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.

Equal heat, unequal harm: rural–urban disparity in heat-related illness and a burden-matched comparison of heat-alert design levers in Korea

Eunjung CHO

ejcho@nigt.re.kr

Green Technology Center

Research Article

Keywords: heat-related illness, heat–health warning systems, alert frequency, forecast verification, standardized morbidity ratio, environmental epidemiology

Posted Date: October 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10841434/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

Heat warnings usually rest on a single temperature threshold, and the design debate has concentrated on two levers: the heat indicator, and localization. How often alerts are issued is rarely placed on the same scale. In a daily panel of 230 Korean municipalities, May–September 2019–2025 (244,430 area-days; 16,927 heat-illness emergency-department reports), standardized morbidity ratios (SMR) were estimated against a weather-only baseline and the three levers compared at a matched alert burden under four objective functions; rural municipalities were those below 500 persons/km². Exposure was near-identical and, if anything, marginally hotter in urban municipalities (mean daily maximum 27.83 vs 27.97°C), yet the rural SMR was 1.98 against 0.87, a ratio of 2.28 (95% CI 1.91–2.68); adding climate normals lowered it to 1.32, a lower bound given their collinearity with rurality. With the alert-day share fixed at 11.7%, percentile-matched thresholds captured 3.6 percentage points (pp) more cases than the single national rule and incidence-matched thresholds 1.2 pp more, though the latter captured 19.1 pp more rural cases. Switching the indicator yielded 0.8 pp, the two correlating at 0.95 daily. Raising the share to 28.7% increased case capture by 31.7 pp and Peirce skill by 17.2 pp, exceeding every lever on all four objectives — a gain in discrimination, not permissiveness. Half of all cases, about 1,160 a year, fall on non-alert days under the national 33°C rule. Alert frequency appears to dominate the debated levers; region-specific thresholds matter mainly on equity objectives, the indicator hardly at all.

1. Introduction

Heat is the weather hazard contributing most to mortality, and its burden is growing as warming and population ageing coincide (Cissé et al., 2022). Annual mortality attributable to non-optimal temperature is substantial, and the heat share of it is rising (Zhao et al., 2021). A considerable fraction of currently observed heat mortality has been attributed to anthropogenic climate change (Vicedo-Cabrera et al., 2021), and net increases in temperature-related excess mortality are projected under high-emission scenarios (Gasparrini et al., 2017; Masselot et al., 2025). In Korea, emergency-department surveillance reports of heat-related illness have risen rapidly, with 2023 reports 80.2% above the previous year (Lee et al., 2024), and domestic projections suggest further increases (Oh et al., 2025).

In most countries heat warnings are issued from a single temperature threshold, a design that embeds the premise that the same temperature implies the same risk. European warning systems differ in trigger and in whether thresholds are percentile-based or absolute, but nearly all apply a small number of criteria nationwide (Casanueva et al., 2019). Debate over which indicator should trigger a heat warning, and over whether thresholds should be localized, has been persistent. Temperature–mortality relationships and optimal temperatures differ markedly across locations (Gasparrini et al., 2015). In a systematic review of heat–health warning systems, Chandra N and Lee (2025) recommended localization from the national to the city scale. County-level thresholds derived from morbidity data in the United States and region-specific thresholds in Japan have likewise been proposed (Phung et al., 2026; Vaidyanathan et al., 2019).

How often heat warnings should be issued, by contrast, has received little attention. Alert frequency is generally treated as a consequence that follows once a threshold is set, rather than as a design choice with its own benefits and costs. Korea illustrates how much this matters. Until 2019, Korean heat advisories and warnings used daily maximum temperatures of 33°C and 35°C. From 2020 the same two values on an apparent-temperature scale were piloted, and this became operational on 15 May 2023. From the 2026 summer season, a reinforced severe heat warning was introduced, together with a region-specific tropical-night advisory (KMA, 2026). The system therefore moved simultaneously along two axes — a change of indicator and a finer spatial resolution — and alert frequency changed as a by-product. That the level of the threshold, and not only its form, is consequential is visible in the outcome data: Alexander et al. (2026) found a quarter of United States heat-related deaths to occur below a Heat Index of 33°C, well under most operational warning thresholds. Evidence on whether refining warning systems improves health outcomes nonetheless remains mixed (Urban et al., 2025; Weinberger et al., 2021).

Determining which of these changes actually conferred a benefit is a forecast-verification problem. The principle that rules must be compared after being matched on event frequency has been standard in that field since Peirce (1884), and frequency-bias-corrected threshold comparison (Hamill, 1999) is its most explicit form. In heat epidemiology, percentile-based local thresholds are a direct expression of the same idea, since a local 90th or 95th percentile imposes by construction the same alert-day share everywhere (Zhang et al., 2012). Alert frequency, however, has only ever been a condition to be matched, never a design choice to be selected.

The single-threshold premise fits particularly poorly when urban and rural areas are compared. The urban heat island might suggest that cities are more vulnerable, yet studies using heat-illness hospitalization or emergency care as the outcome report a rural excess instead (Ahn et al., 2025; Jagai et al., 2017; Mendrinos et al., 2024). The highest risk was likewise found among older adults in China (Hou et al., 2025). The same pattern holds in Korea: Lee et al. (2022) analyzed the relationship between population density and heat mortality risk across 229 districts and reported rural areas to be the most vulnerable. The most recent study using the same emergency-department surveillance data as the present analysis was a national analysis based on 16 observation stations (Park and Kim, 2026). To our knowledge, no study has yet removed weather by generating expected case counts from weather alone — rather than entering weather as a term in a regression — and estimated this disparity nationwide at the municipal level.

This study therefore (i) estimates the residual rural–urban disparity once exposure is equalized and bounds it against a baseline that allows for climatic acclimatization, and (ii) compares three levers — the alert indicator, region-specific differentiation, and alert frequency — jointly at a matched alert burden under four objective functions.

2. Materials and methods

2.1 Study design and data

The study covers 230 Korean municipalities over May–September 2019–2025. The 230 units comprise the 226 basic local governments in force as of 2025, plus Jeju City, Seogwipo City and Sejong Special Self-Governing City, with Gunwi County counted separately before and after its transfer from North Gyeongsang Province to Daegu Metropolitan City on 1 July 2023. All 230 units are referred to below as municipalities.

Because the surveillance data record the district of occurrence only from 2019 onward, all analyses were performed on the 2019–2025 municipal panel. The surveillance system nominally operates from 20 May to 30 September (Lee et al., 2024), but data are supplied from 1 May, so the panel was set to 1 May–30 September. Joining case, weather and population data on this grid yielded 244,430 area-days.

Municipalities were classified as rural where population density fell below 500 persons/km², following the density criterion the OECD regional typology applies to Korea (Brezzi et al., 2011). Area is the land area excluding the marine portion included in the source data (Ministry of Land, Infrastructure and Transport, 2024), and the 2019–2025 mean population was used. National population density in this panel is 515 persons/km², placing the criterion close to the national mean, and the 230 municipalities divide into 117 rural and 113 urban.

The surveillance data are individual-level reports from the Korea Disease Control and Prevention Agency heat-related illness emergency-department surveillance system, released through the national open-data portal. Of 29,762 cumulative reports for 2011–2025, 17,209 reports for 2019–2025 include the district, and the 16,927 (98.4%) whose place names matched the population and weather data were analyzed. Because the system aggregates reports for patients presenting to emergency departments, mild cases are not captured, and reporting propensity and severity selection may differ across regions. Observed counts are therefore a report-based measure of patients reaching emergency care rather than a measure of incidence.

Weather data are daily observations from the Korea Meteorological Administration Automated Synoptic Observing System (ASOS): daily mean, maximum and minimum temperature, daily mean and minimum relative humidity, sunshine duration and precipitation. Ninety-four stations observed the full May–September 2019–2025 period, and each municipality was assigned the station at the shortest great-circle distance from its administrative centroid. The result is that 87 stations serve 230 municipalities, so that on average 2.6 municipalities share a single daily series.

Perceived heat load was computed using the United States National Weather Service Rothfusz heat index (Rothfusz, 1990), a polynomial approximation to the apparent-temperature model of Steadman (1979) including low-humidity (< 13%) and high-humidity (> 85%) adjustments. The primary definition combines daily maximum temperature with daily minimum humidity, since using daily mean humidity would apply a humidity higher than that prevailing at the hour of maximum temperature. Indicator choice affects both which days are classified as heat days and the estimated risk (Heo and Bell, 2019), and

there is no consensus on which index best predicts health outcomes (Heo et al., 2019; Pantavou et al., 2025).

2.2 Age standardization and the baseline model

Indirect standardization using five-year age bands was applied to remove differences in age structure across municipalities (Clayton and Kaldor, 1987; Inskip et al., 1983). Bands were consolidated into five groups (0–14, 15–39, 40–64, 65–74, 75 and over), and national rates were derived internally from the same data (0.3, 2.3, 3.5, 4.3 and 6.8 per 100,000 population per season, respectively). The municipality-year age-standardization ratio is the expectation obtained by applying these national rates to the local age structure, divided by the expectation obtained by applying the national crude rate to total population; it ranged from 0.84 to 1.38 (mean 1.072). The baseline model offset is $log(population \div 100{,}000 \times age-standardization\ ratio)$.

The baseline model generates daily expected case counts from weather alone. Predictors are eight weather variables — daily mean, maximum and minimum temperature, mean and minimum relative humidity, sunshine, precipitation and the heat index — and three seasonality terms — month, day of year and day of week. The learner is gradient boosting with a Poisson objective (LightGBM; Ke et al., 2017), trained on 2019–2023 with early stopping after 60 rounds on 2024–2025 loss. What the baseline is required to remove is exposure, not vulnerability. Population attributes such as socioeconomic variables, population size and municipality fixed effects were therefore excluded.

A weather-only baseline treats all municipalities as sharing a single temperature–response function, so systematic differences in acclimatization remain in the residual. A second baseline was therefore fitted with each municipality's climatological mean daily maximum temperature and mean heat index added as predictors. Both are properties of weather and may legitimately enter a baseline intended to remove exposure. They are, however, strongly collinear with rurality, so a specification including them also removes part of the quantity being estimated. The weather-only baseline is accordingly treated as the primary specification, and the acclimatization specification is read as a lower bound.

Two specifications testing the baseline itself are reported alongside it. One adds 1–3 day lags of daily maximum temperature and the heat index, their three-day moving averages, and the run length of consecutive days at or above 33°C. The other re-estimates expectations out of sample by leave-one-year-out (LOYO) cross-validation, removing the concern that 2024–2025 served both as the early-stopping loss and as the basis of its own expectations. Because the result should not depend on which learner produced the expectations, the same computation was repeated with XGBoost, a Poisson generalized linear model and an LSTM with Poisson loss (Chen and Guestrin, 2016; Hochreiter and Schmidhuber, 1997), and municipality rankings were compared by Spearman correlation.

The standardized morbidity ratio (SMR) is the observed case count divided by the sum of expected cases. For group comparisons, both the arithmetic mean of municipality SMRs and the pooled SMR ($\Sigma O / \Sigma E$) are reported, weighting small municipalities equally and by case count respectively.

Confidence intervals at 95% were obtained from 2,000 bootstrap resamples of municipalities within groups. Because the limitations of SMRs in small-area comparison are well known (Julious et al., 2001), they are used only for relative comparison. Residual disparity is summarized by the negative-binomial overdispersion parameter τ, which is invariant to the scale of the expectations (Ver Hoef and Boveng, 2007).

2.3 Threshold estimation and alert simulation

Municipality thresholds were estimated in two ways. First, the occurrence threshold is the temperature at which the probability of any case reaches 10%, taken from a municipality-specific logistic regression of case occurrence on daily maximum temperature. Second, the incidence threshold is the temperature at which fitted incidence reaches a target R^* (per 100,000 population per day), taken from a Poisson regression of daily case counts with the age-standardisation offset. Each municipality's threshold was bootstrapped 200 times so that the median within-municipality standard error could be compared with the observed between-municipality standard deviation, separating true heterogeneity from estimation error.

The occurrence threshold is confounded by population size. Where population is small, the event "at least one case in a day" is itself rare, so the threshold forms at a higher temperature even at equal vulnerability. The incidence threshold, which removes this arithmetic through the age-standardization offset, was therefore taken as the primary measure, and the occurrence threshold was computed only to demonstrate the confounding.

Rules with different thresholds cannot simply be placed side by side. Under a monotone rule, lowering the trigger necessarily raises capture, so comparing rules with different numbers of alert days measures the permissiveness of the trigger rather than the merit of the rule. Comparisons of the indicator and region-specific levers therefore fix the alert-day share first (Hamill, 1999), at the value the national 33°C daily-maximum rule realizes in this panel; the frequency lever, by construction, varies it.

Because what would have happened under a different rule is not observed, alert status was re-assigned on the same panel according to each rule and the cases each rule would have captured estimated. The simulation compares four rule families across the 229 municipalities remaining after Sejong, which reported no cases, was excluded: (a) a single national threshold on daily maximum temperature, (b) a single national threshold on the heat index, (c) municipality thresholds matched on incidence, and (d) municipality thresholds matched on local percentiles. Because indicator comparison can depend on which rule sets the burden, it was performed in both directions. The alert-day share was also swept from 5% to 40% in 0.5 pp steps to trace the entire burden–capture frontier.

Four objective functions were evaluated for every rule. Case capture is the share of cases occurring on alert days, which an agency minimizing the total number of missed cases would maximize. Incidence-weighted capture weights each case by the reciprocal of its municipality's population, approximating an objective defined on incidence. Rural case capture is capture computed within rural municipalities and is

a direct equity criterion. The Peirce skill score is the hit rate for daily case occurrence minus the false-alarm rate, a standard verification score insensitive to frequency (Peirce, 1884). The standard deviation of realized alert-day shares across municipalities is also reported.

Confidence intervals for all lever contrasts were obtained from 2,000 cluster bootstrap resamples of municipalities drawn with replacement. Because a cluster resample duplicates a municipality's days as a block, each municipality's threshold is invariant to the resample, so fixing thresholds at their full-period estimates is exact; these intervals contain only between-municipality sampling variation. The year dimension was handled separately by two further resampling schemes: a year-cluster bootstrap over the seven summers, and a two-way bootstrap resampling municipalities and years simultaneously. Threshold estimation error was handled by year-level cross-validation.

3. Results and Discussion

3.1 The rural–urban disparity

Table 1 presents municipality characteristics and exposure by rural–urban classification. Exposure was near-identical between rural and urban municipalities and, if anything, marginally hotter in urban areas. Mean daily maximum temperature was 27.83°C rural against 27.97°C urban, a difference of 0.14°C, and the same held in the upper tail that matters for heat-related illness. Days at or above 33°C accounted for 11.73% of rural and 11.77% of urban area-days, and the upper percentiles and the heat index differed no more.

Table 1
Municipality characteristics and exposure by rural–urban classification

Variable	All	Rural (< 500/km²)	Urban (≥ 500/km²)
Municipalities	230	117	113
Observed cases	74 (69)	58 (50)	90 (81)
Age-standardised SMR	1.43 (1.10)	1.98 (1.21)	0.87 (0.60)
Population density (persons/km²)	3,697 (5,831)	128 (111)	7,392 (6,515)
Mean population	223,996 (222,849)	82,653 (78,171)	370,342 (229,638)
Land area (km²)	440 (382)	711 (304)	160 (216)
Urbanised land share (%)	46.5 (43.0)	9.2 (10.7)	85.2 (26.3)
Mean daily maximum temperature (°C)	27.90 (0.98)	27.83 (1.05)	27.97 (0.89)
Mean heat index (°C)□	29.47	29.39	29.55
Days ≥ 33°C (%)	11.75	11.73	11.77
Days ≥ 35°C (%)	3.09	2.87	3.32
95th percentile of daily maximum (°C)	34.30	34.30	34.40
99th percentile of daily maximum (°C)	36.00	35.90	36.10
Municipality threshold, heat index (°C)□	38.21 (3.75)	36.02 (2.94)	40.51 (3.07)

Note. Values are means (standard deviations) where dispersion is meaningful.

□ *Primary heat index, based on daily maximum temperature and daily minimum relative humidity.*

□ *Incidence threshold at the operating point used in the simulation (R* = 0.104), computed for the 229 municipalities (117 rural, 112 urban) remaining after Sejong, for which no slope could be fitted because it reported no cases.*

Against this near-identical exposure, the age-standardized SMR was 1.98 rural and 0.87 urban, a ratio of 2.28. The case-weighted pooled SMR preserved the disparity at 1.66 and 0.82. Rural municipalities hold 18.8% of the national population and account for 40.1% of cases. Under a classification by administrative type, which uses neither population nor area, the SMR was 2.11 in counties (gun), 1.33 in cities (si) and 0.73 in metropolitan districts (gu), giving a county-to-district ratio of 2.91 on mean SMRs and 2.82 on pooled SMRs — larger than the density-based ratio. The ratio was 2.04 under the general

OECD threshold of 150 persons/km², which classifies 84 municipalities as rural. The disparity is therefore unlikely to be an artefact of defining rurality by density.

How much of the disparity remains depends on how climatic acclimatization is handled, with results shown in Table 2. Under the weather-only baseline the ratio was 2.28 in sample and 2.48 out of sample; adding each municipality's climatological mean temperature and mean heat index to the baseline reduced it to 1.32 in sample and 1.34 out of sample, and lowered τ from 1.37 to 0.40. The second specification attributes to acclimatization, that part of the disparity which tracks long-run local climate. Because municipality normals are collinear with rurality, the estimate is reported as 2.28 and the lower bound as 1.32, and the substantive conclusion is that the disparity does not disappear under the strongest weather adjustment available here. Nor does anything depend on the learner. Across LightGBM, XGBoost, a Poisson generalized linear model and a Poisson LSTM, municipality SMR rankings agreed at Spearman ρ ≥ 0.984 and the ratio remained within 2.25–2.38.

Table 2
Rural–urban disparity by baseline specification (n = 230)

Baseline specification	τ²	Rural SMR	Urban SMR	Rural/urban (95% CI)
A. Predictors — eight weather variables, LightGBM				
Primary, in sample	1.37	1.98	0.87	2.28 (1.91–2.68)
Primary, out of sample	1.46	1.99	0.80	2.48 (2.08–2.96)
Climate normals added, in sample	0.40	1.39	1.05	1.32 (1.17–1.50)
Climate normals added, out of sample	0.34	1.35	1.01	1.34 (1.18–1.51)
Lagged exposure added, in sample	1.37	1.98	0.88	2.25 (1.90–2.64)
Lagged exposure added, out of sample	1.39	1.95	0.81	2.42 (1.99–2.89)
B. Learner — eight weather variables, in sample				
XGBoost	1.77	2.17	0.96	2.25 (1.91–2.64)
Poisson generalised linear model	1.91	2.07	0.89	2.34 (1.92–2.86)
LSTM with Poisson loss	2.15	2.28	0.96	2.38 (1.99–2.83)

Under the weather baseline, municipality SMRs ranged from 0.00 to 7.66 with τ² = 1.37 (χ²/df = 37.9). Crude ratios without age standardization ranged from 0.00 to 8.87 with χ²/df = 42.7, so adjusting for age structure reduced overdispersion by only 11%. SMRs declined monotonically with population density (Fig. 1). Mean SMRs by density quintile were 2.40, 1.77, 1.48, 0.86 and 0.64, and the rank correlation with log density (− 0.646) was stronger than that with log population (− 0.588).

3.2 Alert thresholds and the three design levers

At the operating point used in the simulation (R^* = 0.104), the mean heat-index threshold across the 229 municipalities was 38.21°C (SD 3.75), being 36.02°C rural and 40.51°C urban. Bootstrapping each municipality's threshold gave a median within-municipality standard error of 0.73°C (IQR 0.56–1.01), so that approximately 3.68°C of the observed 3.75°C between-municipality standard deviation represents true heterogeneity. Rank correlations between thresholds and log population were + 0.586 on the heat index and + 0.545 on maximum temperature: less populous municipalities reach a given incidence at a lower temperature. In the same data the occurrence threshold showed the opposite correlation, − 0.515, and the rank correlation between the two thresholds was only 0.338, so the two definitions give substantially different orderings. The positive correlation of the incidence threshold reflects the substantive fact that smaller populations reach the same incidence at lower temperatures, whereas the negative correlation of the occurrence threshold is an arithmetic consequence of rare events.

With the alert-day share fixed at 11.7%, the value the national 33°C maximum-temperature rule generates, the four rule families differ by objective function, as shown in Table 3 and Fig. 2. On case capture, incidence-matched municipality thresholds achieved 53.1% against 51.9% for the national rule, a gain of 1.2 pp (95% CI − 2.3 to + 4.8); on the heat index the corresponding figure was 51.2%, or − 0.6 pp. The same incidence-matched rule, however, captured 70.2% of rural cases against 51.1% under the national rule, and 67.5% against 49.8% on the incidence-weighted objective, gains of 19.1 pp (95% CI + 15.0 to + 23.9) and 17.7 pp (+ 14.6 to + 21.5) respectively. Capture is weighted by case count, and the incidence-matched rule shifts alert days toward municipalities with few cases and high incidence. It therefore improves protection substantially where risk is highest while registering no gain at all on objectives that count cases.

Table 3
Performance of the design levers at a matched alert burden

Alert Rule	Objective function			
	Case capture	**Incidence-weighted**	**Rural capture**	**Peirce skill**
Reference — alert-day share 11.72%				
National rule, max. temperature 33°C	51.87 (49.1, 54.5)	49.82 (46.3, 53.1)	51.13 (47.4, 54.8)	37.63 (35.5, 39.8)
Change from reference (pp)				
Indicator: national heat index	+ 0.8 (− 0.9, + 2.7)	+ 0.8 (− 1.1, + 3.1)	+ 0.8 (− 1.7, + 3.9)	+ 0.7 (− 0.7, + 2.2)
Region-specific: percentile-matched	+ 3.6 (+ 1.7, + 5.7)	+ 5.8 (+ 3.1, + 9.2)	+ 2.9 (+ 0.0, + 5.9)	+ 3.5 (+ 1.9, + 5.2)
Region-specific: incidence-matched	+ 1.2 (− 2.3, + 4.8)	+ 17.7 (+ 14.6, + 21.5)	+ 19.1 (+ 15.0, + 23.9)	+ 0.3 (− 2.6, + 3.6)
Alert frequency: 11.72% → 28.69%	+ 31.7 (+ 29.5, + 33.9)	+ 33.5 (+ 30.9, + 36.2)	+ 32.3 (+ 29.4, + 35.2)	+ 17.2 (+ 15.3, + 19.0)
Note: Values are percentages, with 95% cluster bootstrap confidence intervals in parentheses.				

Parameterization matters as much as the objective function. Municipality thresholds matched on local percentiles — each municipality issuing alerts on its own hottest 11.7% of days — captured 3.6 pp more cases than the national rule (95% CI + 1.7 to + 5.7). This is a gain on the indicator the national rule itself favours, and the Peirce skill score also rose by 3.5 pp (+ 1.9 to + 5.2). It achieved this while equalizing the burden almost completely: the standard deviation of realized alert-day shares across municipalities was 0.21 pp, far below the 5.75 pp of the national rule and the 9.77 pp of the incidence-matched rule.

Because temperature is recorded to 0.1°C, a deterministic percentile rule cannot reach the reference burden exactly. Using each municipality's 11.7% quantile directly yields a national realized share of 12.06%, exceeding the reference by 0.34 pp. Where a rule not matched exactly gains 4.7 pp, the exactly matched rule gains 3.6 pp, and the latter is what is reported here.

Switching the indicator, by contrast, did almost nothing. At 11.7%, the national heat-index rule (threshold 37.7°C) captured 52.7%, 0.8 pp above maximum temperature (95% CI − 0.9 to + 2.7); at 28.7%, the 30.5°C maximum-temperature rule captured 84.10% and the 33°C heat-index rule 83.95%, a difference of 0.15 pp. The two indicators correlate at 0.946 daily across the panel, so once alert frequency is matched they select almost the same days. The apparent superiority of the heat index — 84.0% against 51.9% — is a comparison at burdens differing by a factor of 2.4 rather than a property of the indicator.

The lever the field does not debate, alert frequency, is the largest of the three. Holding the indicator and national uniformity fixed and raising only the alert-day share to 28.7% — the value the current apparent-

temperature 33°C criterion realizes of itself — increased case capture by 31.7 pp (95% CI + 29.5 to + 33.9) and the Peirce skill score by 17.2 pp (+ 15.3 to + 19.0). It exceeds every other lever on all four objectives. Even the narrowest margin, on rural capture against the incidence-matched rule, is + 13.2 pp (+ 9.2 to + 17.0). The Peirce skill score distinguishes this from mere permissiveness: a rule that issues more alerts without added discrimination scores zero at any frequency. Here, however, the hit rate rose from 47.4% to 80.7%, a gain of 33.2 pp, while the false-alarm rate rose only from 9.8% to 25.8%, a gain of 16.1 pp (95% CI + 15.6 to + 16.6); the difference of 17.2 pp is precisely the increase in Peirce skill. Even so, under the national 33°C rule 8,147 of the 16,927 cases in the simulation panel (48.1%) occurred on days carrying no alert, and the best-performing rule at the same burden reduced it only to 44.5%.

3.3 Discussion

The central result is that the policy question "do region-specific alert thresholds help?" has no answer independent of two choices analysts usually leave implicit — the objective function, and the parameterization of "region-specific". First, alert frequency is not a condition to be matched but a design lever that can be selected, and it is the largest of the three. Since Peirce (1884), forecast verification has required event frequency to be matched before rules are compared (Hamill, 1999), and percentile thresholds in heat epidemiology impose the same alert frequency everywhere by construction (Zhang et al., 2012). Placing all three on the same burden shows the largest available gain to lie on precisely that dimension. Korea’s transition to apparent temperature in 2020–2023 was a genuine policy change, but these results suggest that its benefit may have derived from the alert frequency it produced rather than from the heat index itself.

Second, which objective function serves as the scoring criterion reverses the verdict on region-specific thresholds. Case capture, a count, is weighted by case number, so a rule that moves alert days toward municipalities with few cases and high incidence can hardly show a gain on it. That is exactly what the incidence-matched rule does, and the + 1.2 pp result is an arithmetic consequence of this rather than evidence that region-specific differentiation failed. The same rule gains 19.1 pp on rural case capture. Which of the two counts as success is a normative choice, and the contribution here lies not in adjudicating it but in making it explicit and quantifying the cost on each side.

Third, how “region-specific” is parameterized matters as much as the objective. Matching municipalities to a common target incidence and matching them to a common local percentile are both “region-specific thresholds”, yet they move differently on case capture. Percentile-based local thresholds have long been used in heat epidemiology precisely because they impose a common alert frequency by construction (Zhang et al., 2012), and these results indicate that it is this property, rather than region-specific differentiation as such, that determines success on case capture.

Fourth, using a threshold as a vulnerability indicator requires first establishing that its definition is invariant to sample size. A threshold defined by the probability of at least one case per day is confounded by population size, because that event is itself rare where population is small. The two definitions correlate at only 0.338 and order municipalities differently.

These contributions notwithstanding, the study has several limitations. First, the simulation assumes perfect foresight and does not represent behavioral response. Observed weather is used, whereas operational alerts are issued on 1–2 day forecasts. All capture figures are therefore biased upward. Thresholds close to the local climatological distribution are more sensitive to forecast error than high absolute thresholds, so the direction of the bias in rule comparison is unknown. The effect of issuing alerts more often is evaluated here only through capture and skill, which measure the cases an alert reached rather than the harm it averted; since the evidence that heat warnings improve health outcomes is itself mixed (Urban et al., 2025; Weinberger et al., 2021), the comparisons here should be read as comparisons of protective opportunity rather than of averted harm. The behavioral response to a changed alert frequency — above all alert fatigue — cannot be observed in these data, and this is the most important qualification attaching to the frequency result.

Second, several biases of undetermined direction remain in the disparity estimate. The denominator is the registered resident population whereas the numerator follows the district of occurrence, so risk may be underestimated in districts with large daytime inflows and overestimated in rural areas with seasonal inflows. The surveillance data record neither severity nor outcome, so it is not possible to test whether rural and urban case series were selected differently on severity — a real possibility where access to emergency care differs, and one that could bias the ratio in either direction. Age standardization used five broad age groups derived internally from the same data, so a finer, sex-stratified, or externally referenced standardization could remove more age structure than the 11% reduction in overdispersion observed here. Exposure measurement error arising from station assignment attenuates the exposure–response relationship and biases the disparity downward. Finally, the acclimatization specification that lowers the ratio to 1.32 overcontrols to an unknown degree, because climate normals are collinear with rurality; this is why that value is reported as a lower bound rather than a second estimate and is not used in any analysis beyond Table 2. None of these biases is therefore described as conservative.

Third, the rules simulated here stand at some distance from operational warnings. Municipality thresholds were estimated from log-linear Poisson regressions with a single predictor — daily maximum temperature or the heat index — so threshold-type or non-monotone exposure–response relationships are not reflected in the threshold definition. Where the target incidence is set also moves threshold levels considerably. At the operating point used here (R^* = 0.104), the estimated threshold exceeds the municipality's observed maximum in only 1.3% of municipalities on either indicator, but raising R^* to 0.50 lifts that share to 48.9% on maximum temperature and 38.4% on the heat index. This is why a low operating point was chosen, and equally why only the ordering across municipalities and the relative comparison between rules — not the absolute level of the thresholds — are treated as interpretable. The construction of the rules stands at a similar distance. Operational issuance involves forecaster judgement and persistence conditions and is made over warning zones, whereas the simulation applies a single-day threshold municipality by municipality. The heat index used here is the NWS Rothfusz formulation, which is not the same function as the KMA apparent temperature, so the 33°C of this study cannot be equated with the KMA's 33°C (Rothfusz, 1990).

4. Conclusions

Under near-identical heat exposure — the same on average and in the upper tail, and marginally hotter in urban areas if anything — rural municipalities in Korea carried 2.3 times the heat-illness risk of urban municipalities. Even in a specification removing each municipality's climate normals as exposure, the ratio did not fall below 1.3. The disparity withstands age standardization, four baseline learners, lagged exposure and out-of-sample validation. Municipality alert thresholds are genuinely heterogeneous, with 96% of the variance in municipality thresholds representing true differences rather than estimation error.

That heterogeneity is real does not determine how it should be handled. At a matched alert burden, incidence-matched thresholds are indistinguishable from the single national rule on case capture but capture 19.1 pp more rural cases, whereas percentile-matched thresholds capture 3.6 pp more cases while equalizing the burden across municipalities almost completely. Which of these counts as an improvement is a normative choice, and reporting a single capture statistic conceals it. Changing the heat indicator is close to immaterial once the burden is matched, because the candidate indicators select almost the same days.

Both levers are small beside the one the field does not contest. Raising the alert-day share from 11.7% to 28.7% — the value Korea's current criterion generates of itself — increased case capture by 31.7 pp and the Peirce skill score by 17.2 pp. This exceeds the two debated levers on all four objectives, at the cost of raising the false-alarm rate from 9.8% to 25.8%.

What no rule at the current alert burden can change is that half of all heat-illness cases — approximately 1,160 per year — occur on days carrying no alert. Raising capture substantially requires issuing alerts more often, and how much more often is a question about alert fatigue, response capacity and the effectiveness of response, not about thresholds. Answering the spatial question properly would require exposure reconstructed from a denser observation network than the station assignment used here, and a simulation based on forecasts rather than observations.

Declarations

Consent to Participate declaration

Not applicable.

Ethics declaration

Not applicable.

Declaration of Competing Interest

The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

All primary data are publicly available. Individual-level reports from the Korea Disease Control and Prevention Agency heat-related illness emergency-department surveillance system are available from the national open-data portal (data.go.kr); daily KMA ASOS observations from the Open MET Data Portal (data.kma.go.kr); and population, household, fiscal, employment and park statistics from the Ministry of the Interior and Safety, the Local Finance Integrated Open System, and KOSIS. Processed datasets and analysis outputs are available from the corresponding author upon reasonable request.

CRediT author statement

E.C.: Conceptualization; Methodology; Software; Formal analysis; Investigation; Data curation; Validation; Visualization; Writing – original draft; Writing – review & editing. E.C. is the sole author and approved the final manuscript.

References

1. Ahn M, Keith L, Brown HE (2025) Rural heat health disparities: Evidence from the U.S. National Emergency Medical Services Information System (NEMSIS). J Clim Change Health 22:100432. https://doi.org/10.1016/j.joclim.2025.100432
2. Alexander J, Cubbin C, Yang Z-L, Optimizing US, Heat Alerts, Weather (2026) Clim Soc 18, 641–653. https://doi.org/10.1175/WCAS-D-25-0079.1
3. Brezzi M, Dijkstra L, Ruiz V (2011) OECD extended regional typology: the economic performance of remote rural regions. OECD Regional Development Working Papers 2011
4. Casanueva A, Burgstall A, Kotlarski S, Messeri A, Morabito M, Flouris AD, Nybo L, Spirig C, Schwierz C (2019) Overview of Existing Heat-Health Warning Systems in Europe. IJERPH 16, 2657. https://doi.org/10.3390/ijerph16152657
5. Chandra N, Lee SVS, J.K.W (2025) A Systematic Review of Heat Health Warning Systems: Enhancing the Framework Towards Effective Health Outcomes. Curr Envir Health Rpt 12:31. https://doi.org/10.1007/s40572-025-00496-5
6. Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Presented at the KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, pp. 785–794. https://doi.org/10.1145/2939672.2939785
7. Cissé G, McLeman R, Adams H, Aldunce P, Bowen K, Campbell-Lendrum D, Clayton S, Ebi KL, Hess J, Huang C others, 2022. Health, wellbeing, and the changing structure of communities. Clim Change

7, 1041–1170

8. Clayton D, Kaldor J (1987) Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. Biometrics 43:671. https://doi.org/10.2307/2532003
9. Gasparrini A, Guo Y, Hashizume M, Lavigne E, Zanobetti A, Schwartz J, Tobias A, Tong S, Rocklöv J, Forsberg B, Leone M, De Sario M, Bell ML, Guo Y-LL, Wu C, Kan H, Yi S-M, De Coelho SZS, Saldiva M, Honda PHN, Kim Y, Armstrong H, B (2015) Mortality risk attributable to high and low ambient temperature: a multicountry observational study. Lancet 386:369–375. https://doi.org/10.1016/S0140-6736(14)62114-0
10. Gasparrini A, Guo Y, Sera F, Vicedo-Cabrera AM, Huber V, Tong S, Coelho DSZS, Nascimento Saldiva M, Lavigne PH, Matus Correa E, Valdes Ortega P, Kan N, Osorio H, Kyselý S, Urban J, Jaakkola A, Ryti JJK, Pascal NRI, Goodman M, Zeka PG, Michelozzi A, Scortichini P, Hashizume M, Honda M, Hurtado-Diaz Y, Cruz MC, Seposo J, Kim X, Tobias H, Iñiguez A, Forsberg C, Åström B, Ragettli DO, Guo MS, Wu YL, Zanobetti C, Schwartz A, Bell J, Dang ML, Van TN, Heaviside DD, Vardoulakis C, Hajat S, Haines S, Armstrong A, B (2017) Projections of temperature-related excess mortality under climate change scenarios. Lancet Planet Health 1:e360–e367. https://doi.org/10.1016/S2542-5196(17)30156-0
11. Hamill TM (1999) Hypothesis Tests for Evaluating Numerical Precipitation Forecasts. Wea Forecast 14:155–167. https://doi.org/10.1175/1520-0434(1999)014%3C0155:HTFENP%3E2.0.CO;2
12. Heo S, Bell ML (2019) Heat waves in South Korea: differences of heat wave characteristics by thermal indices. J Expo Sci Environ Epidemiol 29:790–805. https://doi.org/10.1038/s41370-018-0076-3
13. Heo S, Bell ML, Lee J-T (2019) Comparison of health risks by heat wave definition: Applicability of wet-bulb globe temperature for heat wave criteria. Environ Res 168:158–170. https://doi.org/10.1016/j.envres.2018.09.032
14. Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
15. Hou K, Zhang L, Yang F, Hu W, Xu X, Yao X, Wang Z (2025) Inequality in mortality risk of older adults between rural and urban areas exposure to temperature variations in China. BMC Public Health 25:3000. https://doi.org/10.1186/s12889-025-24088-y
16. Inskip H, Beral V, Fraser P, Haskey J (1983) Methods for age-adjustment of rates. Stat Med 2:455–466. https://doi.org/10.1002/sim.4780020404
17. Jagai JS, Grossman E, Navon L, Sambanis A, Dorevitch S (2017) Hospitalizations for heat-stress illness varies between rural and urban areas: an analysis of Illinois data, 1987–2014. Environ Health 16:38. https://doi.org/10.1186/s12940-017-0245-1
18. Julious S, Nicholl J, George S (2001) Why do we continue to use standardized mortality ratios for small area comparisons? J Public Health 23:40–46. https://doi.org/10.1093/pubmed/23.1.40
19. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R,

Vishwanathan S, Garnett R (eds) Advances in Neural Information Processing Systems. Curran Associates, Inc

20. KMA, (Korea Meteorological Administration) (2026) Reform of the heat-wave advisory system after 18 years and subdivision of special-report zones after 22 years [in Korean]
21. Lee J, An D, Ahn Y (2024) Results of Operating the 2023 Heat-related Illness Surveillance System. Public Health Wkly Rep 17:1421–1431. https://doi.org/10.56786/PHWR.2024.17.34.1
22. Lee W, Choi M, Bell ML, Kang C, Jang J, Song I, Kim Y-O, Ebi K, Kim H (2022) Effects of urbanization on vulnerability to heat-related mortality in urban and rural areas in South Korea: a nationwide district-level time-series study. Int J Epidemiol 51:111–121. https://doi.org/10.1093/ije/dyab148
23. Masselot P, Mistry MN, Rao S, Huber V, Monteiro A, Samoli E, Stafoggia M, de'Donato F, Garcia-Leon D, Ciscar J-C, Feyen L, Schneider A, Katsouyanni K, Vicedo-Cabrera AM, Aunan K, Gasparrini A (2025) Estimating future heat-related and cold-related mortality under climate change, demographic and adaptation scenarios in 854 European cities. Nat Med 31:1294–1302. https://doi.org/10.1038/s41591-024-03452-2
24. Mendrinos A, O'Brien J, Davis M, Baldwin A, Zaitchik BF, Britton A, Mwanja I, Gohlke JM (2024) Association between summertime emergency department visits and maximum daily heat index in rural and non-rural areas of Virginia (2015–2022). Sci Total Environ 948:174753. https://doi.org/10.1016/j.scitotenv.2024.174753
25. Oh J, Lee W, Park J, Kang C, Seo Y, Kim H, Hwang S (2025) Projection of temperature-related severe injuries under climate change: a nationwide study of South Korea. BMC Public Health 25:3516. https://doi.org/10.1186/s12889-025-24844-0
26. Pantavou K, Fillon A, Li L, Maniadis Z, Nikolopoulos GK (2025) Thermal indices for assessing the impact of outdoor thermal environments on human health: a systematic review of epidemiological studies. Int J Biometeorol 69:1843–1866. https://doi.org/10.1007/s00484-025-02948-x
27. Park MS, Kim BW (2026) Consistently higher and steeper apparent temperature–heat-related illness risk among occupational cases in Korea: evidence from national emergency department surveillance. Front Public Health 14:1786564. https://doi.org/10.3389/fpubh.2026.1786564
28. Peirce CS (1884) The numerical measure of the success of predictions. Science ns-4. 453–454. https://doi.org/10.1126/science.ns-4.93.453-a
29. Phung VLH, Honda Y, Oka K, Madaniyazi L, Ng CFS, Tobias A, Hashizume M (2026) Determining Location-Specific Thresholds for Heat Warning Systems to Mitigate Heatstroke Mortality in Japan. Environ Health 4:144–153. https://doi.org/10.1021/envhealth.5c00113
30. Rothfusz L, The heat index equation. NWS Southern Region Technical Attachment, SR/SSD 90 – 23, Fort Worth, Texas, Shitzer A (1990) (2006) Wind-chill-equivalent temperatures: regarding the impact due to the variability of the environmental convective heat transfer coefficient. Int J Biometorol 50, 224–232
31. Steadman RG (1979) The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science. Journal of Applied Meteorology and Climatology

32. Urban A, Huber V, Henry S, Plaza NP, Tušlová L, Dasgupta S, Masselot P, Cvijanovic I, Mistry M, Pascal M, de'Donato F, Di Napoli C, Gosling SN, Kohnová S, Kyselý J, Lüthi S, Pau L-F, Ragettli MS, Ruuhela R, Ryti N, Silva SDNPD, Zemah-Shamir S, Thiery W, Vicedo-Cabrera A-M, Wieczorek J, Sera F, Armstrong B, Gasparrini A On behalf of the PROCLIAS TG 3.11, the MCC Collaborative Research Network, 2025. The effectiveness of heat prevention plans in reducing heat-related mortality across Europe. Environ Res Lett 20, 124071. https://doi.org/10.1088/1748-9326/ae2775
33. Vaidyanathan A, Saha S, Vicedo-Cabrera AM, Gasparrini A, Abdurehman N, Jordan R, Hawkins M, Hess J, Elixhauser A (2019) Assessment of extreme heat and hospitalizations to inform early warning systems. Proc. Natl. Acad. Sci. U.S.A. 116, 5420–5427. https://doi.org/10.1073/pnas.1806393116
34. Ver Hoef JM, Boveng PL, QUASI-POISSON VS (2007) NEGATIVE BINOMIAL REGRESSION: HOW SHOULD WE MODEL OVERDISPERSED COUNT DATA? Ecology 88, 2766–2772. https://doi.org/10.1890/07-0043.1
35. Vicedo-Cabrera AM, Scovronick N, Sera F, Royé D, Schneider R, Tobias A, Astrom C, Guo Y, Honda Y, Hondula DM, Abrutzky R, Tong S, Coelho MDSZS, Saldiva PHN, Lavigne E, Correa PM, Ortega NV, Kan H, Osorio S, Kyselý J, Urban A, Orru H, Indermitte E, Jaakkola JJK, Ryti N, Pascal M, Schneider A, Katsouyanni K, Samoli E, Mayvaneh F, Entezari A, Goodman P, Zeka A, Michelozzi P, de'Donato F, Hashizume M, Alahmad B, Diaz MH, Valencia CDLC, Overcenco A, Houthuijs D, Ameling C, Rao S, Di Ruscio F, Carrasco-Escobar G, Seposo X, Silva S, Madureira J, Holobaca IH, Fratianni S, Acquaotta F, Kim H, Lee W, Iniguez C, Forsberg B, Ragettli MS, Guo YLL, Chen BY, Li S, Armstrong B, Aleman A, Zanobetti A, Schwartz J, Dang TN, Dung DV, Gillett N, Haines A, Mengel M, Huber V, Gasparrini A (2021) The burden of heat-related mortality attributable to recent human-induced climate change. Nat Clim Chang 11:492–500. https://doi.org/10.1038/s41558-021-01058-x
36. Weinberger KR, Wu X, Sun S, Spangler KR, Nori-Sarma A, Schwartz J, Requia W, Sabath BM, Braun D, Zanobetti A, Dominici F, Wellenius GA (2021) Heat warnings, mortality, and hospital admissions among older adults in the United States. Environ Int 157:106834. https://doi.org/10.1016/j.envint.2021.106834
37. Zhang K, Rood RB, Michailidis G, Oswald EM, Schwartz JD, Zanobetti A, Ebi KL, O'Neill MS (2012) Comparing exposure metrics for classifying 'dangerous heat' in heat wave and health warning systems. Environ Int 46:23–29. https://doi.org/10.1016/j.envint.2012.05.001
38. Zhao Q, Guo Y, Ye T, Gasparrini A, Tong S, Overcenco A, Urban A, Schneider A, Entezari A, Vicedo-Cabrera AM, Zanobetti A, Analitis A, Zeka A, Tobias A, Nunes B, Alahmad B, Armstrong B, Forsberg B, Pan S-C, Íñiguez C, Ameling C, De La Valencia C, Åström C, Houthuijs C, Dung D, Royé DV, Indermitte D, Lavigne E, Mayvaneh E, Acquaotta F, de' Donato F, Di Ruscio F, Sera F, Carrasco-Escobar F, Kan G, Orru H, Kim H, Holobaca H, Kyselý I-H, Madureira J, Schwartz J, Jaakkola J, Katsouyanni JJK, Diaz KH, Ragettli M, Hashizume MS, Pascal M, M., Michelozzi P, Matus Correa P, Goodman P, Saldiva N, Abrutzky PH, Osorio R, Rao S, Fratianni S, Dang S, Colistro TN, Huber V, Lee V, Seposo W, Honda X, Guo Y, Bell YL, Li ML, S (2021) Global, regional, and national burden of mortality associated with non-optimal ambient temperatures from 2000 to 2019: a three-stage modelling study. Lancet Planet

Health 5:e415–e425 De Sousa Zanotti Stagliorio Coélho, M., Valdés Ortega, N., Ryti, N., Scovronick, N.,. https://doi.org/10.1016/S2542-5196(21)00081-4

Figures

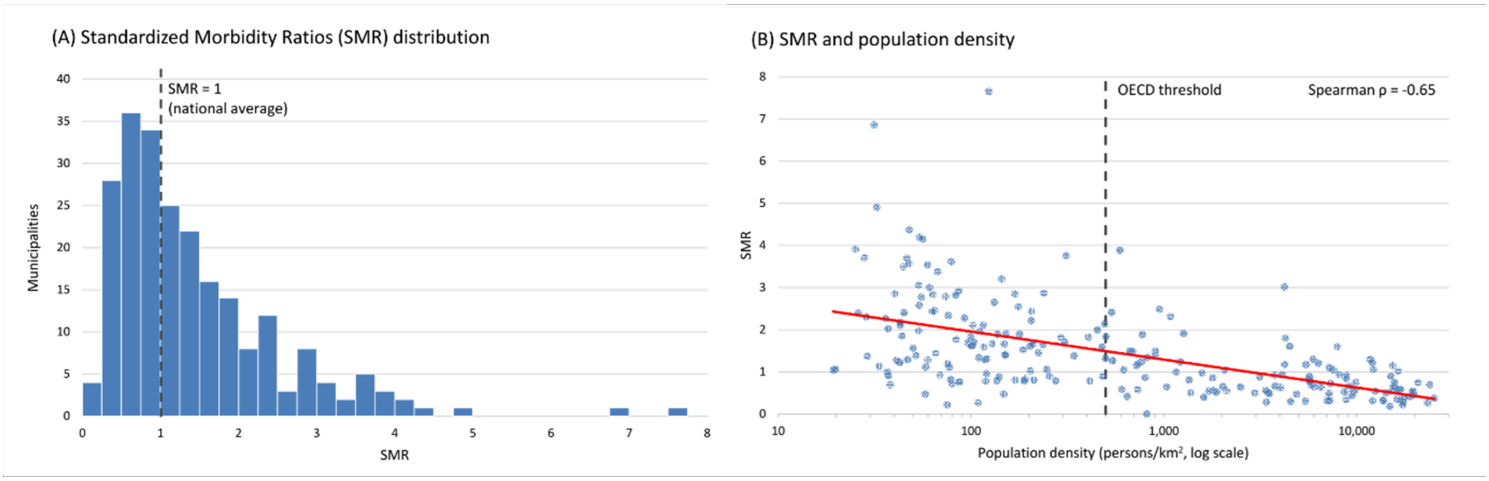


Figure 1

Age-standardized morbidity ratio. (A) Distribution across 230 municipalities, May–September 2019–2025; dashed line, national average. (B) SMR against population density (log scale).

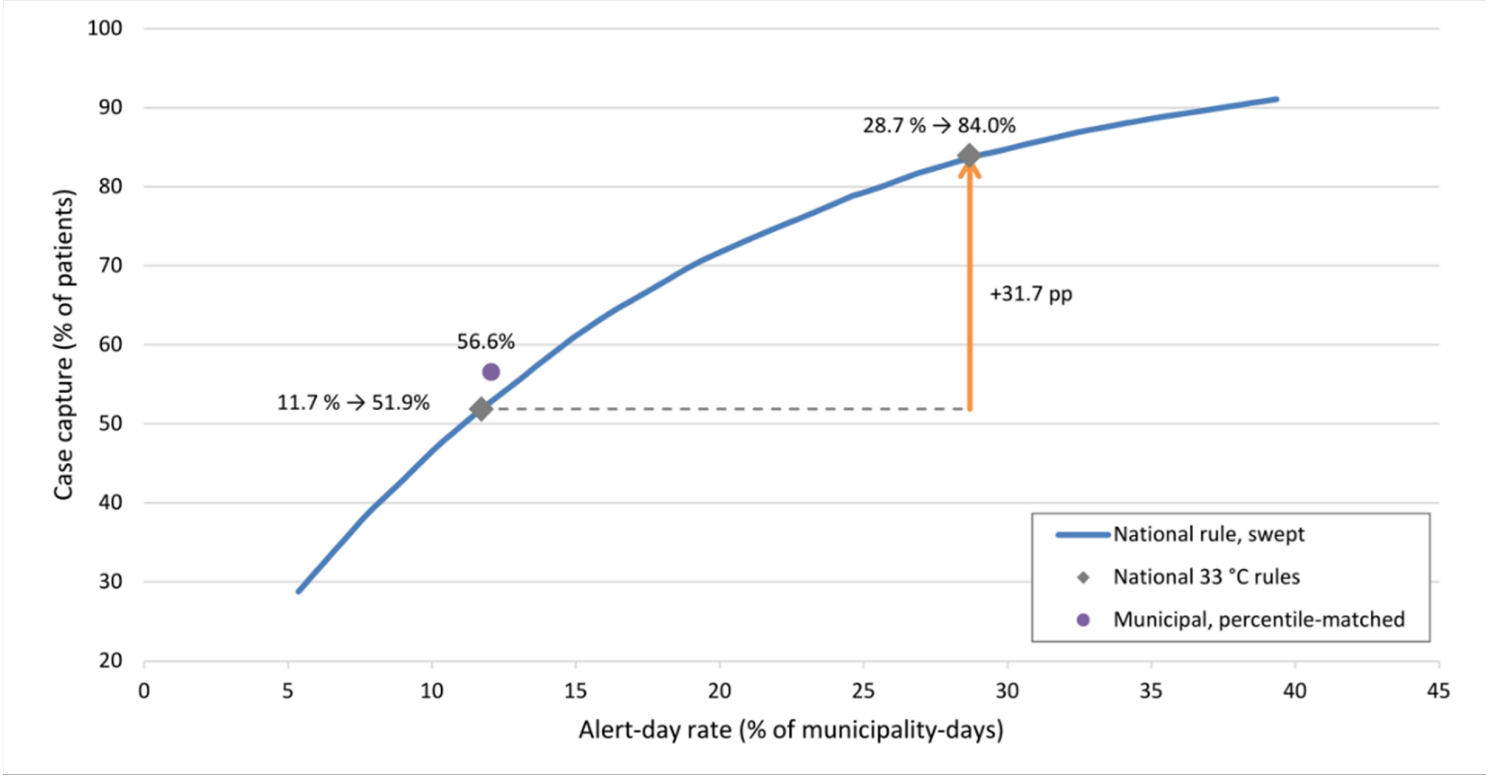


Figure 2

Burden–capture frontier.