste management, and pollution control can improve the public health foundation, but without coordinated hospital location planning, primary care service outlets, physician allocation, service-radius management, and fiscal investment, their contribution to health service supply capacity cannot be fully realized. From the perspective of healthcare delivery, accessibility, and governance, eco-sanitation improvement should be understood as an external condition of the health system rather than a substitute for

strengthening health service capacity.

Policy strategies should therefore be differentiated by city type and linked directly to health service governance actions. For eastern, densely populated, and high-GDP cities, the priority should not be merely to increase the coverage of eco-sanitation facilities, but to optimize the spatial matching among ecological spaces, primary care outlets, and high-level medical resources, and to promote the balanced distribution of medical services from core urban areas to peripheral and population-growth areas. For central and western, small and medium-sized, and fiscally weaker cities, filling gaps in basic health resources should be placed on the same level of importance as eco-sanitation improvement; priority should be given to increasing primary physicians, beds, and health-center service nodes, while improving the quality of wastewater treatment, harmless waste treatment, and green-space provision. For cities with high pollution pressure and carbon emissions, environmental governance, low-carbon transition, and health service expansion should be incorporated into project evaluation simultaneously, with latent health demand, primary healthcare capacity, health workforce allocation, and pollution-control performance treated as joint constraints. At the cross-departmental level, a joint monitoring mechanism for the eco-sanitation environment and health service supply capacity should be established. Green space, wastewater treatment, waste management, pollution control, beds, physicians, and primary care service outlets should be included in a unified performance framework to encourage departments of urban construction, ecology and environment, health, and finance to share data, budget objectives, and service coverage responsibilities.

This study has several limitations. First, health service supply capacity is constructed from beds, physicians, and hospitals/health centers. Although these indicators reflect supply-side health service capacity and show moderate correlation with the number of all health institutions per 10,000 persons, they cannot directly represent quality of care, actual service utilization, or residents' health outcomes. Second, city-level data cannot capture differences in healthcare accessibility across neighborhoods, communities, and population groups. Third, the CEADs carbon-emission data do not fully cover the same years as the main sample, and some health service indicators and mechanism variables contain missing values. Fourth, this study is not a strict natural experimental design, and the results should be interpreted as conditional associations and directional evidence. Fifth, the current data package does not include consistent city-center coordinates, adjacency matrices, or commuting-flow matrices over time, making it difficult to construct a stable spatial weights matrix; therefore, spatial autocorrelation and spatial panel tests such as Moran's I, SAR, or SEM have not yet been conducted. This means that cross-city spatial spillovers remain a potential source of unaddressed bias. Future research could combine resident health surveys, medical insurance claims, geocoded medical institution locations, remote-sensing measures of green-space exposure, urban spatial weights matrices, and quasi-natural experiments to further test the effects of eco-sanitation improvement on health service supply capacity and resident health outcomes.

5 Conclusion

This study does not assume that eco-sanitation improvement is a sufficient condition for enhancing health service supply capacity. Instead, using panel data for Chinese prefecture-level and above cities from 2002 to 2024, it examines whether eco-sanitation improvement can be converted into health service supply capacity and identifies the mismatch mechanisms and boundary conditions in this process. The study constructs indices of urban health service supply capacity, eco-sanitation environment, green infrastructure, municipal sanitation facilities, and pollution pressure, and combines two-way fixed effects models, mechanism tests, moderation analysis, heterogeneity analysis, and interpretable machine learning.

The results show that the eco-sanitation environment, green infrastructure, and municipal sanitation facilities in Chinese cities improved continuously, and pollution pressure generally declined; however, health service supply capacity improved more slowly and was mainly driven by the expansion of beds and licensed physicians rather than by a synchronous increase in the number of medical institutions. The baseline regressions show that the lagged eco-sanitation environment does not exhibit a stable positive average effect on health service supply capacity, indicating that eco-sanitation improvement does not automatically translate into stronger healthcare resource supply capacity. Instead, temporal, spatial, and institutional mismatches exist between the two.

Dimension-specific results show that green infrastructure, municipal sanitation facilities, and pollution pressure play different roles. The positive coefficient of pollution pressure more likely reflects a healthcare resource-allocation response in high-pollution and high-demand cities, rather than any positive role of pollution itself. The mechanism tests do not identify a stable transmission chain of eco-sanitation environment -> urban construction investment -> health service supply capacity. Extended analyses of energy and carbon emissions further indicate that ecological governance, low-carbon transition, and health service resource allocation need to be promoted in a coordinated manner.

Heterogeneity and interpretable machine learning results further show clear city-type differences and nonlinear boundaries in the relationship between the eco-sanitation environment and health service supply capacity. The relationship mainly appears in eastern and central cities, high-GDP cities, large-population cities, and cities on the southeastern side of the Hu Huanyong line. SHAP results show that per capita GDP, population size, industrial structure, and financial development are the main factors explaining differences in health service supply capacity, while eco-sanitation variables provide supplementary explanatory power.

The innovations of this study are as follows. First, it extends the research object from residents' health outcomes to supply-side urban health service capacity, clarifies that the operational definition mainly captures the stock of health resources and service network density, and shifts the focus from positive-effect identification to conversion-relationship identification. Second, it integrates green infrastructure, municipal sanitation facilities, pollution pressure, energy intensity, and carbon emissions into a unified analytical framework, revealing mismatch mechanisms in the conversion from eco-sanitation improvement to health service supply capacity. Third, it combines panel econometric models with interpretable machine learning to characterize urban heterogeneity and nonlinear boundaries while identifying average relationships. Overall, healthy city development should promote coordination between eco-sanitation governance and health service resource allocation, and should avoid equating ecological infrastructure improvement with stronger health service supply capacity.

List of abbreviations

CEADs: Carbon Emission Accounts and Datasets
CO₂: Carbon dioxide
ESE: Eco-sanitation environment index
FE: Fixed effects
GDP: Gross domestic product
GII: Green infrastructure index
GIS: Geographic information system
HSC: Health service supply capacity index
IND2: Share of secondary industry in gross domestic product
LIME: Local Interpretable Model-Agnostic Explanations
MAE: Mean absolute error
MSI: Municipal sanitation facility index
N: Number of observations
NO_x: Nitrogen oxides
PCA: Principal component analysis
PGDP: Per capita gross domestic product
PM_{2.5}: Particulate matter with an aerodynamic diameter of 2.5 µm or less
PPI: Pollution pressure index
R&D: Research and development
R²: Coefficient of determination
RMSE: Root mean squared error
RQ: Research question
SAR: Spatial autoregressive model
SD: Standard deviation
SE: Standard error
SEM: Spatial error model
SHAP: SHapley Additive exPlanations
SO₂: Sulfur dioxide
tce: Tonne of coal equivalent
VIF: Variance inflation factor

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki, where applicable. Ethics approval and informed consent to participate were not required because the study exclusively used aggregated, publicly available city-level statistical data and did not involve human participants, identifiable or individual-level human data, clinical records, human tissue, or biological samples.

Consent for publication

Not applicable.

Availability of data and materials

The processed dataset and supporting materials used in this study are publicly available in the Mendeley Data repository, Version 1, at <https://doi.org/10.17632/hw4j6xv99v.1>[60]. The repository contains the analytical panel data, variable definitions, index weights, model specifications, statistical results, and machine-learning outputs. The original source data were obtained from the official statistical sources described in the Methods section and the source manifest included in the repository.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors gratefully acknowledge the support of the National Natural Science Foundation of China (Grant No. 52578027).

Authors' contributions

Z.P. conceived and designed the study, developed the research framework, constructed the dataset, performed the econometric and machine learning analyses, interpreted the results, and drafted the manuscript. L.W. contributed to data collection and organization, variable checking, interpretation of the empirical findings, and critical revision of the manuscript. Both authors reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgments

The authors wish to express their sincere gratitude to all individuals and institutions that contributed to this study. Special thanks are extended to the data providers whose publicly available statistical materials made this research possible. The authors also deeply appreciate the constructive feedback and support from colleagues and peers during the preparation of this study.

References

- [1] Zhu YG, Jones KC. Urbanisation and health in China. *Lancet*. 2010;376(9737):232-3. doi:10.1016/S0140-6736(10)61142-7.
- [2] Galea S, Vlahov D. Urban health: evidence, challenges, and directions. *Annu Rev Public Health*. 2005;26:341-65. doi:10.1146/annurev.publhealth.26.021304.144708.
- [3] Pinchoff J, Mills CW, Balk D. Urbanization and health: the effects of the built environment on chronic disease risk factors among women in Tanzania. *PLoS One*. 2020;15(11):e0241810. doi:10.1371/journal.pone.0241810.
- [4] Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. *Lancet Glob Health*. 2018;6(11):e1196-e1252. doi:10.1016/S2214-109X(18)30386-3.
- [5] Donabedian A. The quality of care: how can it be assessed? *JAMA*. 1988;260(12):1743-8. doi:10.1001/jama.260.12.1743.
- [6] Green LA, Fryer GE, Yawn BP, Lanier D, Dovey SM. The ecology of medical care revisited. *N Engl J Med*. 2001;344(26):2021-5. doi:10.1056/NEJM200106283442611.
- [7] Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav*. 1995;36(1):1-10. doi:10.2307/2137284.

981 [8] Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Beech R, et al. What
982 does "access to health care" mean? *J Health Serv Res Policy*. 2002;7(3):186-8.
983 doi:10.1258/135581902760082517.

984 [9] Levesque JF, Harris MF, Russell G. Patient-centred access to health care: conceptualising
985 access at the interface of health systems and populations. *Int J Equity Health*. 2013;12:18.
986 doi:10.1186/1475-9276-12-18.

987 [10] Anand S, Bärnighausen T. Human resources and health outcomes: cross-country
988 econometric study. *Lancet*. 2004;364(9445):1603-9. doi:10.1016/S0140-6736(04)17313-3.

989 [11] Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health.
990 *Milbank Q*. 2005;83(3):457-502. doi:10.1111/j.1468-0009.2005.00409.x.

991 [12] Yip W, Hsiao W. Harnessing the privatisation of China's fragmented health-care
992 delivery. *Lancet*. 2014;384(9945):805-18. doi:10.1016/S0140-6736(14)61120-X.

993 [13] Yip W, Fu H, Chen AT, Zhai T, Jian W, Xu R, et al. 10 years of health-care reform in
994 China: progress and gaps in Universal Health Coverage. *Lancet*. 2019;394(10204):1192-204.
995 doi:10.1016/S0140-6736(19)32136-1.

996 [14] Chen J, Lin Z, Li LA, Li J, Wang Y, Pan Y, et al. Ten years of China's new healthcare
997 reform: a longitudinal study on changes in health resources. *BMC Public Health*.
998 2021;21:2272. doi:10.1186/s12889-021-12248-9.

999 [15] Barber RM, Fullman N, Sorensen RJD, Bollyky T, McKee M, Nolte E, et al. Healthcare
1000 Access and Quality Index based on mortality from causes amenable to personal health care in
1001 195 countries and territories, 1990-2015. *Lancet*. 2017;390(10091):231-66.
1002 doi:10.1016/S0140-6736(17)30818-8.

1003 [16] Fullman N, Yearwood J, Abay SM, Abbafati C, Abd-Allah F, Abdela J, et al. Measuring
1004 performance on the Healthcare Access and Quality Index for 195 countries and territories and
1005 selected subnational locations. *Lancet*. 2018;391(10136):2236-71.
1006 doi:10.1016/S0140-6736(18)30994-2.

1007 [17] Lozano R, Fullman N, Mumford JE, Knight M, Barthelemy CM, Abbafati C, et al.
1008 Measuring universal health coverage based on an index of effective coverage of health
1009 services in 204 countries and territories, 1990-2019. *Lancet*. 2020;396(10258):1250-84.
1010 doi:10.1016/S0140-6736(20)30750-9.

1011 [18] Hartig T, Mitchell R, de Vries S, Frumkin H. Nature and health. *Annu Rev Public*
1012 *Health*. 2014;35:207-28. doi:10.1146/annurev-publhealth-032013-182443.

1013 [19] Markevych I, Schoierer J, Hartig T, Chudnovsky A, Hystad P, Dzhambov AM, et al.
1014 Exploring pathways linking greenspace to health: theoretical and methodological guidance.
1015 *Environ Res*. 2017;158:301-17. doi:10.1016/j.envres.2017.06.028.

1016 [20] Twohig-Bennett C, Jones A. The health benefits of the great outdoors: a systematic
 1017 review and meta-analysis of greenspace exposure and health outcomes. *Environ Res.*
 1018 2018;166:628-37. doi:10.1016/j.envres.2018.06.030.

1019 [21] Gascon M, Triguero-Mas M, Martínez D, Dadvand P, Rojas-Rueda D, Plasència A, et al.
 1020 Residential green spaces and mortality: a systematic review. *Environ Int.* 2016;86:60-7.
 1021 doi:10.1016/j.envint.2015.10.013.

1022 [22] James P, Banay RF, Hart JE, Laden F. A review of the health benefits of greenness. *Curr*
 1023 *Epidemiol Rep.* 2015;2:131-42. doi:10.1007/s40471-015-0043-7.

1024 [23] Vienneau D, de Hoogh K, Faeh D, Kaufmann M, Wunderli JM, Röösli M. More than
 1025 clean air and tranquillity: residential green is independently associated with decreasing
 1026 mortality. *Environ Int.* 2017;108:176-84. doi:10.1016/j.envint.2017.08.012.

1027 [24] Tzoulas K, Korpela K, Venn S, Yli-Pelkonen V, Kaźmierczak A, Niemelä J, et al.
 1028 Promoting ecosystem and human health in urban areas using green infrastructure: a literature
 1029 review. *Landsc Urban Plan.* 2007;81(3):167-78. doi:10.1016/j.landurbplan.2007.02.001.

1030 [25] Kondo MC, Fluehr JM, McKeon T, Branas CC. Urban green space and its impact on
 1031 human health. *Int J Environ Res Public Health.* 2018;15(3):445. doi:10.3390/ijerph15030445.

1032 [26] Maas J, Verheij RA, Groenewegen PP, de Vries S, Spreeuwenberg P. Green space,
 1033 urbanity, and health: how strong is the relation? *J Epidemiol Community Health.*
 1034 2006;60(7):587-92. doi:10.1136/jech.2005.043125.

1035 [27] Mitchell R, Popham F. Effect of exposure to natural environment on health inequalities:
 1036 an observational population study. *Lancet.* 2008;372(9650):1655-60.
 1037 doi:10.1016/S0140-6736(08)61689-X.

1038 [28] Dockery DW, Pope CA 3rd, Xu X, Spengler JD, Ware JH, Fay ME, et al. An association
 1039 between air pollution and mortality in six U.S. cities. *N Engl J Med.* 1993;329(24):1753-9.
 1040 doi:10.1056/NEJM199312093292401.

1041 [29] Pope CA 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. Lung cancer,
 1042 cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA.*
 1043 2002;287(9):1132-41. doi:10.1001/jama.287.9.1132.

1044 [30] Hoek G, Krishnan RM, Beelen R, Peters A, Ostro B, Brunekreef B, et al. Long-term air
 1045 pollution exposure and cardio-respiratory mortality: a review. *Environ Health.* 2013;12:43.
 1046 doi:10.1186/1476-069X-12-43.

1047 [31] Brook RD, Rajagopalan S, Pope CA 3rd, Brook JR, Bhatnagar A, Diez-Roux AV, et al.
 1048 Particulate matter air pollution and cardiovascular disease. *Circulation.*
 1049 2010;121(21):2331-78. doi:10.1161/CIR.0b013e3181dbee1.

1050 [32] Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, et al. Estimates and
 1051 25-year trends of the global burden of disease attributable to ambient air pollution. *Lancet.*
 1052 2017;389(10082):1907-18. doi:10.1016/S0140-6736(17)30505-6.

1053 [33] Burnett R, Chen H, Szyszkowicz M, Fann N, Hubbell B, Pope CA 3rd, et al. Global
1054 estimates of mortality associated with long-term exposure to outdoor fine particulate matter.
1055 Proc Natl Acad Sci U S A. 2018;115(38):9592-7. doi:10.1073/pnas.1803222115.

1056 [34] Landrigan PJ, Fuller R, Acosta NJR, Adeyi O, Arnold R, Basu N, et al. The Lancet
1057 Commission on pollution and health. Lancet. 2018;391(10119):462-512.
1058 doi:10.1016/S0140-6736(17)32345-0.

1059 [35] Lelieveld J, Evans JS, Fnais M, Giannadaki D, Pozzer A. The contribution of outdoor air
1060 pollution sources to premature mortality on a global scale. Nature. 2015;525:367-71.
1061 doi:10.1038/nature15371.

1062 [36] Zhang Q, Jiang X, Tong D, Davis SJ, Zhao H, Geng G, et al. Transboundary health
1063 impacts of transported global air pollution and international trade. Nature. 2017;543:705-9.
1064 doi:10.1038/nature21712.

1065 [37] Chen Y, Ebenstein A, Greenstone M, Li H. Evidence on the impact of sustained
1066 exposure to air pollution on life expectancy from China's Huai River policy. Proc Natl Acad
1067 Sci U S A. 2013;110(32):12936-41. doi:10.1073/pnas.1300018110.

1068 [38] Rohde RA, Muller RA. Air pollution in China: mapping of concentrations and sources.
1069 PLoS One. 2015;10(8):e0135749. doi:10.1371/journal.pone.0135749.

1070 [39] Matus K, Nam KM, Selin NE, Lamsal LN, Reilly JM, Paltsev S. Health damages from
1071 air pollution in China. Glob Environ Change. 2012;22(1):55-66.
1072 doi:10.1016/j.gloenvcha.2011.08.006.

1073 [40] Ebenstein A, Fan M, Greenstone M, He G, Zhou M. New evidence on the impact of
1074 sustained exposure to air pollution on life expectancy from China's Huai River Policy. Proc
1075 Natl Acad Sci U S A. 2017;114(39):10384-9. doi:10.1073/pnas.1616784114.

1076 [41] Fewtrell L, Kaufmann RB, Kay D, Enanoria W, Haller L, Colford JM. Water, sanitation,
1077 and hygiene interventions to reduce diarrhoea in less developed countries: a systematic
1078 review and meta-analysis. Lancet Infect Dis. 2005;5(1):42-52.
1079 doi:10.1016/S1473-3099(04)01253-8.

1080 [42] Clasen T, Schmidt WP, Rabie T, Roberts I, Cairncross S. Interventions to improve water
1081 quality for preventing diarrhoea: systematic review and meta-analysis. BMJ. 2007;334:782.
1082 doi:10.1136/bmj.39118.489931.BE.

1083 [43] Mara D, Lane J, Scott B, Trouba D. Sanitation and health. PLoS Med. 2010;7(11):e1000363. doi:10.1371/journal.pmed.1000363.

1084 [44] Bartram J, Cairncross S. Hygiene, sanitation, and water: forgotten foundations of health.
1085 PLoS Med. 2010;7(11):e1000367. doi:10.1371/journal.pmed.1000367.

1086 [45] Prüss-Ustün A, Bartram J, Clasen T, Colford JM, Cumming O, Curtis V, et al. Burden of
1087 disease from inadequate water, sanitation and hygiene in low- and middle-income settings.
1088 Trop Med Int Health. 2014;19(8):894-905. doi:10.1111/tmi.12329.

1089

1090 [46] Montgomery MA, Elimelech M. Water and sanitation in developing countries: including
1091 health in the equation. *Environ Sci Technol*. 2007;41(1):17-24. doi:10.1021/es072435t.

1092 [47] Hutton G, Haller L, Bartram J. Global cost-benefit analysis of water supply and
1093 sanitation interventions. *J Water Health*. 2007;5(4):481-502. doi:10.2166/wh.2007.009.

1094 [48] Penchansky R, Thomas JW. The concept of access: definition and relationship to
1095 consumer satisfaction. *Med Care*. 1981;19(2):127-40.
1096 doi:10.1097/00005650-198102000-00001.

1097 [49] Luo W, Wang F. Measures of spatial accessibility to health care in a GIS environment:
1098 synthesis and a case study in the Chicago region. *Environ Plan B Plan Des*.
1099 2003;30(6):865-84. doi:10.1068/b29120.

1100 [50] Wang F, Luo W. Assessing spatial and nonspatial factors for healthcare access: towards
1101 an integrated approach to defining health professional shortage areas. *Health Place*.
1102 2005;11(2):131-46. doi:10.1016/j.healthplace.2004.02.003.

1103 [51] Luo W, Qi Y. An enhanced two-step floating catchment area method for measuring
1104 spatial accessibility to primary care physicians. *Health Place*. 2009;15(4):1100-7.
1105 doi:10.1016/j.healthplace.2009.06.002.

1106 [52] Shan Y, Guan D, Hubacek K, Zheng B, Davis SJ, Jia L, et al. City-level climate change
1107 mitigation in China. *Sci Adv*. 2018;4(6):eaq0390. doi:10.1126/sciadv.aaq0390.

1108 [53] Shan Y, Huang Q, Guan D, Hubacek K. China CO₂ emission accounts 1997-2015. *Sci*
1109 *Data*. 2018;5:170201. doi:10.1038/sdata.2017.201.

1110 [54] Liu Z, Guan D, Wei W, Davis SJ, Ciais P, Bai J, et al. Reduced carbon emission
1111 estimates from fossil fuel combustion and cement production in China. *Nature*.
1112 2015;524:335-8. doi:10.1038/nature14677.

1113 [55] Mi Z, Meng J, Guan D, Shan Y, Song M, Wei YM, et al. Chinese CO₂ emission flows
1114 have reversed since the global financial crisis. *Nat Commun*. 2017;8:1712.
1115 doi:10.1038/s41467-017-01820-w.

1116 [56] Breiman L. Random forests. *Mach Learn*. 2001;45:5-32. doi:10.1023/A:1010933404324.

1117 [57] Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*.
1118 2001;29(5):1189-232. doi:10.1214/aos/1013203451.

1119 [58] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural*
1120 *Inf Process Syst*. 2017;30:4765-74.

1121 [59] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions
1122 of any classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on*
1123 *Knowledge Discovery and Data Mining*; 2016. p. 1135-44. doi:10.1145/2939672.2939778.

1124 [60] Peng Z. Data for: mismatch between urban eco-sanitation improvement and health
1125 service supply capacity: evidence from Chinese city panel data and interpretable machine
1126 learning [dataset]. Mendeley Data. 2026. Version 1. doi:10.17632/hw4j6xv99v.1.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [HealthServiceCarryingCapacityDatasetandResults.xlsx](#)