Performance of drug sales and primary care encounters data for early detection of influenza-like illness surges in Brazil: a national time series analysis

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

[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] Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

[4] Medicine and Precision Public Health Laboratory (MeSP[2]), Gonçalo Moniz Institute, Oswaldo Cruz Foundation (FIOCRUZ); Salvador, 40296-710, Brazil.

[5] Faculty of Medicine, Federal University of Bahia, Salvador, Brazil

[6] Institute of Mathematics and Statistics of Federal University of Bahia, Salvador, Brazil.

[7] Luiz Coimbra Institute of Graduate and Engineering Research (COPPE), Federal University of Rio de Janeiro (UFRJ); Rio de Janeiro, 21941-859, Brazil.

[8] Bahiana School of Medicine and Public Health (EBMSP); Salvador, 40290-000, Brazil.

*Corresponding author: juliane.oliveira@fiocruz.br.

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

Abstract

Background: Effective pandemic preparedness requires integration and continuous evaluation of multiple data sources, including clinical and alternative data. This nationwide study with a large population evaluated the value of over-the-counter (OTC) drug sales and primary health care (PHC) encounter data for the early detection of surges in respiratory disease–related hospitalization in Brazil.

Methods: From November 2022 to December 2024, we analysed weekly time series data on PHC encounters, OTC drug sales used for influenza-like illness (ILI) treatment, and respiratory diseases-related hospitalizations across all 510 immediate regions in Brazil. Health data were provided by the Brazilian Ministry of Health, and drug sales data were sourced from IQVIA. We applied negative-binomial autoregressive moving average models in the context of Statistical Process Control techniques to identify early warnings in PHC and OTC time series. Surges in respiratory disease–related hospitalizations were defined as periods of weekly counts exceeding a defined threshold, enabling evaluation of each surveillance stream's performance.

Findings: During the study period, Brazil recorded 1.8 million respiratory diseases-related hospitalizations, 51 million PHC encounters and 584 million OTC drug units sold for ILI. In general, all three data shows increases in cases in March, with ILI-associated PHC encounters and hospitalizations peaking in May, and ILI-associated OTC sales drug sales peaking in July. A total of 13,106 and 13,066 warnings were detected in PHC encounters and OTC drug sales time-series, respectively. OTC data anticipated 63.8% of 556 hospitalization surges events 1–3 weeks in advance, detected 6.5% concurrently, and missed 29.7%. PHC data anticipated 54.3%, detected 8.5% concurrently, and missed 37.2%. OTC data showed higher sensitivity (70.3%) but lower specificity (48.8%) compared to PHC data (62.8% and 52.8%, respectively) to anticipate respiratory diseases-related hospitalization surges. Performance varied regionally: both data streams performed best in the Center-West, while OTC underperformed in urban areas and PHC in smaller populations. In 68.8% of regions, at least one stream provided high precision (Pr ≥ 51.8% for PHC encounters data and Pr ≥ 46.8% for OTC drug sales), and in 62.6% of regions, they anticipated at least 60% of hospitalizations surges.

Interpretation: This study demonstrates the reliability and value of using OTC drug sales data, complementing PHC-based surveillance to enhance early detection of respiratory diseases-related hospitalizations surges in a large upper-middle-income country.

Funding: The Rockefeller Foundation; Royal Society; Brazilian National Research Council (CNPq).

Research in context

Evidence before this study

We searched PubMed in February 2025, without language or date restrictions, using the terms ("surveillance" OR "outbreak detection" OR "warning system") AND ("over-the-counter drug sales"). This search yielded 80 peer-reviewed articles published between 1979 and 2025. Following a systematic review, 25 original research articles were identified relevant for the use of OTC drug sales data in syndromic surveillance systems. Most of the studies (23) were published prior to 2015, with limited progress in this field in the last decade. Most of these studies (22) assessed the correlation between OTC drug sales and traditional surveillance data or evaluated outbreak detection performance using various algorithms. Overall, the literature supports OTC drug sales as a promising data source for early outbreak detection, especially given the near real-time availability of such data in many countries. However, these studies report real-world OTC data limited to nine countries (Canada, China, France, Great Britain, Indonesia, Japan, Netherlands, Slovenia, and the United States). Furthemore, they reported constraints related to data length, representativeness, or pharmacy sampling, which limited comprehensive evaluations of timeliness, sensitivity, and specificity. Additionally, challenges were noted in the selection of appropriate drug categories that align with clinical surveillance signals. Notably, despite the potential of OTC data, few studies have reported on the implementation of OTC-based systems either as a standalone data source or fully integrated into national surveillance models.

Added value of this study

To our knowledge, this is the first large-scale study evaluating the performance of both OTC drug sales and PHC encounter data for syndromic surveillance. We assessed their ability to anticipate surges in respiratory diseases-related hospitalizations nationally, providing region-specific guidance for implementation. As one of the most comprehensive evaluations in a low- to middle-income country with high socio-demographic inequalities, our findings offer actionable insights for strengthening pandemic preparedness using existing real-time data streams.

Implications of all the available evidence

Our findings reinforce the promise of OTC drug sales data as a valuable component of syndromic surveillance systems, particularly when integrated with clinical data such as PHC encounters. Given the heterogeneous performance across regions, local tailoring and combined strategies can strengthen early warning systems. These insights are particularly relevant for resource-limited settings aiming to improve their capacity to tempestively detect and respond to emerging public health threats. Since OTC drug sales data are routinely collected in many countries, this surveillance approach can be readily implemented across diverse international settings, extending its usefulness beyond Brazil.

Background

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]. Syndromic surveillance is being increasingly used for early outbreak detection[2,6], enabling the joint use of traditional surveillance sources with alternative, health-related data streams collected and analyzed in near- or real-time[7]. 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[8–11].

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[12]. Despite these attributes, OTC drug sales have been underused for early outbreak detection. Previous studies, mainly conducted in high-income countries [13–17], 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, 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, including PHC encounter data[18]. 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[19].

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 modeling 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 surveillance performance of these data streams.

Methods

Data sources

We analyzed weekly time series data on PHC encounters and OTC drug sales, aggregated across 510 immediate regions in Brazil. Each region represents groupings of municipalities strongly connected to local urban centers, serving population needs for goods, employment, health services, and education (Sup. Fig 1), as defined by the Brazilian Institute of Geography and Statistics (IBGE). The dataset covers the period from November 2022 to December 2024.

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* [9].

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 analyzing medication purchasing trends[19,20]. 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 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 (Text S1).

To assess the ability of these series to anticipate the burden of respiratory diseases-related hospitalizations, we extracted hospitalizations cases, as per the International Classification of Diseases-10 (ICD-10 Chapter X, code range J00-J99) from the Brazilian Hospital Information System (SIH), covering the period from November 2022 to December 2024. 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 Moving Average models that best predict each time-series for each of the 510 immediate regions in Brazil. Using the estimated values obtained by the best fitted model, we applied the Statistical Process Control techniques to establish epidemic thresholds that are used to flag anomalous values in each time series[21,22]. We name the procedures as NegBi-SPC.

To define the NegBi-SPC procedure, let $\square_{\square}$ for $\square = 1, \ldots, \square$, 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 as:

$$\square\square\square(\square_{\square}) = \square_0 + \sum_{\square=1}^{\square} \square_{\square}\square_{\square-\square} + \square_1 \times \square\square\square\square\square + \square_2 \times \square\square\square\square\square\square\square\square\square\square\square\square$$

where $\square_{\square}$ is the mean at time t, $\square_0$ is the intercept, $\square_{\square}$ are the coefficients for the autoregressive terms varying from $\square = 1$ to $\square = 3$, and $\square_{\square}$, $\square = 1, 2$, are terms that capture the temporal trends and seasonal effects of the series considered.

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. 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, control limits were calculated to flag warnings represented by anomalous values as:

$$\square\square\square = \hat{\square}_{\square} - \square\square_{\hat{\square}_{\square}} \text{ and } \square\square\square = \hat{\square}_{\square} + \square\square_{\hat{\square}_{\square}}$$

where $\hat{\square}_{\square}$ is the value estimated by the model at time t, $\square_{\hat{\square}_{\square}}$ is the standard deviation estimate of the response predicted, and $\square$ represents the distance between the control limits.

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 December 2024, plus two standard deviations. Isolated warnings were excluded, while consecutive warnings and those separated by only one week were grouped together to define respiratory diseases-related hospitalization surges (hereafter referred to as hospitalization surges). To assess the robustness of this definition, we added one more alternative criteria for identifying warnings in the hospitalization time series in Text S2.

We then assessed the performance of OTC and PHC time series in anticipating hospitalization surges using the following metrics[21]: Probability of Detection (POD), defined as the probability that an 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 surge weeks, True Negatives (TN) are non-surge weeks without warnings, False Negatives (FN) are surge weeks without warnings, False Positives (FP) are non-surge weeks 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 Pearson correlation coefficients. Additionally, to examine the temporal relationships between each pair of time series within each immediate region, we computed lagged Pearson 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%.

Data Availability

Under the terms of our data use agreements with the Brazilian Ministry of Health and IQVIA, we are not authorized to share the primary datasets with third parties. Researchers interested in accessing these data should submit requests directly to both parties. The codes to ensure reproducibility of our results are available in our GitHub repository[23].

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 December 2024, Brazil recorded 51,052,372 PHC encounters associated with ILI symptoms and 1,803,869 respiratory diseases-related hospitalizations, averaging 464,112 (95% CI: 437,028 – 491,197) PHC encounters and 16,251 (95% CI: 15,381 – 17,121) hospitalizations per week. Simultaneously, monitored pharmacies sold 584,528,064 drug units associated with treatment of ILI symptoms, with a mean weekly sale of 5,313,891 units (95% CI: 5,058,530 – 5,569,253).

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, 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) (Figure S2). However, when analysing trend and seasonality by immediate regions, distinct dynamics emerged based on the presence or absence of trends and seasonality (Table S1).

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 (74.7% of regions), PHC encounters and hospitalizations (95.7%), and OTC drug sales and hospitalizations (66.7%). For PHC-hospitalization pairs, 45.5% of regions showed synchronous correlations (lag 0), while 6.7% exhibited PHC encounters preceding hospitalizations by 1–3 weeks, while 43.5% showed the reverse pattern. Among OTC-hospitalization pairs, 25.9% were synchronous (lag 0), with OTC drug sales preceding hospitalizations by 1–3 weeks in 26.9% of regions and hospitalizations preceding OTC sales in 13.9%. PHC-OTC correlations were

synchronous in 31% of regions, with OTC drug sales leading PHC encounters in 36.1% and PHC encounters leading OTC sales in 7.7%, both considering 1 to 3 weeks anticipation (Table S2).

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.

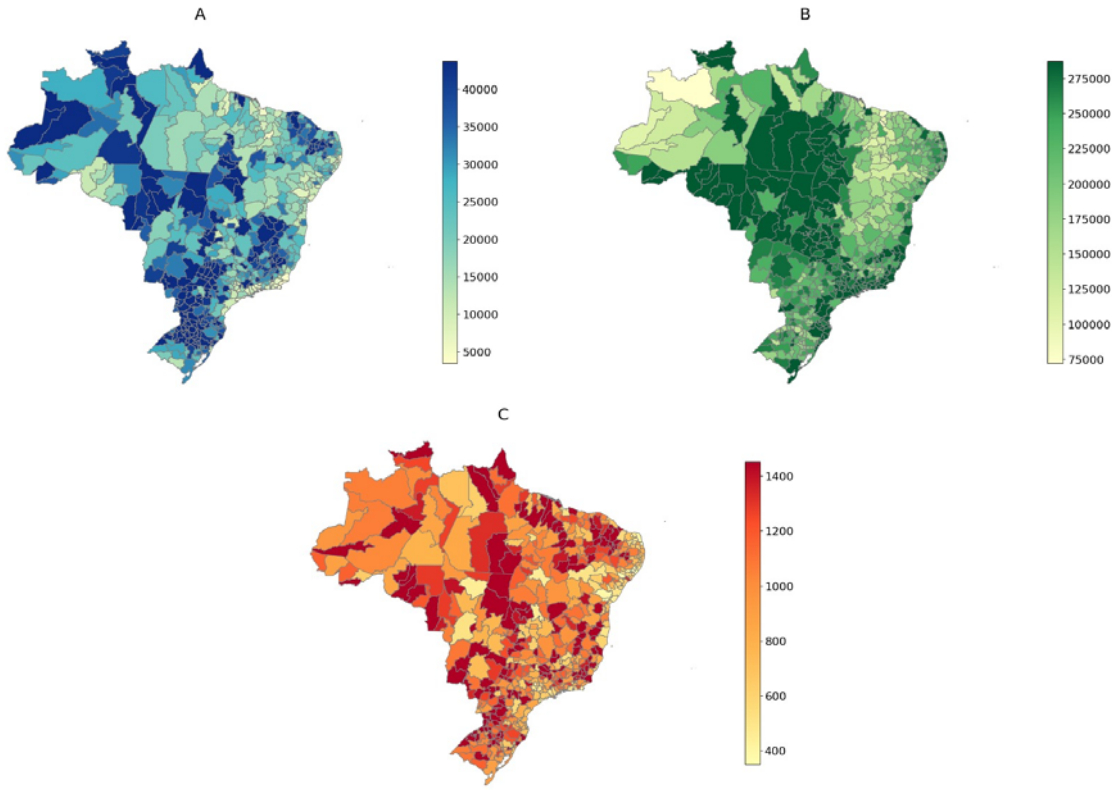

Figure 1: Spatial distribution of (A) ILI-related PHC encounters, (B) OTC drug units sold and (C) respiratory diseases-related hospitalizations across Brazil's immediate regions from November 2022 to December 2024. To enhance comparability, the indicators are presented as rates per 100,000 inhabitants.

There were strong to very strong correlations between regional rates of PHC encounters ($\rho = 0.696$, $p < 0.001$) and OTC drug sales ($\rho = 0.803$, $p < 0.001$) with hospitalizations. Correlations were particularly high in the South and Center-West regions. For PHC encounters, the correlations reached $\rho = 0.837$ ($p < 0.001$) in the South and $\rho = 0.829$ ($p < 0.001$) in the Center-West. Similarly, for OTC drug sales, the correlations were $\rho = 0.874$ ($p < 0.001$) and $\rho = 0.810$ ($p < 0.001$), respectively. OTC drug sales also showed very strong correlation in the North ($\rho = 0.852$, $p < 0.001$), whereas PHC encounters exhibited the weakest correlation in that region ($\rho = 0.596$, $p < 0.001$) (Table S3 and S4).

Overall, we observed a strong correlation between PHC encounters and OTC drug sales ($\rho = 0.642$, p-value < 0.001), with exception of the South regions that showed a moderate correlation ($\rho = 0.463$, p-value < 0.001).

Performance of anomaly detection

In the study period, 556 respiratory diseases-related hospitalization surges were identified in 403 (79%) immediate regions. Accordingly, by applying a NegBi-SPC model to each timeseries, we identified 13,106 warnings in PHC encounters and 13,066 in OTC drug sales data; of those, 10,442 and 10,282 are warnings, in PHC and OTC drug sales respectively, corresponding to the regions showing at least one hospitalizations surges. Warnings/surges emerged in February (epidemiological week 6) for PHC encounters and OTC sales, and in March (epidemiological week 10) for hospitalizations, aligning with expected nationwide seasonal patterns, as shown before. Notably, we also observe an increase of warnings starting in August (epidemiological week 32) for ILI related PHC encounters, followed by a similar pattern in OTC drug sales in September (epidemiological week 36) (Figure 2, Table S4). However, we identified only six hospitalizations surges from August to December of 2023 and 21 surges from August to December of 2024.

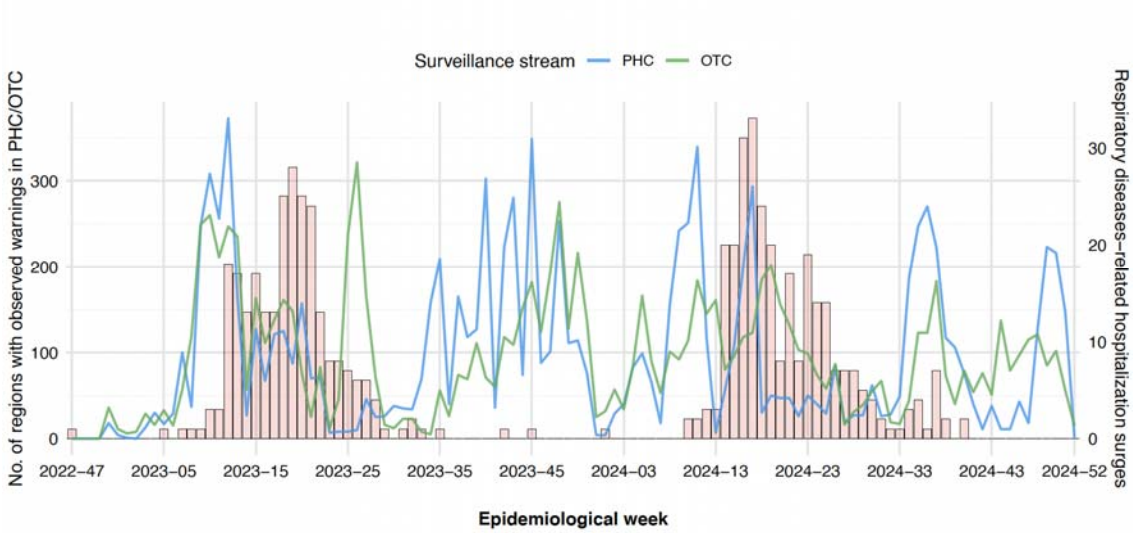

Figure 2: Number of regions with observed warnings in the PHC and OTC time series and/or hospitalization surges, per week, from November 2022 to December 2024. Counts of immediate regions with hospitalization surges in red, and warnings in OTC (green) and PHC (blue). Shaded vertical bands indicate periods with peaks in hospitalization surges.

Table 1 highlights the comparative timeliness performance of PHC and OTC data in anticipating hospitalization surges. Assessing the OTC time series, we could detect 63.8% (355/556) of

hospitalization warnings 1 to 3 weeks in advance, compared with 54.3% (302/556) detected by the PHC encounters data. This represents 53 additional hospitalization surges anticipated by monitoring OTC series in relation to PHC alone. In addition, the PHC series had a slightly higher detection proportion of hospitalization surges within the same week: 8.5% (47/556) versus 6.5% (36/556). However, the proportion of missed hospitalization surges was lower for OTC (29.7% (165/556)) than for PHC (37.2% (207/556)).

Table 1. Timeliness of early warnings detected in PHC and OTC time series in relation to hospitalization surges at the national and regional levels. This table 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. Results are presented for the national level, geographic region, and by population size .

Region / Population size	**PHC detection rate**			**OTC detection rate**		
	Early (1 to 3 weeks)	Synchronous (0 week)	Missed	Early (1 to 3 weeks)	synchronous (0 week)	Missed
Region						
National	54.3% (302/556)	8.5% (47/556)	37.2% (207/556)	63.8% (355/556)	6.5% (36/556)	29.7% (165/556)
North	53.8% (50/93)	9.7% (9/93)	36.6% (34/93)	63.4% (59/93)	6.5% (6/93)	30.1% (28/93)
Northeast	51.7% (91/176)	4.5% (8/176)	43.8% (77/176)	64.2% (113/176)	5.7% (10/176)	30.1% (53/176)
Center-West	58.9% (43/73)	16.4% (12/73)	24.7% (18/73)	65.8% (48/73)	6.8% (5/73)	27.4% (20/73)
Southeast	54.5% (72/132)	9.1% (12/132)	36.4% (48/132)	64.4% (85/132)	6.8% (9/132)	28.8% (38/132)
South	56.1% (46/82)	7.3% (6/82)	36.6% (30/82)	61.0% (50/82)	7.3% (6/82)	31.7% (26/82)

Population-size						
Small	54.2% (83/153)	6.5% (10/153)	39.2% (60/153)	62.1% (95/153)	7.2% (11/153)	30.7% (47/153)
Medium	55.9% (162/290)	8.6% (25/290)	35.5% (103/290)	65.5% (190/290)	6.9% (20/290)	27.6% (80/290)
Large	50.4% (57/113)	10.6% (12/113)	38.9% (44/113)	61.9% (70/113)	4.4% (5/113)	33.6% (38/113)

The timeliness of PHC-based warnings varied across regions and population size. The Center-West region exhibited the highest early detection rates, with 58.9% (43/73) of warnings issued 1 to 3 weeks in advance to hospitalization surges. In contrast, the Northeast had slightly lower PHC timeliness, with 51.7% (91/176) issued 1 to 3 weeks in advance and 43.8% (77/176) of hospitalizations surges missed. Performance also varied by population size, with optimal early detection (1–3 weeks anticipation) in medium-sized regions (55.9%,162/290) while large population regions showed reduced performance, with early anticipation rates of 50.4% (57/113) for PHC data. For OTC-based warnings the differences across regions and between municipalities with different population sizes were less than 5% (Table 2).

By assessing sensitivity (Se) and probability of detection (POD), we verified that OTC drug sales data (Se = 70.3% and POD = 77.3%) outperformed PHC encounters data (Se = 62.8% and POD = 68.9%), indicating improved early warning capabilities. However, OTC data also showed a higher false positive rate (FPR) and lower precision (Pr) across all regions, suggesting potential over-detection (Table 2).

Table 2: Performance metrics for PHC encounters and OTC drug sales data in detecting hospitalization surges across regions and population sizes. The table presents Sensitivity (Se), Specificity (Sp), Probability of Detection (POD), False Positive Rate (FPR), and Precision (Pr) for both PHC and OTC time series.

Region or population-scale	**PHC encounters data**					**OTC drug sales data**				
	Se	**Sp**	**POD**	**FPR**	**Pr**	**Se**	**Sp**	**POD**	**FPR**	**Pr**

Region										
National	62.8%	52.8%	68.9%	47.2%	52.9%	70.3%	48.8%	77.3%	51.2%	49.1%
North	63.4%	50.0%	77.4%	50.0%	50.2%	69.9%	47.5%	81.7%	52.5%	47.8%
Northeast	56.2%	52.4%	59.7%	47.6%	52.4%	69.9%	49.6%	74.4%	50.4%	49.9%
Center-West	75.3%	54.7%	80.8%	45.3%	55.0%	72.6%	48.0%	82.2%	52.0%	48.4%
Southeast	63.6%	53.1%	69.7%	46.9%	53.2%	71.2%	48.3%	78.0%	51.7%	48.5%
South	63.4%	54.2%	67.1%	45.8%	54.3%	68.3%	50.1%	73.2%	49.9%	50.3%
Population-scale*										
Small	60.8%	51.4%	68.0%	48.6%	51.5%	69.3%	47.7%	76.5%	52.3%	48.0%
Medium	64.5%	53.3%	70.7%	46.7%	53.5%	72.4%	48.8%	79.3%	51.2%	49.1%
Large	61.1%	53.2%	65.5%	46.8%	53.2%	66.4%	50.3%	73.5%	40.7%	50.4%

PHC encounters exhibited better performance in the Center-West, achieving the highest Se (75.3%) and POD (80.8%) among all regions, compared with the use of OTC drug sales data. However, OTC data generally outperformed PHC across most regions, with sensitivity ranging from 68.3% (South) to 72.6% (Center-West) and POD between 73.2% to 82.2%. The Northeast showed the weakest PHC performance, with sensitivity (56.2%) and POD (59.7%) substantially lower than other regions.

Regarding population size, medium-sized cities achieved the best overall performance for both PHC and OTC, balancing higher sensitivity (64.5% for PHC, 72.4% for OTC) with lower FPR (46.7% for PHC, 51.2% for OTC). In contrast, PHC data showed lower performance in small population regions, while OTC performance declined in large population areas.

Complementary value of using OTC and PHC

Overall, 164 out of 510 regions (32.2%) demonstrated high precision in both PHC (Pr ≥ 51.8%) and OTC (Pr ≥ 46.8%) data streams. PHC data alone outperformed OTC in 95 regions (18.6%), while

OTC performed better in 92 regions (18.0%). In contrast, both data sources exhibited low precision in 52 regions (10.2%).

Regarding the timeliness score, both PHC and OTC data showed high anticipation in 161 regions (31.6%). OTC outperformed PHC in 95 regions (18.6%), whereas PHC had higher timeliness in 63 regions (12.4%). In 84 regions (16.5%), both data sources showed low anticipation. Due to the absence of hospitalization surges events, precision and timeliness could not be evaluated in 107 regions (21.0%) (Figure 3).

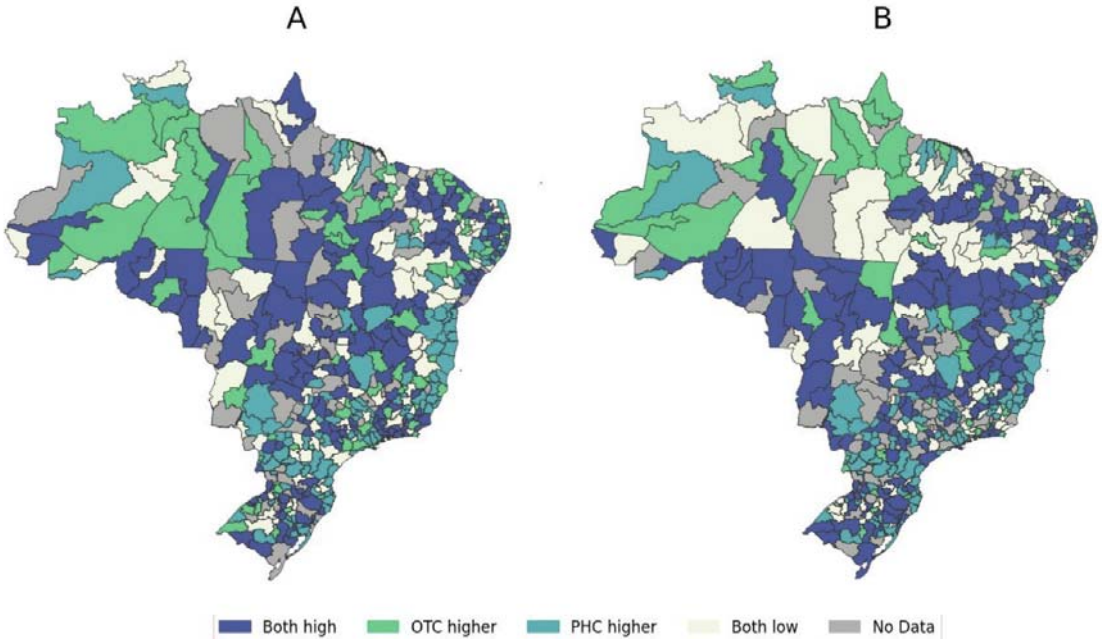

Figure 3: Classification of regions based on A) anticipation and B) precision indicators using PHC encounter and OTC drug sales data. A region was classified as "High" in anticipation if anticipation scores were ≥ 40% for both PHC encounter and OTC drug sales data. For precision, regions with values ≥ 53.4% for PHC and ≥ 49.0% 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 (Text S2). 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 analyzed 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 data shows high sensitivity performance and provides early detection of respiratory diseases-related hospitalization surges, but at the cost of increased false alarms, while PHC data offers greater specificity, albeit with slightly lower sensitivity. 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[15]. 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 LMIC context[2,14,24–26]. 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[2,7]. 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[27]. 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 a 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 potential and quality of Brazil's PHC encounter data for surveillance purposes were previously established by our group[9]. 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 behavior suggests the occurrence of real, albeit mild, outbreaks. ILI is predominantly caused by self-limiting infections that are commonly managed with non-prescription medications and initial clinical consultations in primary care[28]. 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[29]. 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.

PHC data showed variation across Brazilian regions and by municipality population size, particularly in terms of sensitivity and probability of detection. This finding is in alignment with our previous study[9], 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. This variability implies that the effectiveness of PHC data for early outbreak detection may require locally-referred calibration to ensure consistent performance. The performance of OTC data was more stable across regions and population strata, with less variability in sensitivity, specificity, false positive rate, and precision indicating that it may serve as a more uniform and scalable surveillance tool, especially in settings with limited or uneven primary care infrastructure. 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[10]. 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 analyzing 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 50% of surges in respiratory diseases-related hospitalizations were detected using PHC data, and over 60% with 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 21% 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.

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.

Author contributions

Conceptualization: JFO, MB-N, VB; Methodology: JFO, TC-S, PANB, NBS, RLF, FMHSF, VB; Investigation: JFO, TC-S, PANB, 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.

Declaration of interests

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

Data and materials availability

Supplementary Figures 1 and 2

Fig. S1: Map of Brazil's Immediate Geographic Regions

Fig. S2: Temporal trends of influenza-like illness indicators in Brazil and across Brazilian regions

Supplementary Table 1 to 5

Table S1: Summary of immediate regions by presence of statistically significant trend and seasonality.

Table S2: Lagged correlations between PHC encounters, respiratory diseases-related hospitalizations, and OTC drug sales across immediate regions.

Table S3: Total counts and rates (per 100,000 inhabitants) of respiratory diseases-related hospitalization cases, ILI-related PHC encounters, and OTC drug units sold across Brazil and by region.

Table S4: Monthly counts of surges in respiratory diseases-related hospitalizations and warnings in ILI-related PHC encounters and OTC drug sales.

Table S5: Spearman correlation coefficients between respiratory diseases-related hospitalization counts, PHC visits, and OTC drug sales at the national and regional levels.

Supplementary Text 1 and 2

Text S1: OTC medications cluster analysis.

Text S2: Alternative definitions for surges in respiratory diseases-related hospitalizations.

Text S3: Supplementary references.

References

1 Morgan OW, Aguilera X, Ammon A, *et al.* Disease surveillance for the COVID-19 era: time for bold changes. *Lancet* 2021; **397**: 2317–9.

2 Ramos PIP, Marcilio I, Bento AI, *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* 2024; **10**: e47673.

3 Al Knawy B, Adil M, Crooks G, *et al.* The Riyadh Declaration: the role of digital health in fighting pandemics. *Lancet* 2020; **396**: 1537–9.

4 Oliveira JF, Alencar AL, Cunha MCLS, *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* 2024; **6**: e570–9.

5 Oliveira JF, Vasconcelos AO, Alencar AL, *et al.* Balancing human mobility and health care coverage in sentinel surveillance of Brazilian Indigenous areas: Mathematical optimization

approach. *JMIR Public Health Surveill* 2025; **11**: e69048.

6 Smith GE, Elliot AJ, Lake I, *et al.* Syndromic surveillance: two decades experience of sustainable systems - its people not just data! *Epidemiol Infect* 2019; **147**: e101.

7 Triple S Project. Assessment of syndromic surveillance in Europe. *Lancet* 2011; **378**: 1833–4.

8 Sočan M, Erčulj V, Lajovic J. Early detection of influenza-like illness through medication sales. *Cent Eur J Public Health* 2012; **20**: 156–62.

9 Cerqueira-Silva T, Oliveira JF, Oliveira V de A, *et al.* Early warning system using primary health care data in the post-COVID-19 pandemic era: Brazil nationwide case-study. *Cad Saude Publica* 2024; **40**: e00010024.

10 Todd S, Diggle PJ, White PJ, Fearne A, Read JM. The spatiotemporal association of non-prescription retail sales with cases during the 2009 influenza pandemic in Great Britain. *BMJ Open* 2014; **4**: e004869.

11 Dong X, Boulton ML, Carlson B, Montgomery JP, Wells EV. Syndromic surveillance for influenza in Tianjin, China: 2013–14. *J Public Health (Oxf)* 2016; : fdw022.

12 Pivette M, Mueller JE, Crépey P, Bar-Hen A. Drug sales data analysis for outbreak detection of infectious diseases: a systematic literature review. *BMC Infect Dis* 2014; **14**: 604.

13 Sugawara T, Ohkusa Y, Shigematsu M, Taniguchi K, Murata A, Okabe N. An experimental study for syndromic surveillance using OTC sales. *Kansenshogaku Zasshi* 2007; **81**: 235–41.

14 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* 2005; **54**: 47–52.

15 Vergu E, Grais RF, Sarter H, *et al.* Medication sales and syndromic surveillance, France. *Emerg Infect Dis* 2006; **12**: 416–21.

16 Muchaal PK, Parker S, Meganath K, Landry L, Aramini J. Evaluation of a national pharmacy-based syndromic surveillance system. *Can Commun Dis Rep* 2015; **41**: 203–6.

17 Magruder S. Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of human disease. *Johns Hopkins Apl Technical Digest* 2003; **24**: 349–53.

18 Kovacs R, Maia Barreto JO, da Silva EN, *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* 2021; **9**: e331–9.

19 IQVIA Brasil. https://www.iqvia.com/pt-br/locations/brazil (accessed May 5, 2025).

20 Where Brazilians go to shop. https://www.iqvia.com/blogs/2017/11/where-brazilians-go-to-shop (accessed May 5, 2025).

21 Bédubourg G, Le Strat Y. Evaluation and comparison of statistical methods for early temporal detection of outbreaks: A simulation-based study. *PLoS One* 2017; **12**: e0181227.

22 Sena JGS de, Ferreira PH, Fiaccone RL. O Controle Estatístico de Processos como Ferramenta para Controle e Prevenção de Epidemias de Malária na Região da Amazônia Legal. *Braz Jour Biom* 2022; **40**. DOI:10.28951/bjb.v40i1.538.

23 Oliveira JF. Github Directory. https://github.com/Julian-

sun/Preemptive_Surveillance_PHC_OTC (accessed May 9, 2025).

24 Davies GR, Finch RG. Sales of over-the-counter remedies as an early warning system for winter bed crises. *Clin Microbiol Infect* 2003; **9**: 858–63.

25 Socan M, Erculj V, Lajovic J. Early detection of influenza like illness through medication sales. *Cent Eur J Public Health* 2012; **20**: 156–62.

26 Zhang Y, Li L, Dong X, *et al*. Influenza surveillance and incidence in a rural area in China during the 2009/2010 influenza pandemic. *PLoS One* 2014; **9**: e115347.

27 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 (accessed May 12, 2025).

28 Peters S, Rowbotham S, Chisholm A, *et al*. Managing self-limiting respiratory tract infections: a qualitative study of the usefulness of the delayed prescribing strategy. *Br J Gen Pract* 2011; **61**: e579–89.

29 Campos GS, Bandeira AC, Sardi SI. Zika virus outbreak, Bahia, Brazil. *Emerg Infect Dis* 2015; **21**: 1885–6.

30 Najmi A-H, Magruder SF. An adaptive prediction and detection algorithm for multistream syndromic surveillance. *BMC Med Inform Decis Mak* 2005; **5**: 33.