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.

What hospitals do, not how their revenue is composed: service content and the uneven depth of the COVID-19 shock in Zhejiang, China

Yongqiang Lai
Hangzhou Medical College
Ye Li
Hangzhou Normal University
Feiyang Zheng
Hangzhou Medical College
Fangqi Qu
Minzu University of China
Xiyu Zhang
Peking University
Jinqin Jiang
Hangzhou Medical College
Yaming Gu
wstgym@126.com

Hangzhou Medical College

Research Article

Keywords: health system resilience, service disruption, hospital revenue, deferrable care, health policy monitoring, China

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11615326/v1

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

Additional Declarations: No competing interests reported.

Abstract

Background

When hospital revenue falls during a shock, the headline decline conflates two problems that require different responses: patients receiving fewer services, and each service earning less. The COVID-19 shock hit hospital finances unevenly, and why some hospitals lost far more than others is usually sought in their baseline revenue structure.

Methods

Using monthly monitoring reports from public hospitals in Zhejiang Province, China (2016–2022; 30,571 hospital-months), we separate each hospital's revenue change into volume and unit-service-revenue components, and then ask which baseline characteristics predict shock depth.

Results

Revenue fell to an index of 75.56 in February–April 2020 (2019 same month = 100). Almost all of the decline (134.4%) came from fewer services, while revenue per service rose and buffered the fall; in the recovery, positive growth came mainly from revenue per service (80.7%). Baseline revenue structure explains almost none of the cross-hospital variation in shock depth: the mechanical channel through which structure could act accounts for 0.1% of variance, and structural shares add no explanatory power ($\Delta R^2 = 0.001$). What separates hospitals is what they do. Those whose work is largely deferrable care, notably dental, ophthalmic and rehabilitation services, lost far more than those centered on emergency and chronic-disease care.

Conclusions

Resilience monitoring should therefore report the volume and revenue-per-service channels separately, alongside service content: three indicators built from routine reports identify which hospitals are exposed before the next shock arrives, and distinguish a care-backlog problem from a revenue-per-service problem. Reporting only how much revenue fell misses this choice.

1 Background

The COVID-19 pandemic delivered a synchronized shock to health systems, and its most immediate financial manifestation was a sharp fall in hospital revenues. Across ten national health systems, outpatient and inpatient volumes fell by more than 30% in the month of the shock, while revenues fell further and recovered more slowly [1]; responses in 28 countries differed markedly in the depth and duration of service disruption [2]. The shock shifted health-systems research from whether systems can

withstand shocks to how resilience can be measured in routine monitoring data. Progress since has been uneven: of 68 studies measuring resilience in high-income countries, more than 80% used quantitative indicators, but these concentrated on resources and service delivery, touched the recovery and learning stages least, and were poorly standardized [3]. What can be measured and what needs to be monitored are not yet aligned.

In China, the sharpest revenue fluctuation came in February–April 2020, followed by three years of recovery and recurrence. “How much hospital revenues fell” thus became the most cited figure in monitoring reports, fiscal-compensation calculations, and performance assessment. That figure mixes two problems of different nature: whether patients received fewer services, or each service earned less.

The first is a level problem: how much of the decline reflects fewer services, and how much reflects a change in revenue per service? If outpatient visits fall by 30% while revenue falls by 20%, the hospital has experienced both a volume contraction and a relative rise in unit service value, and “revenue fell 20%” alone does not describe the impact on its operations. Separating the two yields a quantity comparable across institutions and over time, which we call the asymmetric gap.

The second is a dispersion problem: under the same shock, why did some providers fall by 40% and others by only 10%? A popular answer is that structure determines resilience: hospitals with a higher drug share, or a higher service-revenue share, are more resistant. Reviews of organizational resilience do list governance, resources and service composition as core elements [4, 5], and resilience indices use structural indicators as rankable proxies [6]. Yet the answer has an arithmetic problem. The decline in total revenue is itself a weighted average of component growth rates, weighted by baseline shares, so “a larger share of a component with a smaller decline yields a smaller total decline” holds without any behavioral mechanism. In such share × common-shock measures, identifying power comes from the shares themselves, and the mechanical association survives controls [7, 8]. Treating it as evidence of structural resilience treats an identity as a finding.

Our argument sits at the intersection of three literatures. The first measures health-system and hospital resilience, moving from conceptual definition to indicator construction and developing a resistance–recovery-speed–recovery-shape tradition [4–6, 9–13], but characterizes resilience along a single dimension and lacks a quantity that separates revenue from volume. The second explains what makes hospitals resistant, mostly by regressing shock magnitude on structural characteristics [14–19]; its problem is the dependent variable, in which the structural variable appears on both sides. The third is the price–quantity separation tradition [20, 21]: monitoring reports give no component price indices, but they give component revenues and volumes, whose ratio is unit-service revenue and permits a restricted separation. We join the three: the tools of the third answer the first's measurement question, and that answer bounds the second's explanatory space.

The paper answers three questions. Q1 (level): of the relative change in total revenue during the shock, how much occurred on the unit-service-revenue side and how much on the volume side, and how does the difference evolve? Q2 (dispersion): can cross-hospital variation in the decline be explained by

baseline revenue structure? Q3 (linkage): does the answer to Q1 constrain the answer to Q2, that is, is the explanatory space available to structural variables narrow to begin with? Q3 is the increment. Structural shares can affect total growth through only two paths: differences in components' own growth rates, or reallocation of shares during the shock. Component growth rates differ mainly on the unit-service-revenue side, which province-wide pricing and payment rules drive and which is highly co-directional across hospitals; a four-step diagnostic finds the reallocation limited and not persistent. The space that “structure determines resilience” requires is therefore narrow arithmetically and narrower empirically, and the negative answer to Q2 is a corollary of the data-generating process rather than a sampling accident.

The paper makes three contributions. First, measurement: it decomposes the decline into the asymmetric gap and the unit-service-revenue–volume decomposition. Second, identification: it breaks the structure-determines-resilience proposition into four testable links (a mechanical-component decomposition, a placebo noise floor, cross-year persistence, and residual explainability) and reaches a negative conclusion, using two one-sided tests (TOST) to distinguish “no effect” from “insufficient power” rather than treating non-significance as a conclusion. Third, method and conclusion are linked: the shock-period decline is dominated by volume, the recovery-period growth by unit-service revenue, and cross-hospital variation in shock depth by service content, with revenue-structure shares playing no role at either end.

2 Study setting, data and variables

2.1 Institutional background

Zhejiang is among China's most developed provinces. In 2022 its GDP was 77,715 billion yuan, GDP per capita 118,496 yuan, and permanent population 65.77 million, with per-capita disposable income of 60,302 yuan and an urban–rural income ratio of 1.90. In June 2021 the central government designated it a demonstration zone for common prosperity and coordinated urban–rural development [22, 23]. Its population is also aged: at the end of 2022, residents aged 60 and above accounted for 20.2% and those aged 65 and above for 14.9% [24], shifting demand toward chronic disease and elderly care.

Zhejiang moved early on public-hospital reform: comprehensive reform in 2010; from 2012 the “two sinkings and two improvements” policy, under which 54 tertiary-A hospitals supported 122 county-level hospitals; and designation as a national comprehensive medical-reform pilot province in 2016 [25]. County-level medical communities were piloted in 11 counties in 2017 and rolled out province-wide in 2018 [26], reaching 161 communities with an in-county consultation rate of 89.54% by the end of the 13th Five-Year Plan (public data of the Zhejiang Provincial Health Commission). Drug and consumable markups were abolished, volume-based procurement introduced, and medical-service prices adjusted throughout 2016–2022 [27, 28].

Three institutional processes bear on our analysis. Payment reform moved from fee-for-service to payment by diagnosis-related groups or disease types, with global budgets and retained surplus inside county-level medical communities, changing both component structure and unit-service revenue. Pricing and markup policy (abolition of drug and consumable markups, volume-based procurement, upward adjustment of technical-service prices) ensures that unit-service revenue does not move with volume. Pandemic containment (temporary service suspensions) depressed volume, while nucleic-acid testing capacity raised component revenues. Because the three overlap, we attribute no change to a single policy; together they define the boundary of our identification, namely that unit-service revenue is driven by a province-wide policy package. Its change therefore contains components we cannot separate, so we use “unit-service-revenue effect” rather than “price effect” (Section 3.2).

2.2 Data source, cleaning, and descriptive statistics

The data are monthly operational monitoring reports of Zhejiang public hospitals, covering 84 months from January 2016 to December 2022 and 39 fields. Hospitals complete them monthly and county and prefecture authorities review them; they are administrative monitoring data. We follow the reporting guidelines for observational studies in declaring definitions and limitations item by item [29–31]. The monetary unit is 10,000 yuan.

Six cleaning rules applied in sequence (criteria, exclusion counts, and classification in Appendix Table A1) yield 30,571 hospital-month observations covering 376 institutions and 92 county units;

Table 1
Descriptive statistics of key variables (full cleaned sample, 2016–2022)

Variable	n	Mean	SD	P25	Median	P75
Total medical revenue (10,000 yuan)	30,571	3,656.13	6,514.92	710.11	1,715.89	3,828.10
Outpatient and emergency visits	30,571	57,703	63,670	19,192	39,685	71,714
Inpatient discharges	30,571	1,875	2,608	420	1,041	2,284
Cost per outpatient visit (yuan)	30,568	247.14	389.71	187.53	223.95	268.47
Cost per inpatient discharge (yuan)	29,153	9,146.99	5,338.29	6,187.15	8,222.23	10,656.36
Drug revenue share (%)	30,569	32.96	11.45	26.6	32.6	39.13
Medical-service revenue share (%)	30,569	33.16	12.7	26.04	30.69	36.69
Test and examination revenue share (%)	30,569	24.48	35.42	18.81	24.35	28.97
Outpatient revenue share (%)	30,569	52.05	41.93	39.11	48.89	61.21

Notes: The sample is all 30,571 cleaned hospital-month observations (376 hospitals, 84 months), neither the balanced panel nor the analysis sample. Monetary amounts are in 10,000 yuan; per-service costs are converted to yuan; shares are percentages; visit counts are integers with thousands separators; other variables are reported to two decimals. Revenue and share variables are right-skewed, so medians and quartiles are the primary readings; the median sum of the three revenue shares is 87.6%, consistent with the “three mutually exclusive components + other” definition.

Table 1 reports levels and dispersion. The mean and dispersion of total revenue set the scale of the dependent variables, and outpatient and inpatient volumes differ by two orders of magnitude, one reason the volume index in Section 3.1 is an equal-weight composite rather than a magnitude sum. The baseline characteristics (drug share, medical-service revenue share, outpatient revenue share) vary far more across institutions than their annual means vary over time, which both permits their use as cross-sectional regressors and limits their explanatory space.

2.3 Variables, samples, and measurement definitions

Revenue covers medical services, drugs, test-and-examination, and consumable-materials cost; volume is outpatient-and-emergency visits and inpatient discharges; cost is medical expenditure. Three definitions fix which figures can be placed together. The composition vector uses only the three mutually exclusive components (drugs, medical services, tests), with consumable-materials cost treated as a cost

concept and the balance recorded as “other.” Nucleic-acid testing revenue in 2022 is converted to a consistent reporting unit (audit in Appendix Table A5), so test-related indicators are reported with and without it, the inclusive measure being primary. Row-level compliance with the additive identity is a report-quality variable, not an admission threshold (rates in Appendix Table A7). Derivations and coverage are in Appendix B.1.

Two samples differ in inclusion criteria, not data quality. The balanced panel (312 hospitals) requires complete observation over 84 months, is used for provincial series, indices, the asymmetric gap and spatial stratification, and must be complete because the indices use the same month of 2019 as base. The analysis sample (371 hospitals) requires at least 10 months in 2019 and 30 months in 2020–2022, and is used for the four resilience elements and the hospital-level and panel regressions. Their cumulative shock depths over February–April 2020 differ by one decimal place with the same sign, so results are insensitive to composition (Appendix A.8).

3 Methods

3.1 Shock window and the asymmetric gap

Calendar noise in monthly year-on-year rates comes mainly from the shifting timing of the Spring Festival. Single-month February revenue growth was + 25.13%, − 1.61% and + 8.28% in 2017–2019, whereas cumulative January–February growth was + 12.86%, + 14.38% and + 7.47%: the single-month measure swings by more than twenty percentage points across years, the cumulative one far less (Appendix Table A4). We therefore use the cumulative February–April 2020 rate as the primary measure, reporting single-month and annual measures only side by side. On direction: February 2020 contained no Spring Festival holiday while the baseline months carry the post-festival low base, which pushes the February rate upward, so the single-month figure understates rather than overstates the collapse.

Using the same month of 2019 as base (= 100), we construct a total-revenue index IR and a volume index IV, and define the asymmetric gap as their difference in index points: Gt = IRt − IVt. The volume index is the equal-weighted average of visits and discharges. We neither add headcounts directly, since their magnitudes differ by two orders, nor weight them by 2019 revenue shares (0.4126 outpatient, 0.5858 inpatient), which would place the object of the test inside the measure. Appendix Table B3 reports revenue-share-weighted and log-difference alternatives; under all four constructions the share of months with a positive gap is 100%. Four details: the base is the same month to absorb seasonality; a zero or missing base leaves the index undefined and the observation is dropped, not imputed; the gap is computed at the provincial level and separately at the hospital level; and it is a difference between indices, not an accounting loss (Appendix B). The gap is a level difference, whereas the decomposition below is a share-weighted logarithmic identity, so the two are reported separately and never converted.

3.2 The unit-service-revenue–volume decomposition

We decompose the monthly logarithmic change in total medical revenue. Let R be total revenue, v outpatient-and-emergency visits, d inpatient discharges, fem cost per outpatient visit (outpatient-and-emergency revenue ÷ v), fip cost per inpatient discharge (inpatient revenue ÷ d), sem and sip the 2019 same-month shares of the two parts, and u a residual term; then:

$$\Delta \ln R_t \approx [s_{em}\, \Delta \ln f_{em,t} + s_{ip}\, \Delta \ln f_{ip,t}] + [s_{em}\, \Delta \ln v_t + s_{ip}\, \Delta \ln d_t] + u_t$$

The bracketed terms are the unit-service-revenue effect and the volume effect. The formula is approximate for two reasons: weights are fixed at baseline shares, which vary within the window and leave an interaction term; and the logarithmic difference is a first-order expansion of the growth rate. The combined residual is reported at 9.7%, a low residual share being necessary for the decomposition to be credible. Decomposition is monthly over 31 months. One definitional statement accompanies the framework: per-service cost is nominal and its rise contains at least four components (active adjustment of medical-service prices, changes in service composition, changes in insurance payment and settlement rules, and pass-through of abolished consumable markups), which we cannot separate without independent price data; hence “unit-service-revenue effect” rather than a single price meaning.

3.3 The four resilience elements and the estimation specification

Whether the structural explanation can hold depends on whether resilience is measured clearly (Appendix Figure A0 summarizes the framework). We model four dimensions. Shock depth is the cumulative revenue year-on-year decline over February–April. Recovery speed is the first month in which the three-month moving average returns to the 2019 monthly average, counted from January 2020 and 0-based; read 1-based, the median would be 9 rather than 8 months. Recovery shape has four mutually exclusive categories (V ≤ 12 months, U 13–24, L > 24, censored) summing to the sample, with non-recovered hospitals listed as censored. Long-term displacement is the change in 2022 average monthly revenue relative to 2019, nominal. We report them side by side rather than as a composite, because weighting requires weights with no empirical basis; reviews likewise caution that resilience indicators must be read with the dynamics of the shock, not as ranking scores [3, 6].

The recovery period begins in June 2020 by definition, not by data boundary. February–May 2020 was still the shock's tail, with May's single-month Δln revenue at − 0.1061; placing May in the shock period would understate depth, and placing it in the recovery period would record “still falling” as “recovering.” Those months are therefore listed separately (Appendix Table B2) and assigned to neither window. The criterion uses the 2019 monthly average as a static baseline; since nominal revenue grew by about 10% a year before the shock (Appendix C), this baseline is easier to reach than a counterfactual extrapolated from the pre-shock trend. Appendix Table B9, row S8, reports a trend baseline, and the two differ substantially, so all recovery conclusions are stated relative to the 2019 monthly average rather than as absolute dates.

The main estimate uses a panel interaction specification:

$$\Delta \ln R_{it} = \alpha_i + \gamma_t + S_t \cdot X_i'\beta + S_t \cdot M_i \cdot \delta + \varepsilon_{it}$$

Here $\Delta \ln R_{it}$ is the monthly year-on-year logarithmic change; S_t a shock dummy equal to 1 in February–April 2020; X_i a vector of time-invariant baseline characteristics (drug share, medical-service revenue share, outpatient revenue share, institution-type dummies); M_i the mechanical component (components' common growth rates substituted into the weighted identity, entered as a regressor); α_i a hospital fixed effect absorbing the main effects of X_i [32]; γ_t a calendar-month fixed effect; and standard errors clustered by hospital [33, 34]. The sample is 13,348 hospital-months (371 hospitals), January 2020 to December 2022. Identification comes entirely from $S_t \times X_i$, so estimates capture relative differences during the shock, and entering the mechanical component directly makes the structural coefficients net of the additive identity.

We apply two one-sided tests (TOST [35, 36]) to the structural variables, to distinguish “no effect” from “insufficient power.” The criterion is that the 90% confidence interval lies entirely within (−Δ, +Δ), so the conclusion depends on the margin. We set Δ = 0.05 against the shock's own magnitude: the mean logarithmic year-on-year rate during the shock is − 0.314, and Δ = 0.05 is 15.9% of it, asking whether a structural variable's effect exceeds about one-sixth of the shock. This is economic, not statistical, significance.

Because resilience has four non-mergeable elements, we also run a two-step incremental test for each: the service-content block first, the revenue-structure block second, with ΔR^2 and the joint test of the structural block as the answer (Section 4.4).

4 Results

4.1 Shock profile and the revenue–volume divergence

Over February–April 2020 the total-revenue index was 75.56, the outpatient-and-emergency index 64.06 and the discharge index 71.79 (2019 same month = 100). For February alone the indices are 58.43, 46.08 and 56.31, so the two measures differ by 17.13, 17.98 and 15.48 percentage points. Month by month, revenue fell 41.08% in February, turned positive from August, and remained negative in 7 of the 12 months; outpatient visits fell 53.46% and turned positive only in November; discharges fell 42.79% and were negative in all 12 months, still − 1.51% in December (Appendix Table A3). By the end of 2020 revenue had returned to and slightly exceeded the previous year's level, while inpatient volume had not.

A: provincial indices of total revenue, outpatient-and-emergency visits and discharges (2019 same month = 100). B: the asymmetric gap and its three-month moving average. C: single-month (February 2020) versus cumulative (February–April) shock depth. D: recovery trajectories. The two reference lines in D differ: the solid line is the recovery criterion (three-month moving average = 100, i.e. return to the 2019 monthly average), the dotted line the lower bound 95 of the sensitivity grid (S1), not a criterion (Appendix Table B9).

Over the 31 months from June 2020 to December 2022 the gap was positive in every month (31/31; sign test $p = 9.31 \times 10^{-10}$, treating months as independent; Newey–West standard errors [37] and bootstrap results in Appendix Table B3), mean 8.1 and median 8.21 percentage points, range 2.77–13.89. We use the raw series; the three-month moving-average median is 7.89 and the two are not interchangeable. The gap widens (slope 0.18 points per month, SE 0.051, t = 3.51),. At hospital level the direction is the same: 274 hospitals (90 county clusters), mean 7.72 (t = 10.16), 83.9% positive.

The ex-ante placebo shifts the construction back one window (base 2016, event 2017): the median gap is 6.15 percentage points, 74.9% of the main result, with a steeper slope. A positive deviation of revenue from volume is therefore a long-run feature of the province's hospitals, not a 2020 effect, and we report no quantity for a pandemic-caused gap: the 8.21 percentage points is the observed level, about three-quarters reproducible before the pandemic (Appendix B.10).

4.2 Window dependence of the decomposition: the volume side in the shock period, the unit-service-revenue side in the recovery

We decompose monthly log changes following Section 3.2 (monthly values in Appendix Table B13, trajectories in Fig. 2; window and level comparison in Table 2; full detail in Appendix Table B14; variance measures in Appendix Table B15). The decomposition is window-dependent, and the two windows give opposite answers.

Shock period (February–April 2020): Δln revenue − 0.2803, unit-service-revenue effect 0.0964, volume effect − 0.3767. By average contribution share, volume accounts for 134.4% and unit-service revenue for − 34.4%; absolute shares exceed 100% because the components have opposite signs. At the hospital level (345 hospitals) the direction is the same: Δln revenue − 0.2541, unit-service revenue 0.0891 (share − 35.1%), volume − 0.3499 (share 137.7%).

Recovery period (June 2020–December 2022, 31 months): mean Δln revenue + 0.087, of which unit-service revenue accounts for 80.7% and volume 9.6%. This decomposes positive growth, not a decline. Including the shock period (35 months, mean 0.0472) gives a unit-service-revenue share of 155%. Windows are never converted into one another, and every share is cited with its window.

Table 2
Window dependence of the decomposition: provincial versus hospital level (unit-service-revenue–volume decomposition)

Window	Level	n	Δln revenue	Unit-service-revenue effect	Volume effect	Share P (%)	Share V (%)
Shock, Feb–Apr 2020 (cum.)	Provincial aggregate (balanced panel)	3	-0.28	0.096	-0.377	-34.4	134.4
Recovery, Jun 2020–Dec 2022 (mean)	Provincial aggregate (balanced panel)	31	0.087	0.07	0.008	80.7	9.6
Shock, Feb–Apr 2020 (cum.)	Hospital level (analysis sample)	345	-0.254	0.089	-0.35	-35.1	137.7
Recovery, Jun 2020–Dec 2022 (mean)	Hospital level (analysis sample)	345	0.041	0.057	-0.035	138.4	-84

Notes: only the four core rows are shown; the full window, level and weighting comparison and the monthly decomposition are in Appendix Tables B14 and B13. The periods are not interchangeable: Δln revenue is − 0.280 in February–April 2020 and + 0.087 from June 2020. A share is effect ÷ Δln revenue, negative when the signs differ, meaning the component buffered rather than caused the change. The windows give opposite answers (V share 134.4% in the shock; P share 80.7% in recovery). Hospital and provincial results agree in sign; the hospital-level recovery volume share is − 84.0% but + 19.9% weighted by baseline size.

One disagreement within the recovery window must be recorded: provincial aggregate and hospital-equal weighting have opposite signs. The volume effect is 0.0083 (share 9.6%) at the provincial aggregate but − 0.0348 (− 84%) at hospital-equal weight; weighted by baseline size it is 0.0145 (19.9%), varying monotonically with size (small − 0.0518 / medium − 0.0302 / large − 0.0224). The disagreement comes from the weighting, not the data: the volume recovery is concentrated in larger hospitals, so the equal-weight figure describes the smaller ones (Appendix Table B14).

One number is easily misread: the unit-service-revenue effect's variance share in the price–volume block is 32.1%, which seems to contradict “highly consistent”. The two measure different things: 32.1% is the dispersion of each hospital's own effect around its mean, whereas “highly consistent” refers to the common province-wide component, which weighted by baseline shares yields the mechanical component with 0.1% variance. Only the latter bears on our conclusion (Appendix Table B15).

The two windows answer “which side matters more” oppositely but lead to the same conclusion: revenue-structure shares have no explanatory space, and the reason lies in variance. The mechanical component is the only channel through which structure can act. Common growth rates are highly

consistent across hospitals (weighted range 8.29 points), so its variance share is only 0.1% (1.6% full year): even if structure fully determined the weights, it would explain under 1%. Differences fall on the volume side (103.5%, offset by covariance − 33.68%).

A: decomposition of the cumulative logarithmic change (unit-service-revenue effect, volume effect, and residual). B: indices of cost per outpatient visit and per inpatient discharge (2019 same month = 100). C: shock depth versus recovery time. D: median gap versus recovery time.

From February–April 2019 to February–April 2020, common growth rates were − 19.72% for medical services, − 22.52% for drugs, − 20.41% for tests and − 28.01% for other: the declines were close across categories.

Returning to the gap: the median unit outpatient fee index is 110.1 (2019 = 100) and the unit inpatient fee index 105.8. The revenue index exceeds the outpatient-visit index by 3.73 percentage points (t = 3.84) and the discharge index by 11.7 points (t = 9.31; balanced panel ∩ named counties, n = 274, 90 county clusters, equal weights, CR1 standard errors). The inpatient side drives the lag: discharges were negative in all 12 months of 2020 while unit inpatient fees rose.

One boundary on the “buffering” interpretation: Section 3.2 noted four components in per-service cost change, of which “change in service composition” most likely reflects a passive rise in severe cases as mild cases deferred care. That is a behavioral and epidemiological component, not a province-wide rule component. If it is non-trivial, part of the shock-period buffer is a case-mix artifact, which weakens the premise that unit-service revenue is driven by unified rules and therefore affects the interpretation, not the values, of the Section 4.2 decomposition. We lack case-mix, severity and referral-flow data to separate it (Section 5.4).

4.3 Can revenue structure explain cross-hospital differences: a four-step diagnostic

Existing studies take a regression of decline on structural indicators as evidence of resilience, but, as Section 4.2 shows, the decline is arithmetically a weighted average of component growth rates. We use a four-step diagnostic: structural differences can be called a source of resilience only if all four steps pass (Appendix Tables B4–B8).

Four alternatives must be ruled out. First, a falsification check: report quality correlates with the absolute residual at only 0.018, and residual SDs are similar where identities hold (172 hospitals, 19.13) and where they deviate (199, 13.3). Second, institution-type power is concentrated in specialty outpatient care: dropping dental hospitals cuts it from 16.7% to 2.4%, and the structural variables' correlations with the residual (0.101, − 0.174, − 0.27) follow the step-four gradient. Third, the correlation pattern (Table 3): the two revenue shares correlate with total change at 0.016 and 0.04, whereas the outpatient share is − 0.265 and correlates more with volume (− 0.209) than with unit-service revenue (− 0.121). Fourth, specification sensitivity (Appendix Table B18): raw shares are collinear with the constant, inflating

standard errors threefold, and under the normalized composition the joint test turns significant (p = 0.0136), so we do not rest on it.

Table 3
Correlations of baseline revenue structure with the decomposition components

Window	Baseline structural variable	Total change ΔlnR	Unit-service-revenue effect P	Volume effect V	Residual u
Shock (Feb–Apr 2020, cum.)	Baseline drug revenue share	0.016	-0.071	0.045	0.089
Shock (Feb–Apr 2020, cum.)	Baseline medical-service revenue share	0.04	-0.107	0.095	0.035
Shock (Feb–Apr 2020, cum.)	Baseline outpatient revenue share	-0.265	-0.121	-0.209	0.142
Shock (Feb–Apr 2020, cum.)	Baseline size, ln(2019 revenue)	-0.181	0.145	-0.236	-0.194
Recovery (Jun 2020–Dec 2022, mean)	Baseline drug revenue share	-0.054	-0.06	-0.017	-0.025
Recovery (Jun 2020–Dec 2022, mean)	Baseline medical-service revenue share	0.113	0.013	0.09	0.07
Recovery (Jun 2020–Dec 2022, mean)	Baseline outpatient revenue share	0.04	0.306	-0.132	0.101
Recovery (Jun 2020–Dec 2022, mean)	Baseline size, ln(2019 revenue)	0.055	-0.054	0.114	-0.237

Notes: hospital level (345 hospitals), Pearson correlations. The three revenue shares correlate with total change below 0.06 in absolute value; the outpatient share, a service-content measure, correlates at − 0.265. Correlations are not regression coefficients and carry no controls (controls in Table 4; specification sensitivity in Appendix Table B18).

A–C: monthly coefficients for total medical revenue, outpatient-and-emergency visits, and inpatient discharges (in multiples of the exposure standard deviation; base period January 2020). D: comparison under group (prefecture-city) and trend controls.

4.4 Correlates: service content, not revenue structure

We first report the four elements (Appendix Table C1; Fig. 1D). Shock depth across 371 hospitals: mean − 21.52%, median − 21.74%, SD 16.28, 10–90 range − 35.63% to − 6.39%, extremes − 94.5% and + 122.4%. Recovery speed: median 8 months. Recovery shape: 283 V (76.3%), 65 U (17.5%), 8 L (2.2%), 15 censored

(4%). Long-term displacement: median + 7.87%. A level caveat: February–April mean rates are − 30.18% for visits (median − 32%) and − 21.52% for revenue, so volume fell more deeply; these are hospital-equal-weighted and not comparable to the provincial index of 64.06.

Main estimates are in Table 4. In M5 the outpatient-share interaction is − 0.7163 (− 0.0716 per 10 percentage points, $p < 0.001$): hospitals with a higher outpatient share fell more, consistent with exposure to deferrable services. Institution-type interactions are jointly significant ($F = 7.19$, $p = 1.38\times10^{-5}$), strongest for dental hospitals (− 1.65) and positive for maternal-and-child hospitals (+ 0.11); the drug share is 0.5063 ($p = 0.083$) and the medical-service share − 0.2246 ($p = 0.539$). Shares are mechanically collinear, so single-share coefficients are compared only in magnitude: the medical-service coefficient moves from − 1.401 to − 0.225 across M2–M5 with no stable sign. The outpatient coefficient does not depend on dental hospitals: excluding them (363 hospitals) gives − 0.0735 ($p < 0.001$), almost identical to − 0.0716 (Appendix Table B10).

This coefficient is the only positive finding the paper rests on. Institution type is inferred from name keywords and is a proxy; the outpatient share is not, being taken directly from report fields. Its significance has two channels we cannot separate, exposure to deferrable services and differences in price adjustment between outpatient and inpatient care; we document the association only (Section 5.4). The gradient does not depend on institution-type composition: within general hospitals ($n = 207$), each 10 points is associated with − 3.002 points (SE 0.662, $p < 0.001$). Its level drifts (within-hospital SD 2.03; 2016–2019 correlation 0.565) but its ranking is stable (0.937; 76.4% same quartile), so it supports rank-based screening, not thresholds.

Table 4
Main estimates of the panel interaction specification (Shock × baseline characteristics; dependent variable = monthly logarithmic year-on-year change)

Spec.	Variable	Coef. (SE)	95% CI	p	per 10 pp	per 1 SD	R^2
M1	Shock × drug share	1.067 (0.424)	[0.237, 1.898]	0.012	0.1067	0.1234	0.214
M2	Shock × drug share	0.169 (0.268)	[-0.357, 0.694]	0.529	0.0169	0.0195	0.225
M2	Shock × medical-service revenue share	-1.401 (0.446)	[-2.275, -0.528]	0.002	-0.1401	-0.1584	0.225
M3	Shock × drug share	0.776 (0.368)	[0.054, 1.497]	0.035	0.0776	0.0897	0.244
M3	Shock × medical-service revenue share	-0.888 (0.318)	[-1.512, -0.264]	0.005	-0.0888	-0.1004	0.244
M3	Shock × outpatient revenue share	-0.862 (0.214)	[-1.281, -0.442]	< 0.001	-0.0862	-0.1598	0.244
M4	Shock × drug share	1.073 (0.312)	[0.461, 1.684]	< 0.001	0.1073	0.124	0.281
M4	Shock × medical-service revenue share	0.718 (0.421)	[-0.108, 1.544]	0.088	0.0718	0.0811	0.281
M4	Shock × outpatient revenue share	-0.549 (0.140)	[-0.824, -0.274]	< 0.001	-0.0549	-0.1019	0.281
M4	Shock × institution type (dental)	-1.790 (0.551)	[-2.871, -0.709]	0.001	—	—	0.281
M4	Shock × institution type (maternal & child)	0.260 (0.076)	[0.110, 0.409]	< 0.001	—	—	0.281
M4	Shock × institution type	0.064 (0.048)	[-0.030, 0.158]	0.181	—	—	0.281

Notes: M1 drug share only; M2 adds medical-service share; M3 outpatient share; M4 institution type; M5 mechanical component and size. All use n = 13,348, so no n column; R^2 by specification. Shock = 1 in February–April 2020; hospital fixed effects absorb time-invariant characteristics including prefecture city, calendar-month fixed effects absorb seasonality; standard errors clustered by hospital. Base type is general hospitals; joint test in M5 F = 7.19 ($p = 1.38\times10^{-5}$). The “per 10 pp” column applies only to the three shares (0–1 scale); dummies and log size are marked “—”. Shares are collinear, so single-share coefficients are compared only in magnitude. Mechanical component SD 0.51 points, so one SD is about 7.35 points. “Per 1 SD” rescales by each variable's SD: drug + 0.0585, medical service − 0.0254, outpatient − 0.1328, mechanical + 0.0735, size − 0.0701.

Spec.	Variable	Coef. (SE)	95% CI	p	per 10 pp	per 1 SD	R^2
	(other specialty)						
M4	Shock × institution type (TCM)	-0.001 (0.027)	[-0.054, 0.052]	0.967	—	—	0.281
M5	Shock × drug share	0.506 (0.292)	[-0.067, 1.079]	0.083	0.0506	0.0585	0.286
M5	Shock × medical-service revenue share	-0.225 (0.366)	[-0.941, 0.492]	0.539	-0.0225	-0.0254	0.286
M5	Shock × outpatient revenue share	-0.716 (0.118)	[-0.947, -0.485]	< 0.001	-0.0716	-0.1328	0.286
M5	Shock × institution type (dental)	-1.650 (0.508)	[-2.645, -0.654]	0.001	—	—	0.286
M5	Shock × institution type (maternal & child)	0.110 (0.070)	[-0.027, 0.248]	0.116	—	—	0.286
M5	Shock × institution type (other specialty)	0.067 (0.042)	[-0.015, 0.149]	0.108	—	—	0.286
M5	Shock × institution type (TCM)	0.021 (0.028)	[-0.033, 0.075]	0.45	—	—	0.286
M5	Shock × mechanical component	14.405 (7.959)	[-1.194, 30.004]	0.07	—	0.0735	0.286
M5	Shock × size (log)	-0.055 (0.019)	[-0.092, -0.018]	0.003	—	-0.0701	0.286

Notes: M1 drug share only; M2 adds medical-service share; M3 outpatient share; M4 institution type; M5 mechanical component and size. All use n = 13,348, so no n column; R^2 by specification. Shock = 1 in February–April 2020; hospital fixed effects absorb time-invariant characteristics including prefecture city, calendar-month fixed effects absorb seasonality; standard errors clustered by hospital. Base type is general hospitals; joint test in M5 $F = 7.19$ ($p = 1.38\times10^{-5}$). The “per 10 pp” column applies only to the three shares (0–1 scale); dummies and log size are marked “—”. Shares are collinear, so single-share coefficients are compared only in magnitude. Mechanical component SD 0.51 points, so one SD is about 7.35 points. “Per 1 SD” rescales by each variable's SD: drug + 0.0585, medical service − 0.0254, outpatient − 0.1328, mechanical + 0.0735, size − 0.0701.

Whether structural variables add explanatory power for the four elements is answered by a two-step test (service content first, revenue structure second; Appendix Table B17). For shock depth, ΔR^2 is 0.001 (joint $p = 0.9029$), no incremental power; for recovery speed 0.024–0.036 ($p \le 0.0051$); for recovery shape 0.028 ($p < 0.001$); for long-term displacement 0.029 ($p = 0.0796$). We report these as they are and limit

the explanatory conclusion to shock depth, with no mechanism claim for the time dimension. Replacing the two raw shares with the normalized three-way composition gives $\Delta R^2 = 0.0018$ (joint $p = 0.8521$) for shock depth, so the negative conclusion does not depend on this choice (Appendix Table B18; Appendix Table B17, last row).

The equivalence tests came out opposite to expectation (Table 5). At $\Delta = 0.05$ none of the three 90% intervals lies within the equivalence interval, so none is declared equivalent, including the two structural variables. Widening to $\Delta = 0.10$ makes all three declarable, but would also equivalence away the outpatient share, whose effect is non-zero. We therefore claim no equivalence to zero; the reportable result is two-sided: the drug share is 0.0506 (per 10 points, $p = 0.0833$), not significantly different from zero, with an effect no larger than Δ not ruled out; the medical-service share is − 0.0225 ($p = 0.5389$). Point estimates are identical across four inference measures; only standard errors change (Appendix Table B16).

Table 5
Equivalence tests (TOST): separating “not significant” from “no effect” (per 10 percentage points)

Variable	Point estimate	95% CI	90% CI	Two-sided p	Δ = 0.05	Δ = 0.10
Drug revenue share	0.0506	[-0.0067, + 0.1079]	[+ 0.0025, + 0.0987]	0.083	Not equivalent	Equivalent
Medical-service revenue share	-0.0225	[-0.0941, + 0.0492]	[-0.0826, + 0.0377]	0.539	Not equivalent	Equivalent
Outpatient revenue share (demonstration)	-0.0716	[-0.0947, -0.0485]	[-0.0910, -0.0522]	< 0.001	Not equivalent	Equivalent

Notes: the 90% confidence interval must lie entirely within (−Δ, +Δ) before $|\beta| < \Delta$ is declared. The reference scale is the shock-period mean log year-on-year rate, − 0.314, so $\Delta = 0.05$ is 15.9% of it. The outpatient share falls outside the interval at $\Delta = 0.05$ yet is declared equivalent at $\Delta = 0.10$, showing that the procedure responds differentially. Full results are in Appendix Table B11.

4.5 Spatial heterogeneity and robustness

Details of the S1–S9 checks are in Appendix Tables B9 and B10 and Appendix B.11. Four results carry the argument: across 27 recovery-threshold cells the shock–recovery correlation stays negative (− 0.555 to − 0.167); the three nucleic-acid measures give 8.21, 7.65 and 9.02 percentage points; truncating the recovery period to June 2022 turns the provincial volume effect from 0.0083 to − 0.0053 and cuts median recovery from 8 to 7 months; and a trend baseline reclassifies recovery shape substantially. Recovery conclusions are therefore stated relative to the 2019 monthly average, and the direction of the volume component in the recovery window is the least robust result.

Within-province variation can be large. Across 91 county units (Appendix D), median county shock depth is 78.5, interquartile range 73.4–82.1, range 27.7–108.4 (Fig. 4; Appendix Table D1). The max/min ratio

of 3.92 is sensitive to unit size: with at least 3 hospitals it falls to 1.76, and at least 5 to 1.34; the 13 single-hospital units contribute most of the range. Between-city variance is 17.8%, not a major source. County shock depth and recovery speed correlate at r = − 0.35 (R^2 = 0.122), subject to mean reversion [38], so only the direction is reported.

A: distribution of county shock depth. B: county shock depth versus recovery time. C: dual-axis comparison of the two. D: county shock depth versus the 2022/2019 retention rate.

County labels follow the prefecture in which a hospital sits. Under county grouping the event-study contemporaneous coefficient is − 0.39 (t = − 3.09), with leave-one-out cross-check 0.821; under prefecture grouping it falls to − 0.051 (t = − 1.2) and loses significance. County segmentation therefore carries between-group information that city segmentation smooths away; we use counties and list this limit on identification precision in Section 5.4.

5 Discussion

5.1 Main findings and mechanism

The results summarize in three steps. Level: the revenue index fell to 75.56 in February–April 2020, almost entirely through volume (share 134.4%), while unit-service revenue was positive (0.0964) and buffered; in the recovery (31 months, mean Δln revenue + 0.087) unit-service revenue dominated (80.7%). Dispersion: cross-hospital differences cannot be explained by baseline revenue structure; the mechanical component carries almost no information, the shock-year dispersion falls within the placebo range, residuals show no persistence, and there is no incremental power for shock depth. Service content explains the variation. Link: the only channel for structural variables is the mechanical component, variance share 0.1%. This pattern, a larger quantity decline with prices rising, matches multi-country evidence [1, 39–43].

Existing studies often cite “hospitals with a higher drug share fell less” as evidence of structural resilience [4–6],. Our first-step diagnostic reproduces the phenomenon (correlations from + 0.777 to − 0.91 across windows). But total decline is itself a weighted function of the structural variables, so the association holds arithmetically without a behavioral mechanism. Our variance decomposition shows cross-hospital differences fall mainly on the volume side (103.5%), with the unit-service-revenue side smaller (32.1%), so structural variables depend on the side where differences are smaller. They carry information, but not the kind that identifies vulnerable hospitals. The negative conclusion concerns cross-sectional explanatory power, not whether structure changes: in provincial aggregates the medical-service share rose from 27.14% to 35.18% and the drug share fell from 34.99% to 27.82%, consistent with pricing policy. The two are different propositions, and conflating them is one reason structure-determines-resilience conclusions keep recurring.

The stable correlate is the outpatient revenue share. Two explanations compete, a behavioral exposure structure and an institutional pace of price adjustment, but both reduce to the fact that the shock

compresses demand and the depth depends on whether a service can be deferred [44, 45]. Hospitals dominated by elective services (dental, ophthalmic, rehabilitation) almost stopped in February–April (mean shock depth for dental hospitals − 62.5%, baseline outpatient revenue share 99.5%), while hospitals with emergency and chronic-refill services were less affected. For monitoring, the outpatient revenue share and the elective share within outpatient care are ex-ante indicators readable from the reports. A screening indicator must correlate with shock-period decline and be observable at baseline; revenue-structure indicators satisfy the second but not the first: they have no stable association with shock-period decline, and the two one-sided tests cannot show equivalence in either direction, so neither direction supports their use in screening.

5.2 Relation to the literature

The disagreement lies in how the dependent variable is constructed. On structural shares, existing studies regress decline on shares and report significant coefficients [4, 5]; we strip the mechanical component first. On measuring the shock, most use volume or revenue alone [39, 42]; we build both indices and take their difference. On equivalence, common practice reads no effect from $p > 0.05$; two one-sided tests separate no effect from insufficient power [35, 36]. On baseline testing, most treat the post-shock gap as the event effect [1, 40]; we test levels ex ante and find three-quarters of the gap reproducible [46, 47]. On unit-service revenue, the literature speaks of a price effect [20, 21]; without price data we assign no single meaning.

5.3 Policy and monitoring implications

First, monitoring should shift from revenue decline to the asymmetric gap. monitoring should report the gap (level), the decomposition (source) and service-content composition (ex-ante vulnerability). Second, revenue-structure shares should not be used to screen vulnerability: their explanatory power for shock depth lies in the noise range, and their incremental power for recovery time is about $\Delta R^2 \approx 0.036$. The recommendation is conditional: the outpatient share ranks stably (0.937) and holds within general hospitals, but remains a proxy for elective care, and tier and case mix are unobservable (Section 5.4); until official codes exist it should not enter allocation rules. The inpatient side needs its own window: the deviation of revenue relative to discharges (11.7 points) exceeds that relative to visits (3.73), and discharges were negative in all 12 months of 2020. Third, the cost-to-revenue ratio deserves attention beyond the shock: medical expenditure over total revenue was already 1.0435 in 2019 (median 1.0224), with 56.3% of hospital-months above 1; it rose to 1.0988 (2020), 1.0624 (2021) and 1.0826 (2022), and only 27.5% of hospitals were at or below their 2019 level in 2022. Total revenue excludes fiscal appropriations, so this describes medical services, not hospital losses. Fourth, two data suggestions: the 2022 nucleic-acid field has both unit misreporting (10 rows corrected) and unclear attribution (12 rows retained), which require different treatment; and the reports need separate fields for elective and emergency volumes, per-service cost by component, and fiscal compensation versus insurance settlement.

5.4 Limitations

The estimates are associations under a single-period shock, not causal effects. First, the shock was synchronous province-wide, with no unexposed control group and one episode, so the comparison describes moderation across hospitals and cannot identify the shock's size. Second, the event study shows that the coefficients in the 8 pre-shock periods are not jointly zero ($F = 2.62$, $p = 0.0125$ for revenue; $F = 4.14$, $p = 3\times10^{-4}$ for discharges), so the path is reported as an observed trend without causal interpretation [46–50]. Third, hospital tier, category and governing level are not provided; type and spatial attribution are proxies, and the service-content interpretation cannot be separated from the explanation that tertiary hospitals treat more complex cases. Fourth, conclusions are limited to shock depth: incremental power for recovery speed, shape and displacement is ΔR^2 0.024–0.036 (p 0.0796 for displacement).

6 Conclusions

This paper separates the pandemic's revenue decline into two questions and shows that the answers constrain each other. The shock-period decline came almost entirely from volume (share 134.4%), while unit-service revenue dominated the positive growth of the recovery (80.7%). Cross-hospital differences in shock depth cannot be explained by revenue structure; service content explains them, and the variance share of the only channel through which structure could act is 0.1%. Monitoring should therefore report the asymmetric gap, the decomposition, and service-content composition, not revenue-structure shares.

Generalisability differs by layer. The measurement conclusions apply to any system that collects component revenues alongside component volumes; the explanatory conclusions assume unified pricing and payment rules and may not hold where prices are market-determined; the spatial conclusions describe this province and are not extrapolated.

Abbreviations

AI
artificial intelligence
CPC
Communist Party of China
CR1
cluster-robust standard errors
COVID-19
coronavirus disease 2019
GDP
gross domestic product
IR
total revenue index
IV
volume index

SD
standard deviation
SE
standard error
TOST
two one-sided tests.

Declarations

Ethics approval and consent to participate:

This study used de-identified, institution-level aggregated administrative monitoring data. It involved no human participants, no patient-level records and no individually identifiable information. The study was not submitted to an ethics committee or institutional review board, because no such approval was required, and no informed consent to participate was applicable. Under Article 32, item 2, of the Ethical Review Measures for Life Sciences and Medical Research Involving Humans, issued by the National Health Commission of China in 2023, ethical review may be waived for research using anonymized information or data. All analyses were performed in accordance with relevant guidelines and regulations.

Consent for publication:

Not applicable.

Competing interests:

The authors declare that they have no competing interests.

Funding:

This study was supported by the Zhejiang Provincial Soft Science Research Program, project “Evaluation of Medical Insurance Payment Reform in County-level Medical Communities Based on SPO Theory: A Case Study of Deqing County” (No. 2025C35003), principal investigator Yaming Gu. The funder had no role in study design, data collection, analysis, interpretation, or the decision to submit the manuscript for publication.

Author Contribution

Yongqiang Lai: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization, Project administration. Ye Li: Conceptualization, Methodology, Supervision, Writing – review & editing. Feiyang Zheng: Investigation, Data curation, Writing – review & editing. Fangqi Qu: Formal analysis, Writing – review & editing. Xiyu Zhang: Investigation, Writing – review & editing. Jinqin Jiang: Resources, Supervision, Writing – review & editing. Yaming Gu: Conceptualization, Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Acknowledgements:

Not applicable.

Use of generative AI: Generative AI tools were used only to assist with language editing and translation of parts of the text. All other work was carried out by the authors, who take full responsibility for the content of the article.

Availability of data and materials:

The raw administrative monitoring records cannot be redistributed by the authors; access is by application to the data provider (Appendix E). All derived variables, code lists and estimation outputs needed to reproduce the reported results are available from the corresponding author on reasonable request.

References

1. Arsenault C, Gage A, Kim MK, Kapoor NR, Akweongo P, Amponsah F, et al. COVID-19 and resilience of healthcare systems in ten countries. Nat Med. 2022;28:1314–24. 10.1038/s41591-022-01750-1.
2. Haldane V, De Foo C, Abdalla SM, Jung A-S, Tan M, Wu S, et al. Health systems resilience in managing the COVID-19 pandemic: lessons from 28 countries. Nat Med. 2021;27:964–80. 10.1038/s41591-021-01381-y.
3. Fleming P, O'Donoghue C, Almirall-Sanchez A, Mockler D, Keegan C, Cylus J, et al. Metrics and indicators used to assess health system resilience in response to shocks to health systems in high income countries—A systematic review. Health Policy. 2022;126:1195–205. 10.1016/j.healthpol.2022.10.001.
4. Barasa E, Mbau R, Gilson L. What is resilience and how can it be nurtured? A systematic review of empirical literature on organizational resilience. Int J Health Policy Manag. 2018;7:491–503. 10.15171/ijhpm.2018.06.
5. Biddle L, Wahedi K, Bozorgmehr K. Health system resilience: a literature review of empirical research. Health Policy Plan. 2020;35:1084–109. 10.1093/heapol/czaa032.

6. Kruk ME, Ling EJ, Bitton A, et al. Building resilient health systems: a proposal for a resilience index. BMJ. 2017;357:j2323. 10.1136/bmj.j2323.
7. Goldsmith-Pinkham P, Sorkin I, Swift H. Bartik instruments: what, when, why, and how. Am Econ Rev. 2020;110:2586–624. 10.1257/aer.20181047.
8. Borusyak K, Hull P, Jaravel X. Quasi-experimental shift-share research designs. Rev Econ Stud. 2022;89:181–213. 10.1093/restud/rdab030.
9. Holling CS. Resilience and stability of ecological systems. Annu Rev Ecol Syst. 1973;4:1–23. 10.1146/annurev.es.04.110173.000245.
10. Kruk ME, Myers M, Varpilah ST, Dahn BT. What is a resilient health system? Lessons from Ebola. Lancet. 2015;385:1910–2. 10.1016/S0140-6736(15)60755-3.
11. Barasa EW, Cloete K, Gilson L. From bouncing back, to nurturing emergence: reframing the concept of resilience in health systems strengthening. Health Policy Plan. 2017;32:iii91–4. 10.1093/heapol/czx118.
12. Blanchet K, Nam SL, Ramalingam B, Pozo-Martin F. Governance and capacity to manage resilience of health systems: towards a new conceptual framework. Int J Health Policy Manag. 2017;6:431–5. 10.15171/ijhpm.2017.36.
13. Turenne CP, Gautier L, Degroote S, Guillard E, Chabrol F, Ridde V. Conceptual analysis of health systems resilience: a scoping review. Soc Sci Med. 2019;232:168–80. 10.1016/j.socscimed.2019.04.020.
14. Barber SL, Borowitz M, Bekedam H, Ma J. The hospital of the future in China: China's reform of public hospitals and trends from industrialized countries. Health Policy Plan. 2014;29:367–78. 10.1093/heapol/czt023.
15. Yip W, Hsiao W. China's health care reform: a tentative assessment. China Econ Rev. 2009;20:613–9.
16. Yip W, Fu H, Chen AT, et al. 10 years of health-care reform in China: progress and gaps in Universal Health Coverage. Lancet. 2019;394:1192–204. 10.1016/S0140-6736(19)32136-1.
17. Meng Q, Mills A, Wang L, Han Q. What can we learn from China's health system reform? BMJ. 2019;365:l2349. 10.1136/bmj.l2349.
18. Yi H, Miller G, Zhang L, Li S, Rozelle S. Intended and unintended consequences of China's zero markup drug policy. Health Aff. 2015;34:1391–8. 10.1377/hlthaff.2014.1114.
19. Yuan J, Lu ZK, Xiong X, Jiang B. Lowering drug prices and enhancing pharmaceutical affordability: an analysis of the national volume-based procurement (NVBP) effect in China. BMJ Glob Health. 2021;6:e005519. 10.1136/bmjgh-2021-005519.
20. Newhouse JP. Medical care costs: how much welfare loss? J Econ Perspect. 1992;6:3–21. 10.1257/jep.6.3.3.
21. Baumol WJ. Macroeconomics of unbalanced growth: the anatomy of urban crisis. Am Econ Rev. 1967;57:415–26.

22. Zhejiang Provincial Bureau of Statistics. Statistical communique on the national economic and social development of Zhejiang Province in 2022. Hangzhou: Zhejiang Provincial Bureau of Statistics; 2023.
23. CPC Central Committee and State Council. Opinions on supporting Zhejiang Province in pursuing high-quality development and building a demonstration zone for common prosperity. Beijing: People's Publishing House; 2021.
24. Zhejiang Provincial Department of Civil Affairs. Reply to Proposal No. 631 of the First Session of the 13th Zhejiang Provincial Committee of the Chinese People's Political Consultative Conference. Zhejiang Provincial Department of Civil Affairs, Hangzhou. 2023.
25. Zhejiang Provincial People's Government. Comprehensive pilot programme for deepening the reform of the pharmaceutical and healthcare system (Zhe Zheng Fa [2016] No. 19). Hangzhou. 2016.
26. Zheng Q, Zhou L. Does integrated care reshape service utilization patterns? Evidence from China's compact county-level medical community on utilization, cost, and equity. BMC Public Health. 2026. 10.1186/s12889-026-28882-0.
27. Zhejiang Provincial Medical Security Bureau, Zhejiang Provincial Health Commission. Notice on issuing the medical service price reform plan for provincial public hospitals in Zhejiang Province (Zhe Yi Bao Lian Fa [2019] No. 8). Hangzhou. 2019.
28. Zhejiang Provincial Medical Security Bureau, Zhejiang Provincial Health Commission. Zhejiang Provincial Department of Finance, Zhejiang Provincial Administration for Market Regulation. Implementation opinions on establishing a dynamic adjustment mechanism for medical service prices (Zhe Yi Bao Lian Fa [2021] 20). Hangzhou. 2021.
29. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370:1453–7. 10.1016/S0140-6736(07)61602-X.
30. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med. 2015;12:e1001885. 10.1371/journal.pmed.1001885.
31. National Health Commission. 2023 national monitoring and analysis of the performance of tertiary public hospitals (Guo Wei Yi Zheng Han [2025] No. 55). Beijing: National Health Commission; 2025.
32. Frisch R, Waugh FV. Partial time regressions as compared with individual trends. Econometrica. 1933;1:387–401.
33. Cameron AC, Gelbach JB, Miller DL. Robust inference with multiway clustering. J Bus Econ Stat. 2011;29:238–49.
34. Cameron AC, Miller DL. A practitioner's guide to cluster-robust inference. J Hum Resour. 2015;50:317–72.
35. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm. 1987;15:657–80. 10.1007/BF01068419.

36. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Soc Psychol Personal Sci. 2017;8:355–62. 10.1177/1948550617697177.
37. Newey WK, West KD. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica. 1987;55:703–8.
38. Galton F. Regression towards mediocrity in hereditary stature. J Anthropol Inst G B Irel. 1886;15:246–63.
39. Moynihan R, Sanders S, Michaleff ZA, Scott AM, Clark J, To EJ, et al. Impact of COVID-19 pandemic on utilisation of healthcare services: a systematic review. BMJ Open. 2021;11:e045343. 10.1136/bmjopen-2020-045343.
40. Rhodes JH, Santos T, Young GJ. Hospitals' uneven recovery from the COVID-19 pandemic. Health Aff Sch. 2023;1:qxad034. 10.1093/haschl/qxad034.
41. Wang Y, Bai G, Anderson G. COVID-19 and hospital financial viability in the US. JAMA Health Forum. 2022;3:e221018. 10.1001/jamahealthforum.2022.1018.
42. Chen Y, Cai M, Li Z, Lin X, Wang L. Impacts of the COVID-19 pandemic on public hospitals of different levels: six-month evidence from Shanghai, China. Risk Manag Healthc Policy. 2021;14:3635–51. 10.2147/RMHP.S314604.
43. Park JY, Kim TH, Jang SY, Lee SG. Impact of COVID-19 on the profitability of general hospitals in Korea. J Prev Med Public Health. 2026;59:46–55. 10.3961/jpmph.25.303.
44. Mehrotra A, Chernew M, Linetsky D, Hatch H, Cutler D. The impact of the COVID-19 pandemic on outpatient visits: practices are adapting to the new normal. New York: The Commonwealth Fund; 2020. 10.26099/2v5t-9y63.
45. Stuart B. How the COVID-19 pandemic has affected provision of elective services: the challenges ahead. Health Affairs Forefr. 2020. 10.1377/forefront.20201006.263687.
46. Callaway B, Sant'Anna PHC. Difference-in-differences with multiple time periods. J Econom. 2021b;225:200–30. 10.1016/j.jeconom.2020.12.001.
47. Rambachan A, Roth J. A more credible approach to parallel trends. Rev Econ Stud. 2023;90:2555–91. 10.1093/restud/rdad018.
48. Callaway B, Goodman-Bacon A, Sant'Anna PHC. Difference-in-differences with a continuous treatment. arXiv:2107.02637. 2021a.
49. Sun L, Abraham S. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. J Econom. 2021;225:175–99. 10.1016/j.jeconom.2020.09.006.
50. Roth J, Sant'Anna PHC, Bilinski A, Poe J. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. J Econom. 2023;235:2218–44. 10.1016/j.jeconom.2023.03.008.

Figures

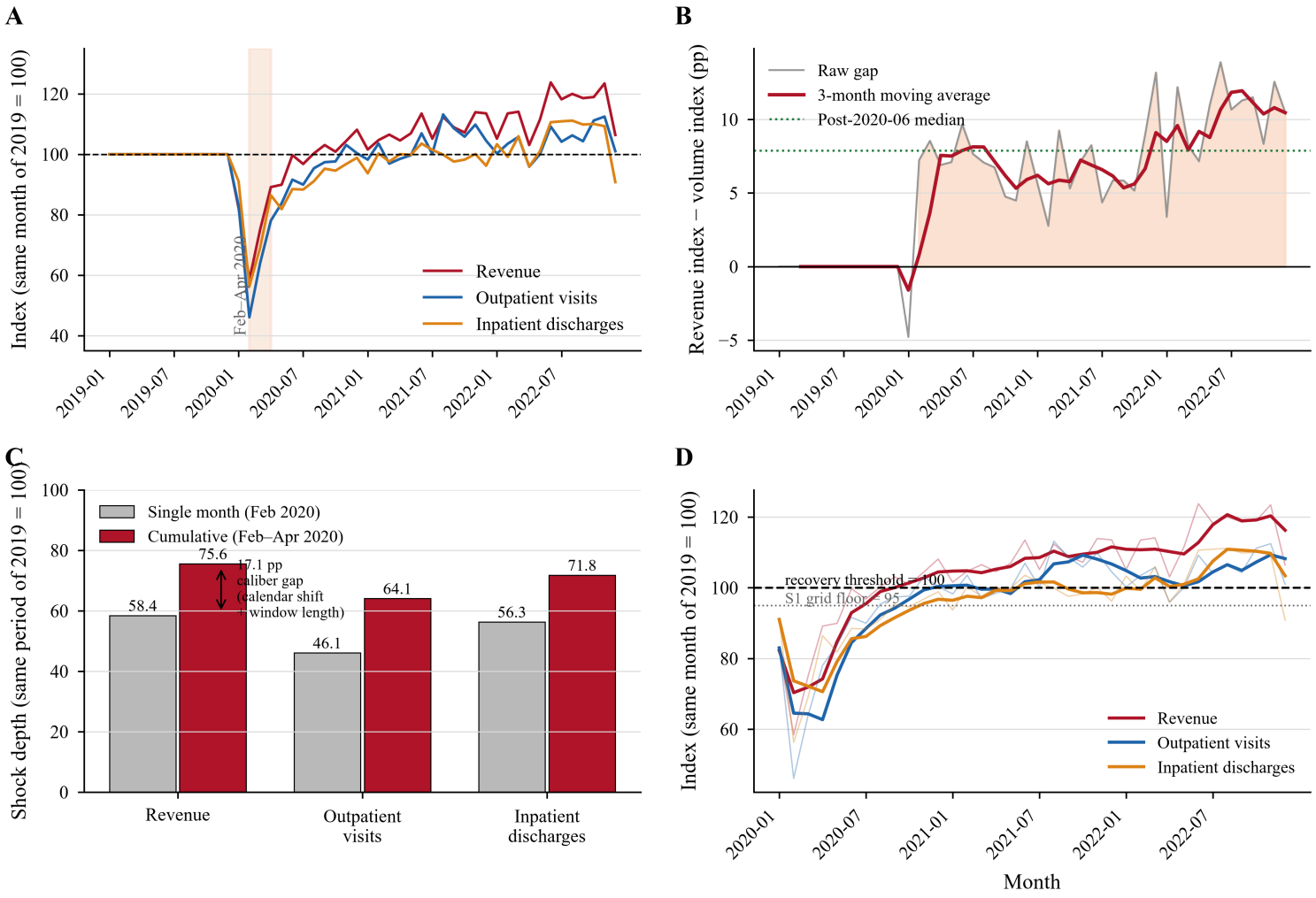


Figure 1

Shock profile, the asymmetric gap, and recovery trajectories.

A: provincial indices of total revenue, outpatient-and-emergency visits and discharges (2019 same month = 100). B: the asymmetric gap and its three-month moving average. C: single-month (February 2020) versus cumulative (February–April) shock depth. D: recovery trajectories. The two reference lines in D differ: the solid line is the recovery criterion (three-month moving average = 100, i.e. return to the 2019 monthly average), the dotted line the lower bound 95 of the sensitivity grid (S1), not a criterion (Appendix Table B9).

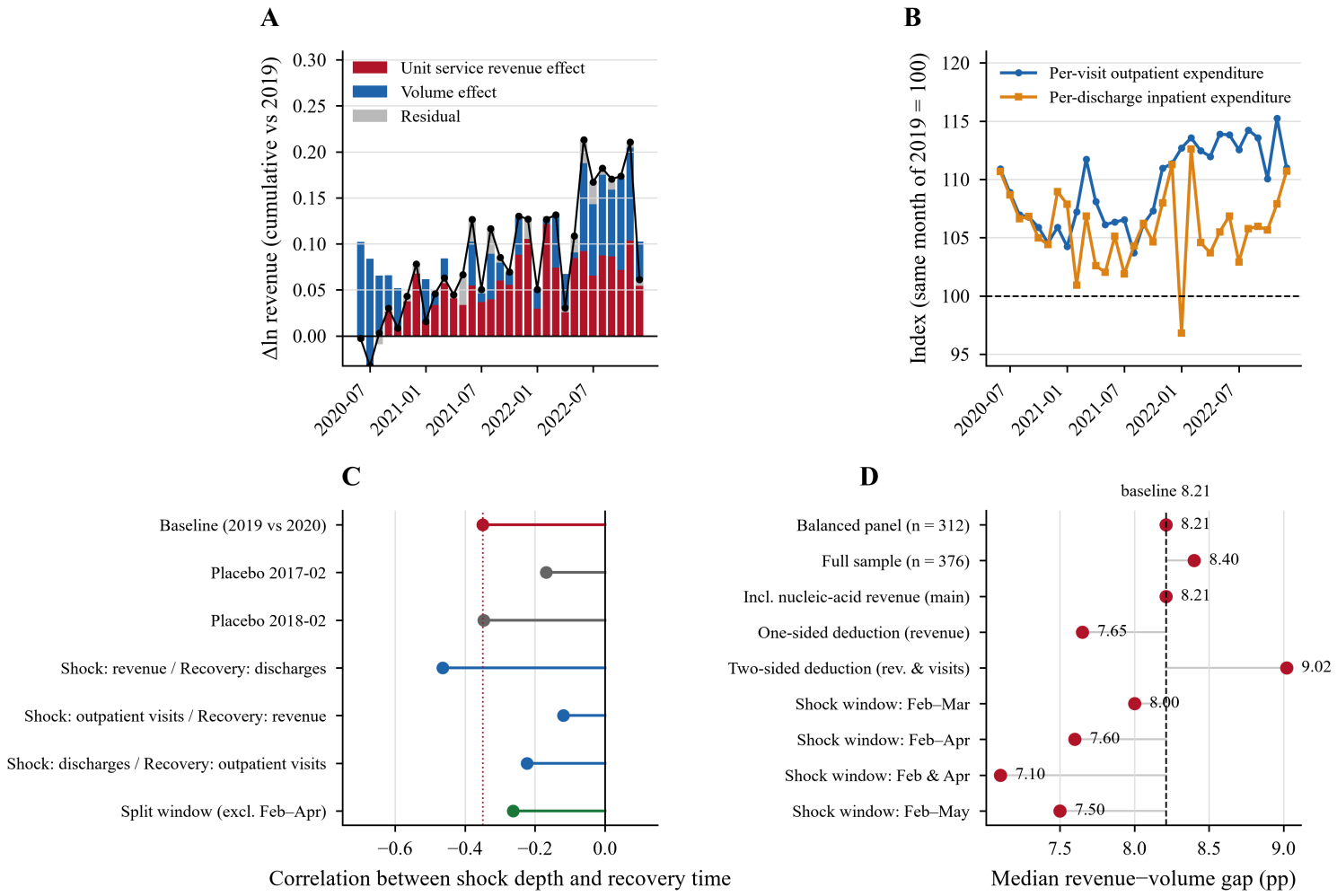


Figure 2

Path decomposition of the revenue–volume divergence.

A: decomposition of the cumulative logarithmic change (unit-service-revenue effect, volume effect, and residual). B: indices of cost per outpatient visit and per inpatient discharge (2019 same month = 100). C: shock depth versus recovery time. D: median gap versus recovery time.

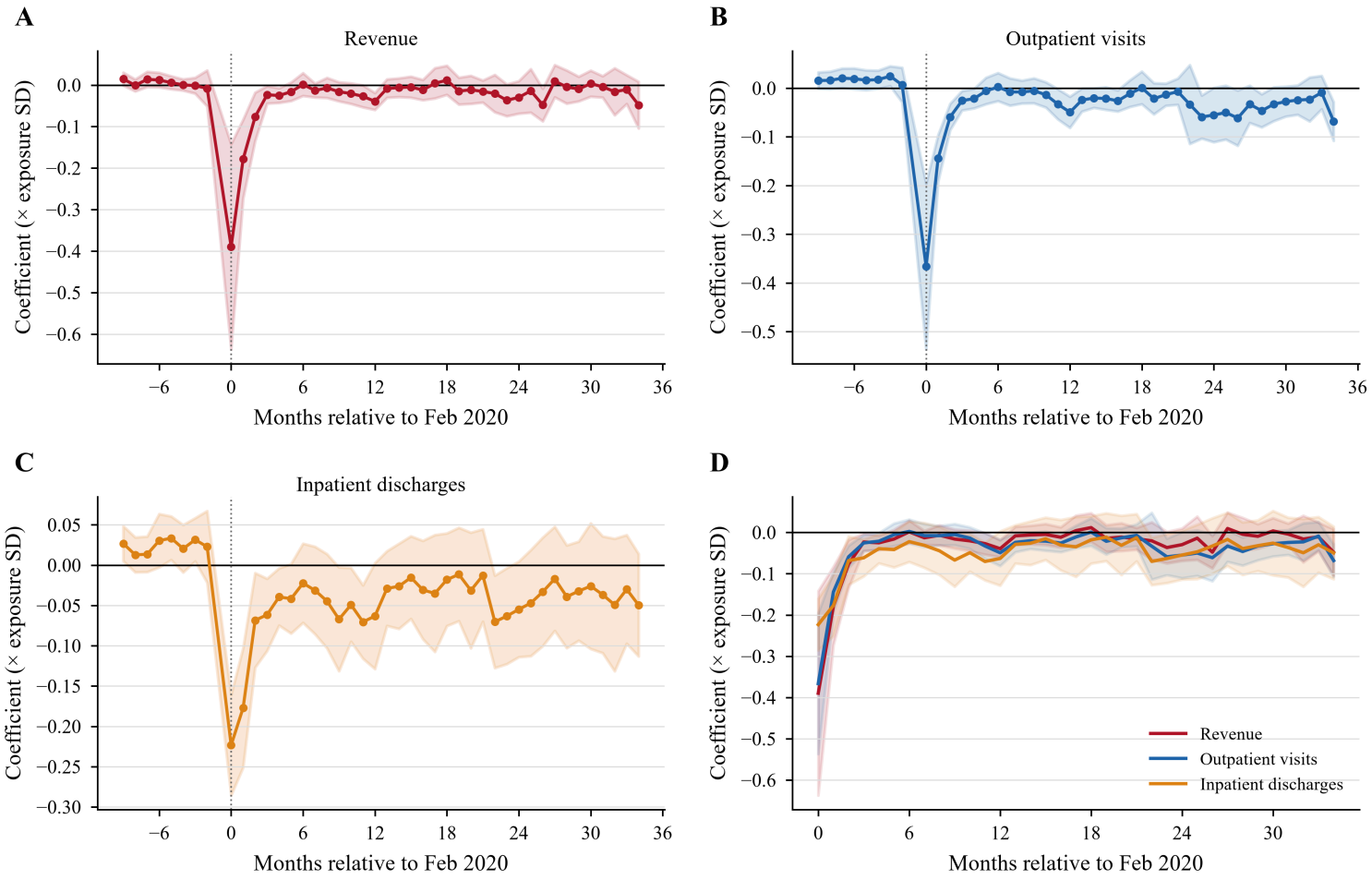


Figure 3

Event study: dynamic effects of the exposure structure.

A–C: monthly coefficients for total medical revenue, outpatient-and-emergency visits, and inpatient discharges (in multiples of the exposure standard deviation; base period January 2020). D: comparison under group (prefecture-city) and trend controls.

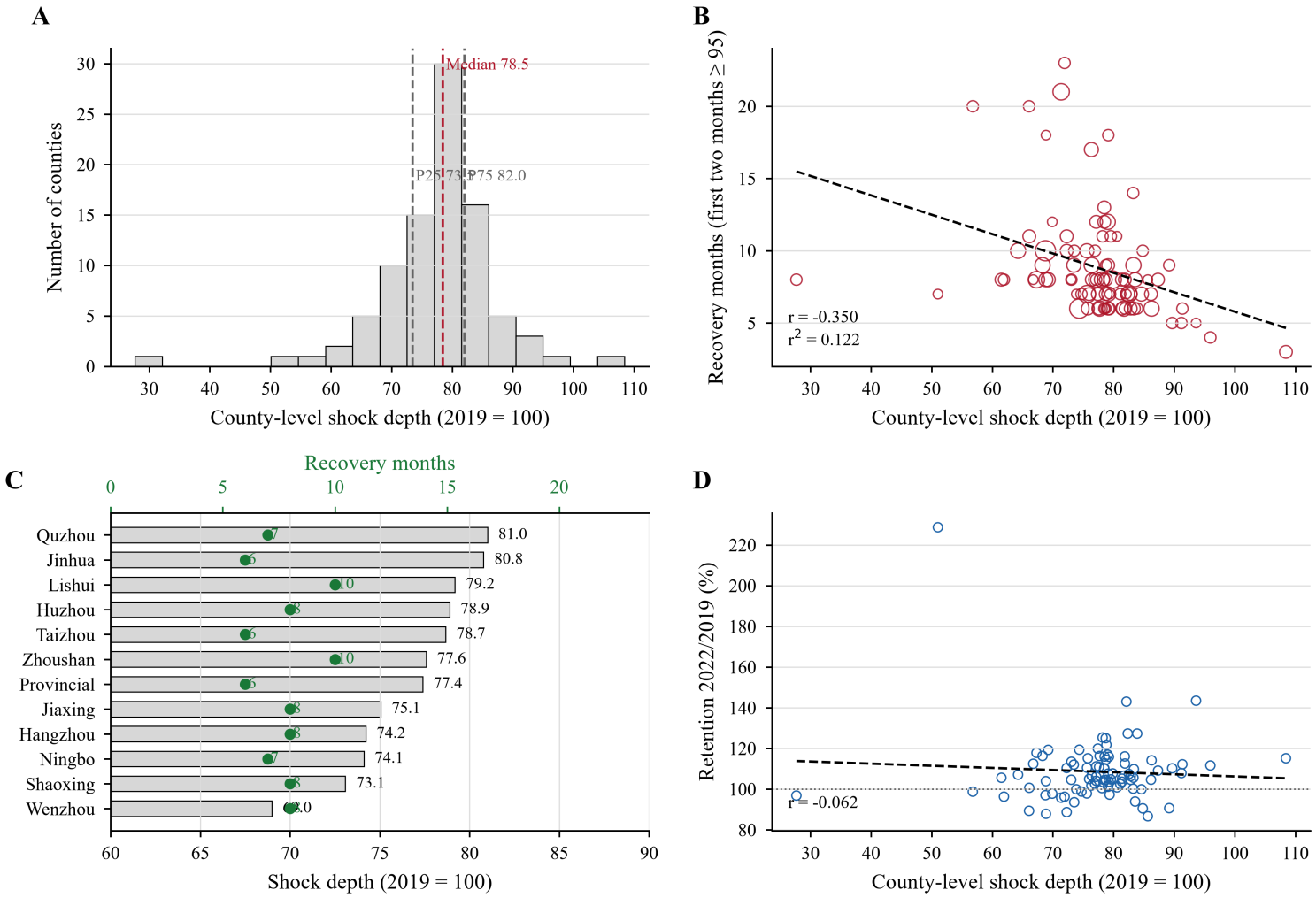


Figure 4

Spatial heterogeneity: county shock depth, recovery, and retention.

A: distribution of county shock depth. B: county shock depth versus recovery time. C: dual-axis comparison of the two. D: county shock depth versus the 2022/2019 retention rate.

Supplementary Files

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

- AppendixFigureA2.png
- AppendixFigureB5.png
- AppendixFigure1.png
- AppendixFigureA3.png
- AppendixFigureA4.png
- AppendixFigureB4.png
- AppendixFigureB3.png

- AppendixFigureB2.png
- AppendixFigureA1.png
- AppendixFigureC3.png
- AppendixFigureC1.png
- AppendixFigureB1.png
- AppendixFigureC2.png
- AppendixFigureB6.png
- Appendix.docx