Anticipating influenza-like illness outbreaks via syndromic surveillance using over-the-counter drug sales and primary health care data

Juliane F. Oliveira[1,2,3*], Thiago Cerqueira-Silva[4,5], Pedro Anibal N. Brito[6], Nívea B. da Silva[7], Rosemeire L. Fiaconne[7], Maria Célia S. L. Cunha[8], Gerson G. Cunha[8], Ray B. Miranda[8], Lucas M. Dias[8], Fábio M. H. S. Filho[1], Gerson O. Penna[9], Pablo Ivan P. Ramos[1,3], Izabel Marcilio[1,10], Manoel Barral-Netto[1,3,5,&], Viviane S. Boaventura[3,5,6,&*]

[1]Center for Data and Knowledge Integration for Health (CIDACS), Gonçalo Moniz Institute, Oswaldo Cruz Foundation (FIOCRUZ); Salvador, 41745-715, Brazil.
[2] Centre of Mathematics of the University of Porto (CMUP), Department of Mathematics, Porto, Portugal
[3] Digi-Saude National Institute of Science and Technology in Digital Health (DigiSaude-INCT)
[4] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom
[5] Medicine and Precision Public Health Laboratory (MeSP[2]), Gonçalo Moniz Institute, Oswaldo Cruz Foundation (FIOCRUZ); Salvador, 40296-710, Brazil.
[6] Faculty of Medicine, Federal University of Bahia, Salvador, Brazil
[7] Institute of Mathematics and Statistics of Federal University of Bahia, Salvador, Brazil.
[8] Luiz Coimbra Institute of Graduate and Engineering Research (COPPE), Federal University of Rio de Janeiro (UFRJ); Rio de Janeiro, 21941-859, Brazil.
[9] Tropical Medicine Centre, University of Brasília, Fiocruz School of Government; Brasília, 70910-900, Brazil.
[10] Bahiana School of Medicine and Public Health (EBMSP); Salvador, 40290-000, Brazil.

*Corresponding author:
juliane.oliveira@fiocruz.br (Juliane F Oliveira),
viviane.boaventura@fiocruz.br (Viviane S. Boaventura)

[&]These authors contributed equally to this work and are co-senior authors.

Abstract

Effective pandemic preparedness relies on integrating diverse data sources for early outbreak detection. This study assessed whether over-the-counter (OTC) drug sales and primary health care (PHC) data could anticipate surges in respiratory-disease–related hospitalizations in Brazil. From November 2022 to June 2025, we analysed weekly time-series across 510 regions using a negative binomial autoregressive model within Statistical Process Control techniques. OTC data anticipated 56.6% of 746 hospitalization surges 1–3 weeks in advance, detected 9.5% concurrently, and missed 33.9%. PHC data anticipated 59.5%, detected 10.3% concurrently, and missed 30.2%. PHC data showed higher sensitivity and specificity than OTC (69.8% vs. 66.1%, and 49.5% vs. 47.8%). Performance varied regionally, and in 76.7% of regions, at least one stream showed high precision. These findings support the value of OTC drug sales, alongside PHC data, as early indicators of hospitalization surges in respiratory illness surveillance.

Introduction

Early outbreak detection remains a critical challenge in public health[1–3]. Timely and accurate identification of unusual patterns that indicate the emergence or re-emergence of infectious diseases is essential for triggering rapid intervention, mitigating disease spread, and protecting vulnerable populations[4,5]. Traditional health surveillance refers to the systematic collection, analysis, and interpretation of structured health data, usually relying on predefined case definitions, for monitoring trends and estimating disease burden[6]. However, traditional surveillance methods generally present with timeliness and coverage limitations, and do not prioritize early warning[7]. To address this gap, syndromic surveillance is being increasingly used for early outbreak detection[2,8], enabling the joint use of traditional surveillance sources with alternative, health-related data streams collected and analysed in near- or real-time[9]. In this context, data from over-the-counter (OTC) drug sales and administrative data of primary health care (PHC) encounters have been listed among the data streams with reported capabilities to improve outbreak detection[10–13].

OTC drug sales data offer an accessible and real-time alternative for syndromic surveillance. They are widely available across countries, require minimal integration into the formal health systems, and capture consumer behaviour that often precedes health-seeking in formal medical settings[14]. Despite these attributes, OTC drug sales have been underused for early outbreak detection. Previous studies, mainly conducted in high-income countries [15–19], confirm the feasibility of using drug sales to identify respiratory disease outbreaks. However, the majority were published before 2015, and few have conducted real-time or prospective evaluations of OTC data. Key limitations include short observation periods, less than two years in most cases, which are not sufficient to assess seasonal patterns of disease outbreaks. Other reported limitations involve the absence of standardized drug classification systems, and restricted population or pharmacy coverage. Furthermore, few studies have systematically compared OTC data with other real-world, pre-hospital syndromic data sources such as PHC encounters.

Brazil presents a compelling case for advancing multi-source syndromic surveillance. As one of the world's largest and most populous countries, Brazil faces substantial heterogeneity in disease dynamics, healthcare access, and data infrastructure. The country's public Unified Health System (SUS) provides universal coverage while progressively incorporating digital health reporting. For instance, the recording of every publicly funded PHC encounter – that includes the date, municipality, and reason of encounter – is mandatory for resource allocation, resulting in a comprehensive and consistent national PHC information dataset[20]. In parallel, Brazil's extensive private pharmacy network, comprising over 100,000 stores nationwide, generates high-frequency OTC drug sales data across a broad range of geographic and demographic contexts, which can be leveraged for health surveillance in complement to the data collected at the primary care level[21].

This study leverages this unique setting to evaluate the potential of OTC drug sales and PHC encounter data for early detection of respiratory diseases-related hospitalization surges. We aim to: (i) detect warnings in each data stream using robust time-series modelling and anomaly detection methods; (ii) assess their capabilities to anticipate hospitalizations surges; and (iii) explore contextual factors, such as geographic characteristics, that may influence the early warning performance of these data streams.

Role of the funding source

The funders had no role in the study design, data collection, analysis, interpretation, writing of this report, or decision to submit the paper for publication.

Results

Data overview and descriptive analysis

Between November 2022 and June 2025, Brazil recorded 62,289,087 PHC encounters associated with ILI symptoms and 2,294,329 respiratory diseases-related hospitalizations, averaging 458,008 (95% CI: 432,605 – 483,411) PHC encounters and 16,870 (95% CI: 16,155 – 17,585) hospitalizations per week. Simultaneously, monitored pharmacies sold 689,297,990 drug units associated with treatment of ILI symptoms, with a mean weekly sale of 5,068,368 units (95% CI: 4,792,123 – 5,344,612).

At the national level, the number of PHC encounters, OTC drug sales and hospital admissions showed no significant trends in the study period. However, significant seasonal patterns were observed and maintained over the years, with all three data streams rising in March (epidemiological week [EW] 10). Records of PHC encounters and hospitalizations peaked in May (from EW 18 to 22), while OTC drug sales peaked later in July (from EW 27 to 31) (Supplementary Figure 2). However, when analysing trend and seasonality by immediate regions, distinct dynamics emerged based on the presence or absence of trends and seasonality (Supplementary Table 1).

Lagged cross-correlation analysis revealed strong temporal associations among surveillance streams across Brazilian immediate regions. Moderate to very strong correlations were observed between PHC encounters and OTC drug sales (78.6% of regions), PHC encounters and hospitalizations (94.7%), and OTC drug sales and hospitalizations (63.9%). For PHC-hospitalization pairs, 35.7% of regions showed synchronous correlations (lag 0), while 11.4% exhibited PHC encounters preceding hospitalizations by 1–3 weeks, while 47.7% showed the reverse pattern. Among OTC-hospitalization pairs, 16.3% were synchronous (lag 0), with OTC drug sales preceding hospitalizations by 1–3 weeks in 32.0% of regions and hospitalizations preceding OTC sales in 15.6%. PHC-OTC correlations were synchronous in 22.2% of regions, with OTC drug sales leading PHC encounters in 47.1% and PHC encounters leading OTC sales in 9.5%, both considering 1 to 3 weeks anticipation (Supplementary Table 2).

Lastly, to compare the spatial distribution of these counts across immediate regions, Figure 1 shows the number of cases per 100,000 inhabitants for PHC encounters, OTC drug sales, and hospitalizations over the study period (Supplementary Table 3).

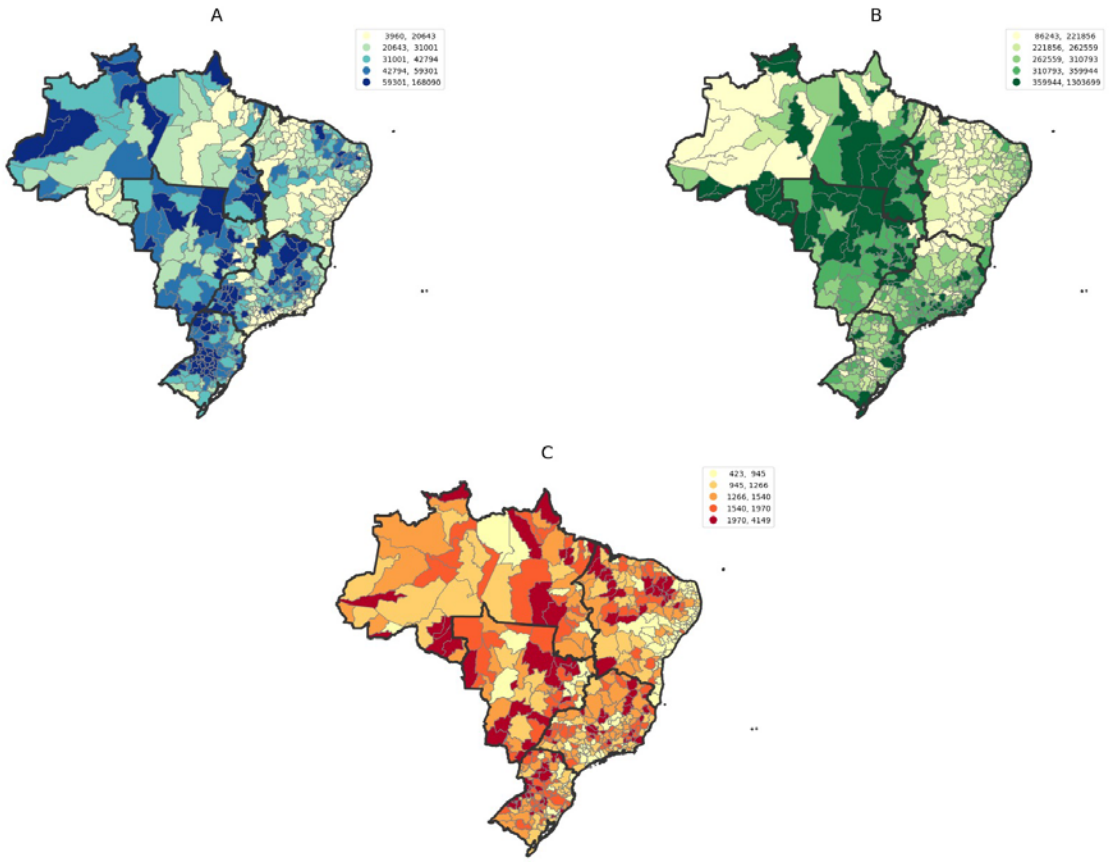


Figure 1: Spatial distribution of (A) ILI-related PHC encounters, (B) OTC drug units sold and (C) respiratory diseases-related hospitalizations across Brazil's immediate regions from November 2022 to June 2025. To enhance comparability, the indicators are presented as rates per 100,000 inhabitants.

There were strong to very strong correlations between regional rates of PHC encounters ($\rho = 0.704$, $p < 0.001$) and OTC drug sales ($\rho = 0.81$, $p < 0.001$) with hospitalizations. Correlations were particularly high in the South and Center-West regions. For PHC encounters, the correlations reached $\rho = 0.836$ ($p < 0.001$) in the South and $\rho = 0.810$ ($p < 0.001$) in the Center-West. Similarly, for OTC drug sales, the correlations were $\rho = 0.873$ ($p < 0.001$) and $\rho = 0.809$ ($p < 0.001$), respectively. OTC drug sales also showed very strong correlation in the North ($\rho = 0.869$, $p < 0.001$), whereas PHC encounters exhibited the weakest correlation in that region ($\rho = 0.618$, $p < 0.001$) (Supplementary Table 4).

Overall, we observed a strong correlation between PHC encounters and OTC drug sales ($\rho = 0.646$, p-value < 0.001), with exception of the Southeast regions that showed a moderate correlation ($\rho = 0.462$, p-value < 0.001).

Performance of anomaly detection

In the study period, 746 respiratory diseases-related hospitalization surges were identified in 469 (92.0%) of 510 immediate regions, of which 41 (8.0%) showed no hospitalization surges. Accordingly, by applying the NegBi-SPC model to each time series, we identified 13,106 warnings in PHC encounters and 13,066 in OTC drug sales. Of these, 10,442 (PHC) and 10,282 (OTC) occurred

in regions that experienced at least one hospitalization surge (Supplementary Note 3 and Supplementary Data).

Warnings/surges emerged in February (EW 7) for PHC encounters and OTC sales, and in March (EW 10) for hospitalizations, aligning with expected nationwide seasonal patterns, as shown before. Notably, we also observed an increase in warnings starting in August (EW 32) for PHC encounters, followed by a similar pattern in OTC drug sales in September (EW 36) (Figure 2, Supplementary Table 5). However, we identified only five hospitalizations surges from August to December of 2023 and 12 surges from August to December of 2024.

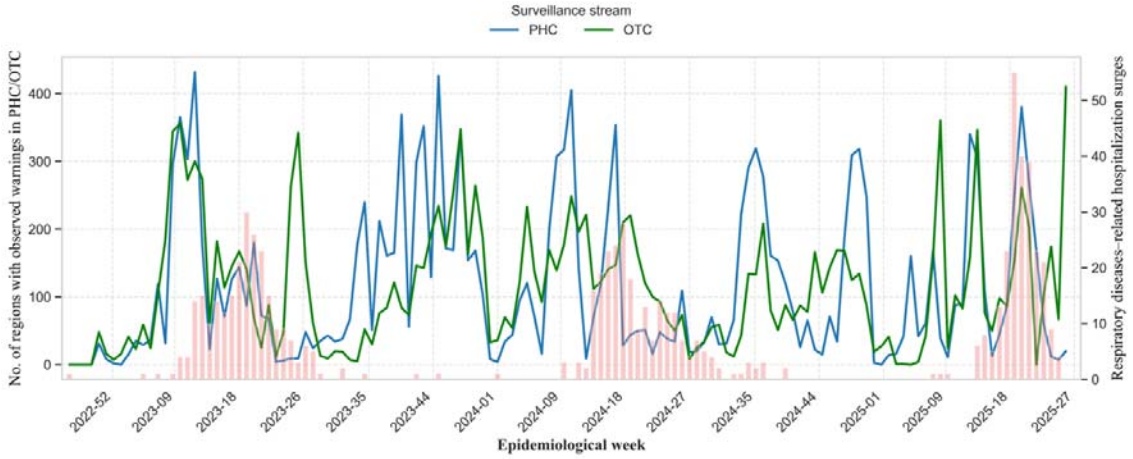


Figure 2: Number of regions with observed warnings in the PHC and OTC time series and/or hospitalization surges, per week, from November 2022 to June 2025. Counts of immediate regions with hospitalization surges in red, and warnings in OTC (green) and PHC (blue).

Figure 3 highlights the comparative timeliness performance of PHC and OTC data in anticipating hospitalization surges. Assessing the OTC time series, we could detect 56.6% (422/746) of hospitalization warnings 1 to 3 weeks in advance, compared with 59.5% (444/746) detected by the PHC encounters data. This represents 22 additional hospitalization surges anticipated by monitoring PHC series in relation to OTC alone. In addition, the PHC series had a slightly higher detection proportion of hospitalization surges within the same week: 10.3% (77/746) versus 9.5% (71/746). Also, the proportion of missed hospitalization surges was lower for PHC (30.2% (225/746)) than for OTC (33.9% (253/746)).

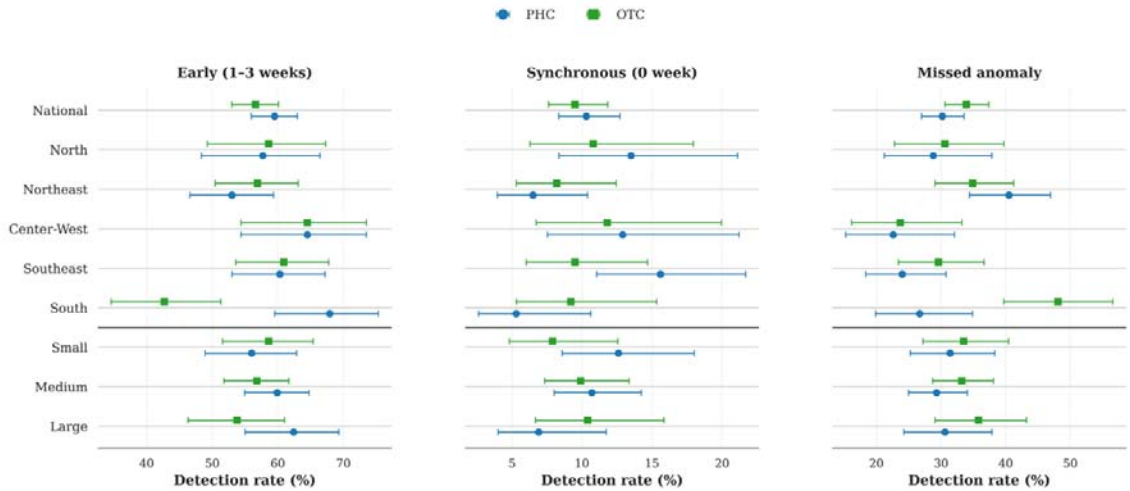


Figure 3: Timeliness of early warnings detected in PHC and OTC time series in relation to hospitalization surges at the national and regional levels. This figure summarizes the proportion of hospitalization surges anticipated (1 to 3 weeks early), detected in the same week (synchronous), or missed based on PHC and OTC series, along with their corresponding 95% confidence intervals (CIs). Results are presented for the national level, geographic region, and by population size.

The timeliness of PHC-based warnings varied across regions and population size. The Center-West region exhibited the highest early detection rates, with 64.5% (60/93) of warnings issued 1 to 3 weeks in advance to hospitalization surges. In contrast, the Northeast had slightly lower PHC timeliness, with 53.0% (123/232) issued 1 to 3 weeks in advance and 40.5% (94/232) of hospitalizations surges missed. Performance also varied by population size, with optimal early detection (1–3 weeks anticipation) in large-sized regions 62.4% (108/173) while small population regions showed reduced performance, with early anticipation rates of 56.0% (107/191) for PHC data. For OTC-based warnings the differences across regions between municipalities with different population sizes were less than 5% (Figure 4), with exception of South and large-sized regions.

By assessing sensitivity and probability of detection, we verified that PHC encounters data (Se = 69.8% and POD = 75.7%) slightly outperformed OTC drug sales data (Se = 66.1% and POD = 74.4%), indicating improved early warning capabilities during the study period. OTC data also showed a higher false positive rate (FPR = 52.2% for OTC vs FPR = 50.5% for PHC) and lower precision (Pr = 48.0% for OTC vs Pr = 49.8% for PHC) across all regions, suggesting potential over-detection (Figure 4).

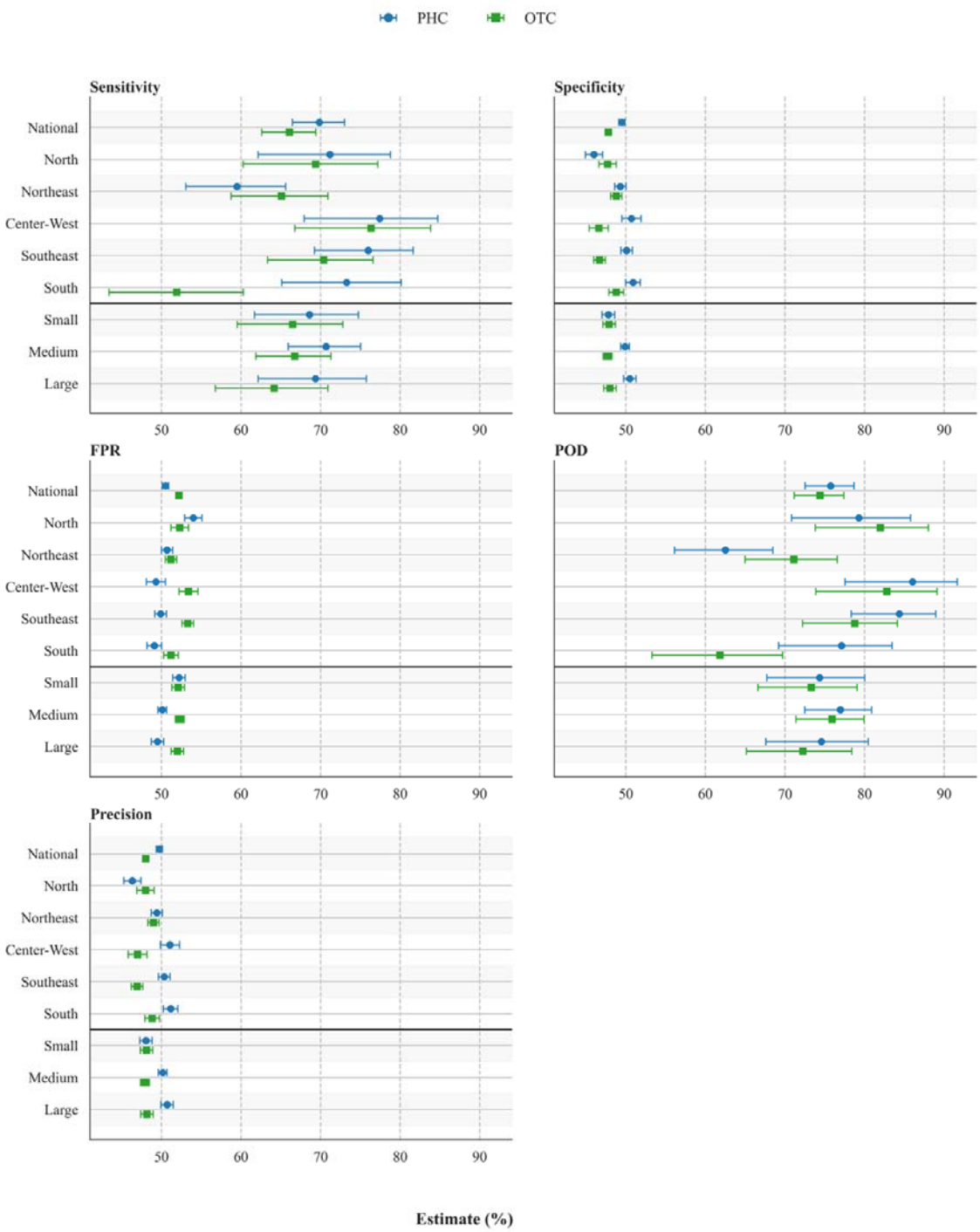


Figure 4: Performance metrics for PHC encounters and OTC drug sales data in detecting hospitalization surges across regions and population sizes. The figure presents sensitivity, specificity, probability of detection (POD), false positive rate (FPR), and precision for both PHC and OTC time series, along with their corresponding 95% CIs.

Both PHC encounters and OTC drug sales data streams exhibited better performance in the Center-West, achieving the highest Se (77.4% vs 76.3%) and POD (86.0% vs 82.8%) among all regions. Overall, PHC data generally outperformed OTC across most regions during the study period, with sensitivity ranging from 59.5% (Northeast) to 77.4% (Center-West) and POD between 62.5%

(Northeast) to 86.0% (Center-West). The Northeast (Se = 59.5% and POD = 62.5% for PHC) and the South (Se = 51.9% and POD = 61.8% for OTC) showed the weakest performance for the PHC and OTC streams, respectively.

Regarding population size, medium-sized cities achieved the best overall performance for both PHC and OTC (Se = 70.7% for PHC, Se = 66.8% for OTC). In contrast, PHC data showed lower performance in small population regions (Se = 68.6% and FPR = 52.2%), while OTC performance declined in large population areas (Se = 64.2% and FPR = 52.0%).

Complementary value of using OTC and PHC

Overall, 143 (28.0%) out of 510 regions demonstrated high precision in both PHC (Pr ≥ 50.0%) and OTC (Pr ≥ 47.8%) data streams. OTC data alone outperformed PHC in 125 (24.5%) regions, while PHC performed better in 123 (24.1%) regions. In contrast, both data sources exhibited low precision or no data in 119 (23.3%) regions.

Regarding the timeliness score, both PHC and OTC data showed high anticipation in 202 (43.1%) regions. PHC outperformed OTC in 101 (21.5%) regions, whereas OTC had higher timeliness in 74 (15.8%) regions. In 133 (28.4%) regions, both data sources showed low anticipation or no data. Due to the absence of hospitalization surges, precision and timeliness could not be evaluated in 41 (8.0%) regions, represented in our results as no dada category (Figure 5).

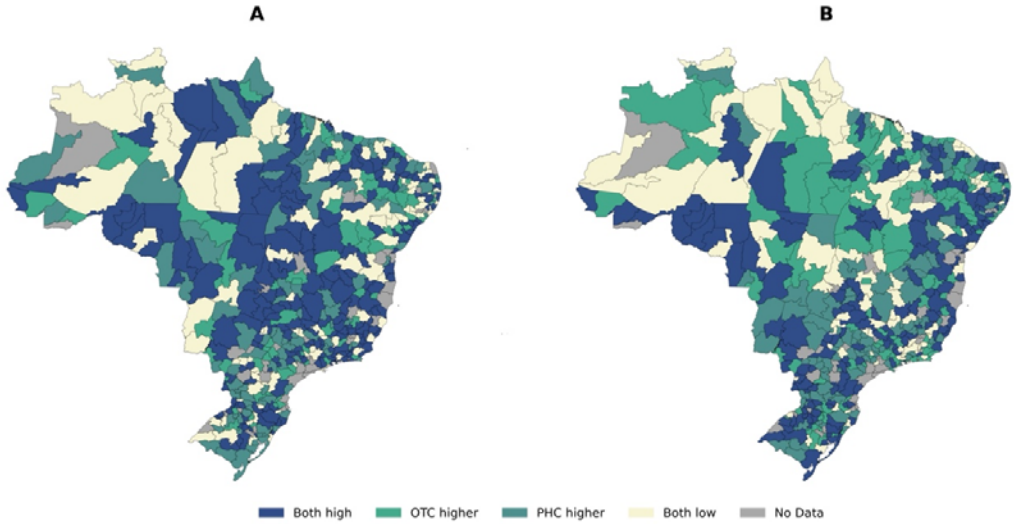


Figure 5: Classification of regions based on A) anticipation and B) precision indicators using PHC encounter and OTC drug sales data. A region was classified as “High” in anticipation, if anticipation scores were ≥ 60% for both PHC encounter and OTC drug sales data. For precision, regions with values ≥ 50.0% for PHC and ≥ 47.8% for OTC (median thresholds) were also considered “High”.

Effect of alternative anomaly definition in hospitalizations time-series

We assessed the robustness of our findings by modifying the threshold criteria used to define hospitalization surges (Supplementary Note 2). Overall, the findings remain broadly consistent under the new definition. However, some shifts are observed, particularly in the timeliness patterns by population size, and in performance metrics when analysed by both macro-regions and population

strata. Notably, the fundamental value of using PHC and OTC data for outbreak anticipation, individually or in conjunction, persisted across threshold definitions.

Discussion

This study provides novel evidence that ILI-associated OTC drug sales can serve as early indicators of surges of respiratory diseases-related hospitalizations, providing 1–3 weeks of advance warning. We also observed that integrating ILI-associated OTC sales and PHC encounter data enhances the system's capacity to anticipate these hospitalization surges. The use of OTC and PHC data show a good sensitivity performance and provide early detection of respiratory diseases-related hospitalization surges, but at the cost of increased false alarms. The core conclusion that PHC and OTC data are valuable for anticipating outbreaks held true regardless of how the threshold of hospitalization surges was defined. This complementary use of data streams with distinct operational strengths enables balanced surveillance systems capable of both timeliness detection and reliable warning, supporting more effective public health responses to emerging respiratory outbreaks.

While syndromic surveillance using multiple data sources with varied methodologies may enhance ILI outbreak trend detection, few studies have directly compared their effectiveness. In France, previous research suggested that drug sales were more sensitive and provided timelier signals than sentinel surveillance systems[17]. The high performance of OTC sales data observed in our analysis in a large, continental country like Brazil supports its potential use in areas where PHC data are unavailable. In these two countries OTC sales data proved highly effective as an independent surveillance tool. Previous studies in high-income countries also supported the utility of OTC drug sales for syndromic surveillance while our results extend this evidence to a low- and middle-income countries (LMIC) context[2,16,22–24]. As OTC data are routinely collected for commercial purposes, they usually provide high coverage of pharmaceutical sales, offering a representative proxy of medication consumption nationwide and creating no additional burden on the health system[2,9]. This is especially vital for rapidly extending surveillance in regions with limited public health infrastructure. These areas, often disproportionately affected by respiratory disease outbreaks due to fragile health systems, could benefit significantly from an early-warning system that requires no additional data collection burden. Using OTC sale for syndromic surveillance offers scalable, cost-effective solutions aligned with WHO pandemic preparedness goals, as set forth by the recently agreed proposal for a Pandemic Agreement[25]. Integrating such information from multiple countries could bridge surveillance gaps, enhancing early detection and targeted responses for respiratory infections and other communicable diseases.

Our findings indicate strong spatial and temporal alignment between OTC drug sales, PHC encounters, and hospitalizations, confirming that these routinely collected data sources capture epidemiologically meaningful disease fluctuations in Brazil. The warnings in PHC and OTC data also captured regular increases in ILI activity not associated with rise in hospitalizations. While non-confirmatory, concurrent increases in both medical visit patterns and self-medication behaviour suggests the occurrence of real, albeit mild, outbreaks. These increases in warnings from PHC encounters, which are sometimes preceded or followed by rises in warnings from OTC drug sales, reflect the fact that ILI is predominantly caused by self-limiting infections commonly managed with OTC medications. As a result, individuals often self-manage symptoms using OTC drugs and, consequently, do not seek additional care[26]. Detecting these early and less severe events is also important for guiding public health actions, as they may still place significant demand on primary care and serve as early indicators of emerging threats[27,28]. Interestingly, during out-of-season surges of

respiratory diseases-related hospitalization periods, PHC warnings tend to precede those from OTC data, suggesting that OTC data are more likely to anticipate the onset of severe cases rather than the earliest signals of mild illness in the community. These findings underscore the need to evaluate the surveillance potential of these indicators using additional data sources, such as emergency unit admissions, to fully characterize the system's performance across the disease severity spectrum. Finally, we also observe spikes in OTC warnings following hospitalization surges. One explanation for these spikes is increased drug consumption driven by promotional campaigns carried out by the pharmaceutical industry [29].

PHC and OTC data showed variation across Brazilian regions and by population strata, particularly in terms of timelines, sensitivity and probability of detection. This variability implies that the effectiveness of PHC and OTC data for early outbreak detection may require locally-referred calibration to ensure consistent performance. This finding is in alignment with our previous study[11], which suggested the performance of PHC-based surveillance is context-dependent, potentially influenced by disparities in health system access and reporting practices across regions of different sizes. However, for both OTC and PHC data, the POD was lower in more populous regions. This is probable because in these urban settings early signs of illness may be diluted making it more difficult to detect surges[12]. These results suggest the importance of enhancing data granularity to improve early detection performance.

In practice, daily decision-making requires clarity on which data stream (PHC or OTC) offers more reliable early warning signals. The only study to directly explore this question, conducted by Najmi et al.[30], suggests that analysing each data stream independently performs better for detecting early signs of ILI than combining both in a unified bivariate model. Here, we derived region-specific indicators that assess both timeliness and precision to guide local health managers in tailoring surveillance strategies. Our findings highlight Brazil's highly heterogeneous landscape, where some regions benefit from both data sources, while others display complementary strengths.

Our study has strengths and limitations. More than 60% of surges in respiratory diseases-related hospitalizations were detected using PHC and OTC data—remarkable results considering the use of real-world, highly volatile time series in a setting without an integrated surveillance system. This performance was achieved through a national, integrative analysis in a continentally-sized LMIC, using routinely collected data with broad geographic coverage. However, while hospitalizations were used as the gold standard, this may underestimate outbreaks caused by highly transmissible but less severe pathogens. We were also unable to evaluate the performance of PHC and OTC data in 8.0% of regions due to the absence of hospitalization surges indicative of outbreak periods. A more comprehensive evaluation requires standardized, granular outbreak definitions from health authorities.

In conclusion, this work demonstrated that integrating OTC sales with primary-care encounters sharpens early-warning capability. These findings support the use of OTC data as a robust, scalable, low-burden alternative for strengthening early warning systems in resource-limited settings and for broader application across countries aiming to improve epidemic preparedness. Longer series and pathogen-specific analyses (e.g., arboviruses, diarrhoeal diseases) will clarify how generalisable this approach is beyond respiratory syndromes.

Methods

Data sources

We analysed weekly time series data on OTC drug sales and PHC encounters, aggregated across 510 immediate regions in Brazil. Each region represents groupings of municipalities strongly connected to local urban centres, serving population needs for goods, employment, health services, and education (Supplementary Figure 1), as defined by the Brazilian Institute of Geography and Statistics (IBGE). The dataset covers the period from November 2022 to June 2025.

We used OTC sales data from IQVIA, a global health data and technology company whose coverage exceeds 95% of the Brazilian pharmaceutical retail market, providing a robust source for analysing medication purchasing trends[21,31]. The data includes the weekly number of medication units sold, covering 77 active substances and 54 drug classes categorized according to the Anatomical Therapeutic Chemical (ATC4) classification system. We filtered the data to include only drug classes frequently used for Influenza-like illness (ILI) treatment and applied a cluster analysis to group medications with similar temporal patterns, in accordance with known ILI OTC drug sales peaks in state capitals. Based on this analysis, six drugs were selected for surveillance: Ascorbic Acid, Chlorphenamine Maleate, Phenylephrine, Paracetamol, Loratadine, and Benzydamine Hydrochloride (Supplementary Note 1).

PHC data were extracted from the Brazilian Health Information System for Primary Care (SISAB), a database that registers data on all publicly-funded PHC encounters in Brazil, coded by either the International Classification of Diseases (ICD-10) or the International Classification of Primary Care (ICPC-2). ILI related encounters were filtered to construct an ILI-specific time series following the approach of Cerqueira-Silva et al [11].

To assess the ability of these series to anticipate the burden of respiratory diseases-related hospitalizations, we extracted hospitalizations cases, as per the ICD-10 (Chapter X - Diseases of the respiratory system, code range J00-J99) from the Brazilian Hospital Information System (SIH), covering the period from November 2022 to June 2025. Hospitalization data was used as the gold standard for evaluating the performance of each data stream.

We also used 2024 IBGE population estimates to classify regions by population size: small (≤ 117,789 inhabitants), medium (117,790 – 324,768 inhabitants), and large (> 324,768 inhabitants).

Statistical Model

To evaluate the performance of OTC drug sales and PHC encounter data in providing early warnings of surges in respiratory disease–related hospitalization, we employed a series of negative binomial autoregressive models that best predict each time-series for each of the 510 immediate regions in Brazil. Negative binomial models are well suited for counting data that exhibit overdispersion, which is a key feature of our time series. Using the estimated values obtained by the best fitted model, we applied the Statistical Process Control technique to establish epidemic thresholds that are used to flag anomalous values in each time series[32,33]. We name the procedure as NegBi-SPC.

To define the NegBi-SPC procedure, let y_t for $t = 1, \ldots, n$, denote the weekly (four windows) moving average of counts of OTC or PHC encounters, modeled as a negative binomial distribution. The NegBi-SPC model is defined in equation (1):

$$log(\mu_t) = \beta_0 + \sum_{i=1}^{3} \beta_i y_{t-i} + \gamma t + \sum_{k=1}^{K} \left(a_k \cos \frac{2k\pi t}{N} + b_k \sin \frac{2k\pi t}{N} \right) \qquad (1)$$

where μ_t is the mean at time t, β_0 is the intercept, β_i are the coefficients for the autoregressive terms varying from $i = 1$ to 3. The coefficient γ captures the long-term trend in the series, indicating whether the mean level is increasing or decreasing over the study period. The summation of sine and cosine terms represents the seasonal component (a periodic pattern that repeats at regular intervals within the time series) with period N = 52 weeks, $\{a_k, b_k\}$ Fourier coefficients for the k-th harmonic, and K denotes the maximum number of harmonics included to adequately capture the seasonal structure of the time series.

Prior to including trend and seasonality terms, we first fit a negative binomial regression model including only the trend. We then assessed its statistical significance (p-value < 0.05) as a criterion for trend term inclusion in subsequent models. The seasonality component was evaluated using the Friedman test and included if found significant. When the test indicated significant seasonality, we applied the scipy.fft module to decompose the series and extract the relevant K Fourier coefficients for inclusion into the model. For each time series, we tested multiple model specifications, including combinations of trend, seasonality, and lagged terms. Models were estimated via maximum likelihood estimation, and the best-fitting model was selected based on the Akaike Information Criterion (AIC).

After model selection, lower control limits (LCL) and upper control limits (UCL) were calculated to flag anomalous values as described by equations (2) and (3):

$$LCL = \hat{y}_t - \omega \sigma_{\hat{y}_t} \qquad (2)$$
$$UCL = \hat{y}_t + \omega \sigma_{\hat{y}_t} \qquad (3)$$

where $\hat{y}_t$ is the value estimated by the model at time t, $\sigma_{\hat{y}_t}$ is the standard deviation estimate of the response predicted, and ω represents the distance between the control limits. Control limits is a statistical indicator on control charts to monitor the expected range of variation of the process. If data points fall outside the control limits or show unusual patterns, it signals that the process may need attention or adjustment. Throughout the text, we refer to anomalies found in OTC or PHC time series as warnings.

Performances of early detection of respiratory diseases-related hospitalization surges

To evaluate the performance of early outbreak detection, anomalies in respiratory diseases-related hospitalizations were defined as weekly counts exceeding the historical median from November 2022 to June 2025, plus two standard deviations. An isolated anomaly was defined as a single week with an anomaly, which was not followed or preceded by any other anomaly within a period of three weeks. Isolated anomalies were excluded, as they are more likely to represent transient fluctuations than a consistent signal of an increase in hospitalization. On the other hand, consecutive anomalies were defined as two or more anomalous weeks occurring consecutively or separated by only one non-anomaly week. Consecutive anomalies were grouped together marking periods of sustained respiratory diseases-related hospitalization. The starting weeks of these periods are named as

hospitalization surges. To assess the robustness of this definition, we added one more alternative criterion for identifying surges in the hospitalization time series in Supplementary Note 2.

We then assessed the performance of OTC and PHC time series in anticipating hospitalization surges using the following metrics[32]: Probability of Detection (POD), defined as the probability that a warning is generated at least once between the start and the end of an hospitalization surge; False Positive Rate (FPR), as the proportion of weeks corresponding to an alarm in the absence of a surge; Sensitivity (Se) and Specificity (Sp) are defined as Se = TP/(TP+FN) and Sp = TN/(TN+FP), where True Positives (TP) are warnings associated with hospitalization surge, True Negatives (TN) are weeks without warnings and hospitalization surges, False Negatives (FN) are weeks with hospitalization surge and without associated warnings, False Positives (FP) are weeks without hospitalization surges and with warnings. Lastly, we defined Precision (Pr) as Pr = (TP + TN) / (TP + TN + FP + FN). Immediate regions that did not present hospitalization surges during the study period were not included in this analysis.

Lastly, we conducted descriptive analyses to complement our main results. To explore how the proportions of PHC encounters, OTC drug sales, and hospitalizations vary across regions, we calculated Spearman correlation coefficients. Additionally, to examine the temporal relationships between each pair of time series within each immediate region, we computed lagged Spearman cross-correlations. The strength of correlation was classified as follows: very weak (ρ = 0.00–0.19), weak (ρ = 0.20–0.39), moderate (ρ = 0.40–0.59), strong (ρ = 0.60–0.79), and very strong (ρ = 0.80–1.00).

Complementary value of using OTC and PHC

To assess the complementary value of OTC and PHC data sources for detecting hospitalization surges, we calculated two metrics for each immediate region: precision, as previously described; and a timeliness score, defined as the proportion of warnings occurring either during or up to three weeks prior to a hospitalization surge, relative to the total number of hospitalization surges in that region. A region was classified as having high precision if its precision value was above the median for all regions, and as having high timeliness performance if its timeliness score was higher than 60%.

Acknowledgments
This study is part of the Alert-Early System of Outbreaks with Pandemic Potential (ÆSOP, http://aesop.health), an initiative under development by Brazil's Fundação Oswaldo Cruz (Fiocruz) and the Federal University of Rio de Janeiro with financial support from the Rockefeller Foundation's Health Initiative (Grant 2023-PPI-007 awarded to MB-N). MB-N, PIPR, and VB are CNPq Research Fellows. TC-S acknowledges funding from the Royal Society (NIF\R1\231435). The excellent project management and administrative support from Fundação de Apoio à Fiocruz (FIOTEC) is greatly appreciated. We acknowledge Meiruze Sousa Freitas for the strategic support provided during the development of this work. We acknowledge the great support of ACESSA (https://acessa.org.br/) and IQVIA for procuring and granting access to drug sales data.

Data Availability
The data that support the findings of this study are available from the Brazilian Ministry of Health and IQVIA but restrictions apply to the availability of these data, which were used under licence for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Brazilian Ministry of Health and IQVIA.

Code Availability

All generated data and codes to ensure reproducibility of our results are available in our GitHub repository[34].

Declaration of interests

Authors declare that they have no financial or personal relationships that could be construed as competing interests.

Author contributions

Conceptualization: JFO, MB-N, VB; Methodology: JFO, TC-S, PANB, NBS, RLF, FMHSF, VB; Investigation: JFO, TC-S, PANB, GOP, IM, PIPR; Visualization: JFO, MCSLC, GGC, RBM, LMD, PIPR; Funding acquisition: MB-N; Project administration: IM, MB-N, PIPR; Supervision: MB-N; Writing – original draft: JFO, VB; Writing – review & editing: all authors.

References

1. Morgan, O. W. et al. Disease surveillance for the COVID-19 era: time for bold changes. Lancet 397, 2317–2319 (2021).

2. Ramos, P. I. P. et al. Combining digital and molecular approaches using health and alternate data sources in a next-generation surveillance system for anticipating outbreaks of pandemic potential. JMIR Public Health Surveill. 10, e47673 (2024).

3. Al Knawy, B. et al. The Riyadh Declaration: the role of digital health in fighting pandemics. Lancet 396, 1537–1539 (2020).

4. Oliveira, J. F. et al. Human mobility patterns in Brazil to inform sampling sites for early pathogen detection and routes of spread: a network modelling and validation study. Lancet Digit. Health 6, e570–e579 (2024).

5. Oliveira, J. F. et al. Balancing human mobility and health care coverage in sentinel surveillance of Brazilian Indigenous areas: Mathematical optimization approach. JMIR Public Health Surveill. 11, e69048 (2025).

6. Teutsch, S. M. & Thacker, S. B. Planning a public health surveillance system. Epidemiol. Bull. 16, 1–6 (1995).

7. Ganser, I., Thiébaut, R. & Buckeridge, D. L. Global variations in event-based surveillance for disease outbreak detection: Time series analysis. JMIR Public Health Surveill. 8, e36211 (2022).

8. Smith, G. E. et al. Syndromic surveillance: two decades experience of sustainable systems - its people not just data! Epidemiol. Infect. 147, e101 (2019).

9. Triple S Project. Assessment of syndromic surveillance in Europe. Lancet 378, 1833–1834 (2011).

10. Sočan, M., Erčulj, V. & Lajovic, J. Early detection of influenza-like illness through medication sales. Cent Eur J Public Health 20, 156–162 (2012).

11. Cerqueira-Silva, T. et al. Early warning system using primary health care data in the post-COVID-19 pandemic era: Brazil nationwide case-study. Cad. Saude Publica 40, e00010024 (2024).

12. Todd, S., Diggle, P. J., White, P. J., Fearne, A. & Read, J. M. The spatiotemporal association of non-prescription retail sales with cases during the 2009 influenza pandemic in Great Britain. BMJ Open 4, e004869 (2014).

13. Dong, X., Boulton, M. L., Carlson, B., Montgomery, J. P. & Wells, E. V. Syndromic surveillance for influenza in Tianjin, China: 2013–14. J. Public Health (Oxf.) fdw022 (2016).

14. Pivette, M., Mueller, J. E., Crépey, P. & Bar-Hen, A. Drug sales data analysis for outbreak detection of infectious diseases: a systematic literature review. BMC Infect. Dis. 14, 604 (2014).

15. Sugawara T. et al. An experimental study for syndromic surveillance using OTC sales. Kansenshogaku Zasshi 81, 235–241 (2007).
16. Ohkusa, Y., Shigematsu, M., Taniguchi, K. & Okabe, N. Experimental surveillance using data on sales of over-the-counter medications--Japan, November 2003-April 2004. MMWR Suppl. 54, 47–52 (2005).
17. Vergu, E. et al. Medication sales and syndromic surveillance, France. Emerg. Infect. Dis. 12, 416–421 (2006).
18. Muchaal, P. K., Parker, S., Meganath, K., Landry, L. & Aramini, J. Evaluation of a national pharmacy-based syndromic surveillance system. Can. Commun. Dis. Rep. 41, 203–206 (2015).
19. Magruder, S. Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of human disease. Johns Hopkins Apl Technical Digest 24, 349–353 (2003).
20. Kovacs, R. et al. Socioeconomic inequalities in the quality of primary care under Brazil's national pay-for-performance programme: a longitudinal study of family health teams. Lancet Glob. Health 9, e331–e339 (2021).
21. IQVIA Brasil. https://www.iqvia.com/pt-br/locations/brazil.
22. Davies, G. R. & Finch, R. G. Sales of over-the-counter remedies as an early warning system for winter bed crises. Clin. Microbiol. Infect. 9, 858–863 (2003).
23. Socan, M., Erculj, V. & Lajovic, J. Early detection of influenza like illness through medication sales. Cent. Eur. J. Public Health 20, 156–162 (2012).
24. Zhang, Y. et al. Influenza surveillance and incidence in a rural area in China during the 2009/2010 influenza pandemic. PLoS One 9, e115347 (2014).
25. WHO Member States conclude negotiations and make significant progress on draft pandemic agreement. https://www.who.int/news/item/16-04-2025-who-member-states-conclude-negotiations-and-make-significant-progress-on-draft-pandemic-agreement.
26. Peters, S. et al. Managing self-limiting respiratory tract infections: a qualitative study of the usefulness of the delayed prescribing strategy. Br. J. Gen. Pract. 61, e579-89 (2011).
27. Campos, G. S., Bandeira, A. C. & Sardi, S. I. Zika virus outbreak, Bahia, Brazil. Emerg. Infect. Dis. 21, 1885–1886 (2015).
28. Oliveira, J. F. et al. Interdependence between confirmed and discarded cases of dengue, chikungunya and Zika viruses in Brazil: A multivariate time-series analysis. PLoS One 15, e0228347 (2020).
29. Fulone, I. et al. Pharmaceutical policies: effects of policies regulating drug marketing. Cochrane Database Syst. Rev. 6, CD013780 (2023).
30. Najmi, A.-H. & Magruder, S. F. An adaptive prediction and detection algorithm for multistream syndromic surveillance. BMC Med. Inform. Decis. Mak. 5, 33 (2005).
31. Where Brazilians go to shop. https://www.iqvia.com/blogs/2017/11/where-brazilians-go-to-shop.
32. Bédubourg, G. & Le Strat, Y. Evaluation and comparison of statistical methods for early temporal detection of outbreaks: A simulation-based study. PLoS One 12, e0181227 (2017).
33. Sena, J. G. S. de, Ferreira, P. H. & Fiaccone, R. L. O Controle Estatístico de Processos como Ferramenta para Controle e Prevenção de Epidemias de Malária na Região da Amazônia Legal. Braz. Jour. Biom. 40, (2022).
34. Preemptive_Surveillance_PHC_OTC: Preemptive Surveillance: Using Anomaly Detection in Primary Health Care and Drug Data to Track Influenza-Like Illness in Brazil. (Github).

Figure 1: Spatial distribution of (A) ILI-related PHC encounters, (B) OTC drug units sold and (C) respiratory diseases-related hospitalizations across Brazil's immediate regions from November 2022 to June 2025. To enhance comparability, the indicators are presented as rates per 100,000 inhabitants.

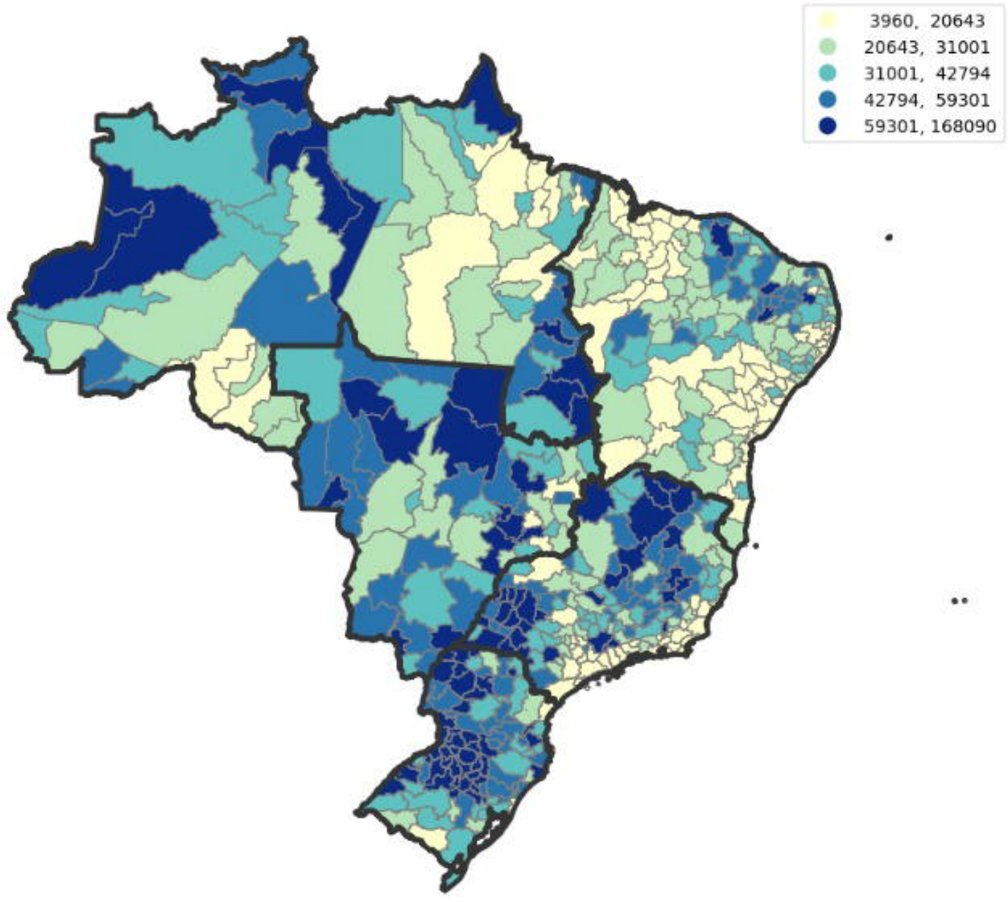
Figure 2: Number of regions with observed warnings in the PHC and OTC time series and/or hospitalization surges, per week, from November 2022 to June 2025. Counts of immediate regions with hospitalization surges in red, and warnings in OTC (green) and PHC (blue).

Figure 3: Timeliness of early warnings detected in PHC and OTC time series in relation to hospitalization surges at the national and regional levels. This figure summarizes the proportion of hospitalization surges anticipated (1 to 3 weeks early), detected in the same week (synchronous), or missed based on PHC and OTC series, along with their corresponding 95% confidence intervals (CIs). Results are presented for the national level, geographic region, and by population size.

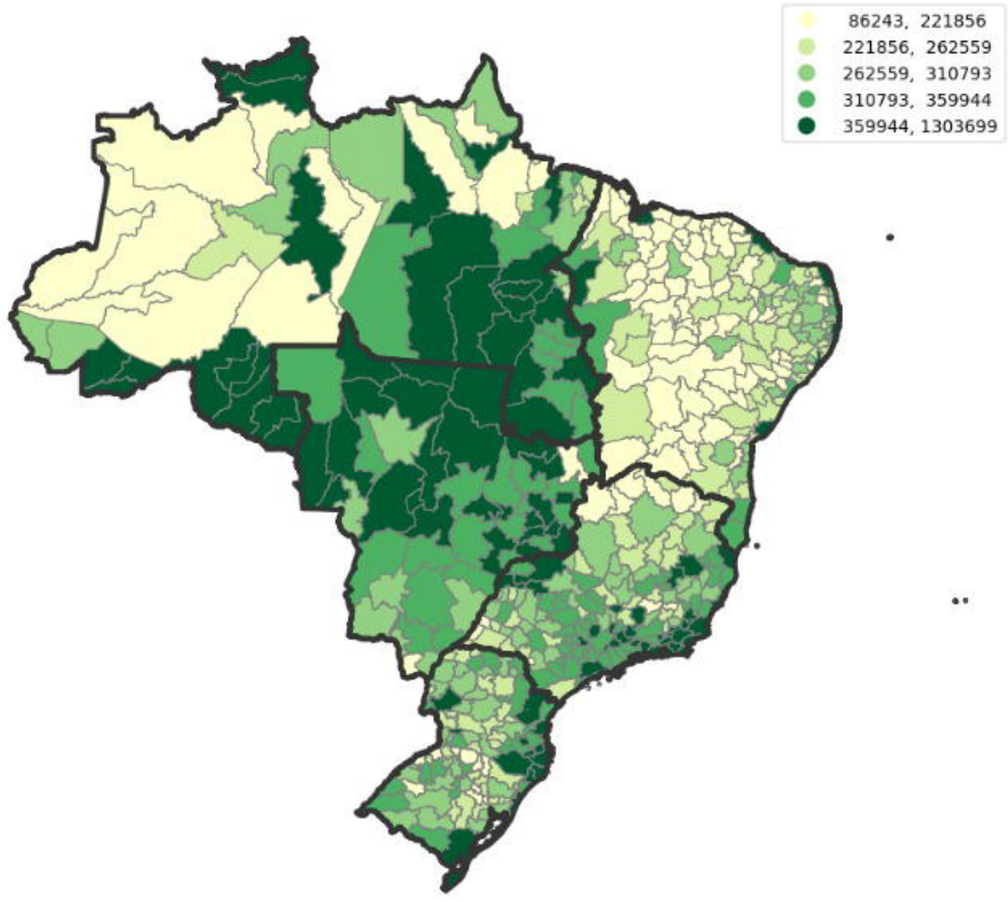
Figure 4: Performance metrics for PHC encounters and OTC drug sales data in detecting hospitalization surges across regions and population sizes. The figure presents sensitivity, specificity, probability of detection (POD), false positive rate (FPR), and precision for both PHC and OTC time series, along with their corresponding 95% CIs.

Figure 5: Classification of regions based on A) anticipation and B) precision indicators using PHC encounter and OTC drug sales data. A region was classified as “High” in anticipation, if anticipation scores were ≥ 60% for both PHC encounter and OTC drug sales data. For precision, regions with values ≥ 50.0% for PHC and ≥ 47.8% for OTC (median thresholds) were also considered “High”.

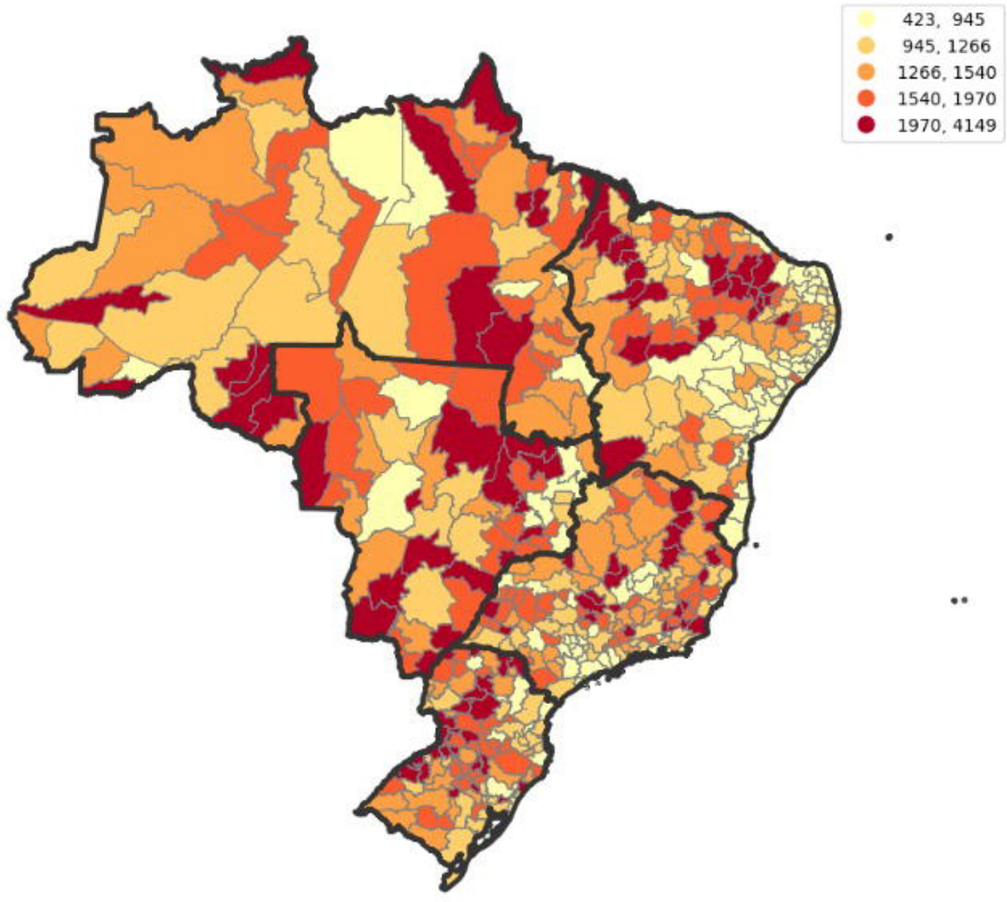
A

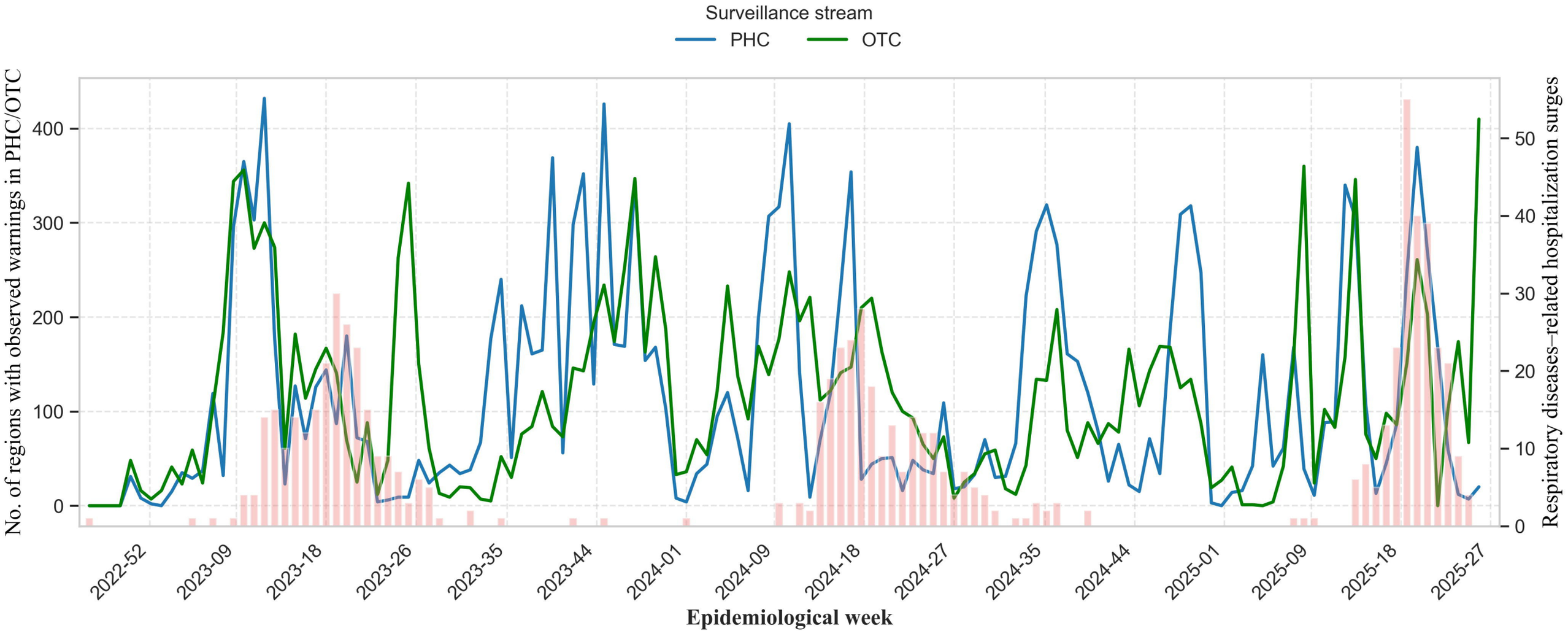

3960, 20643
20643, 31001
31001, 42794
42794, 59301
59301, 168090

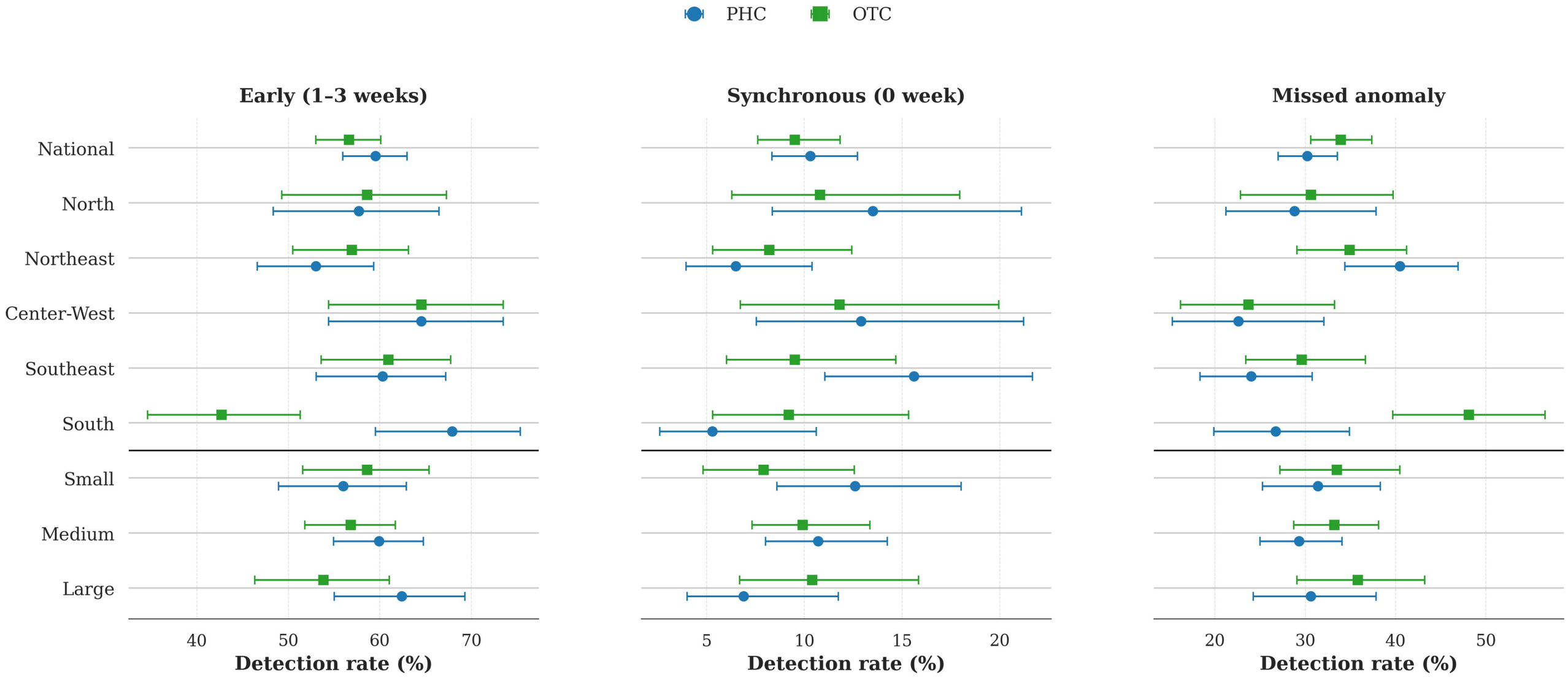
B

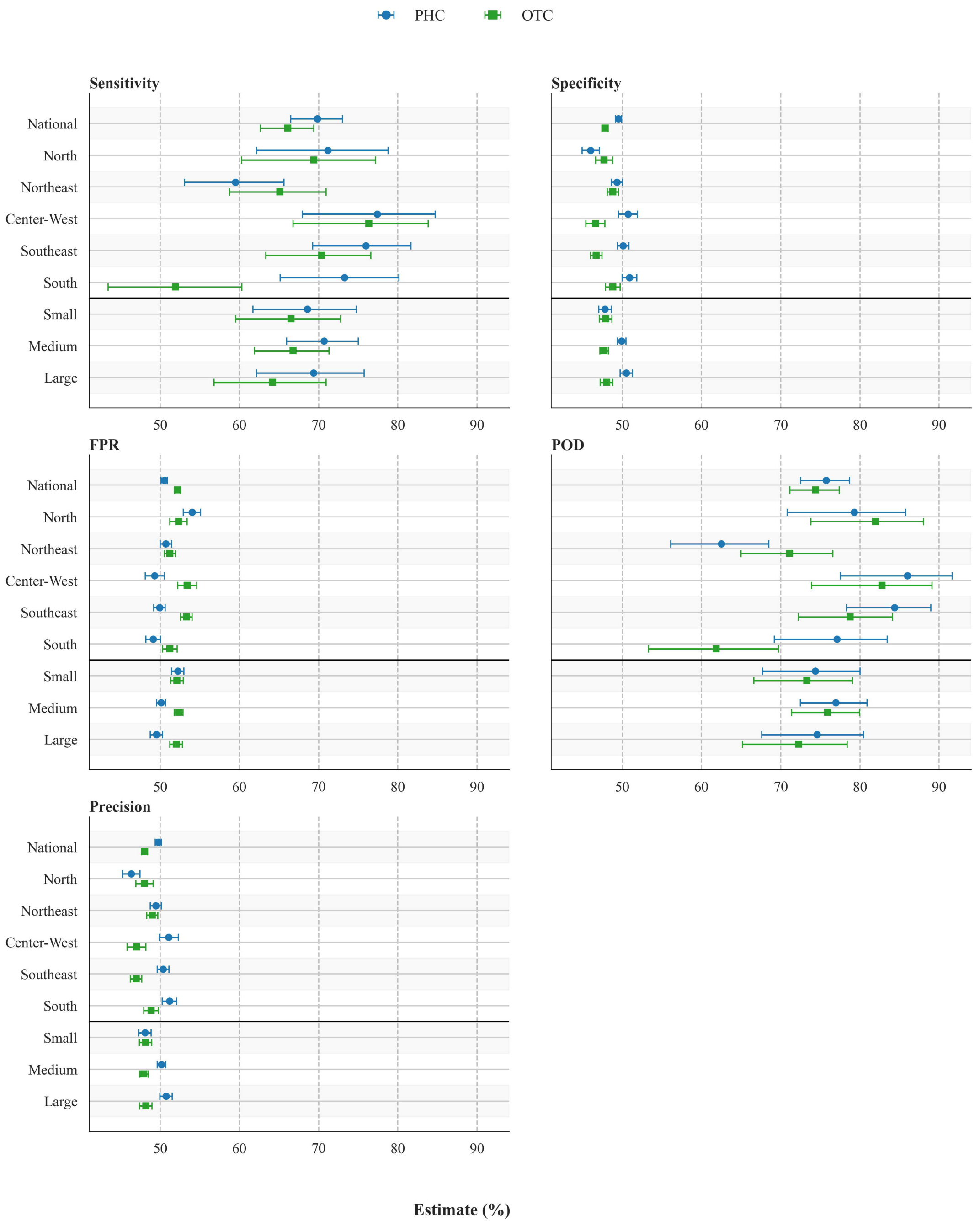

86243, 221856
221856, 262559
262559, 310793
310793, 359944
359944, 1303699

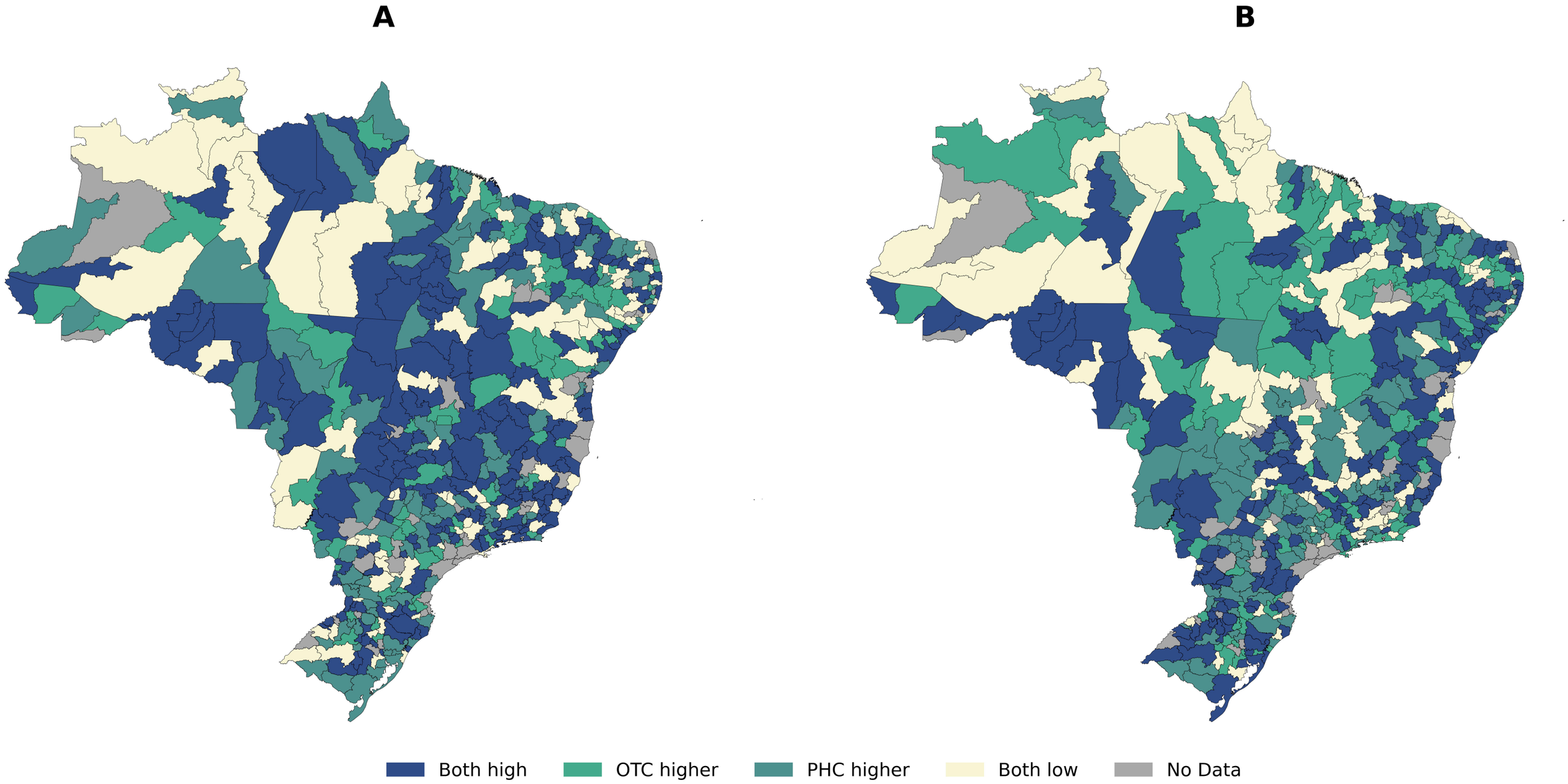
C


423, 945
945, 1266
1266, 1540
1540, 1970
1970, 4149


Surveillance stream
PHC
OTC
No. of regions with observed warnings in PHC/OTC
Respiratory diseases–related hospitalization surges
0
100
200
300
400
0
10
20
30
40
50
2022-52
2023-09
2023-18
2023-26
2023-35
2023-44
2024-01
2024-09
2024-18
2024-27
2024-35
2024-44
2025-01
2025-09
2025-18
2025-27
Epidemiological week


PHC
OTC
Early (1–3 weeks)
Synchronous (0 week)
Missed anomaly
National
North
Northeast
Center-West
Southeast
South
Small
Medium
Large
40
50
60
70
5
10
15
20
20
30
40
50
Detection rate (%)
Detection rate (%)
Detection rate (%)


PHC
OTC
Sensitivity
Specificity
FPR
POD
Precision
National
North
Northeast
Center-West
Southeast
South
Small
Medium
Large
50
60
70
80
90
Estimate (%)


A
B
Both high
OTC higher
PHC higher
Both low
No Data