Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Urban-rural residence as the dominant source of inequality in the effectiveness of China's multi-tiered medical security system: an individual-level concentration index decomposition of CHARLS, 2011-2020

Yilong Wang
Northwest Minzu University
Yuxuan Wang
Northwest Minzu University
Aihong He
hah631020@163.com
Lanzhou Bowen College of Science and Technology

Research Article

Keywords: multi-tiered medical security, health equity, concentration index, catastrophic health expenditure, financial protection, middle-aged and older adults, China

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11483094/v1

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

Additional Declarations: No competing interests reported.

Urban-rural residence as the dominant source of inequality in the effectiveness of China's multi-tiered medical security system: an individual-level concentration index decomposition of CHARLS, 2011-2020

Yilong Wang[1], Yuxuan Wang[1], Aihong He[2,*]

1.School of Management, Northwest Minzu University, Lanzhou, Gansu, China

2. School of Economics and Management, Lanzhou Bowen College of Science and Technology, Lanzhou, Gansu, China

*Corresponding author: Aihong He, School of Economics and Management, Lanzhou Bowen College of Science and Technology, Lanzhou, Gansu, China. Email: hah631020@163.com

Abstract

Background: China has built a multi-tiered medical security system to protect its rapidly ageing population from medical costs, but the evaluations available rest on provincial indicators. Aggregates of that kind cannot show how protection is distributed among individuals. We therefore measured the effectiveness of the system for middle-aged and older Chinese adults and traced where the inequality in it originates.

Methods: We used five waves of the China Health and Retirement Longitudinal Study (2011-2020), comprising 93953 person-wave observations on 25390 respondents aged 45 and above. We constructed an individual-level index of multi-tiered medical security effectiveness from 11 indicators covering four dimensions, namely adequacy, equity, synergy and safety. Weights combined the Analytic Hierarchy Process with the entropy weight method. We measured inequality in the index with the concentration index, ranked by household consumption per capita, and decomposed it by the Wagstaff method into the contributions of urban-rural residence, education, macro-region and other determinants. We estimated determinants by

pooled and two-way fixed-effects models, with standard errors clustered at the individual level.

Results: Effectiveness rose from 27.52 in 2011 to 28.79 in 2020. Safety scored highest throughout, at 0.89 or above. Adequacy was intermediate, at 0.24-0.27. Equity (0.12-0.13) and synergy (0.09-0.11) remained low. Synergy grew fastest (17.2%) and equity slowest (3.8%). Inequality was pro-rich in every wave, and the concentration index rose from 0.020 in 2011 to 0.0240 in 2020. Urban-rural residence contributed 29.1% of the pooled concentration index and education 16.4%, whereas macro-region contributed -0.1%. Urban residents scored 3.02 to 4.29 points above rural residents. Household consumption per capita was positively associated with effectiveness in urban and eastern subsamples, but not in rural ones.

Conclusions: Inequality in the effectiveness of multi-tiered medical security among middle-aged and older Chinese adults is an urban-rural and individual phenomenon, not a regional one. Narrowing it calls for urban-rural equalisation and stronger supplementary and floor protection, not reallocation across regions. The finding also carries a general lesson for health equity measurement. Jurisdictional aggregates average away the within-jurisdiction variation that dominates at the individual level.

Keywords: multi-tiered medical security; health equity; concentration index; catastrophic health expenditure; financial protection; middle-aged and older adults; China

Background

China's population is ageing faster than that of almost any other large economy. At the end of 2025, 323.4 million people, or 23.0% of the national population, were aged 60 or above [1], and chronic disease, disability and long-term care needs are rising in step with the cost of medical services [2]. Whether the health financing system converts this frequent and income-heavy expenditure risk into manageable out-of-pocket (OOP) payments or into impoverishment is therefore a central test of system performance [3, 4], and one that concerns the distribution of protection as

much as its average level, since a system can raise average coverage while leaving the least advantaged households behind.

China has responded by building a multi-tiered medical security system. A basic medical insurance layer covers the urban employee and urban-rural resident schemes plus government medical care. Above it sit a medical assistance layer providing a means-tested floor for the poor, and a supplementary layer of critical illness insurance, employer schemes and commercial health insurance. The basic tier has expanded in stages since 1998 and its rural and urban resident schemes were merged from 2016 onwards [5, 6]. The supplementary tier is voluntary, so its reach depends most directly on a household's ability to pay [7, 8].

Recent policy has set out the objective of moving from fragmented programmes to an integrated architecture. The 2020 Opinions on Deepening the Reform of the Medical Security System [9] and the 2021 14th Five-Year Plan for Universal Medical Security [10] both identified unbalanced and inadequate development as the core problem, and the 2026 15th Five-Year Plan commits to fully establishing the multi-tiered system within five years [11]. Because the tiers differ in statutory status, funding source and coverage depth, a multi-tiered system can score well on one dimension while performing poorly on another, so a framework that reports only an aggregate figure conceals the imbalances that matter for policy [12, 13].

Yet the system is widely characterised as established in form but deficient in function. Basic insurance carries an overwhelming share of the burden, since enrolment exceeds 95% yet OOP spending still accounted for 27.3% of total health expenditure in 2023 [14], well above most OECD countries [15, 16]. The supplementary tiers remain thin, with commercial health insurance reimbursing only about 5% of total health expenditure [17, 18]. Protection gradients also persist across the urban-rural divide and across regions, and institutional segmentation has tended to reproduce existing inequalities [19, 20].

Multi-tiered design is intrinsic to modern social protection, because layered design and reallocated responsibility can align the roles of government, market, society and the individual [19, 21]. Liu et al. [8] show that a multitiered coverage

system strengthens financial protection for low-income households while the tiers above and below basic insurance remain thin, and Yan et al. [22], Doshmangir et al. [19] and Wang et al. [23] document how fragmentation across insurance schemes reproduces inequitable access and uneven financial burden.

Measurement practice offers two influential frameworks. Donabedian's structure-process-outcome model emphasises the instrumental rationality of system performance [24, 25], while the World Health Organization's framework and the multi-dimensional assessments that followed emphasise equity, efficiency and quality, with the trade-off between the last two now the central evaluative criterion [3, 26, 27]. For a multi-tiered system, whose tiers are complements rather than substitutes, evaluation has to be multi-dimensional by construction and must capture whether the layers actually work together [12, 20].

Two strands of empirical work are relevant. The first measures financial protection directly, typically through catastrophic health expenditure (CHE) and impoverishment due to OOP payments [28, 29]. CHARLS-based studies have documented rising CHE among Chinese households with hypertension between 2011 and 2018 and persistent pro-rich inequity in utilisation [30, 31]; CHE risk varies systematically with socioeconomic status [32, 33] and is elevated among older adults living alone with cognitive impairment [34, 35]. The second strand constructs composite indices of health system performance. Chen et al. [36] build an 11-indicator provincial index covering basic insurance, supplementary insurance and medical assistance, weighting it by AHP and the entropy method and analysing it with the Dagum Gini coefficient. Zhang, Chu and Zhao [20] develop a 37-indicator framework organised around adequacy, equity, synergy and safety for China's 31 provinces; Su et al. [13] assess the equity and efficiency of provincial health resource allocation; and Zhao and Chen [12] assess the equity and efficiency of the medical and health service system.

Evidence on the effects of multi-tiering is also accumulating. Dong et al. [18] find that supplementary private health insurance reduces vulnerability to expected poverty and catastrophic health expenditure among Chinese households. Zhang et al.

[37] find that integrating urban and rural resident insurance reshapes healthcare utilisation while a pro-rich gradient in utilisation inequality persists, and Wang and Qin [38] show that insurance integration raises the level of medical security and improves the health of rural older adults even as pro-rich health inequality remains. Multi-tiering therefore matters for both cost protection and health, but its benefits are unevenly distributed.

Three gaps limit the usefulness of this literature for policy. First, the unit of analysis is predominantly the province, and provincial indices capture the supply-side capacity of a jurisdiction rather than how protection is distributed across individuals or how that distribution changes with age [20, 36]. Second, evaluation frameworks tend to emphasise either institutional structure or financial burden, and rarely apply a complete multi-dimensional framework at the individual level [8, 12]. Third, evidence on how effectiveness evolves over time, and on what determines it, remains thin [37, 38].

This study addresses these gaps using five waves of CHARLS (2011, 2013, 2015, 2018 and 2020). It makes three contributions. First, we construct an individual-level, trackable and decomposable indicator system for multi-tiered medical security effectiveness, connecting macro-institutional logic to micro-level circumstances. Second, we combine AHP with the entropy weight method so that weights reflect both expert judgement and observed variation, validate the index against an entropy-weighted TOPSIS specification, and measure inequality in it by the concentration index, decomposed by the Wagstaff method, which treats region as a covariate rather than as the organising partition. Third, we characterise the time path of effectiveness, its urban-rural and regional heterogeneity, and its determinants within a two-way fixed-effects framework, and show that the conversion of household economic capacity into protection is context-dependent.

We adapt the adequacy-equity-synergy-safety framework of Zhang, Chu and Zhao [20] to the individual level. Adequacy is captured through basic insurance enrolment, supplementary or commercial coverage, the reimbursement ratio and the inverse OOP burden; equity through access to medical assistance and urban-rural

fairness; synergy through the number of tiers held and inter-tier linkage, defined as simultaneous coverage by basic and supplementary schemes; and safety through the absence of catastrophic expenditure, household affordability and continuity of coverage. Effectiveness is generated jointly by individual endowments, household resources and institutional context, and medical need acts on it through service utilisation [8, 20].

This study therefore examines how the effectiveness of the multi-tiered system varies with age and socioeconomic position, where the inequality in it originates, and whether its health gains operate through a reduced OOP burden. Age is expected to follow an inverted-U profile, education and household economic status should be positively associated with effectiveness and more strongly so in urban and better-developed areas, and we expect the urban-rural divide to be the principal source of inequality, with rural residents scoring lower. We test separately whether the OOP share mediates the association between effectiveness and health.

Methods

Study design and setting

This is a longitudinal observational study of middle-aged and older adults in China, based on a nationally representative household panel, reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for cohort studies. The completed STROBE checklist is provided as Additional file 1. No large language model was used to generate the data, the analysis or the interpretation of results; automated tools were used only for language editing.

Data source and study population

This study uses five waves of CHARLS (2011, 2013, 2015, 2018 and 2020), a nationally representative longitudinal survey of Chinese adults aged 45 and above conducted by the National School of Development at Peking University using multi-stage stratified probability-proportional-to-size sampling. It covers 150 counties and

450 communities across 28 provinces, and its baseline matches the 2010 census closely on age and sex [39, 40].

Respondents aged 45 and above were retained, and observations missing key demographic variables were excluded. The final analytic sample comprises 93953 person-wave observations on 25390 respondents, an average of 3.70 waves each, distributed as 17060 (2011), 18095 (2013), 19989 (2015), 19542 (2018) and 19267 (2020). Of the respondents, 10604 (41.8%) were observed in all five waves, 12381 (48.8%) in two to four waves and 2405 (9.5%) in one wave only. This unbalanced structure is a feature of an ageing survey rather than a defect, and attrition is correlated with health and with the outcomes of interest, which constrains the estimators that can be used.

Indicator system

Table 1 gives the symbol, definition and weight of every indicator. All 11 indicators are constructed from individual-level CHARLS information, are oriented so that higher values indicate greater effectiveness, and are min-max normalised to the unit interval before weighting. Table 2 reports the descriptive statistics of the composite index, its four dimensions and the covariates of the main specification.

Table 1 Indicator system, definitions and weights

Dimension	Indicator	Symbol	Definition	AHP weight	Entropy weight	Combined weight
Adequacy (A)	Basic insurance enrolment	A1	Equals 1 if the respondent is enrolled in any basic medical insurance scheme (urban employee, urban resident, new rural cooperative, urban-rural resident, or government medical care), otherwise 0	0.1400	0.0052	0.0726
	Supplementary or commercial coverage	A2	Equals 1 if the respondent is enrolled in a supplementary or commercial medical insurance scheme, otherwise 0	0.0900	0.2268	0.1584
	Reimbursement ratio	A3	Reimbursed expenditure divided by total medical expenditure, among respondents with positive expenditure	0.1000	0.0709	0.0855
	Inverse OOP burden	A4	One minus the ratio of out-of-pocket expenditure to total medical expenditure, among respondents with	0.0700	0.0656	0.0678

			positive expenditure			
	Access to medical assistance	B2	Equals 1 if the household receives medical assistance, otherwise 0	0.1000	0.3640	0.2320
	Urban-rural fairness	B3	Equals 1 for insured urban residents, 0.5 for insured rural residents, and 0 for the uninsured	0.1000	0.0091	0.0546
Synergy (C)	Number of tiers held	C1	The number of tiers held, divided by two	0.0600	0.0055	0.0327
	Inter-tier linkage	C2	Equals 1 if basic and supplementary cover are held simultaneously, otherwise 0	0.0900	0.2374	0.1637
Safety (D)	Absence of catastrophic expenditure	D1	One minus an indicator for out-of-pocket payments exceeding 40% of household consumption per capita	0.0900	0.0033	0.0467
	Household affordability	D2	One minus the ratio of out-of-pocket expenditure to household consumption per capita, truncated at one	0.1000	0.0065	0.0532
	Continuity of coverage	D3	One minus the within-person standard deviation of insurance status across the five waves	0.0600	0.0057	0.0328

Note: all 11 indicators are min-max normalised to the unit interval before weighting, and indicators measuring a burden rather than a benefit are inverted at construction so that a higher value unambiguously means better protection. Weights combine expert judgement (AHP) and observed variation (entropy) following Chen et al. [36], and the consistency ratio of the AHP judgement matrix is below 0.1. The outcome variable, the multi-tiered medical security effectiveness index, is the weighted aggregate of the 11 indicators, min-max rescaled to 0 to 100.

Table 2 Descriptive statistics

Variables	Obs	Imputed (n)	Mean	S.D.	Min	Median	Max
Effectiveness index (MME, 0-100)	93953	0	28.2350	8.5357	0.0000	27.4306	100.0000
Adequacy (A)	93953	0	0.2542	0.1327	0.0001	0.1890	1.0001
Equity (B)	93953	0	0.1252	0.0713	0.0001	0.0953	1.0001
Synergy (C)	93953	0	0.1002	0.1433	0.0001	0.0834	1.0001
Safety (D)	93953	0	0.8983	0.1861	0.0001	1.0001	1.0001
Age (years)	93953	0	61.1678	10.1399	45.0000	60.0000	120.0000
Education (1-4)	93953	105	2.0203	1.0597	1.0000	2.0000	4.0000
Rural residence (0/1)	93953	0	0.5964	0.4906	0.0000	1.0000	1.0000
Marital status (0/1)	93953	74	0.8571	0.3500	0.0000	1.0000	1.0000
Log household consumption per capita	93953	23829	8.9571	0.8184	1.6422	8.9976	14.2210
Medical need (0-2)	93953	4442	0.8026	0.5004	0.0000	1.0000	2.0000

Note: statistics are computed on the estimation sample used in every regression model, which comprises 93953 person-wave observations, so the number of observations is identical for every variable tabulated. The Imputed (n) column gives the number of values supplied by imputation rather than observed. Sub-index maxima slightly above one reflect rounding in the min-max normalisation. Two further variables enter the analysis but are not tabulated here: self-rated health (88239 observations), used only in the mediation analysis

estimated on complete cases, and the chronic-disease indicator (90531 observations), used only as a component of medical need. Additional file 4 (Table S3) reports the missingness of every variable by survey wave. The effectiveness index is rescaled to 0-100 and the four dimension sub-indices are on the 0-1 scale.

Four features of the indicator system are not visible from Table 1. First, the reimbursement ratio and the inverse OOP burden are defined only for respondents reporting positive medical expenditure, and how the remaining respondents are treated is tested in the robustness checks. Second, urban-rural fairness is set to 1 for insured urban residents, 0.5 for insured rural residents and 0 for the uninsured, so that it is directly sensitive to the urban-rural gradient in entitlement, while inter-tier linkage is non-zero only when two tiers are held simultaneously and therefore carries the joint-operation content of the framework. Third, the three burden-type indicators are inverted at construction, so that every indicator enters the index with a positive orientation. Fourth, universal entitlement was specified in an earlier version as a separate equity indicator, but it is defined identically to basic insurance enrolment and is therefore perfectly collinear with it; retaining both counted the same information twice and mechanically inflated the equity dimension, so it was dropped and its AHP weight was reallocated evenly to the two remaining equity indicators.

Measurement of the effectiveness index

Combined weighting. To avoid the subjectivity of purely expert weights and the data-dependence of purely objective weights, AHP and the entropy weight method are combined, following Chen et al. [36].

The entropy method assigns larger weights to indicators with greater variation. Let x'_{ij} denote the normalised value of indicator j for individual i. The proportion of individual i in the total for indicator j, the entropy of indicator j, its degree of divergence and its entropy weight are, in turn,

$$p_{ij} = \frac{x'_{ij}}{\sum_{i=1}^{n} x'_{ij}},\quad e_j = -\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij}\ln p_{ij},\quad d_j = 1 - e_j,\quad j_{ent}^{W} = \frac{d_j}{\sum_{j=1}^{m} d_j} \qquad (1)$$

AHP weights were derived from a 1-9 scale expert judgement matrix that passed the consistency test (consistency ratio < 0.1). The combined weight, which averages the expert and entropy weights and re-normalises them to sum to one, is

$$W_j = \frac{j_{ahp}^W + j_{ent}^W}{\sum_{j=1}^m (j_{ahp}^W + j_{ent}^W)} \quad (2)$$

Three indicators dominate the weighting, namely access to medical assistance (0.2320), inter-tier linkage (0.1637) and supplementary or commercial coverage (0.1584), which together account for 55.4% of the total. That the two highest-weighted indicators concern the floor and the supplementary tier is itself informative, because the principal constraint on effectiveness is not the basic layer, whose coverage is close to saturation, but the tiers above and below it. Consistent with this, basic insurance enrolment (0.0726) and continuity of coverage (0.0328) receive small weights because their near-universal coverage leaves little variation to explain.

The composite index is a weighted linear aggregate of the normalised indicator values,

$$MME_i = \sum_{j=1}^m W_j \cdot x'_{ij} \quad (3)$$

rescaled by min-max normalisation to the 0-100 range to ensure comparability across waves. Dimension sub-indices are computed analogously within each dimension.

Validation. We re-estimated the index using entropy-weighted TOPSIS, in which the relative closeness to the ideal solution is

$$C_i = \frac{i_-^D}{i_+^D + i_-^D} \quad (4)$$

where i_+^D and i_-^D are the Euclidean distances from individual i to the positive and negative ideal solutions respectively.

Measurement and decomposition of socioeconomic inequality

Inequality in the effectiveness index is measured with the concentration index, the standard measure of income-related inequality in health and health care [41, 42], computed on the individual-level score ranked by household consumption per capita.

The unit of observation is therefore the respondent, and no aggregation to provinces is performed. The concentration index is

$$CI = \frac{2}{\mu} \operatorname{cov}(y_i, R_i) \tag{5}$$

where y_i is the effectiveness index of individual i, μ is its mean, and R_i is the fractional rank of the individual in the distribution of household consumption per capita. The index runs from -1 to 1. A positive value means that effectiveness is concentrated among better-off respondents, a negative value means the reverse, and zero means perfect equality.

The concentration index is then decomposed by the method of Wagstaff, van Doorslaer and Watanabe [43], which writes it as the sum of the contributions of its determinants.

$$CI = \sum_k \frac{\beta_k \bar{x}_k}{\mu} CI_k + \frac{GC_\varepsilon}{\mu} \tag{6}$$

Here β_k is the coefficient on covariate x_k in a linear regression of the effectiveness index on the covariates, and CI_k is its own concentration index, with the bar denoting the covariate mean. The final term is the generalised concentration index of the residual. The covariates are age, age squared, education, marital status, rural residence, chronic disease and macro-region. Insurance enrolment and supplementary coverage are excluded from the decomposition because they enter the index by construction, so retaining them would make part of the decomposition circular.

Region enters Eq. (6) as a covariate rather than as a partition of the sample, so the share of inequality attributable to macro-region is estimated alongside the shares attributable to urban-rural residence and education, keeping the whole analysis at the individual level.

Covariates

The dependent variable is the individual-level multi-tiered medical security effectiveness index, rescaled to 0-100, where higher values indicate that a respondent's combination of basic, supplementary and floor protection delivers more adequate, more equitably distributed, better coordinated and safer financial protection.

Age enters both linearly and as a squared term so that the inverted-U pattern can be tested directly. Education is ordered by the highest level of schooling completed, marital status is binary, household consumption per capita enters in logarithms as a proxy for permanent income less sensitive to transitory shocks than reported income, rural residence distinguishes respondents in rural communities from those in urban ones, and medical need is a composite of diagnosed chronic conditions and functional limitation.

Basic medical insurance enrolment enters the specification both as an input to the index and, in a separate specification, as a regressor; supplementary or commercial coverage enters analogously. Because enrolment in the basic tier is close to universal, its coefficient should be read as the contribution of the basic tier to the level of the index rather than as a contrast between insured and uninsured respondents. Urban-rural status, three macro-regions and four age groups (45-54, 55-64, 65-74 and 75 years and above) define the heterogeneity analysis; we use macro-regions rather than sampled provinces because CHARLS is nationally representative but not at the provincial level. For the mediation test, self-rated health is the outcome and the OOP ratio the candidate mediator.

Handling missing expenditure data in 2020

CHARLS 2020 did not administer the health care and insurance module, so total medical expenditure, OOP expenditure, reimbursement and the derived variables are entirely missing in that wave. Four of the 11 indicators depend on them, so a third of the indicator system is affected.

We adopted a three-part strategy. First, for the main analysis we imputed the missing cost variables by carrying forward each individual's most recent observed value, using the province-year median where no prior value existed; this preserves sample completeness and cross-wave comparability and reflects the persistence of insurance arrangements and cost-sharing structures within individuals [44]. Second, as a robustness check we re-imputed the 2020 values using a two-part model [45, 46], comprising a logit for the probability of any medical expenditure followed by

estimates of the reimbursement ratio, the OOP ratio and the log of OOP expenditure among those with positive expenditure.

$$P(Y_i > 0 \mid Z_i) = \Lambda(\gamma Z_i), \quad E(Y_i \mid Y_i > 0, Z_i) = \theta Z_i \tag{7}$$

where $\Lambda(\cdot)$ is the logistic distribution function and Z_i is a vector of covariates observed in 2020, namely inpatient and outpatient utilisation, chronic conditions, insurance enrolment, self-rated health, household consumption per capita, demographics and region. The five estimated equations, reported in Additional file 2 (Table S1), show that the probability of any medical expenditure and the reimbursement ratio both rise with inpatient utilisation, chronic conditions and basic insurance enrolment. The OOP ratio falls with inpatient utilisation and enrolment, and rises with rural residence. Because the predicted 2020 values use only information actually collected in 2020, the imputation avoids using future values of the outcome.

Third, the entire index was re-estimated on the four waves with complete expenditure data (2011-2018), with the weights re-derived, to test whether the findings depend on the 2020 wave at all.

Missing covariate data

Coverage of the covariates was high but not complete. Household consumption per capita was the most affected variable, missing for 23829 person-wave observations (25.4%), spread across all five waves because the household expenditure module was not administered to every household in every wave. Education was missing for 105 observations, marital status for 74, and the two items that together make up medical need for 4442; age and rural residence were complete.

Restricting the analysis to complete records would have removed 26598 observations and, because non-response to the expenditure module is correlated with socioeconomic status, would have tilted the sample towards better-off households. We therefore imputed household consumption per capita at the wave-specific median, education at the overall median, and marital status and the two components of medical need at zero. Self-rated health was not imputed, because it enters only the mediation analysis, which is estimated on complete cases. Table 2 reports the number of

imputed values for every variable. Because imputing a quarter of a key covariate is a strong assumption, we also re-estimated the baseline specification on complete cases as a robustness check (Additional file 6, Table S5).

Statistical analysis

To identify determinants, the following specification is estimated.

$$MME_{it} = \alpha + \beta_1 Age_{it} + \beta_2 it_2^{Age} + \gamma X_{it} + \mu_i + \lambda_t + \varepsilon_{it} \quad (8)$$

where i indexes respondents and t waves, μ_i are individual fixed effects absorbing time-invariant heterogeneity, λ_t are wave fixed effects absorbing macro-level policy and cyclical shocks, and X_{it} is a vector of controls comprising education, marital status, rural residence, log household consumption per capita and medical need. Standard errors are clustered at the respondent level throughout.

Two specifications of Eq. (8) are reported, for an identification reason that is easy to overlook. Age advances mechanically with the survey wave, since a respondent interviewed in 2011 and again in 2020 is nine years older, so individual effects absorb the year of birth and wave effects absorb the survey year. Only 0.48% of the standard deviation of age survives the two-way transformation, against 29.0% after individual demeaning alone, leaving almost no independent variation from which to recover an age profile. Column (1) is accordingly a pooled model with wave fixed effects, which identifies the age profile from cross-sectional and within-person variation combined, at the cost of leaving cohort composition uncontrolled. Column (2) is the two-way fixed-effects model, which removes time-invariant individual heterogeneity and common time shocks; age and its square are not separately identified in it.

The choice between random and fixed effects is settled by a Mundlak test, which augments the pooled model with the within-respondent means of the time-varying regressors and tests them jointly. The group means are jointly significant ($F = 259.81$, $p < 0.001$), so the random-effects assumption is rejected. A joint test of the individual and wave effects against the pooled model likewise rejects the pooled specification ($F(25{,}393, 68{,}556) = 4.14$, $p < 0.001$).

Two variables in the framework are definitional components of the outcome, and both require care. Basic medical insurance enrolment enters five of the eleven indicators, which together carry 35.6% of the weight, so regressing the composite index on enrolment recovers a coefficient that is largely definitional rather than behavioural; for that specification we constructed a net index, reweighting the six indicators that do not depend on enrolment to sum to one. The out-of-pocket ratio raises the same problem for the mediation analysis, because three indicators are derived from out-of-pocket spending and carry 16.8% of the weight, so we use a second index that excludes them.

To test whether the OOP share mediates the association between effectiveness and health, a three-step mediation model is estimated with self-rated health as the outcome and the OOP ratio as the candidate mediator.

$$Health_{it} = a_0 + a_1 MME_{it} + \delta_1 X_{it} + u_{it} \quad (9)$$

$$OOP_{it} = b_0 + b_1 MME_{it} + \delta_2 X_{it} + v_{it} \quad (10)$$

$$Health_{it} = c_0 + c_1 MME_{it} + c_2 OOP_{it} + \delta_3 X_{it} + w_{it} \quad (11)$$

The indirect effect is the product $b_1 c_2$, evaluated with the Sobel test and compared with the direct effect c_1. When the two have opposite signs, the pattern is a suppression effect rather than classical mediation. Eqs. (9) to (11) are estimated by ordinary least squares with wave fixed effects and standard errors clustered at the respondent level, on the 88239 observations for which self-rated health is observed. Effectiveness is measured by the index that excludes the three spending-derived indicators, so that the association between effectiveness and the out-of-pocket ratio is not built into the construction of either variable.

All analyses were conducted in Stata. Two-sided tests were used throughout, and statistical significance is reported at the 1%, 5% and 10% levels.

Results

Trends in effectiveness, 2011-2020

Figure 1 and Table 3 report the evolution of the index over the decade. The trend is not monotonic, and the four dimensions diverge sharply.

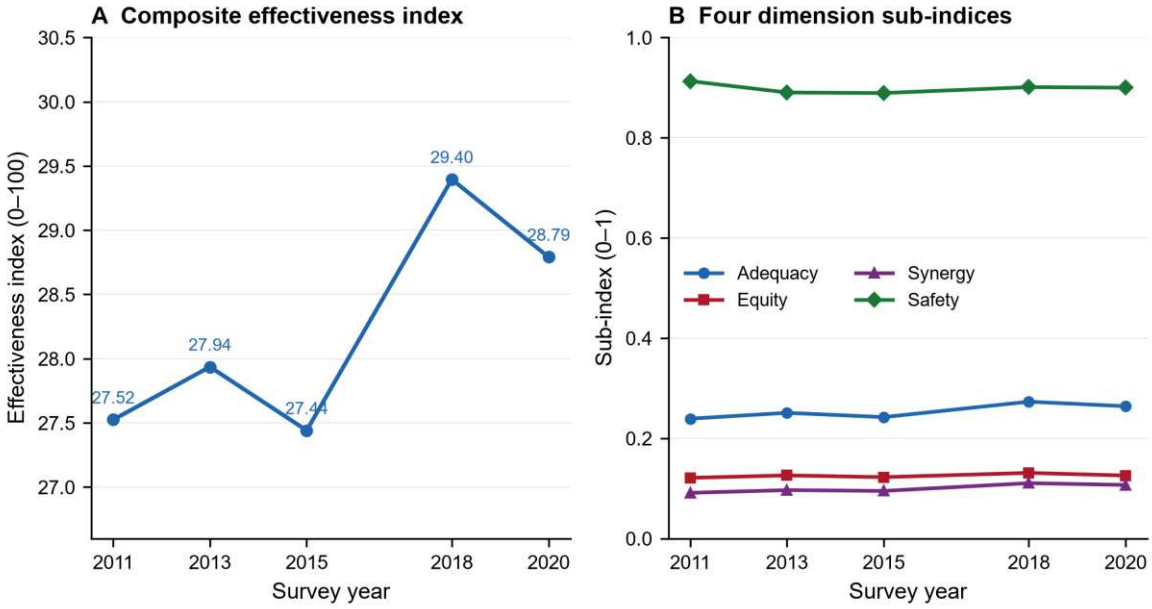


Fig. 1 Effectiveness of China's multi-tiered medical security system, 2011-2020. **(A)** Composite effectiveness index (0-100). **(B)** Sub-indices for the four dimensions, namely adequacy, equity, synergy and safety, each on a 0-1 scale. Source: authors' calculations from CHARLS 2011-2020.

Table 3 Evolution of the effectiveness index and its four dimensions, 2011-2020

Year	Observations	Composite index	Adequacy	Equity	Synergy	Safety	Growth (%)
2011	17060	27.52	0.2391	0.1210	0.0911	0.9130	—
2013	18095	27.94	0.2508	0.1259	0.0965	0.8903	+1.53
2015	19989	27.44	0.2422	0.1224	0.0949	0.8892	−1.79
2018	19542	29.40	0.2730	0.1309	0.1104	0.9010	+7.14
2020	19267	28.79	0.2640	0.1256	0.1068	0.8999	−2.07

Note: the composite index is min-max rescaled to 0-100 and is comparable across waves. The four dimension sub-indices are on the 0-1 scale. Growth is the year-on-year percentage change in the composite index.

The index increased from 27.52 in 2011 to 28.79 in 2020, a cumulative gain of 4.6%, and peaked at 29.40 in 2018 (Figure 1A), a trajectory that aligns with the deepening of reform after 2012, as urban and rural resident insurance were integrated nationwide from 2016 and the National Healthcare Security Administration was

established alongside centralised drug procurement in 2018. Two episodes of decline deserve attention. The 1.79% fall in 2015 coincided with a higher rural share in that wave and with adjustments to outpatient reimbursement policy, and the 2.07% fall in 2020 is consistent with suppressed service utilisation during the COVID-19 pandemic.

The four dimensions diverged sharply (Figure 1B and Table 3). Safety remained high throughout, at 0.89 or above, indicating that the system is comparatively effective at shielding households from catastrophic expenditure. Adequacy occupied the middle range, at 0.24-0.27, while equity (0.12-0.13) and synergy (0.09-0.11) remained persistently low. This pattern substantiates the characterisation of the system as established in form but deficient in function. Synergy started from the lowest base but grew fastest, rising from 0.0911 to 0.1068 (17.2%), whereas equity improved least (3.8%), confirming it as the weakest area.

Inequality in effectiveness and its sources

Inequality in effectiveness was pro-rich and widened over the decade (Figure 2A and Table 4). The concentration index was positive in every wave, rising from 0.0199 (standard error 0.0052) in 2011 to 0.0240 (0.0051) in 2020, with a temporary dip to 0.0168 in 2015. A positive value means that respondents better off in household consumption per capita also enjoy more effective protection, so the system is not neutral with respect to living standards. The 2015 dip coincides with the transition period of urban-rural insurance integration, and the renewed rise after 2018 coincides with the consolidation of centralised procurement and the 2020 reform blueprint.

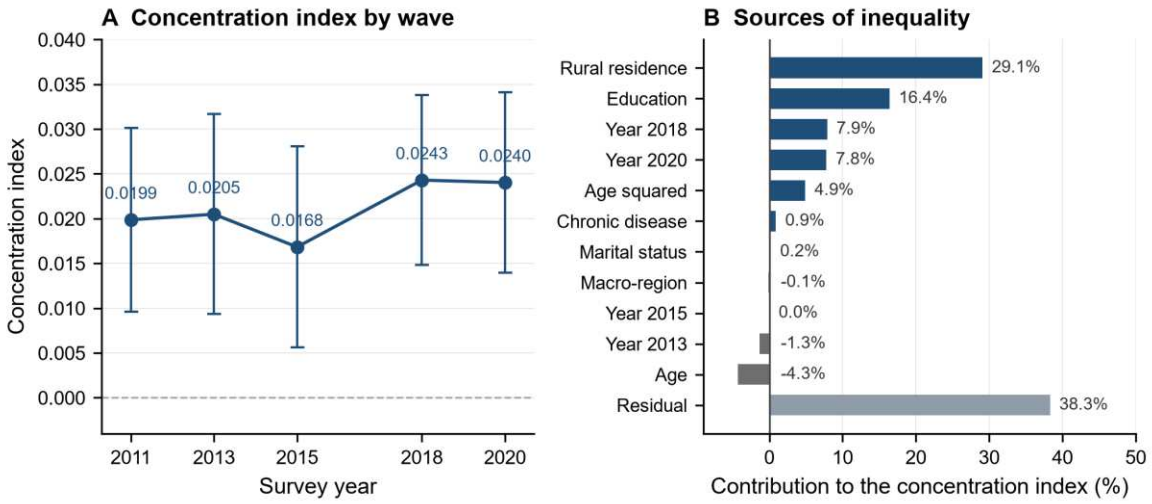


Fig. 2 Socioeconomic inequality in multi-tiered medical security effectiveness. **(A)** Concentration index of the effectiveness index by wave, with 95% confidence intervals. **(B)** Contribution of each determinant to the pooled concentration index, as a percentage of the total. Source: authors' calculations from CHARLS 2011-2020.

The decomposition shows where that inequality originates, and the answer is not where a regional framing would put it (Table 4). Urban-rural residence is the largest single contributor, at 29.1% of the concentration index, followed by education at 16.4%. Macro-region contributes −0.1%, essentially nothing, because once household living standards, urban-rural residence and education are taken into account, knowing which macro-region a respondent lives in says almost nothing about their level of protection. The residual accounts for 38.3%.

Three further features reinforce this reading. The three regional means are close in every wave, spanning 0.21 to 0.68 index points on a scale whose standard deviation is 8.54, so the urban-rural gap is several times larger than any regional gap. Inequality is also uneven across the four dimensions (Table 4), being highest for synergy (0.0684) and declining through equity (0.0496) and adequacy (0.0317) to almost zero for safety (0.0037). Across subgroups it is higher in urban than in rural areas (0.0255 versus 0.0109), and highest in the east (0.0289), intermediate in the centre (0.0260) and lowest in the west (0.0179).

Inequality in multi-tiered medical security effectiveness among middle-aged and older adults is therefore an individual and urban-rural phenomenon rather than a regional one. The contrast with the provincial literature is a matter of resolution rather than substance, because provincial indices are built from jurisdictional aggregates, so variation among individuals inside a province is averaged away before any inequality measure is computed and only variation between provinces survives [13, 36]. Working at the individual level retains the within-province variation, and that variation turns out to dominate. Because inequality is therefore driven by urban-rural residence and education rather than by region, narrowing it requires urban-rural equalisation and better targeting of the supplementary tiers.

Table 4 Concentration index of the effectiveness index and its decomposition by determinant

Panel A. Concentration index by wave

Year	Observations	Concentration index	Standard error	95% confidence interval
2011	13429	0.0199	0.0052	0.0096 to 0.0301
2013	11073	0.0205	0.0057	0.0093 to 0.0317
2015	11131	0.0168	0.0057	0.0056 to 0.0281
2018	15617	0.0243	0.0048	0.0148 to 0.0338
2020	13868	0.0240	0.0051	0.0140 to 0.0341

Panel B. Decomposition of the pooled concentration index

Source	Contribution to the concentration index (%)
Rural residence	29.1
Education	16.4
Year effects	14.4
Chronic disease	0.9
Age, linear and squared	0.6
Marital status	0.2
Macro-region	−0.1
Residual	38.3
Total	100.0

Panel C. Concentration index by dimension and by subgroup

Grouping	Concentration index	Standard error
Synergy	0.0684	0.0041
Adequacy	0.0317	0.0026
Equity	0.0496	0.0024
Safety	0.0037	0.0023
Urban	0.0255	0.0038
Rural	0.0109	0.0030
Eastern region	0.0289	0.0041
Central region	0.0260	0.0040
Western region	0.0179	0.0041

Note: the concentration index is computed on the individual-level effectiveness index ranked by household consumption per capita, so no aggregation to provinces is involved; a positive value indicates that effectiveness is concentrated among better-off respondents. Panel A reports the index wave by wave. Panel B reports the Wagstaff decomposition of the index pooled across waves, with covariates age, age squared, education, marital status, rural residence, chronic disease and macro-region, year effects captured by wave indicators, and contributions expressed as percentages of the pooled concentration index of 0.0244 that sum to 100; insurance enrolment and supplementary coverage are excluded because they enter the index by construction. Panel C reports the index separately for the four dimensions and for urban-rural and regional subgroups. The analytic sample is smaller than the full estimation sample, 65118 person-wave observations against 93953, because the ranking variable and the other covariates of the decomposition are missing for part of every wave.

The distribution of the index also compressed over time (Figure 3). The P90/P10 ratio narrowed from 1.79 in 2011 to 1.58 in 2020. One episode deviates. The tenth percentile fell sharply to 15.46 in 2015, indicating deterioration at the bottom of the distribution, before recovering after 2018.

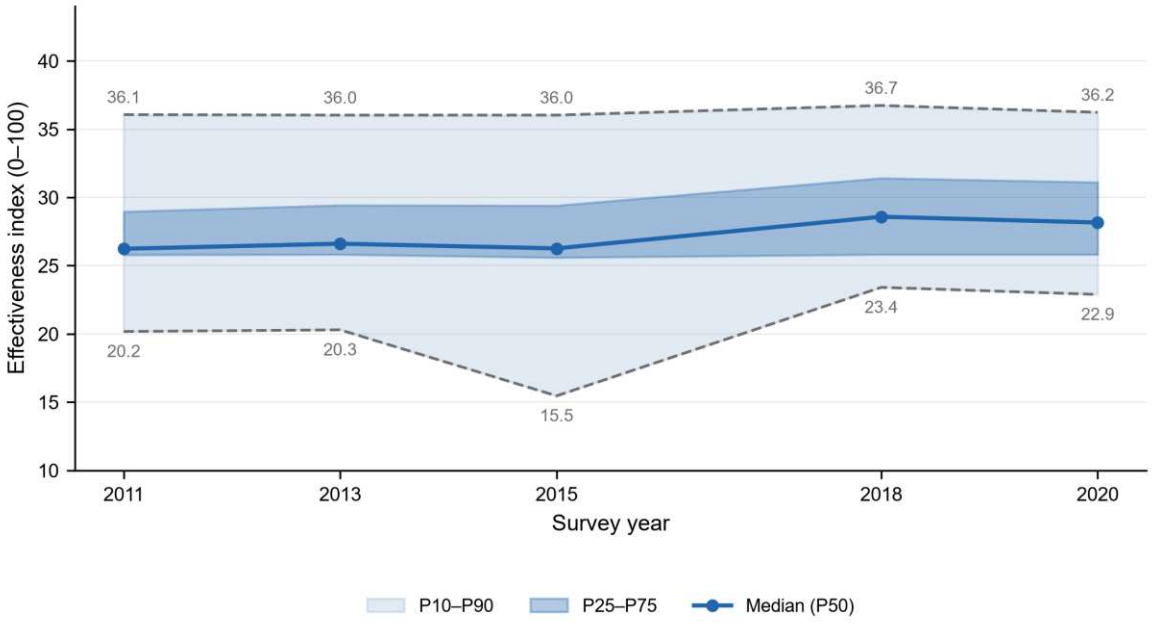


Fig. 3 Distribution of the effectiveness index, 2011-2020. Shaded bands show the P10-P90 and P25-P75 ranges, the solid line shows the median (P50), and dashed lines

show the tenth and ninetieth percentiles. Source: authors' calculations from CHARLS 2011-2020.

Urban-rural and age gradients

The urban-rural gap is the most persistent equity weakness (Figure 4A and Table 5). Urban residents scored consistently 3.02 to 4.29 points higher, with the gap widening to 4.29 in 2018 and remaining at 4.10 in 2020. Decomposed by dimension, the gap is concentrated in equity (urban 0.17-0.19 versus rural 0.09) and synergy (urban 0.09-0.12 versus rural 0.09-0.10), while the safety dimension differs little. This is consistent with Wang and Qin [38].

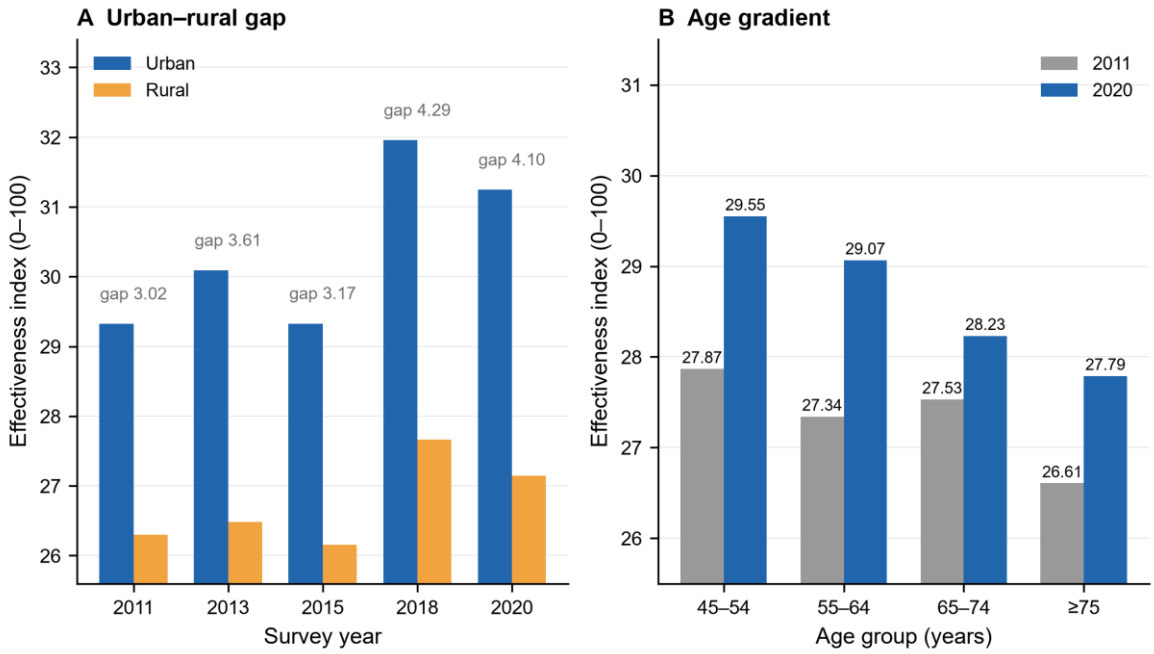


Fig. 4 Subgroup gradients in multi-tiered medical security effectiveness. **(A)** Urban-rural gap, 2011-2020, with values above the bars showing the urban-minus-rural difference in index points. **(B)** Age gradient in 2011 and 2020. Source: authors' calculations from CHARLS 2011-2020.

An age gradient is also evident (Figure 4B and Table 5). In 2011 the index was 27.87 for those aged 45-54 and 26.61 for those aged 75 and above, a gap of 1.26, which by 2020 had widened to 1.76. Older adults face higher chronic disease prevalence and greater medical need, yet their lower incomes and narrower coverage leave them less able to realise protection.

Table 5 Effectiveness by urban-rural status, region and age group, 2011-2020

Group	2011	2013	2015	2018	2020	Change
			Panel A. Urban-rural			
Urban	29.32	30.09	29.33	31.96	31.24	+1.92
Rural	26.30	26.48	26.16	27.67	27.14	+0.84
Urban-rural gap	3.02	3.61	3.17	4.29	4.10	+1.08
			Panel B. Age group			
45-54	27.87	28.24	27.66	30.18	29.55	+1.68
55-64	27.34	27.93	27.57	29.37	29.07	+1.73
65-74	27.53	27.89	27.38	28.83	28.23	+0.70
75+	26.61	26.79	26.22	28.44	27.79	+1.18
Oldest-youngest gap	1.26	1.45	1.44	1.74	1.76	+0.50

Note: values are the mean effectiveness index within each group. The gap is the difference between the higher- and lower-scoring group, in index points. Change is the difference between 2020 and 2011.

Determinants of effectiveness

Table 6 reports three specifications. Column (1) is the pooled model with wave fixed effects, which carries the age profile. Column (2) is the two-way fixed-effects model, which identifies the time-varying covariates after removing time-invariant individual heterogeneity. Column (3) repeats the pooled specification with the net index as the dependent variable and basic insurance enrolment added as a regressor.

The age profile follows an inverted U in the specification that can identify it. The linear term is positive (0.0651, $p < 0.1$) and the squared term negative (−0.0006, $p < 0.05$), so the predicted index peaks at 53.3 years and declines thereafter, with the marginal effect of an additional year falling from +0.010 index points at age 45 to −0.039 at 85. The inverted-U pattern is therefore supported, with a qualification we did not anticipate. The peak occurs well below the sample mean age of 61.

Education and household economic status are positively associated with effectiveness, and the two behave differently once individual heterogeneity is

removed. In the pooled specification the education coefficient is 0.7461 ($p < 0.01$) and the elasticity of household consumption per capita is 0.6535 ($p < 0.01$). In the two-way fixed-effects specification the education coefficient falls to −0.0092 and is insignificant, as expected because education is fixed in adulthood, while the consumption elasticity remains positive and significant at 0.1564 ($p < 0.01$). The gap between the two consumption estimates, 0.6535 against 0.1564, measures how much of the cross-sectional economic gradient reflects stable differences between respondents rather than changing circumstances.

Marriage is protective in both specifications, with married respondents scoring 0.9306 points higher in the pooled model ($p < 0.01$) and 0.6347 points higher in the fixed-effects model ($p < 0.01$), consistent with the caregiving and risk-pooling functions of the household. Rural residence is negatively associated with effectiveness (−3.0347, $p < 0.01$), supporting the expectation that the urban-rural divide is a principal source of inequality, but the coefficient is not identified in the fixed-effects specification because rural residence is fixed for most respondents. Medical need is positively associated with effectiveness in the pooled specification (0.3604, $p < 0.01$) but not in the fixed-effects specification (0.0585, standard error 0.0844).

Basic insurance enrolment shows no association with the parts of the index that do not depend on it. In column (3) the coefficient is −0.0074 with a standard error of 0.1177 ($p = 0.95$), the opposite of the conclusion a naive specification would support. Regressing the full index on enrolment yields 15.64, but that figure is an artefact, because enrolment enters five of the eleven indicators, which together carry 35.6% of the weight. Once the definitional overlap is removed, enrolment status carries no additional information about protection, which is unsurprising given that 92.4% of observations are enrolled.

Table 6 Determinants of multi-tiered medical security effectiveness

Variable	(1) Pooled OLS	(2) Two-way fixed effects	(3) Net index + insurance
Age	0.0651*		0.0319
	(0.0389)		(0.0411)
Age squared	−0.0006**		0.0000

	(0.0003)		(0.0003)
Education	0.7461***	−0.0092	0.6704***
	(0.0430)	(0.0802)	(0.0459)
Marital status	0.9306***	0.6347***	0.1373
	(0.1199)	(0.1552)	(0.1306)
Rural residence	−3.0347***		−1.2339***
	(0.0876)		(0.0939)
Log household consumption per capita	0.6535***	0.1564***	0.7196***
	(0.0419)	(0.0352)	(0.0435)
Medical need	0.3604***	0.0585	0.0427
	(0.0728)	(0.0844)	(0.0741)
Basic insurance enrolment			−0.0074
			(0.1177)
Wave fixed effects	Yes	Yes	Yes
Individual fixed effects	No	Yes	No
Observations	93953	93953	93953
R-squared	0.0694	0.0006	0.0277

Note: individual-clustered standard errors are in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Column (1) is a pooled model with wave fixed effects. Column (2) removes individual and wave effects by alternating projection. Column (3) repeats column (1) with the net index as the dependent variable. Age and its square are not reported in column (2) because they are not separately identified once both sets of effects are included, since only 0.48% of the standard deviation of age survives the two-way transformation. Rural residence is absorbed by the individual fixed effects. The net index in column (3) reweights the five indicators that do not depend on basic insurance enrolment and rescales to 0-100. The random-effects specification is rejected by a Mundlak test ($F = 259.81$, $p < 0.001$).

Heterogeneity by urban-rural status, region and age group

Additional file 3 (Table S2) reports heterogeneity by urban-rural status, region and age group, estimated separately within each subsample. The association between household economic status and effectiveness is significant in urban and eastern subsamples, borderline in western ones and absent in rural ones. The coefficient on log household consumption is 0.4011 ($p < 0.01$) among urban respondents and 0.0327 (standard error 0.0387) among rural ones. Marriage is protective and significant in the urban, rural, eastern and central subsamples and among respondents aged 55 to 64, where it reaches 1.0699 ($p < 0.01$).

The contrast in the consumption coefficient is the substantive result, because it reveals the context-dependence of the conversion of economic capacity into protection. Where insurance supply and service accessibility are adequate, higher income translates into supplementary coverage and better access to quality care; where they are not, higher income alone does not raise protection. This is the micro-

level counterpart of the pro-rich pattern documented by Wang and Qin [38], and it implies that raising rural incomes alone will not improve protection.

Mediation analysis

Table 7 reports the mediation test, using the index that excludes the three spending-derived indicators so that the out-of-pocket ratio is not a component of the effectiveness measure. Effectiveness is associated with a lower out-of-pocket ratio (step 2 coefficient −0.0094, $p < 0.01$), but its association with self-rated health is close to zero in the first step (0.0002, standard error 0.0004) and only becomes positive once the out-of-pocket ratio is controlled for (0.0023, $p < 0.01$). The Sobel statistic is −13.00 and the indirect and direct effects have opposite signs, so the pattern is a suppression effect rather than classical mediation, in that a higher effectiveness score accompanies a lower out-of-pocket share, which is in turn associated with worse self-rated health, plausibly because it reflects limited service use rather than generous protection.

The magnitudes counsel against reading much into the mechanism. The indirect effect of −0.0022 and the direct effect of 0.0023 on a five-point health scale almost cancel, and the total effect is indistinguishable from zero. For comparison, the same equations estimated with the full index, which contains the out-of-pocket ratio among its components, yield an indirect effect of −0.0046 and a direct effect of 0.0079, roughly twice as large in each direction, so a substantial part of the apparent mediation in that specification was definitional. The mediation expectation is not supported.

Table 7 Mediation test with the OOP ratio as candidate mediator

Step	Outcome	Predictor	Observed coefficient	Standard error	P-value
Step 1	Self-rated health	Effectiveness index	0.0002	0.0004	0.68
Step 2	OOP ratio	Effectiveness index	−0.0094	0.0002	<0.01
Step 3	Self-rated health	Effectiveness index	0.0023	0.0005	<0.01
Step 3	Self-rated health	OOP ratio	0.2296	0.0169	<0.01
Sobel test	—	Indirect effect $(b_1\ c_2)$	−0.0022	—	<0.01
Sobel test	—	Direct effect	0.0023	—	<0.01

(c_1)

Note: the effectiveness index excludes the three spending-derived indicators, so that the out-of-pocket ratio is not a component of it. The indirect effect is the product of the step-2 coefficient on the effectiveness index and the step-3 coefficient on the out-of-pocket ratio. The Sobel test statistic is −13.00 and the indirect and direct effects have opposite signs, indicating a suppression effect rather than classical mediation. Models control for age, age squared, education, marital status, rural residence, log household consumption per capita, medical need and wave fixed effects, with standard errors clustered at the respondent level. Observations, 88239. For comparison, the same equations estimated with the full index give 0.0033, −0.0141, 0.0079 and 0.3239 for the four coefficients, an indirect effect of −0.0046 and a Sobel statistic of −17.32.

Robustness checks

Six regression-based robustness checks are reported in Additional file 6 (Table S5). All the regressions use the two-way fixed-effects specification, with age and its square excluded because they are not separately identified, and the coefficients of interest are those on marriage and household consumption.

The pattern is informative. Marriage and consumption remain positive and significant in three of the five regression checks, namely replacing the weighted index with entropy-weighted TOPSIS closeness, excluding the 2020 wave whose expenditure variables required imputation, and re-estimating on complete cases with no imputation. The last is the sharpest test of the imputation, and its reproduction of the baseline results shows that the 25.4% of observations with imputed household consumption do not drive the findings. The two remaining checks are less reassuring. In the balanced panel observed in all five waves, the marriage coefficient falls to 0.4168 ($p < 0.05$) and the consumption elasticity to 0.0713 ($p < 0.1$); with the explanatory variables lagged one wave, which probes reverse causality most directly, neither remains significant. The associations are therefore stable with respect to the measurement of the outcome and the treatment of the 2020 wave and of missing covariates, but not with respect to the two checks that most directly probe selection and reverse causality. This is unsurprising given two features of the design. The within-respondent variation available to identify the fixed-effects coefficients is small,

and the balanced panel is a selected subsample. The associations in Table 6 are therefore best read as conditional associations holding in the full sample.

Four checks on the construction of the index itself are reported in Additional file 5 (Table S4), and none changes the substantive results. Re-deriving the weights on the four waves with complete expenditure data reproduces the index to within 0.10 index points in every year. Replacing carry-forward imputation with the two-part model raises the 2020 index from 28.79 to 31.23 without altering the direction of the trend. Applying the consistent convention that non-spenders receive a reimbursement ratio of zero and an inverse burden of one raises the level by roughly 3.5 points equally in every wave. Clustering standard errors at the province-wave level, which yields 139 clusters, leaves the significance of every coefficient in Table 6 unchanged.

Discussion

Main findings

The results support five conclusions. First, effectiveness improved but remains structurally unbalanced. The index rose from 27.52 to 28.79 between 2011 and 2020, peaking in 2018. The configuration of high safety, intermediate adequacy and low equity and synergy indicates that the system performs well at preventing catastrophic expenditure, its most visible function, while the supplementary tiers and inter-tier coordination remain underdeveloped. Second, the additive effects of multi-tiering are emerging but not yet sufficient. Synergy started from the lowest base yet grew fastest (17.2%), whereas equity improved least (3.8%).

Third, inequality in effectiveness is not a regional phenomenon. The concentration index was positive in every wave, and its decomposition shows that macro-region contributed −0.1%, whereas urban-rural residence contributed 29.1% and education 16.4%. The relevant fault line is therefore urban-rural rather than eastern-central-western, and it cuts through every region rather than separating them. Fourth, the urban-rural gap is the most persistent equity weakness, at 3.02 to 4.29 points. Because economic capacity converts into protection in urban and eastern contexts but much less so in rural and western ones, income growth alone will not

close it. Fifth, household resources matter and the age profile follows an inverted U. Household consumption per capita (0.1564) and marriage (0.6347) are positively associated with effectiveness once individual heterogeneity is removed, while rural residence is negatively associated (−3.0347). The age profile peaks at 53.3 years, earlier than expected. Basic insurance enrolment, once the definitional overlap with the index is removed, shows no additional association with protection.

Comparison with previous studies

The results are consistent with the financial-protection literature. Studies of catastrophic health expenditure in China have found that CHE risk varies systematically with socioeconomic status [32, 33], and that older adults living alone with cognitive impairment face distinctly elevated risk [34, 35]. The present results point the same way. The safety dimension is the highest-scoring and the most stable of the four, whereas the equity dimension is the weakest. Protection against catastrophic cost has improved, but the distribution of that protection has not improved at the same rate.

The contrast with provincial-level composite indices is informative, and recent evidence published in this journal reinforces the point. Wang et al. [47] applied a triple decomposition to provincial-level health inequalities in China and found that differences between provinces dominated, which is the expected result when the unit of analysis is the province. Provincial frameworks that combine AHP and entropy weighting and evaluate jurisdictions rather than individuals [13, 36] necessarily identify disparity between jurisdictions. At the individual level the regional contribution to inequality falls to −0.1%, while urban-rural residence and education together account for more than two fifths of it. This is not a contradiction but a consequence of the unit of analysis. Aggregating to the province averages away the urban-rural and socioeconomic differences that exist inside each province, and it is precisely those differences that dominate at the individual level. Province-level rankings can therefore be an unreliable guide to where protection is weakest.

The multi-dimensional configuration found here also qualifies a common reading of the adequacy-equity-synergy-safety framework [20]. Provincial applications of that

framework, and related assessments of the equity and efficiency of provincial health resource allocation [13, 37], tend to treat the dimensions as moving together. At the individual level they diverge sharply, with safety high, adequacy intermediate, and equity and synergy low. That divergence matters. It shows that a system can be effective at its most visible function, shielding households from catastrophic cost, while the tiers above and below the basic layer remain weakly developed. The results are also consistent with evidence that supplementary coverage reduces household catastrophic expenditure, particularly among low-asset households [8, 38], and that integrating urban and rural resident insurance reshapes utilisation without removing the pro-rich gradient [37]. The weight structure explains why. The tier that would moderate income-related gaps is the tier with the least penetration, so it carries a large weight in the entropy component yet contributes little to the level of protection. Taken together, these comparisons suggest that the binding constraint on China's multi-tiered medical security system is not the level of protection in aggregate, but the way protection is distributed and joined up across tiers.

Implications for policy

Adequacy depends on keeping the basic tier solvent and broad. Universal enrolment has to be maintained while contribution and benefit adjustment mechanisms are optimised and outpatient reimbursement is broadened through pooled-funding reform. Reform of this kind also needs transitional arrangements that protect benefit levels, together with a rules-based mechanism linking reimbursement rates to economic development and to the capacity of the fund.

The supplementary and floor tiers, which carry the largest weights, require particular attention. Access to medical assistance (weight 0.2320) and inter-tier linkage (0.1637) are the highest-weighted indicators, so under-coverage at the floor and supplementary levels is the binding constraint. Active identification of assistance-eligible households through insurance administrative data, rather than reliance on applications, would improve targeting. Inclusive commercial health insurance should be developed in a regulated manner, with gradual relaxation of the age limits and pre-existing condition exclusions that disadvantage older adults. Genuine inter-tier

coordination should be built in parallel, integrating basic insurance, critical illness insurance and medical assistance through a shared information platform and one-stop settlement, so that orderly linkage avoids both gaps in coverage and duplicate compensation. The urban-rural gap, finally, should be closed from within regions, because between-region differences are negligible. Substantive integration of urban and rural resident insurance, with higher contribution subsidies and benefit levels for rural residents, is central. Because income growth does not automatically translate into protection in rural areas, supply-side measures such as strengthening primary care capacity and extending insurance supply into rural markets are needed alongside income support. Policy should also attend to the oldest and the poorest, who have improved least and face a widening age gradient. Long-term care insurance, strengthened medical and long-term care protection, fiscal transfers, medical mutual aid and charitable medical assistance in rural and less-developed areas all have a part to play.

Relevance beyond China

The finding that region is not the organising axis of inequality has implications for how health equity is measured and acted on elsewhere. Many middle-income countries are consolidating fragmented health financing schemes, and must decide whether to target inter-regional disparities or within-region inequalities. The two imply different allocations of a limited fiscal envelope. When inequality is measured on jurisdictional aggregates, inter-regional gaps are mechanically salient, because within-jurisdiction variation has already been averaged away. China's experience also illustrates a sequencing lesson. Universal basic coverage is achievable and does deliver protection against catastrophic cost, but it does not by itself equalise protection. The tiers that would equalise outcomes are the means-tested floor and voluntary supplementary coverage, and these are precisely the tiers that require the most administrative capacity to target and regulate. Countries that have reached near-universal basic coverage should therefore expect the marginal returns to further equity to come from the tiers above and below the basic layer rather than from the basic layer itself.

Limitations

Several limitations should be noted. First, age is not separately identified in the two-way fixed-effects specification. Age advances with the survey wave and only 0.48% of its standard deviation survives the two-way transformation, so the age profile is estimated from the pooled specification with wave fixed effects, which leaves cohort composition uncontrolled. Separating age from period and cohort would require either a birth-cohort design or a specification that replaces wave effects with a flexible function of age. A five-wave panel spanning nine years supports neither.

Second, the fixed-effects models explain very little of the within-respondent variation in effectiveness, since the within-R^2 is 0.0006. The coefficients are therefore estimated imprecisely, which is why the lagged specification and the balanced panel do not reproduce the baseline results. Much of the within-person variation remains unexplained. Unobserved factors such as local insurance supply, provider behaviour and health shocks are plausible candidates, and CHARLS does not observe them at the required level of detail.

Third, CHARLS 2020 did not collect the health care and insurance module, so cost-related indicators required imputation. The trend, the four-dimension pattern and the inequality findings are stable across four different treatments. The level of the index and the identification of the peak year are, however, sensitive to how respondents with zero medical expenditure are treated. Future work using CHARLS 2023 or other surveys with expenditure modules could resolve this. Household consumption per capita was also missing for 25.4% of observations and was imputed at the wave-specific median. The complete-case specification reproduces the substantive results, but median imputation compresses the variance of the imputed variable, and multiple imputation by chained equations would be preferable.

Fourth, supplementary coverage is captured by a single binary item, because CHARLS does not record the benefit depth or generosity of that tier. The indicator system can therefore identify who holds supplementary cover, but not how much protection it buys.

Fifth, this study estimates associations rather than causal effects. The fixed-effects design absorbs time-invariant confounders, and the lagged specification probes reverse causality. Unobserved time-varying confounding cannot, however, be excluded, and the lagged specification does not preserve the significance of the consumption and marriage coefficients. Future work could exploit the 2016 urban-rural insurance integration as a quasi-experimental shock, and could nest individual-level data within provincial macro data.

Conclusions

China's multi-tiered medical security system has made measurable progress. Effectiveness rose over 2011-2020, and the system performs well at protecting middle-aged and older adults from catastrophic health expenditure. Progress has nevertheless been uneven. Equity and synergy remain the weakest dimensions, and the urban-rural divide accounts for the overwhelming share of inequality, while regional differences account for almost none of it. Household economic capacity translates into protection only where insurance supply and service accessibility permit. Completing the multi-tiered system will therefore require consolidating the basic tier, activating the supplementary and floor tiers, building genuine inter-tier coordination, and targeting the urban-rural gap that currently defines the problem. The methodological implication is broader than the Chinese case, because measuring inequality on individuals rather than on jurisdictions changes which disparities are visible, and where policy should look.

Abbreviations

AHP, Analytic Hierarchy Process; CHARLS, China Health and Retirement Longitudinal Study; CHE, catastrophic health expenditure; CI, concentration index; OLS, ordinary least squares; OOP, out-of-pocket; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology; TOPSIS, Technique for Order Preference by Similarity to an Ideal Solution.

Declarations

Ethics approval and consent to participate

CHARLS was approved by the Biomedical Ethics Review Committee of Peking University (IRB00001052-11015), and all respondents provided written informed consent at each wave. This study analysed publicly available, de-identified secondary data, so no additional ethical approval was required.

Consent for publication

Not applicable.

Availability of data and materials

The CHARLS data are publicly available from the China Center for Health Development Studies at Peking University upon application (http://charls.pku.edu.cn). The analysis code (Stata do-files) that reproduces every table and figure is available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Soft Science Special Project of Gansu Basic Research Plan (Grant No. 26JRZA025), the Fundamental Research Funds for the Central Universities of Northwest Minzu University (Grant No. 31920250127), the Postgraduate Education and Teaching Reform Funds of Northwest Minzu University (Grant No. 2026JGYB13).

Authors' contributions

Yilong Wang conceived and designed the study, constructed the individual-level indicator system, performed all statistical analyses, interpreted the results and drafted the manuscript. Aihong He conceptualised and supervised the study, secured the funding, guided the methodological design and the interpretation of the findings, and critically revised the manuscript for important intellectual content. Xu Zhang assisted with data curation and the literature review. Yuxuan Wang assisted with data cleaning and the preparation of the tables and figures. All authors read and approved the final

manuscript. Yilong Wang is the guarantor of this work and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Acknowledgements

The authors thank the China Center for Health Development Studies at Peking University for making the CHARLS data publicly available.

References

1. Ministry of Civil Affairs of the People's Republic of China, China National Committee on Ageing. 2025 national report on the development of ageing undertakings. Beijing: Ministry of Civil Affairs; 2026.

2. National Health Commission of the People's Republic of China. Statistical bulletin on the development of health care in China 2023. Beijing: National Health Commission; 2024.

3. Lozano R, Fullman N, Mumford JE, Knight M, Barthelemy CM, Abbafati C, et al. Measuring universal health coverage based on an index of effective coverage of health services in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1250-84. doi: 10.1016/S0140-6736(20)30750-9

4. Kumah E, Ankomah SE, Fusheini A, Sarpong EK, Anyimadu E, Quist A, et al. Frameworks for health systems performance assessment: how comprehensive is Ghana's holistic assessment tool? Glob Health Res Policy. 2020;5:10. doi: 10.1186/s41256-020-00139-2

5. Zhou Q, Eggleston KN, Liu GG. Health insurance and subjective well-being: evidence from integrating medical insurance across urban and rural areas in China. Health Policy Plan. 2024;39(6):564-82. doi: 10.1093/heapol/czae031

6. General Office of the State Council. The 14th five-year plan for universal medical security. Beijing: State Council; 2021.

7. China Insurance Association. 2023 annual report on commercial health insurance operations. Beijing: China Insurance Association; 2024.

8. Liu G, Zhou Z, Ren Y, Zhai X, Zhuang Y, Wang Z, et al. Strengthening financial protection through multitiered health coverage: evidence from 667 055 low-income households in China. BMJ Glob Health. 2025;10(9):e019440. doi: 10.1136/bmjgh-2025-019440

9. Central Committee of the Communist Party of China, State Council. Opinions on deepening the reform of the medical security system. Beijing: Central Committee of the Communist Party of China; 2020.

10. National Healthcare Security Administration. The 15th five-year plan for universal medical security. Beijing: National Healthcare Security Administration; 2026.

11. National Healthcare Security Administration. Annual statistical bulletin on the development of medical security 2023. Beijing: National Healthcare Security Administration; 2024. (in Chinese)

12. Zhao N, Chen K. Equity and efficiency of medical and health service system in China. BMC Health Serv Res. 2023;23:33. doi: 10.1186/s12913-023-09025-2

13. Su W, Du L, Fan Y, Wang P. Equity and efficiency of public hospitals' health resource allocation in Guangdong Province, China. Int J Equity Health. 2022;21:138. doi: 10.1186/s12939-022-01741-1

14. Organisation for Economic Co-operation and Development. Health at a glance 2023: OECD indicators. Paris: OECD Publishing; 2023.

15. Wagstaff A, van Doorslaer E. Catastrophe and impoverishment in paying for health care: with applications to Vietnam 1993-1998. Health Econ. 2003;12:921-34. doi: 10.1002/hec.776

16. Xu K, Evans DB, Kawabata K, Zeramdini R, Klavus J, Murray CJL. Household catastrophic health expenditure: a multicountry analysis. Lancet. 2003;362:111-7. doi: 10.1016/S0140-6736(03)13861-5

17. China Banking and Insurance Regulatory Commission. Notice on regulating the development of inclusive commercial health insurance. Beijing: CBIRC; 2021.

18. Dong S, Ma J, Yu Z, Li P, Sun J, Li J. Can supplementary private health insurance reduce vulnerability to expected poverty and catastrophic health expenditure in China? Int J Equity Health. 2025;24:288. doi: 10.1186/s12939-025-02670-5

19. Doshmangir L, Bazyar M, Rashidian A, Gordeev VS. Iran health insurance system in transition: equity concerns and steps to achieve universal health coverage. Int J Equity Health. 2021;20:37. doi: 10.1186/s12939-020-01372-4

20. Zhang Z, Chu F, Zhao S. Evaluating China's multi-tiered medical security system: evidence from a multi-dimensional measure of balanced and adequate development. Comp Econ Soc Syst. 2025;6:102-18.

21. Reynolds MM. Health power resources theory: a relational approach to the study of health inequalities. J Health Soc Behav. 2021;62(4):493-511. doi: 10.1177/00221465211025963

22. Yan X, Liu Y, Cai M, Liu Q, Xie X, Rao K. Trends in disparities in healthcare utilisation between and within health insurances in China between 2008 and 2018: a repeated cross-sectional study. Int J Equity Health. 2022;21:30. doi: 10.1186/s12939-022-01633-4

23. Wang J, Zhu H, Liu H, Wu K, Zhang X, Zhao M, et al. Can the reform of integrating health insurance reduce inequity in catastrophic health expenditure? Evidence from China. Int J Equity Health. 2020;19:49. doi: 10.1186/s12939-020-1145-5

24. Donabedian A. Evaluating the quality of medical care. Milbank Mem Fund Q. 1966;44:166-206. doi: 10.2307/3348969

25. Donabedian A. The quality of care: how can it be assessed? JAMA. 1988;260:1743-8. doi: 10.1001/jama.1988.03410120089033

26. Tang D, Zhang Z, Wang X, Bu T, Feng Q. Equity-efficiency trade-offs in healthcare system reform: the perspective of tiered diagnosis and treatment in China. China World Econ. 2025;33(3):150-96. doi: 10.1111/cwe.12592

27. Jacques O, Noel A. Welfare state decommodification and population health. PLoS One. 2022;17(8):e0272698. doi: 10.1371/journal.pone.0272698

28. Wagstaff A, Flores G, Hsu J, Smitz MF, Chepynoga K, Buisman LR, et al. Progress on catastrophic health spending in 133 countries: a retrospective observational study. Lancet Glob Health. 2018;6:e169-79. doi: 10.1016/S2214-109X(17)30429-1

29. Wagstaff A, Flores G, Smitz MF, Hsu J, Chepynoga K, Eozenou P. Progress on impoverishing health spending in 122 countries: a retrospective observational study. Lancet Glob Health. 2018;6:e180-92. doi: 10.1016/S2214-109X(17)30486-2

30. Tang H, Li M, Liu LZ, Zhou Y, Liu X. Changing inequity in health service utilization and financial burden among patients with hypertension in China: evidence from China Health and Retirement Longitudinal Study (CHARLS), 2011-2018. Int J Equity Health. 2023;22:246. doi: 10.1186/s12939-023-02062-7

31. Zhang Y, Coyte PC. Inequality of opportunity in healthcare expenditures in China before and after the 2009 health reform. Health Econ Rev. 2020;10:39. doi: 10.1186/s13561-020-00294-9

32. Li X, Mohanty I, Zhai T, Chai P, Niyonsenga T. Catastrophic health expenditure and its association with socioeconomic status in China: evidence from the 2011-2018 China Health and Retirement Longitudinal Study. Int J Equity Health. 2023;22:194. doi: 10.1186/s12939-023-02008-z

33. Zhang A, Nikoloski Z, Mossialos E. Does health insurance reduce out-of-pocket expenditure? Heterogeneity among China's middle-aged and elderly. Soc Sci Med. 2017;190:11-9. doi: 10.1016/j.socscimed.2017.08.005

34. Li C, Jin S, Cao X, Han L, Sun N, Allore H, et al. Catastrophic health expenditure among Chinese adults living alone with cognitive impairment: findings from the CHARLS. BMC Geriatr. 2022;22:640. doi: 10.1186/s12877-022-03341-8

35. Fu X, Sun N, Xu F, Li J, Tang Q, He J, et al. Influencing factors of catastrophic health expenditure of farmers in poor rural areas in China: a cross-sectional study. Int J Environ Res Public Health. 2021;18:12360. doi: 10.3390/ijerph182312360

36. Chen H, Wang Z, Xiao Q, et al. Levels of multi-tiered medical security in China and their spatiotemporal differentiation. Mod Prev Med. 2024;51:3563-9. (in Chinese)

37. Zhang L, Chen R, Fang Y. Effects of urban and rural resident basic medical insurance on healthcare utilization inequality in China. Int J Public Health. 2023;68:1605521. doi: 10.3389/ijph.2023.1605521

38. Wang X, Qin C. Medical insurance benefits and health inequality: evidence from rural China. Front Public Health. 2024;12:1363764. doi: 10.3389/fpubh.2024.1363764

39. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol. 2014;43:61-8. doi: 10.1093/ije/dys203

40. Zhao Y, Strauss J, Chen X, Wang Y, Gong J, Meng Q, et al. China Health and Retirement Longitudinal Study wave 4 user's guide. Beijing: National School of Development, Peking University; 2020.

41. Wagstaff A, Paci P, van Doorslaer E. On the measurement of inequalities in health. Soc Sci Med. 1991;33:545-57. doi: 10.1016/0277-9536(91)90212-U

42. O'Donnell O, van Doorslaer E, Wagstaff A, Lindelow M. *Analyzing health equity using household survey data: a guide to techniques and their implementation*. Washington, DC: World Bank; 2008.

43. Wagstaff A, van Doorslaer E, Watanabe N. On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. J Econom. 2003;112:207-23. doi: 10.1016/S0304-4076(02)00161-6

44. Little RJA, Rubin DB. Statistical analysis with missing data. 3rd ed. Hoboken, NJ: Wiley; 2019.

45. Duan N, Manning WG, Morris CN, Newhouse JP. A comparison of alternative models for the demand for medical care. J Bus Econ Stat. 1983;1:115-26. doi: 10.2307/1391852

46. Manning WG, Duan N, Rogers WH. Monte Carlo evidence on the choice between sample selection and two-part models. J Econom. 1987;35:59-82. doi: 10.1016/0304-4076(87)90081-9

47. Wang Q, Zhang J, Xu Z, Yin P, Zhou M, Yang L, et al. Evolving trends, regional differences, determinants, and disease sources of provincial-level health inequalities in China 1990-2019: a temporal convergence and novel triple decomposition analysis. Int J Equity Health. 2024;23:203. doi: 10.1186/s12939-024-02283-4

Additional files

Additional file 1: STROBE checklist for cohort studies. Completed checklist itemising where each element of the STROBE statement is addressed in the manuscript. (DOCX)

Additional file 2: Table S1. Two-part model for imputing the missing 2020 expenditure variables. Estimated coefficients of the five equations, with z-statistics for the logit part and t-statistics for the four linear parts. (DOCX)

Additional file 3: Table S2. Heterogeneity by urban-rural status, region and age group. Each row is a separate two-way fixed-effects regression on the indicated subsample, with individual and wave effects removed by alternating projection and standard errors clustered at the individual level. (DOCX)

Additional file 4: Table S3. Missingness of the covariates by survey wave. Number and percentage of person-wave observations with a missing value for each covariate, reported separately for 2011, 2013, 2015, 2018 and 2020. (DOCX)

Additional file 5: Table S4. Robustness of the effectiveness index to the construction of the index. Re-derived weights on the four waves with complete expenditure data, the two-part re-imputation of the 2020 values, the treatment of respondents with zero medical expenditure, and standard errors clustered at the province-wave level. (DOCX)

Additional file 6: Table S5. Robustness of the coefficient estimates. Six two-way fixed-effects specifications, namely baseline, entropy-weighted TOPSIS instead of the weighted index, exclusion of the 2020 wave, the balanced panel of respondents observed in all five waves, complete cases with no imputation, and regressors lagged one wave. Coefficients on education, marital status, log household consumption per capita and medical need, with standard errors clustered at the individual level. (DOCX)

Supplementary Files

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

- Additionalfile1.docx
- Additionalfile3.docx
- Additionalfile4.docx
- Additionalfile5.docx
- Additionalfile6.docx
- Additionalfile2.docx