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.

Depopulation masks persistent geographical inequality in physician supply: an exact decomposition of Japan's physician maldistribution index between its 2024 and 2026 publication rounds

Taeuk Kang
Sungshin Women's University Woonjung Green Campus
Sookja Choi
sjchoi2u@cau.ac.kr

Red Cross College of Nursing, Chung-Ang University

Research Article

Keywords: Health equity, Geographical inequality, Physician maldistribution, Health workforce, Density indicators, Depopulation, Modifiable areal unit problem, Japan

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11481394/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

Geographical inequality in physician supply is monitored almost everywhere with density indicators, and these indicators determine which areas are designated underserved. In depopulating societies their numerator and denominator move in opposite directions, so an indicator can improve without any change in deployment. We quantified this for Japan’s national physician maldistribution index.

Methods

The index equals the standardised physician count divided by the product of population and a standardised utilisation ratio, which permits an exact logarithmic decomposition into supply, population-size and demand-structure effects. Using the official component data for 47 prefectures and 330 secondary medical areas at two publication rounds (January 2024 and April 2026), we decomposed observed change, constructed counterfactual indices holding population constant, and compared dispersion and β-convergence. Sensitivity analyses included a headcount–hours split of the supply effect, an exact decomposition of change in the utilisation ratio, population-weighted dispersion and a prefecture-clustered bootstrap.

Results

The identity reproduced all 377 published values exactly. Nationally, 43.0% of the rise from 255.6 to 266.8 was attributable to population decline; across medical areas the share was 67.3%. Among the 16 highest-index prefectures the mean supply effect was zero, reflecting headcount rather than the hours adjustment. Observed convergence among medical areas (β = −0.085, t = − 5.74) was largely absent in the counterfactual (β = −0.025, t = − 1.72; cluster-bootstrap 95% interval − 0.065 to + 0.008; difference from the observed coefficient − 0.060, 95% interval − 0.068 to − 0.048). In the smallest population quintile the index rose by 10% although the standardised physician count fell, and population decline accounted for 83% of the rise.

Conclusions

Under depopulation, density indicators can register improving equity even when the dispersion of physician supply, net of population change, is essentially unchanged; the discrepancy is greatest at the spatial scale at which scarcity is officially designated. Equity monitoring should report dispersion alongside levels, publish decomposed indicator change, and anchor essential-service indicators on capacity rather than on shrinking utilisation denominators.

Background

Geographical inequality in access to physicians is among the most persistent forms of health inequity in systems that guarantee universal coverage [1, 2]. Where contributions are compulsory but services are locally absent, the legitimacy of the system itself is placed in question [3]; yet because where a physician lives and works is a matter of individual liberty, direct administrative assignment is rarely available. Health systems therefore rely on indicators: numerical measures that identify which areas are underserved, trigger official designations, allocate training quotas and report whether inequality is narrowing [4, 5]. The validity of equity monitoring thus depends on what these indicators can and cannot register.

Almost all such indicators are densities: they divide a count of physicians by a measure of the population to be served, sometimes adjusting the numerator for age, sex and working hours and the denominator for expected utilisation, and they underpin equity assessments of physician distribution in Britain, the United States, Portugal, China and Japan [5–9]. The construction is intuitive, comparable across areas and easily communicated [10, 11]. It also has a property that has attracted little attention in the health services literature: in a shrinking population the denominator falls on its own, so that where depopulation is fastest the indicator improves fastest, whether or not a single additional physician has arrived. If such improvement is reported as progress toward equity, the areas losing population most rapidly will appear to be catching up at the very moment their services are contracting. Conversely, an area that loses physicians may nonetheless post an improving score. Because designation thresholds are typically relative — the lowest third of areas are classified as scarce [12] — the number of designated areas is fixed by construction and cannot register system-wide improvement or deterioration at all.

Japan provides an unusually clean setting in which to quantify this. Since 2019 the Ministry of Health, Labour and Welfare has published a physician maldistribution index for all 47 prefectures and all secondary medical areas, defined by an explicit multiplicative formula, and has released the full component data underlying every published value [13]. The index determines which areas are officially designated physician-scarce and is used in allocating temporary medical school admission quotas, the regional-quota programme that binds graduates to underserved practice [14]. Japan’s population has been declining since 2008, and its regional divergence is extreme: some secondary medical areas lost more than 7% of their residents in the four years covered here [15, 16].

Between January 2024 and August 2025 a ministerial review committee met eleven times to decide the future of these quotas, and its verbatim transcripts record an unusually explicit methodological debate [17]. At the second session a health economist member observed that because the index is essentially population-adjusted physician density, and because urban populations are stable while physician numbers rise whereas peripheral populations fall while physician numbers stagnate, both ends of the distribution would improve “even if nothing is done in policy terms” (authors’ translation); the member concluded that what must be monitored is not the level of each index but the gap between areas. In

subsequent sessions, however, the evidence presented to the committee was largely framed in terms of levels and growth rates.

A substantial literature has tracked the geographical distribution of physicians in Japan, using density and Gini measures, and has concluded that although underserved prefectures and areas improved their scores, inter-regional disparity persisted [9, 18–24]. To our knowledge, however, no study has decomposed indicator change into its numerator and denominator components, nor constructed a counterfactual in which population is held constant. This is the gap addressed here.

We ask three questions. First, how does the observed change in Japan's physician maldistribution index decompose into supply, population-size and demand-structure effects, and does the decomposition differ by baseline maldistribution? Second, does the convergence observed between 2024 and 2026 survive when population change is removed? Third, does the answer depend on the spatial scale of analysis?

Methods

Setting and Indicator Definition

Japan's physician maldistribution index (*ishi henzai shihyō*) is defined as

$$I = M^* / [\,(P / 100{,}000) \times S\,] \quad (1)$$

where M* is the standardised physician count — physicians working in medical facilities, weighted by sex- and age-specific mean working hours relative to the national mean — P is the resident population, and S is the standardised utilisation ratio, which adjusts expected demand for the local age–sex structure and for inpatient cross-boundary flows and daytime population. The national value of S is unity by construction. Because S is re-normalised to unity in each round and its inputs are periodically updated, a change in S between rounds can reflect revisions to those inputs as well as demographic change; we decompose it accordingly.

The index is calculated for 47 prefectures (tertiary medical areas) and for the secondary medical areas that are the operative unit of Japanese health planning. Areas in the lowest third of the national distribution are designated physician-scarce; for the plans beginning in fiscal year 2027, areas in the middle third that fall in the highest decile of a remoteness measure published with the April 2026 round will also be designated [12]. Designation governs eligibility for several policy instruments, and prefecture-level status governs the allocation of temporary admission quotas.

Data

We used the official component datasets published alongside the two most recent index rounds: the January 2024 release, based on the 2020 physician statistics and the population of 1 January 2021, and

the April 2026 release, based on the 2024 physician statistics, adjusted using an individual survey conducted by the Ministry in 2026, and the population of 1 January 2025 [13]. The physician and population windows are therefore both four years long and nearly synchronous (31 December 2020 to 31 December 2024, and 1 January 2021 to 1 January 2025). Each dataset reports, for every unit, the published index together with M*, P and S, permitting exact reconstruction. The formula was identical in the two releases, but several inputs to S were updated: national utilisation rates were re-estimated from the 2023 rather than the 2017 Patient Survey, inpatient and outpatient patient-flow coefficients were re-surveyed by prefectures, and the weight converting outpatient demand into physician demand fell from 0.178 to 0.152 following an updated national supply–demand projection. The sex- and age-specific working-hours ratios came from the same 2022 survey in both rounds, rescaled by a common factor of about 1.008, so that the national working-hours weight equals one in each round. We also used the published component dataset for the obstetric physician maldistribution index (January 2024 and April 2026 rounds), whose denominator is the number of deliveries rather than the population.

Units were matched across rounds by medical-area code and name. Both releases report the 330 secondary medical areas of the 8th Medical Plan; all 330 matched, as did all 47 prefectures, and no unit was excluded.

The deliberative context was characterised from the eleven verbatim transcripts of the ministerial review committee (January 2024 to August 2025) [17], which were read in full by the first author; passages concerning the construction, interpretation or use of the index were extracted and summarised using a document-analysis approach [25, 26]. Quotations are the authors' translations from Japanese. This qualitative component served only to generate the propositions tested quantitatively and is not used as evidence for the quantitative findings. Speakers are identified by role rather than by name.

Exact Decomposition

Because the index is a ratio of products, taking logarithms yields an exact additive decomposition:

$$\Delta \log I = \Delta \log M^* - \Delta \log P - \Delta \log S \quad (2)$$

We label the three right-hand terms the supply effect, the population-size effect (positive when population falls) and the demand-structure effect. The decomposition is an identity rather than an approximation; we verified this by reconstructing every published value. Contribution shares are each term divided by Δ log I. Where Δ log I approaches zero the share is unstable, and we therefore report tertile means rather than means of ratios where relevant.

Because the standardised count is itself the product of the headcount of physicians in medical facilities (R) and a working-hours weight ($w = M^*/R$), the supply effect can be decomposed a second time as $\Delta \log M^* = \Delta \log R + \Delta \log w$. This separates changes in the number of physicians from changes in the age–sex composition that drives the hours adjustment, and allows a null supply effect to be attributed to one or the other.

Counterfactual Construction

Two counterfactuals were constructed. The first holds the denominator at baseline, $I_{CF(P)} = M^*_1 / [(P_0/100{,}000) \times S_1]$, answering what the 2026 index would have been had population not changed. The second holds the numerator at baseline, $I_{CF(M)} = M^*_0 / [(P_1/100{,}000) \times S_1]$, answering what the index would have shown had no physician been added. Comparing the distribution of $I_{CF(P)}$ with the observed 2024 and 2026 distributions isolates the contribution of depopulation to any apparent convergence.

Both counterfactuals are accounting rather than behavioural constructs. They assume that the location of physicians in 2024 would have been the same had population not changed, and they make no claim about how physicians would have responded to a different demographic path. The population-fixed counterfactual therefore answers an attribution question — how much of the observed change in the index, and in its dispersion, arises from the denominator — rather than a causal one.

Dispersion and Convergence

Dispersion was summarised by the coefficient of variation, the Gini coefficient, the Theil index, the absolute range and the maximum-to-minimum ratio. All measures treat each area as one observation (unweighted), because designation is made area by area irrespective of population size; as a sensitivity analysis we also computed population-weighted coefficients of variation, Gini coefficients and Theil indices, weighting each area by its population in the relevant round (2021 for the baseline and the counterfactual, 2025 for the observed 2026 distribution). The Theil index belongs to the class of additively decomposable measures [27] and was decomposed into between-prefecture and within-prefecture components for the medical-area distribution. σ-convergence was assessed by change in dispersion; β-convergence [28] by regressing $\Delta \log I$ on $\log I_0$, with the coefficient reported with its t statistic and, because the index distribution is right-skewed, replicated using Spearman rank correlation. Because medical areas nested within the same prefecture are not independent, 95% confidence intervals for β, for the difference between observed and counterfactual β, and for the proportional change in dispersion were obtained by a cluster bootstrap that resampled prefectures with replacement (B = 2,000; percentile intervals). Because β-convergence estimated from noisy small-area measures is susceptible to regression to the mean, σ-convergence was treated as the primary criterion and β-convergence as supporting evidence.

Sensitivity Analyses

Seven analyses tested robustness. (i) Because the inputs to S were revised between rounds, we decomposed the demand-structure effect exactly into four components — age–sex population structure, national utilisation rates, the outpatient demand weight and patient-flow coefficients — using Shapley weights over all sixteen combinations of 2024 and 2026 inputs, and recomputed dispersion and convergence with S held at its 2024 value. (ii) We repeated the analysis using a crude density —

physicians in medical facilities per 100,000 population, with no standardisation of either numerator or denominator. (iii) We excluded the seven Tokyo ward-area units and, separately, trimmed the baseline distribution by excluding units below the fifth or above the ninety-fifth percentile. (iv) We examined the numerator-fixed counterfactual. (v) We replicated convergence tests using rank statistics. (vi) We decomposed dispersion into between- and within-prefecture components under each scenario. (vii) We recomputed dispersion with population weights. Analyses were performed in Python 3.13 (pandas, NumPy, SciPy); figures were produced with Matplotlib.

Use of large language models

A large language model (Claude, Anthropic) was used to assist with drafting analysis code, structuring the manuscript and language editing. All data processing, analyses, results and interpretations were independently verified by the authors, who take full responsibility for the content; the model is not an author.

Results

Validation of the Decomposition

Reconstructing Eq. (1) reproduced the published index for all 377 units in each round with a maximum absolute difference of 0.00 at the published precision of one decimal place. This confirms that the published components are internally consistent, so that the identity in Eq. (2) can be applied without approximation error.

The Index Rises at Every Level of Aggregation

Nationally the index rose from 239.8 in the 2019 round (computed under a partially different formula and shown for context only) to 255.6 in January 2024 and 266.8 in April 2026, an increase of 11.3% over seven years. Between the two rounds analysed here, 45 of 47 prefectures rose, with a mean increase of 5.3%; the exceptions were Kumamoto (−0.2%) and Kyoto (0.0%).

Disparity did not move in a single direction. The gap between the highest and lowest prefecture widened from 160.1 in 2019 to 171.4 in 2024 — the figure repeatedly cited by prefectural representatives during the committee's deliberations — and then narrowed to 164.2 in 2026. Between 2024 and 2026 all dispersion measures fell: for prefectures the coefficient of variation declined by 10.6% and the Gini coefficient by 10.4%; for medical areas the corresponding figures were 7.0% and 5.1%. β-convergence was strongly significant at both scales (prefectures $\beta = -0.112$, $t = -4.89$; medical areas $\beta = -0.085$, $t = -5.74$). On conventional reading, Japan's physician distribution converged.

Where the Improvement Came From

The decomposition shows that most of it came from the denominator (Table 1). Nationally, of a Δ log I of 0.0431, the supply effect contributed 0.0246 (57.0%) and the population effect 0.0185 (43.0%): the standardised physician count rose from 323,700 to 331,748 while the population fell from 126.65 to 124.33 million.

At subnational level the population effect dominated. Across prefectures the contributions (ratios of mean effects) were 38.6% supply and 61.6% population; across medical areas, 22.8% supply, 67.3% population and 9.9% demand structure. The pattern by baseline tertile differs instructively between the two scales. Among prefectures the population contribution was 55.3% in the lowest tertile, 49.6% in the middle and 99.9% in the highest: in the sixteen prefectures with most physicians the mean supply effect was exactly zero, and the entire recorded improvement was denominator-driven. Among medical areas the contribution was essentially uniform — 66.2%, 69.8% and 65.8% — indicating a systematic property of the measure rather than a feature of particular places, although in absolute terms the population effect in the lowest tertile (0.0582) was more than twice that in the highest (0.0245).

The null supply effect among high-index prefectures is not an artefact of standardisation. Splitting the supply effect into headcount and hours components shows that the working-hours weight contributed at most ±0.0015 in any tertile at either scale. Among the sixteen high-index prefectures the headcount effect was +0.0010 and the hours effect −0.0010; physician numbers rose in eight of these prefectures and fell in eight.

Table 1. Decomposition of change in the physician maldistribution index, 2024 to 2026.

Unit and baseline tertile	n	Δ log I	Supply	Population	Demand structure	Population contribution
National	1	0.0431	0.0246	0.0185	0.0000	43.0%
Prefectures, all	47	0.0510	0.0197	0.0314	−0.0001	61.6%
Lowest tertile	16	0.0679	0.0289	0.0375	0.0014	55.3%
Middle tertile	15	0.0573	0.0308	0.0284	−0.0019	49.6%
Highest tertile	16	0.0281	0.0000	0.0281	0.0000	99.9%
Medical areas, all	330	0.0623	0.0142	0.0420	0.0062	67.3%
Lowest tertile	110	0.0879	0.0086	0.0582	0.0211	66.2%
Middle tertile	110	0.0618	0.0141	0.0431	0.0046	69.8%
Highest tertile	110	0.0373	0.0199	0.0245	−0.0071	65.8%

Effects are mean logarithmic changes; the population effect is positive when population falls. The three effects sum to Δ log I by construction. Tertiles are defined on the 2024 index.

Prefecture-level detail sharpens the point. In Kumamoto the standardised physician count fell by 4.1% and in Kyoto by 2.0% over the four years, yet the index was essentially unchanged in both; the headcount fell by 4.1% and 1.9% respectively, and the working-hours weight changed by less than 0.1%. In Okayama and Ishikawa physician numbers fell by 2.7% and 1.3% respectively while the index rose. Tokushima, Tottori and Akita showed the same pattern, the standardised physician count falling by 3.2%, 2.0% and 1.8% while the index rose by 1.6%, 3.1% and 6.1%. In each case the population effect offset, or more than offset, a contraction in supply. Conversely, the largest single increase — Fukushima, +16.9% — was genuinely supply-led, the supply effect contributing 64%; the headcount rose by 10.2%, an increase also recorded in the national physician statistics.

Convergence Does Not Survive the Counterfactual

Holding population at its 2021 value and recomputing the 2026 index yields the central result (Table 2, Fig. 1). Among medical areas the observed coefficient of variation fell from 0.3375 to 0.3137, a decline of 7.0%; in the counterfactual it was 0.3373, a change of −0.1%. The Gini coefficient, which fell by 5.1% in the observed series, rose by 0.8% in the counterfactual. The absolute range, 682.0 in 2024 and 649.5 in 2026 as observed, was 700.1 in the counterfactual — wider than at baseline.

Table 2. Observed and counterfactual dispersion of the physician maldistribution index.

Scenario	Mean	CV	Gini	Theil	Range	Max/min
Prefectures (n = 47)						
2024 observed	247.9	0.1565	0.0872	0.0119	171.4	1.94
2026 observed	260.3	0.1400	0.0781	0.0096	164.2	1.85
2026 counterfactual	252.5	0.1475	0.0811	0.0106	179.8	1.98
Medical areas (n = 330)						
2024 observed	214.4	0.3375	0.1671	0.0488	682.0	7.33
2026 observed	227.2	0.3137	0.1586	0.0432	649.5	6.29
2026 counterfactual	219.0	0.3373	0.1684	0.0491	700.1	7.22

The counterfactual holds population at its 1 January 2021 value while applying the 2026 standardised physician count and utilisation ratio. CV, coefficient of variation.

β-convergence behaves likewise. The observed coefficient of −0.085 (t = −5.74) falls to −0.025 (t = −1.72) in the counterfactual, losing conventional significance; the rank-based test gives ρ = −0.061 (p = 0.27). At medical-area level, therefore, most of the convergence observed between 2024 and 2026 is attributable to differential population loss: the counterfactual coefficient is less than a third of the observed coefficient and is not statistically distinguishable from zero.

Cluster-bootstrap intervals are consistent with this. The observed β of −0.085 has a 95% interval of −0.122 to −0.050, whereas the counterfactual interval, −0.065 to +0.008, includes zero, although its lower bound does not exclude a moderate degree of genuine convergence. The difference between the two coefficients is −0.060 (95% interval −0.068 to −0.048). The observed 7.0% fall in the coefficient of variation has an interval of −9.5% to −2.8%; the counterfactual interval, −3.0% to +3.0%, is centred on zero. The difference between observed and counterfactual change in the coefficient of variation is −7.0 percentage points (95% interval −8.2 to −4.8), so the contribution of population change to the apparent convergence remains statistically distinguishable from zero after allowing for clustering within prefectures.

At prefecture level the picture is more qualified. Dispersion still fell in the counterfactual, though by roughly half as much (coefficient of variation −5.8% against −10.6% observed), and β remained significant (−0.069, t = −2.96; ρ = −0.435, p = 0.002). Notably, however, the counterfactual absolute range (179.8) exceeded not only the observed 2026 value (164.2) but also the 2024 baseline (171.4). Some genuine convergence occurred between prefectures; little or none is detectable between medical areas.

The Theil decomposition locates the effect precisely. Observed total inequality among medical areas fell from 0.0488 to 0.0432, with both the between-prefecture component (0.0126 to 0.0105) and the within-prefecture component (0.0362 to 0.0326) declining. In the counterfactual all three quantities return almost exactly to baseline (0.0491, 0.0127 and 0.0364). Virtually none of the reduction in either component is therefore attributable to changes in physician deployment. The within-prefecture share remained between 74% and 76% in every scenario, confirming that three-quarters of the inequality among medical areas lies inside prefectures rather than between them.

Scale Dependence

The population contribution rises as spatial units become finer (Fig. 2): it is 43.0% nationally, 61.6% across prefectures and 67.3% across medical areas. Within medical areas the gradient is steeper still when they are ranked by population size: in the smallest quintile population decline accounts for 83% of the rise in the index while the supply effect is negative (−12%), the standardised physician count having fallen in 38 of the 66 areas; the population contribution then falls to 78%, 73%, 50% and 20% in the successively larger quintiles, and the supply contribution reaches 97% in the largest. Mean population change over the four years ranged from −7.6% in the smallest quintile to −0.7% in the largest. Thirty-seven of the 66 areas in the smallest quintile (56%) lie in the lowest tertile of the 2026 index and are therefore designated physician-scarce; they make up a third of the 110 areas designated under the tertile rule, and in them the mean supply effect was −0.017 against a population effect of +0.086. The denominator effect is a property of small and shrinking denominators, and it is therefore most severe in the small, depopulating areas that are of greatest concern for access.

Sensitivity Analyses

Table 3 summarises the seven sensitivity analyses together with the headcount–hours split and the cluster bootstrap reported above. The demand-structure effect among medical areas arose almost entirely from revised patient-flow coefficients rather than from demographic change. In the lowest tertile its mean of +0.0211 comprised +0.0191 from patient-flow coefficients, +0.0033 from updated national utilisation rates, +0.0002 from the outpatient weight and −0.0014 from change in age–sex structure; ageing thus raised expected demand and lowered the index, and six areas whose inpatient-flow coefficient was revised sharply downward accounted for half of the tertile's net flow effect. Holding S at its 2024 value strengthens rather than weakens the central result: a pure-supply index, applying the 2026 physician count to the 2021 population and the 2024 utilisation ratio, diverged among medical areas (coefficient of variation +2.1%, Gini coefficient +3.4%, β = +0.008), whereas among prefectures it still converged (coefficient of variation −5.4%). Freezing physician numbers entirely shows that the index would still have risen in 41 of 47 prefectures and 262 of 330 medical areas, by a mean of 3.2% and 5.2% respectively.

Replacing the index with a crude, wholly unstandardised density reproduces the structure among medical areas — population contribution 73.4%, and a counterfactual coefficient of variation 1.8% higher than baseline — but not among prefectures, where the counterfactual decline (−9.3%) matches the observed decline (−8.8%). Standardisation is therefore not the source of the finding at the operative spatial scale, although prefecture-level conclusions are sensitive to it.

Results are not driven by outliers. Excluding the seven Tokyo ward areas leaves the observed and counterfactual β essentially unchanged (−0.085 and −0.028). Trimming at the fifth and ninety-fifth percentiles attenuates the observed convergence (β = −0.058, t = −3.0) while driving the counterfactual coefficient to essentially zero (β = +0.002, t = +0.1), so the contrast between the two series becomes cleaner rather than weaker.

Population weighting changes the magnitudes but not the direction of the findings. Among medical areas the population-weighted Gini coefficient fell by 5.3% in the observed series and by 1.4% in the counterfactual (cluster-bootstrap 95% interval −4.2% to +2.0%), and the weighted Theil index by 10.9% and 2.8% respectively. About three-quarters of the observed decline in weighted inequality is therefore attributable to population change, and the observed-minus-counterfactual difference in the weighted Gini coefficient (−3.9 percentage points, 95% interval −5.2 to −2.7) excludes zero. Among prefectures the weighted Gini coefficient fell by 8.9% observed and by 4.5% in the counterfactual, mirroring the unweighted result. With weights, the within-prefecture share of the Theil index among medical areas was 70% to 73%.

Table 3. Sensitivity analyses.

Analysis	Specification	Result
Utilisation ratio	Shapley decomposition of Δ log S; index recomputed with S held at its 2024 value	Lowest medical-area tertile: demand effect +0.0211, of which patient flows +0.0191 and age–sex structure −0.0014. S fixed: medical-area CV +2.1%, β +0.008; prefecture CV −5.4%
Numerator-fixed counterfactual	2020 physician count (2024 round) applied to 2025 population	Index still rises in 41/47 prefectures and 262/330 medical areas (mean +3.2% and +5.2%)
Crude density	Facility-based physicians per 100,000, no standardisation	Medical areas: population contribution 73.4%; counterfactual CV +1.8% vs. baseline. Prefectures: counterfactual CV −9.3% vs. observed −8.8%
Outlier exclusion	Seven Tokyo ward areas removed (n = 323)	Observed β −0.085 (t = −5.3); counterfactual β −0.028 (t = −1.7)
Trimming	Units below the 5th or above the 95th baseline percentile excluded (n = 296)	Observed β −0.058 (t = −3.0); counterfactual β +0.002 (t = +0.1)
Rank-based convergence	Spearman correlation of Δ log I with log I_0	Medical areas: observed ρ = −0.258 ($p < 0.001$); counterfactual ρ = −0.061 ($p = 0.27$)
Headcount versus hours	Supply effect split into Δ log headcount and Δ log hours weight	Hours-weight effect within ±0.0015 in every tertile; high-index prefectures: headcount +0.0010, hours −0.0010
Population weighting	CV, Gini and Theil weighted by area population	Medical areas: weighted Gini −5.3% observed vs −1.4% counterfactual (difference −3.9 pp, 95% CI −5.2 to −2.7); weighted Theil −10.9% vs −2.8%. Prefectures: weighted Gini −8.9% vs −4.5%
Cluster bootstrap	Prefectures resampled with replacement, B = 2,000, percentile intervals	Observed β 95% CI −0.122 to −0.050; counterfactual −0.065 to +0.008; difference −0.068 to −0.048. Observed ΔCV −9.5% to −2.8%; counterfactual −3.0% to +3.0%

Consequences of Relative Thresholds

Because designation uses fixed tertiles of the national distribution, the cut-offs move with the distribution. For medical areas the upper cut-off rose from 217.6 to 236.0 (+8.5%) and the lower from 179.3 to 193.7 (+8.0%) between the two rounds; for prefectures the corresponding rises were 3.7% and 6.7%. Thirty-eight of 47 prefectures exceeded the 2024 scarcity threshold by 2026, yet 16 fell below the new one. The number of areas designated physician-scarce cannot change through system-wide improvement or deterioration.

The remoteness measure introduced in the 2026 round partially addresses a different limitation. Applying the published rule, 110 medical areas fall in the lowest tertile and a further 11 are designated through the remoteness criterion, giving 121 designated areas (36.7%). The eleven added have a mean index of 210.9, essentially identical to the middle-tertile mean of 211.9: they are unremarkable on density but geographically isolated. The Pearson correlation between the index and the remoteness measure is only −0.233, indicating that density captures little of geographical access [10,29].

Small Denominators: The Obstetric Index

The obstetric physician maldistribution index shares the same architecture with a different denominator: standardised delivery-attending physicians per 1,000 deliveries. Because births are falling far faster than total population, the denominator problem is likely to be magnified over time. We examine it both across the two rounds and cross-sectionally. Nationally the index stands at 13.39 (9,318 standardised physicians against 695,800 adjusted deliveries).

Among the 256 perinatal medical areas for which the index is defined, the relationship with denominator size is stark (Fig. 3). In the quartile with fewest deliveries — a mean of 270 per year, attended by a mean of 4.4 physicians — the mean index is 22.65, well above the national value, with a coefficient of variation of 1.087. In the quartile with most deliveries (7,270 deliveries, 99.8 physicians) the mean index is 13.98 with a coefficient of variation of 0.348.

Across rounds, the national obstetric index rose from 10.50 to 13.39 (+27.6%) although the standardised number of delivery-attending physicians was essentially unchanged (−0.1%): the whole of the rise reflects the fall in adjusted annual deliveries from 888,500 to 695,800 (−21.7%), whose reference month advanced six years (September 2017 to September 2023) while the physician count advanced four. Of the 254 perinatal medical areas matched by code and name, 251 had delivery-attending physicians in both rounds; in these the index rose in 218 (87%), including 108 in which the physician count fell. Of the remaining three areas, one had no delivery-attending physician in either round; in the other two, Rumoi (Hokkaido) and Noto-hokubu (Ishikawa), the last delivery-attending physicians left and the index fell to zero, the only cases in which it registered the loss.

The extreme cases make the mechanism unmistakable. The Tokyo island area records the highest obstetric index in Japan, 161.84, on the basis of two delivery-attending physicians and twelve deliveries a year. Oki in Shimane, Kami-Gotō in Nagasaki and Takahashi–Niimi in Okayama each record indices between 60 and 90 on the basis of a single physician and twelve deliveries. In these areas a ratio of physicians to deliveries reads as generous provision, yet it conveys nothing about whether a delivery service that requires continuous cover can be sustained by one or two physicians.

Discussion

Principal Findings

Japan's physician maldistribution index improved between 2024 and 2026 at every level of aggregation, and by conventional measures the distribution converged. An exact decomposition shows that the improvement was largely produced by the denominator. Nationally 43% of the increase came from population decline; across medical areas the figure was 67%. Among the sixteen prefectures with the most physicians, the mean supply effect was zero, and in several prefectures the index held steady or rose while physician numbers fell. The null supply effect reflects physician headcount, not the working-hours adjustment.

The convergence is likewise predominantly a denominator phenomenon at the scale that matters. Holding population constant, the coefficient of variation among medical areas changed by − 0.1%, the Gini coefficient rose, the absolute range widened beyond its 2024 value, and β-convergence lost significance on parametric, rank-based and cluster-bootstrap tests. Between prefectures, some genuine convergence survives. The contrast between the two scales is itself a finding: the population contribution rises steeply as units get smaller, reaching 83% in the smallest population quintile of medical areas, where the physician count actually declined.

Two further results follow from indicator design rather than from depopulation as such. Because scarcity is designated by fixed tertiles, the designation thresholds rose by 8% between the two rounds and the number of designated areas cannot respond to system-wide change. And because the obstetric index divides physicians by deliveries, areas where obstetric care is disappearing post the highest scores in the country — two physicians and twelve annual deliveries yield an index twelve times the national value.

Relation to Existing Knowledge

That density indicators depend on their denominator is arithmetically obvious, and the risk of ecological inference from aggregate measures has been recognised since Robinson's work on ecological correlations [30] and Simpson's on aggregation [31]. The dependence of results on the zoning and scale of areal units — the modifiable areal unit problem [32, 33] — is likewise long established. What has not previously been quantified, to our knowledge, is the magnitude of the denominator contribution in a national workforce-monitoring system, or its dependence on spatial scale. Earlier analyses of physician distribution in Japan and elsewhere have documented improving density alongside persistent disparity [4, 5, 20, 21]; our decomposition explains how both statements can be true simultaneously, and shows that most of the apparent narrowing among medical areas in the most recent round cannot be attributed to changes in physician deployment.

We emphasise what the decomposition does and does not imply. When population falls while the number of physicians is unchanged, the number of physicians per resident does rise, and remaining residents may experience shorter waits or smaller patient panels. The population effect is therefore not a measurement error in the narrow sense. Our argument concerns attribution and interpretation: improvement produced by the denominator cannot be credited to workforce policy, and in small areas it does not indicate that services subject to minimum staffing requirements, such as round-the-clock

obstetric or emergency cover, have become more sustainable. Depopulation is, moreover, accompanied by ageing, which raises need per resident; the utilisation ratio captures this only through expected utilisation by age and sex.

The finding also speaks to a broader methodological debate in health workforce planning. Reviews of workforce planning models have long noted that needs-based approaches depend on normative judgements while utilisation-based approaches extrapolate existing patterns and therefore embed unmet need [34–37]. Our results identify a third and less discussed vulnerability that is specific to monitoring rather than projection: an indicator can register improvement while the underlying service contracts, provided the reference population contracts faster. In a depopulating society this is not an edge case but the modal situation in peripheral areas.

Japan's own deliberations anticipated the argument without testing it. The observation that both ends of the distribution would improve without any policy intervention, and that the gap rather than the level should be monitored, was placed on the record at the committee's second session in February 2024. Evidence subsequently presented to the committee nevertheless continued, for the most part, to feature levels and growth rates. Our analysis suggests that the objection was not merely plausible but quantitatively important at the operative spatial scale.

Implications for Equity Monitoring

The findings bear directly on how geographical health equity is assessed. An equity indicator is meant to answer whether systematic differences between population groups are narrowing. The density-based index examined here answered that question affirmatively for 2024–2026, but at the spatial scale that governs designation the affirmative answer did not survive removal of differential population loss: net of population change, dispersion among medical areas was essentially unchanged. Because the denominator shrinks fastest in the least-served places, the indicator tends to report progress toward equity precisely where services are contracting. This can be described as a form of measurement inequity, in that the instrument is least informative for the populations it is designed to protect, and the mechanism will operate in any depopulating setting that monitors provider distribution with population denominators.

Implications for Indicator Design

Four practical implications follow. First, monitoring should report dispersion alongside levels. The level of a density indicator in a shrinking population is not interpretable as a measure of adequacy; the distribution of that indicator across areas is. Reporting both, with the level clearly labelled as relative, would prevent the most common misreading.

Second, indicator change should be decomposed and published in decomposed form. The computation is trivial where the indicator has a multiplicative structure, requires no additional data collection, and converts an ambiguous number into two interpretable ones. A monitoring report stating that an area's

index rose by 6%, of which 5 points came from population loss and 1 from physician gain, is far harder to misread than one reporting only the 6%.

Third, threshold rules deserve scrutiny. Relative designation guarantees a fixed count of scarce areas and is therefore incapable of registering aggregate progress or deterioration; the general risk that a measure ceases to inform once it becomes a target is well documented [38]. A hybrid rule — relative ranking for prioritisation, absolute capacity criteria for designation — would restore that capacity. Japan's introduction of a remoteness measure in 2026 is a step in this direction, and our finding that the measure is only weakly correlated with density ($r = -0.233$) indicates how much geographical access the density indicator was missing.

Fourth, indicators for essential services should not be anchored on a utilisation denominator that collapses with the service. The obstetric index illustrates the failure mode in its purest form. For functions that must be sustained regardless of volume — obstetric care, emergency surgery, out-of-hours cover — a capacity-based measure is required: whether a rota can be staffed, whether a minimum case volume for competence is met, whether the service exists at all. Volume–outcome relationships imply that such thresholds exist and are measurable [39, 40].

Implications beyond Japan

The mechanism is not Japanese. It operates wherever a density indicator is applied to a population that is shrinking unevenly, which now describes much of East Asia, Southern and Eastern Europe, and rural regions across the high-income world [41–43]. Korea is an instructive case: an amendment to the Framework Act on Health and Medical Services, promulgated in April 2025, established a statutory health workforce supply projection committee whose projections inform medical school admission quotas from the 2027 academic year [44]. Korea's fertility rate is the lowest in the world, and its regional demographic divergence is at least as pronounced as Japan's. Because its regional indicators are still being designed, the choices identified here — dispersion rather than level, published decomposition, hybrid thresholds, capacity-based essential-service measures — remain available at no cost. They are considerably more expensive to retrofit.

More generally, the study is a reminder that measurement instruments are not neutral with respect to the phenomena they measure. An indicator that improves as a region empties will, if used to evaluate policy, systematically direct attention away from the places losing services fastest.

Strengths and Limitations

The principal strength is that the decomposition is an identity rather than a model. It reproduces every published value exactly, and requires no assumptions about functional form; the decomposition itself is therefore not subject to specification error, although the counterfactuals rest on the accounting assumption stated in the Methods. The analysis uses complete official data for all units rather than a sample, and the counterfactual is constructed from the same published components.

Several limitations qualify the findings. First, the inputs to the utilisation ratio were revised between rounds, so part of the demand-structure effect reflects data revision rather than change in need; we have decomposed it and shown that holding the ratio fixed strengthens the medical-area result, but the published index itself conflates the two. The 2026 physician count also incorporates a ministry survey without a 2020 counterpart, which may introduce small discontinuities for individual areas.

Second, only two publication rounds carry full component data in comparable form, so the analysis covers a single interval. Earlier rounds used a partially different formula and are not directly comparable; extending the series would require recomputation from primary statistics. Third, the obstetric comparison spans unequal windows (six years for deliveries, four for physicians) and describes the behaviour of the published indicator rather than a synchronised change. Fourth, prefecture-level conclusions proved sensitive to standardisation, whereas medical-area conclusions did not; we have reported this divergence rather than selecting the more favourable specification. Fifth, the qualitative material is drawn from a single advisory body's public record and was used only to generate propositions, not to support the quantitative results.

Sixth, the analysis is ecological and descriptive. The counterfactuals assume that physician location is independent of population change, whereas physicians may themselves leave depopulating areas. Seventh, the primary dispersion measures were unweighted, giving each area equal weight regardless of population; population-weighted measures address a different question — inequality experienced by residents rather than inequality across planning units — and with them about three-quarters, rather than virtually all, of the observed decline among medical areas is attributable to population change. Eighth, the index relates physicians to residents and does not capture travel time, outreach, telemedicine or cross-boundary use beyond the inpatient-flow adjustment in the utilisation ratio.

Finally, the study measures what the indicator does, not what the underlying distribution of physicians does. We do not claim that Japan's physician distribution failed to improve in any respect; we claim that the improvement recorded by this indicator cannot be read as evidence of improved deployment, and that at medical-area level little of the recorded convergence can be so read.

Conclusions

Density-based physician maldistribution indicators improve as populations shrink, and the effect is largest where populations are smallest and falling fastest. In Japan between 2024 and 2026, 43% of the national improvement and two-thirds of the improvement across secondary medical areas came from the denominator; when population is held constant the apparent convergence among medical areas largely disappears. Because the artefact intensifies at exactly the spatial scale at which scarcity is officially designated, and because relative thresholds fix the number of designated areas by construction, the instruments used to monitor maldistribution in depopulating societies are least reliable where they matter most.

The remedy is neither complicated nor costly. Report dispersion alongside levels; publish indicator change decomposed into supply and population components; combine relative ranking with absolute capacity criteria for designation; and anchor essential-service indicators on capacity rather than on utilisation denominators that vanish with the service. Countries now building regional workforce monitoring systems can adopt these conventions at the outset. Doing so is a precondition for equity monitoring that reports progress only where progress has occurred.

Declarations

Ethics approval and consent to participate

The Institutional Review Board of Sungshin Women's University reviewed the study and granted an exemption from full review (SSWUIRB-2026-006). The study analyses publicly available aggregate statistics and published public records; no human participants or individual-level data were involved, and consent to participate was therefore not required.

Consent for publication

Not applicable.

Clinical trial number

Not applicable.

Availability of data and materials

All data analysed in this study are publicly available: the component datasets underlying the physician maldistribution indices from the Ministry of Health, Labour and Welfare, Japan [13], and the verbatim committee minutes from the same Ministry [17].

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Authors' contributions

TK conceived and designed the study, obtained and analysed the data, and drafted the manuscript. SC contributed to the study design and the interpretation of the findings, critically revised the manuscript for important intellectual content, and supervised the work. Both authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Whitehead M. The concepts and principles of equity and health. Int J Health Serv. 1992;22(3):429–45.
2. Braveman P, Gruskin S. Defining equity in health. J Epidemiol Community Health. 2003;57(4):254–8.
3. Ikegami N, Yoo BK, Hashimoto H, Matsumoto M, Ogata H, Babazono A, et al. Japanese universal health coverage: evolution, achievements, and challenges. Lancet. 2011;378(9796):1106–15.
4. Hann M, Gravelle H. The maldistribution of general practitioners in England and Wales: 1974–2003. Br J Gen Pract. 2004;54(509):894–8.
5. Horev T, Pesis-Katz I, Mukamel DB. Trends in geographic disparities in allocation of health care resources in the US. Health Policy. 2004;68(2):223–32.
6. Gravelle H, Sutton M. Inequality in the geographical distribution of general practitioners in England and Wales 1974–1995. J Health Serv Res Policy. 2001;6(1):6–13.
7. Correia I, Veiga P. Geographic distribution of physicians in Portugal. Eur J Health Econ. 2010;11(4):383–93.
8. Anand S, Fan VY, Zhang J, Zhang L, Ke Y, Dong Z, et al. China's human resources for health: quantity, quality, and distribution. Lancet. 2008;372(9651):1774–81.
9. Hara K, Otsubo T, Kunisawa S, Imanaka Y. Examining sufficiency and equity in the geographic distribution of physicians in Japan: a longitudinal study. BMJ Open. 2017;7(3):e013922.
10. Guagliardo MF. Spatial accessibility of primary care: concepts, methods and challenges. Int J Health Geogr. 2004;3:3.
11. Luo W, Wang F. Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. Environ Plann B. 2003;30(6):865–84.
12. Ministry of Health, Labour and, Welfare J. Guidelines for the formulation of physician securing plans (Ishi kakuho keikaku sakutei gaidorain), 8th Medical Plan, latter period. Tokyo: MHLW; 2026. https://www.mhlw.go.jp/content/001718579.pdf. Accessed 25 Oct 2026.

13. Ministry of Health. Labour and Welfare, Japan. Physician maldistribution index and related component datasets (January 2024 and April 2026 releases). https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryou/kinkyu/index.html. Accessed 16 Sep 2026.
14. Matsumoto M, Inoue K, Kajii E. Policy implications of a financial incentive programme to retain a physician workforce in underserved Japanese rural areas. Soc Sci Med. 2010;71(4):667–71.
15. Statistics Bureau of Japan, Ministry of Internal Affairs and Communications. Population estimates. https://www.stat.go.jp/english/data/jinsui/index.html. Accessed 16 Sep 2026.
16. National Institute of Population and Social Security Research. Regional population projections for Japan: 2020–2050. Tokyo: IPSS; 2023.
17. Ministry of Health. Labour and Welfare, Japan. Review Committee on Measures against Physician Maldistribution through the Physician Training Process: verbatim minutes, sessions 1–11. https://www.mhlw.go.jp/stf/shingi/other-isei_127303_00001.html. Accessed 16 Sep 2026.
18. Kobayashi Y, Takaki H. Geographic distribution of physicians in Japan. Lancet. 1992;340(8832):1391–3.
19. Inoue K, Matsumoto M, Toyokawa S, Kobayashi Y. Transition of physician distribution (1980–2002) in Japan and factors predicting future rural practice. Rural Remote Health. 2009;9(2):1070.
20. Toyabe S. Trend in geographic distribution of physicians in Japan. Int J Equity Health. 2009;8:5.
21. Tanihara S, Kobayashi Y, Une H, Kawachi I. Urbanization and physician maldistribution: a longitudinal study in Japan. BMC Health Serv Res. 2011;11:260.
22. Matsumoto M, Inoue K, Farmer J, Inada H, Kajii E. Geographic distribution of primary care physicians in Japan and Britain. Health Place. 2010;16(1):164–6.
23. Matsumoto M, Inoue K, Bowman R, Noguchi S, Toyokawa S, Kajii E. Geographical distributions of physicians in Japan and US: impact of healthcare system on physician dispersal pattern. Health Policy. 2010;96(3):255–61.
24. Sakai R, Wang W, Yamaguchi N, Tamura H, Goto R, Kawachi I. The impact of Japan's 2004 postgraduate training program on intra-prefectural distribution of pediatricians in Japan. PLoS ONE. 2013;8(10):e77045.
25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101.
26. Bowen GA. Document analysis as a qualitative research method. Qual Res J. 2009;9(2):27–40.
27. Shorrocks AF. The class of additively decomposable inequality measures. Econometrica. 1980;48(3):613–25.
28. Barro RJ, Sala-i-Martin X. Convergence J Polit Econ. 1992;100(2):223–51.
29. Wang F. Measurement, optimization, and impact of health care accessibility: a methodological review. Ann Assoc Am Geogr. 2012;102(5):1104–12.
30. Robinson WS. Ecological correlations and the behavior of individuals. Am Sociol Rev. 1950;15(3):351–7.

31. Simpson EH. The interpretation of interaction in contingency tables. J R Stat Soc Ser B. 1951;13(2):238–41.
32. Fotheringham AS, Wong DWS. The modifiable areal unit problem in multivariate statistical analysis. Environ Plann A. 1991;23(7):1025–44.
33. Jelinski DE, Wu J. The modifiable areal unit problem and implications for landscape ecology. Landsc Ecol. 1996;11(3):129–40.
34. Roberfroid D, Leonard C, Stordeur S. Physician supply forecast: better than peering in a crystal ball? Hum Resour Health. 2009;7:10.
35. Birch S, Kephart G, Tomblin-Murphy G, O'Brien-Pallas L, Alder R, MacKenzie A. Human resources planning and the production of health: a needs-based analytical framework. Can Public Policy. 2007;33(Suppl):S1–16.
36. Grover A, Niecko-Najjum LM. Building a health care workforce for the future: more physicians, professional reforms, and technological advances. Health Aff (Millwood). 2013;32(11):1922–7.
37. Lopes MA, Almeida ÁS, Almada-Lobo B. Handling healthcare workforce planning with care: where do we stand? Hum Resour Health. 2015;13:38.
38. Bevan G, Hood C. What's measured is what matters: targets and gaming in the English public health care system. Public Adm. 2006;84(3):517–38.
39. Birkmeyer JD, Siewers AE, Finlayson EVA, Stukel TA, Lucas FL, Batista I, et al. Hospital volume and surgical mortality in the United States. N Engl J Med. 2002;346(15):1128–37.
40. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. Ann Intern Med. 2002;137(6):511–20.
41. Kaufman BG, Thomas SR, Randolph RK, Perry JR, Thompson KW, Holmes GM, et al. The rising rate of rural hospital closures. J Rural Health. 2016;32(1):35–43.
42. United Nations. Department of Economic and Social Affairs, Population Division. World Population Prospects 2024. New York: United Nations; 2024.
43. Hospers GJ. Policy responses to urban shrinkage: from growth thinking to civic engagement. Eur Plan Stud. 2014;22(7):1507–23.
44. Framework Act on Health and Medical Services. (Republic of Korea), partially amended and promulgated 17 April 2025. Korean Law Information Center. https://www.law.go.kr. Accessed 16 Sep 2026.

Figures

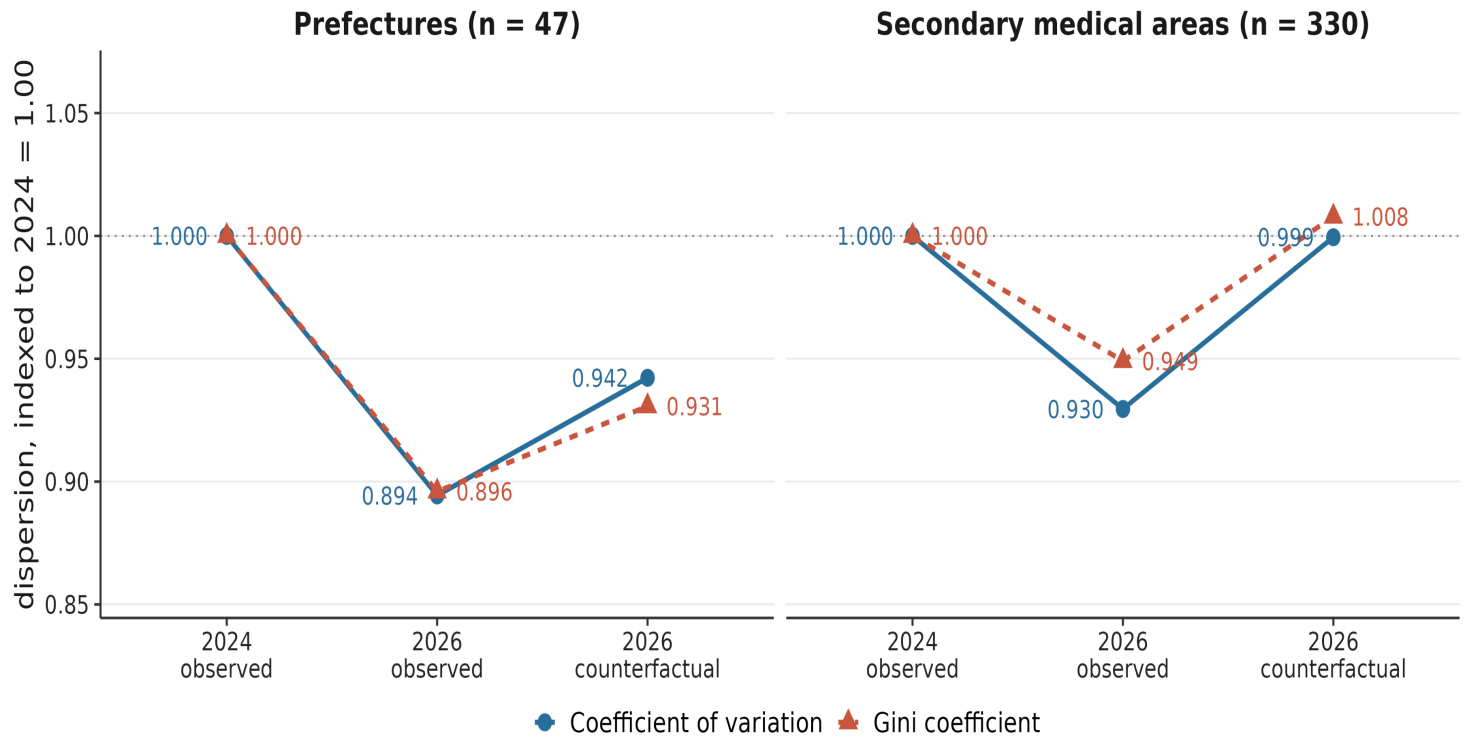


Figure 1

Observed and counterfactual dispersion of the index, indexed to 2024.

Coefficient of variation and Gini coefficient for prefectures (left) and secondary medical areas (right) under the observed 2024 and 2026 distributions and under the counterfactual that applies the 2026 standardised physician count to the 2021 population. Observed convergence largely disappears among medical areas once population change is removed.

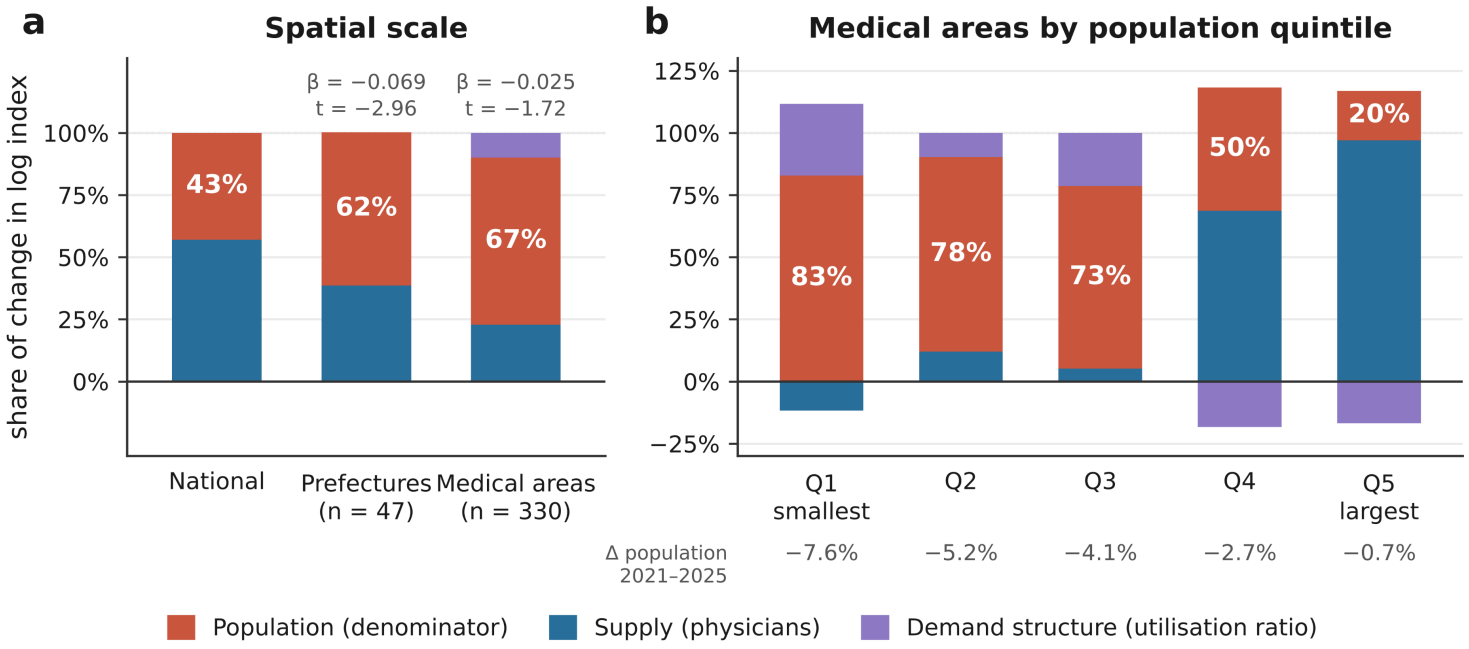


Figure 2

Scale dependence of the denominator effect.

(a) Contributions of the supply, population and demand-structure effects to the mean change in the log index, nationally and at two spatial scales; β and t give the counterfactual β-convergence coefficient. (b) The same contributions by quintile of medical-area population, with mean population change 2021–2025 beneath each quintile. Negative supply shares indicate that the standardised physician count fell on average.

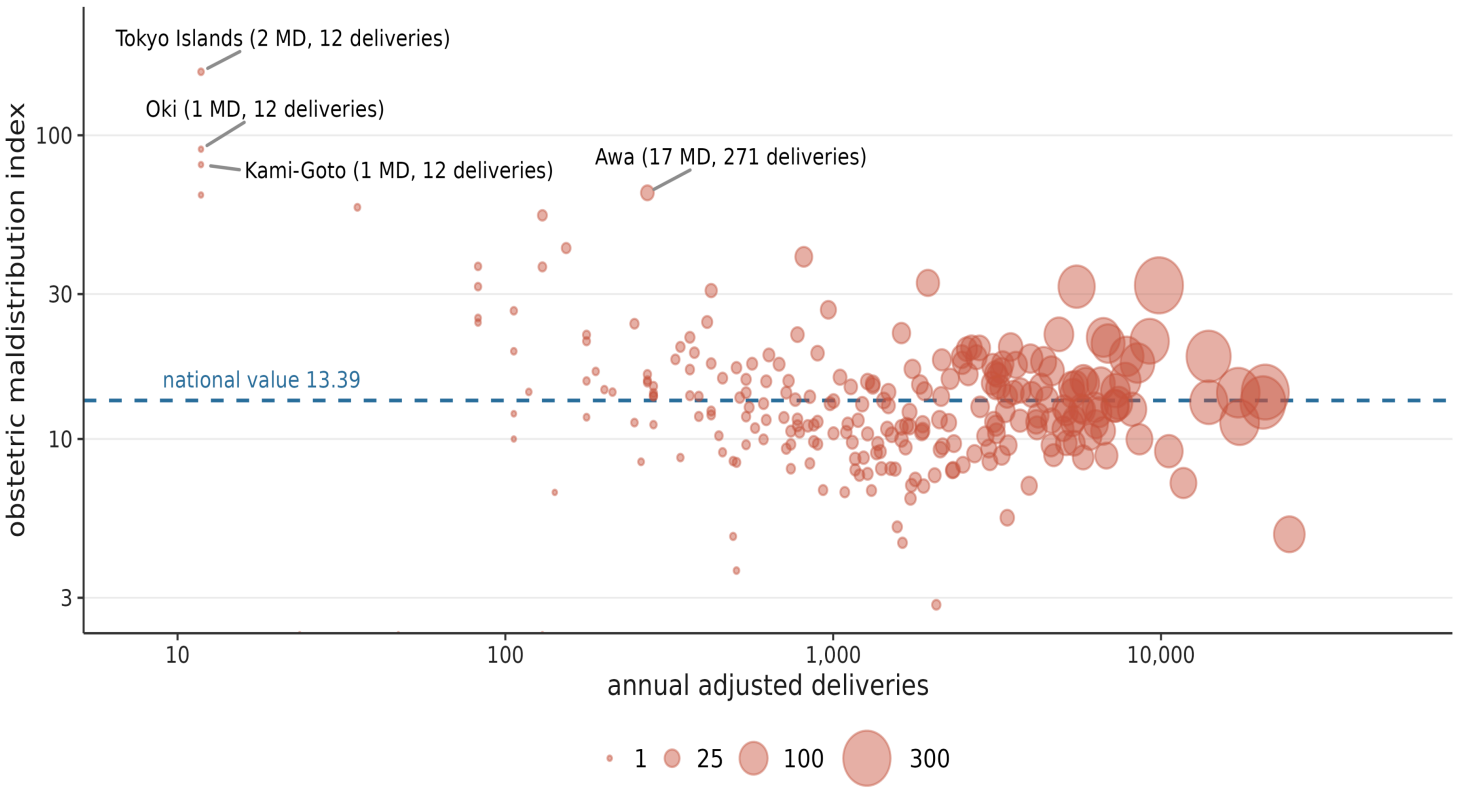


Figure 3

Obstetric maldistribution index against annual deliveries by perinatal medical area. Both axes are logarithmic; bubble size denotes the number of delivery-attending physicians; the dashed line marks the national value. The four highest-index areas are labelled with physician and delivery counts.