

Original Paper

# Impact of Primary Health Care Data Quality on Infectious Disease Surveillance in Brazil: Case Study

Pilar Tavares Veras Florentino<sup>1</sup>, PhD; Juracy Bertoldo Junior<sup>1</sup>, MSc; George Caique Gouveia Barbosa<sup>1</sup>, MSc; Thiago Cerqueira-Silva<sup>2</sup>, PhD; Vinicius de Araújo Oliveira<sup>1,3</sup>, MD; Marcio Henrique de Oliveira Garcia<sup>4</sup>, MSc; Gerson Oliveira Penna<sup>5,6</sup>, PhD; Viviane Boaventura<sup>7</sup>, PhD; Pablo Ivan Pereira Ramos<sup>1</sup>, PhD; Manoel Barral-Netto<sup>1,7</sup>, PhD; Izabel Marcilio<sup>1,8</sup>, PhD

<sup>1</sup>Centro de Integração de Dados e Conhecimento em Saúde (CIDACS), Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil

<sup>2</sup>Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, United Kingdom

<sup>3</sup>Secretaria de Atenção Primária, Ministério da Saúde, Brasília, Brazil

<sup>4</sup>Secretaria de Vigilância em Saúde e Ambiente, Ministério da Saúde, Brasília, Brazil

<sup>5</sup>Escola Fiocruz de Governo, Fundação Oswaldo Cruz (Fiocruz), Rio de Janeiro, Brazil

<sup>6</sup>Núcleo de Medicina Tropical, Universidade de Brasília, Brasília, Brazil

<sup>7</sup>Laboratório de Medicina e Saúde Pública de Precisão, Instituto Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, Brazil

<sup>8</sup>Departamento de Epidemiologia, Escola Bahiana de Medicina e Saúde Pública, Salvador, Brazil

## Corresponding Author:

Pilar Tavares Veras Florentino, PhD  
Centro de Integração de Dados e Conhecimento em Saúde (CIDACS)  
Instituto Gonçalo Moniz, Fundação Oswaldo Cruz  
R. Mundo, 121 - sala 315 - Trobogy  
Salvador, 41745-715  
Brazil  
Phone: 55 7131762357  
Email: [pilar.veras@fiocruz.br](mailto:pilar.veras@fiocruz.br)

## Abstract

**Background:** The increase in emerging and re-emerging infectious disease outbreaks underscores the need for robust early warning systems (EWSs) to guide mitigation and response measures. Administrative health care databases provide valuable epidemiological insights without imposing additional burdens on health services. However, these datasets are primarily collected for operational use, making data quality assessment essential to ensure an accurate interpretation of epidemiological analysis. This study focuses on the development and implementation of a data quality index (DQI) for surveillance integrated into an EWS for influenza-like illness (ILI) outbreaks using Brazil's a nationwide Primary Health Care (PHC) dataset.

**Objective:** We aimed to evaluate the impact of data completeness and timeliness on the performance of an EWS for ILI outbreaks and establish optimal thresholds for a suitable DQI, thereby improving the accuracy of outbreak detection and supporting public health surveillance.

**Methods:** A composite DQI was established to measure the completeness and timeliness of PHC data from the Brazilian National Information System on Primary Health Care. Completeness was defined as the proportion of weeks within an 8-week rolling window with any register of encounters. Timeliness was calculated as the interval between the date of encounter and its corresponding registry in the information system. The backfilled PHC dataset served as the gold standard to evaluate the impact of varying data quality levels from the weekly updated real-time PHC dataset on the EWS for ILI outbreaks across 5570 Brazilian municipalities from October 10, 2023, to March 10, 2024.

**Results:** During the study period, the backfilled dataset recorded 198,335,762 ILI-related encounters, averaging 8,623,294 encounters per week. The EWS detected a median of 4 (IQR 2-5) ILI outbreak warnings per municipality using the backfilled dataset. Using the real-time dataset, 12,538 (65%) warnings were concordant with the backfilled dataset. Our analysis revealed that 100% completeness yielded 76.7% concordant warnings, while 80% timeliness resulted in at least 50% concordant warnings. These thresholds were considered optimal for a suitable DQI. Restricting the analysis to municipalities with a suitable DQI increased concordant warnings to 80.4%. A median of 71% (IQR 54%-71.9%) of municipalities met the suitable DQI threshold weekly. Municipalities with  $\geq 60\%$  of weeks achieving a suitable DQI demonstrated the highest concordance between backfilled and real-time datasets, with those achieving  $\geq 80\%$  of weeks showing 82.3% concordance.

**Conclusions:** Our findings highlight the critical role of data quality in improving the EWS' performance based on PHC data for detecting ILI outbreaks. The proposed framework for real-time DQI monitoring is a practical approach and can be adapted to other surveillance systems, providing insights for similar implementations. We demonstrate that optimal completeness and timeliness of data significantly impact the EWS' ability to detect ILI outbreaks. Continuous monitoring and improvement of data quality should remain a priority to strengthen the reliability and effectiveness of surveillance systems.

*JMIR Public Health Surveill* 2025;11:e67050; doi: [10.2196/67050](https://doi.org/10.2196/67050)

**Keywords:** primary health care; data quality; infectious disease surveillance; Brazil; early warning system

## Introduction

In recent decades, the world has witnessed an unprecedented surge of emerging and re-emerging infectious disease outbreaks, underscoring the need for stronger early warning systems (EWSs) [1,2]. The widespread and growing use of electronic health records (EHRs) has heightened the demand for automated processes in disease surveillance [3,4].

Systematic monitoring of administrative health care databases provides valuable epidemiological insights [5,6]. Importantly, using administrative data for health surveillance avoids the overburden of surveillance teams while ensuring timeliness, as no duplication of registry is required [3]. This cost-efficient approach enhances the ability to detect outbreaks, particularly in low-resource settings, thus contributing to global security [3].

However, an effective automated EWS based on administrative datasets requires that a real-time data quality assessment algorithm is set within the EWS pipeline. Since administrative data are primarily collected for operational purposes, assessing their quality is crucial to accurate interpretation of epidemiological analysis [7,8]. A systematic review on EHR data quality assessment studies found 14 articles describing dedicated data quality programs deployed in real-world settings, while only 4 produced results generally applicable in diverse settings. Ozonze et al [9] suggest there is an absence of comprehensive tools for facilitating reliable and consistent data quality assessments.

Moreover, despite existing methods for evaluating the quality of administrative health data, including EHR data quality assessment [8] and indicators for specific programs such as the Data Quality Audit and the Data Quality Self-Assessment for immunization data [10], there remains a gap in applying similar methods to data used for health surveillance. Although metrics for assessing the quality of surveillance systems are well established [3], to the best of our knowledge, these have not been applied to evaluate administrative data when used for epidemiological surveillance purposes.

This paper describes the development and implementation of a data quality index (DQI) to assess the quality of administrative data used in epidemiological surveillance systems. We focus on applying the DQI to nationwide Brazilian primary health care (PHC) administrative data integrated into an EWS for influenza-like illness (ILI) outbreaks. The study compares the EWS performance across different DQI levels, addressing a critical gap in current

research by establishing metrics that ensure accurate and timely outbreak detection while leveraging the cost-efficiency of administrative health databases.

## Methods

### Study Design

We developed and implemented a data quality assessment algorithm within *ÆSOP* (Alert-Early System of Outbreaks with Pandemic Potential), a previously validated EWS [11]. This EWS applies aberration detection algorithms, such as the Early Aberration Reporting System (C2) [12], to a time series consisting of weekly counts of ILI-related PHC encounters per municipality, aiming at the early detection of outbreaks. To assess the data quality of the PHC data stream, we established the composite indicator DQI to measure the completeness and timeliness of the data. Using the backfilled PHC dataset as a gold standard, we evaluated the impact of data quality in the EWS' performance using different levels of data quality of the weekly updated real-time PHC dataset across all 5570 Brazilian municipalities from October 10, 2023, to March 10, 2024.

### Data Source

Brazil is an upper middle-income country with approximately 212.6 million people living in 5570 municipalities [13], and we included all ILI-related PHC encounters occurring during the study period in our analysis. We analyzed data from the Brazilian Unified Health System (SUS), which stands as one of the largest public health systems globally, providing comprehensive and universal health care to the entire population. The effective management of SUS relies on diverse information systems, among which the Brazilian National Information System on Primary Health Care (SISAB [Sistema de Informação em Saúde para a Atenção Básica]) plays a crucial role. SISAB is a hierarchical, decentralized information system maintained and managed by the Ministry of Health (MoH), and harbors data on all publicly funded PHC encounters in the country. Data registration is mandatory for the allocation of financial resources from the federal to the municipal level. All encounters are coded by the *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) or the *International Classification of Primary Care* (ICPC-2).

According to the MoH's guidelines, municipalities are requested to update the system at least on a monthly basis, with a window of 4 months for amendments following each monthly submission. This operational guideline aligns

with the SISAB's purpose of informing decision-making for the management of the PHC system in the country. However, the EWS uses weekly updates of the SISAB database to detect ILI outbreaks. Therefore, this real-time, weekly updated dataset may present incompleteness and a temporal lag between the dates of encounter and data registration into the system (Figure S1 in [Multimedia Appendix 1](#)).

## The DQI

We defined the dimensions of completeness and timeliness to develop quantitative indicators for monitoring data quality in the EWS. Completeness is one of the most commonly used dimensions in data quality assessment and may be defined as the proportion of data filled with values for each attribute or entity in the database, while timeliness can be defined as the availability of data for decision-making, measured by the time interval between the occurrence of the measured event and its capture in an information system [14].

In our study, completeness refers to the proportion of weeks in each 8-week rolling window with any register of a PHC encounter. The indicator is measured as a fraction, with the numerator ranging from 0 to 8, and the denominator is 8 weeks, which is expressed as a percentage. Timeliness refers to the time interval, in number of weeks, between the date of the PHC encounter and its registry in the database. The indicator is represented by the proportion of registries occurring in 2 weeks or less from the PHC encounter in the same 8-week rolling window.

As it is recommended that the diverse quality dimensions should be collectively analyzed for a more comprehensive evaluation of data quality [15], we combined the 2 selected indicators in a composite measure, named DQI. The DQI is assessed weekly, for each municipality, once the PHC data are updated into the EWS pipeline.

## Impact of DQI on the EWS' Performance

To decide on the minimum required threshold of completeness and timeliness to derive trustworthy results with the EWS, we applied the EWS algorithm to the retrospectively gathered, backfilled PHC dataset. We compared the results to those obtained when applying the EWS to the weekly updated, herein named real-time PHC dataset (Figure S1 in [Multimedia Appendix 1](#)). Using the backfilled dataset as a reference, we calculated the proportion of concordant warnings detected in the real-time dataset. Accordingly, the DQI is expressed as either "suitable" or "unsuitable" when the minimum threshold of both completeness and timeliness is reached, indicating that the data quality may not be adequate for reliable EWS outputs.

Analyses were performed using Python (version 3.9) and R (version 4.3.1) software. The database's description and the scripts are available on GitHub [16].

## Ethical Considerations

The study protocol and procedures were reviewed and approved by the Ethical Review Board of Oswaldo

Cruz Foundation – Fiocruz Bahia (protocol CAAE 61444122.0.0000.0040).

Data on publicly funded PHC encounters were collected and compiled by the MoH for funding reasons. No consent was needed for data collection at this administrative level. For this study, we accessed an aggregated database consisting of the number of encounters per epidemiological week, per municipality, and per diagnostic code. The accessed database has no information at the individual level, and given that this study involves secondary analysis of existing deidentified data and does not involve direct interaction with human participants, it is classified as exempt from the requirement for informed consent under applicable ethical guidelines.

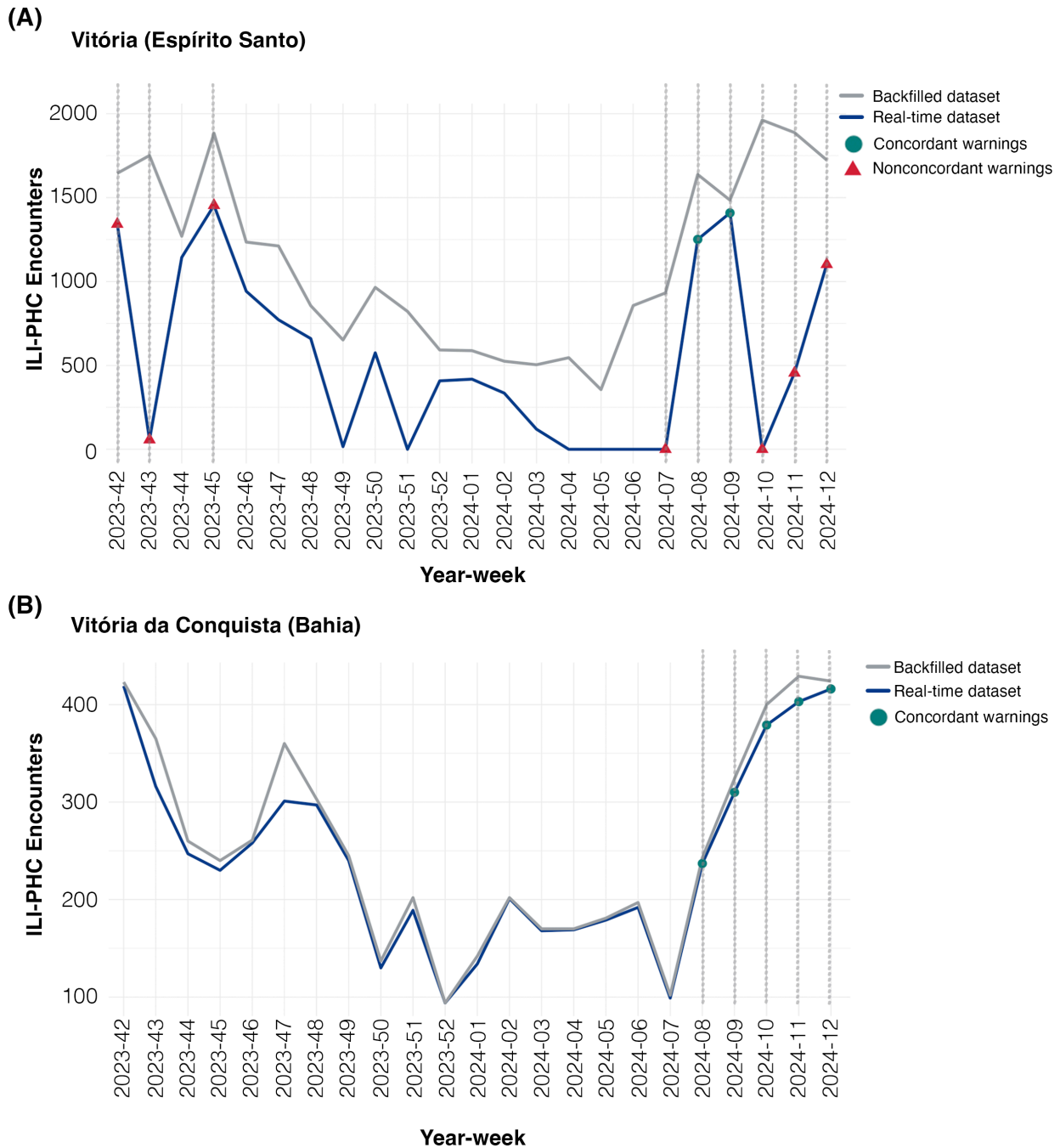
## Results

There were 198,335,762 recorded ILI-related encounters in the backfilled PHC dataset, which corresponds to an average of 8,623,294 encounters per week between October 10, 2023, and March 10, 2024. Using the backfilled dataset, the EWS detected a median of 4 (IQR 2-5) warnings of ILI outbreaks per municipality in the study period.

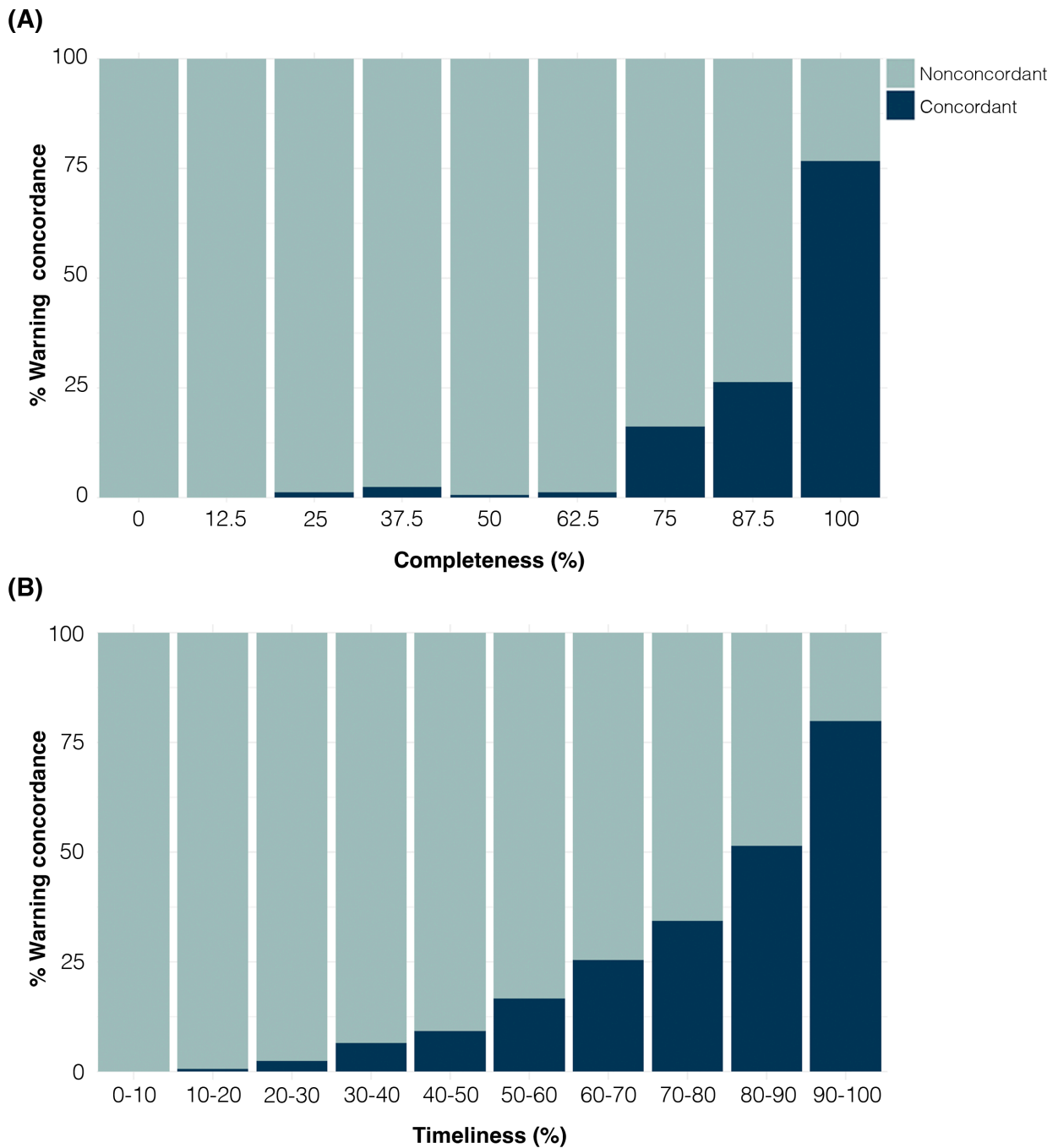
Figure 1 illustrates the impact of the DQI on the ability of the EWS to correctly identify potential ILI outbreaks. Using the real-time dataset, the EWS detected 12,538 (65%) warnings of ILI-outbreaks that were concordant with warnings detected in the backfilled dataset (Table S1 in [Multimedia Appendix 1](#)). The proportion of concordant warnings detected in the real-time dataset, based on different levels of completeness (Figure 2A and Table S1 in [Multimedia Appendix 1](#)) and timeliness (Figure 2B and Table S1 in [Multimedia Appendix 1](#)), indicated that 100% completeness and a minimum of 80% timeliness yielded the highest percentage of concordant warnings. Therefore, these values were established as the thresholds for grading the DQI as suitable or unsuitable for the EWS. Restricting the EWS analysis to municipalities with a suitable DQI, the proportion of warnings for ILI outbreaks concordant to the backfilled dataset increased to 80.4% (Table S2 in [Multimedia Appendix 1](#)). We found a median of 71% (IQR 54%-71.9%) of Brazilian municipalities with a suitable DQI per week in the study period (Figure 3A and Table S2 in [Multimedia Appendix 1](#)).

Additionally, we analyzed concordant warnings by grouping municipalities based on the proportion of weeks in which they exhibited a suitable DQI ( $\leq 20\%$ , 20%-40%, 40%-60%, 60%-80%, and  $\geq 80\%$ ). Our findings revealed that municipalities with over 60% of weeks featuring a suitable DQI had the highest proportion of concordant warnings between the backfilled and real-time datasets (Figure 3B, Table S2 in [Multimedia Appendix 1](#)).

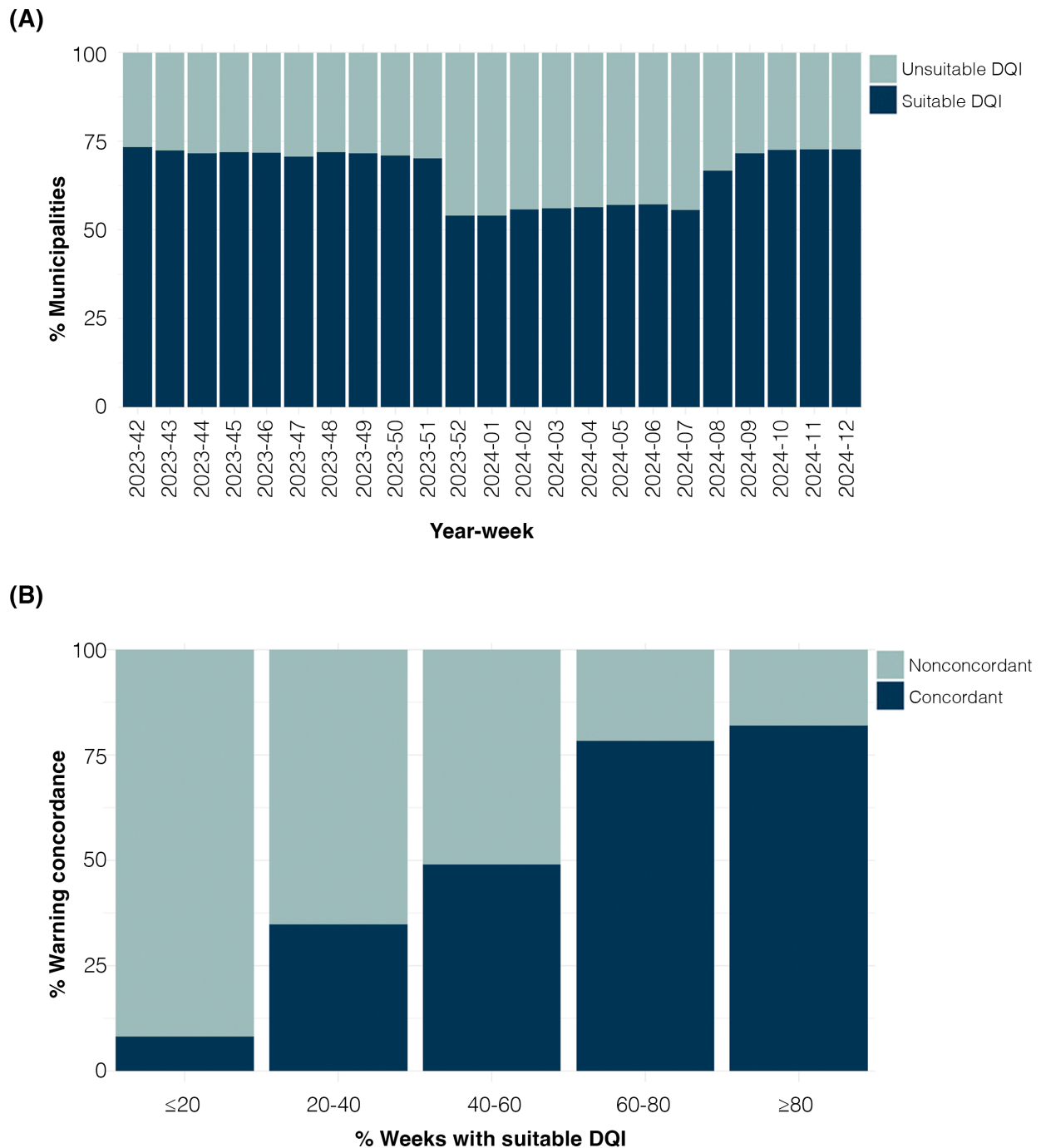
**Figure 1.** Primary Health Care encounters due to influenza-like illness per week in (A) Sao Paulo (municipality with a suitable data quality index in less than 60% of the 23 weeks in study period) and (B) Vitoria da Conquista (municipality with a suitable data quality index for over 80% of the study period). Plots show backfilled (gray line) and real-time (green line) PHC datasets for influenza-like illness (ILI) encounters. Vertical dashed lines show all detected warnings with the Early Aberration Reporting System (EARS), red triangles show nonconcordant warnings between backfield and real-time datasets, and blue circles show concordant warnings between them. ILI: influenza-like illness; PHC: Primary Health Care.



**Figure 2.** Proportion of concordant outbreak warnings detected in backfilled and real-time Brazilian Primary Health Care datasets. Outbreak warnings generated by the Early Aberration Reporting System (EARS) were identified in both backfilled and real-time datasets. Concordant warnings, detected in both datasets within the same week, are represented in dark blue, while nonconcordant warnings, identified only in the backfilled dataset, are shown in light blue. The analysis considers the proportion of concordant and nonconcordant warnings based on real-time dataset. (A) Completeness: the percentage of records from the real-time dataset in each 8-week rolling window (ranging from 0% to 100%) and (B) timeliness: the proportion of records registered within 2 weeks or less of the PHC encounter, measured within the same 8-week rolling window (ranging from 0% to 100%).



**Figure 3.** Proportion of Brazilian municipalities with suitable data quality index (DQI) and concordant warnings over time. (A) Weekly analysis of the proportion of municipalities with a suitable DQI from epidemiological week 42 of 2023 to week 12 of 2024. (B) Proportion of concordant warnings (dark blue), identified in both datasets within the same week, and nonconcordant warnings (light blue), detected only in the backfilled dataset. The analysis is based on the proportion of weeks with a suitable DQI in the real-time dataset from Brazilian municipalities.



## Discussion

### Principal Findings

Our study highlights the critical role of data quality in the performance of the EWS for infectious disease surveillance using PHC data. In addition, we provide a practical approach for monitoring data quality in real time, which can be adapted to other settings and data types. Our findings revealed that municipalities with over 60% of weeks featuring a suitable DQI had the highest proportion of concordant warnings between the backfilled and real-time datasets. Introducing the

DQI as an algorithm integrated into the EWS can guide data management practices and inform decision-making processes.

Similar to our findings, a recent systematic review of the effectiveness of EWS found that the improvement of data is pivotal for emergency department–based surveillance [17]. However, efforts for automatization of data quality assessment are typically scattered [9], and the literature on the operationalization of data quality assessment remains scarce. A study on data quality assessment for public health information systems found a lack of systematic procedures for quality assessment. While quality assessment of quantitative

data generally used descriptive surveys, the authors argued about the importance of systematic scientific data quality assessment [18]. To the best of our knowledge, this is the first publication to assess the importance of integrating data quality monitoring into an EWS.

Fulcher et al [19] demonstrated how administrative health data were successfully used to implement a syndromic surveillance system during the COVID-19 pandemic. However, the process of cleaning data and handling missed data was carried out by a dedicated analyst once the updated database became available [19]. We anticipate that the framework for a data quality assessment integrated to the EWS pipeline presented here can be adapted to other surveillance systems and can provide insights for similar implementations.

Using a retrospectively gathered, backfilled PHC dataset, we evaluated the EWS based on optimal data quality conditions. However, administrative data usually exhibit incompleteness and delays, and the EWS should be capable of detecting outbreaks using the available dataset in real time. Our analysis revealed that high levels of completeness (100%) and timeliness (at least 80%) are necessary to achieve the highest proportion of concordant warnings between backfilled and real-time datasets. Additionally, our results indicate that even incremental data quality improvements substantially enhance the EWS' performance. Achieving such high standards may pose challenges, particularly in low-resource settings that potentially face limitations due to infrastructure such as unreliable internet connectivity and insufficient computer power. Despite these challenges, we found a weekly median of 71% of Brazilian municipalities achieving the threshold for a suitable DQI for the EWS. This result suggests that a significant proportion of municipalities met the minimum threshold for data quality even in constrained settings.

In this study, we used the SUS database, which covers approximately 75% of the Brazilian population, with great granularity, reaching underserved rural and remote regions [20]. This approach allowed us to assess the performance of

the EWS across different regions and health service contexts. However, these findings may not be directly applicable to other countries. It is likely that the use of the EWS in different health system structures and data management practices will need adjustments and may require distinct data quality requirements [8].

Another limitation of this study is that we could not access other dimensions of data quality. Specifically, we could not access the accuracy of registers in the PHC dataset. Accuracy represents the extent to which the data are free of error and reliable [8,9]. We worked with aggregated, secondary data, and did not have access to the complete EHRs, which precluded us from verifying whether the diagnostic codes in the database accurately reflected patients' main clinical problems. It is our perspective that evaluating the accuracy of the *ICD-10* and *ICPC-2* is of great importance. However, given the large numbers of PHC encounters registered weekly, misclassifications of the reason of encounter are likely to be nondifferential. Additionally, syndromic surveillance systems are designed to operate effectively even with some level of imprecision, as their primary purpose is to detect patterns and trends rather than to provide definitive diagnoses.

## Conclusion

Our findings demonstrate that implementing a robust and integrated DQI analysis can significantly enhance the EWS' ability to detect ILI outbreaks, contributing to better public health outcomes and ultimately to global health security. Beyond contributing to the existing literature on EWS, this study highlights the importance of systematic data quality assessment. Continuous monitoring and improvement of data quality should be prioritized to ensure the reliability and effectiveness of surveillance systems. Additionally, our study suggests that similar frameworks can be adapted to different contexts. As health systems increasingly use digital health data for decision-making, our approach represents a model for integrating data quality monitoring into surveillance systems, ultimately enhancing the capacity to detect and respond to infectious disease outbreaks effectively.

## Acknowledgments

This study was funded by the Rockefeller Foundation's Health Initiative (grant 2023-PPI-007 awarded to MBN). MBN, PIPR, and VB are research fellows from the National Council for Scientific and Technological Development (CNPq, Brazil). TCS acknowledges funding from the Royal Society (NIFR1\231435). The funders did not interfere in the analysis, interpretation, or decision to submit the manuscript for publication. This study is part of the Alert-Early System of Outbreaks with Pandemic Potential (*ÆSOP*), an initiative under development by Brazil's Fundação Oswaldo Cruz (Fiocruz) and the Federal University of Rio de Janeiro, and financially supported by Rockefeller Foundation's Health Initiative. Figure S1 in [Multimedia Appendix 1](#) was created with Biorender.com and was published with permission. Generative artificial intelligence was not used for ideation or any part of the study design. Additionally, it was not used for reference searches. Its use was limited to grammatical revisions in the manuscript, with the prompt, "Please check for readability and possible grammatical mistakes" (a transcript of the conversation with the chatbot is provided in [Multimedia Appendix 2](#)).

## Data Availability

Our agreement with the Brazilian Ministry of Health (MoH) for accessing the referenced databases patently denies authorization of access to any third parties. All requests to access these databases must be addressed to the Brazilian MoH.

## Authors' Contributions

Conceptualization: PTVF, IM, and MBN

Data acquisition: VdAO

Data curation and processing: VdAO, JBJ, and GCGB

Formal analysis: PTVF

Script verification: JBJ, GCGB, and TCS

Study design: PTVF, JBJ, GCGB, TCS, VdAO, MHdOG, GOP, VB, PIPR, MBN, and IM

Writing—original draft: PTVF and IM

Writing—review and editing: GOP, TCS, VB, MHdOG, VdAO, and PIPR

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Additional material.

[\[DOCX File \(Microsoft Word File\), 142 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Conversation with the chatbot for grammatical revision.

[\[PDF File \(Adobe File\), 563 KB-Multimedia Appendix 2\]](#)

## References

1. Baker RE, Mahmud AS, Miller IF, et al. Infectious disease in an era of global change. *Nat Rev Microbiol*. Apr 2022;20(4):193-205. [doi: [10.1038/s41579-021-00639-z](https://doi.org/10.1038/s41579-021-00639-z)] [Medline: [34646006](https://pubmed.ncbi.nlm.nih.gov/34646006/)]
2. Morgan OW, Aguilera X, Ammon A, et al. Disease surveillance for the COVID-19 era: time for bold changes. *The Lancet*. Jun 19, 2021;397(10292):2317-2319. [doi: [10.1016/S0140-6736\(21\)01096-5](https://doi.org/10.1016/S0140-6736(21)01096-5)] [Medline: [34000258](https://pubmed.ncbi.nlm.nih.gov/34000258/)]
3. Data quality monitoring and surveillance system evaluation - a handbook of methods and applications. European Centre for Disease Prevention and Control. 2014. URL: <https://www.ecdc.europa.eu/sites/default/files/media/en/publications/Publications/Data-quality-monitoring-surveillance-system-evaluation-Sept-2014.pdf> [Accessed 2025-02-06]
4. 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*. Jan 9, 2024;10:e47673. [doi: [10.2196/47673](https://doi.org/10.2196/47673)] [Medline: [38194263](https://pubmed.ncbi.nlm.nih.gov/38194263/)]
5. Ulrich EH, So G, Zappitelli M, Chanchlani R. A review on the application and limitations of administrative health care data for the study of acute kidney injury epidemiology and outcomes in children. *Front Pediatr*. 2021;9:742888. [doi: [10.3389/fped.2021.742888](https://doi.org/10.3389/fped.2021.742888)] [Medline: [34778133](https://pubmed.ncbi.nlm.nih.gov/34778133/)]
6. Tassi MF, le Meur N, Stéfic K, Grammatico-Guillon L. Performance of French medico-administrative databases in epidemiology of infectious diseases: a scoping review. *Front Public Health*. 2023;11:1161550. [doi: [10.3389/fpubh.2023.1161550](https://doi.org/10.3389/fpubh.2023.1161550)] [Medline: [37250067](https://pubmed.ncbi.nlm.nih.gov/37250067/)]
7. Shaw RJ, Harron KL, Pescarini JM, et al. Biases arising from linked administrative data for epidemiological research: a conceptual framework from registration to analyses. *Eur J Epidemiol*. Dec 2022;37(12):1215-1224. [doi: [10.1007/s10654-022-00934-w](https://doi.org/10.1007/s10654-022-00934-w)] [Medline: [36333542](https://pubmed.ncbi.nlm.nih.gov/36333542/)]
8. Lewis AE, Weiskopf N, Abrams ZB, et al. Electronic health record data quality assessment and tools: a systematic review. *J Am Med Inform Assoc*. Sep 25, 2023;30(10):1730-1740. [doi: [10.1093/jamia/ocad120](https://doi.org/10.1093/jamia/ocad120)] [Medline: [37390812](https://pubmed.ncbi.nlm.nih.gov/37390812/)]
9. Ozonze O, Scott PJ, Hopgood AA. Automating electronic health record data quality assessment. *J Med Syst*. Feb 13, 2023;47(1):23. [doi: [10.1007/s10916-022-01892-2](https://doi.org/10.1007/s10916-022-01892-2)] [Medline: [36781551](https://pubmed.ncbi.nlm.nih.gov/36781551/)]
10. Bloland P, MacNeil A. Defining & assessing the quality, usability, and utilization of immunization data. *BMC Public Health*. Apr 4, 2019;19(1):380. [doi: [10.1186/s12889-019-6709-1](https://doi.org/10.1186/s12889-019-6709-1)] [Medline: [30947703](https://pubmed.ncbi.nlm.nih.gov/30947703/)]
11. Cerqueira-Silva T, Oliveira JF, de Araújo Oliveira V, 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(11):e00010024. [doi: [10.1590/0102-311XEN010024](https://doi.org/10.1590/0102-311XEN010024)] [Medline: [39775767](https://pubmed.ncbi.nlm.nih.gov/39775767/)]
12. Zhu Y, Wang W, Atrubin D, Wu Y. Initial evaluation of the early aberration reporting system--Florida. *MMWR Suppl*. Aug 26, 2005;54:123-130. [Medline: [16177703](https://pubmed.ncbi.nlm.nih.gov/16177703/)]
13. 2022 population census: main results. Brazilian Institute of Geography and Statistics. 2022. URL: <https://www.ibge.gov.br/en/statistics/social/health/22836-2022-census-3.html> [Accessed 2024-12-18]
14. Fox C, Levitin A, Redman T. The notion of data and its quality dimensions. *Inf Process Manag*. Jan 1994;30(1):9-19. [doi: [10.1016/0306-4573\(94\)90020-5](https://doi.org/10.1016/0306-4573(94)90020-5)]
15. Hassenstein MJ, Vanella P. Data quality—concepts and problems. *Encyclopedia*. Feb 2022;2(1):498-510. [doi: [10.3390/encyclopedia2010032](https://doi.org/10.3390/encyclopedia2010032)]



16. AESOP-data-documentation. GitHub. URL: <https://github.com/cidacslab/AESOP-Data-Documentation/tree/main/DataPipeline> [Accessed 2025-02-06]
17. Meckawy R, Stuckler D, Mehta A, Al-Ahdal T, Doebbeling BN. Effectiveness of early warning systems in the detection of infectious diseases outbreaks: a systematic review. *BMC Public Health*. Nov 29, 2022;22(1):2216. [doi: [10.1186/s12889-022-14625-4](https://doi.org/10.1186/s12889-022-14625-4)] [Medline: [36447171](https://pubmed.ncbi.nlm.nih.gov/36447171/)]
18. Chen H, Hailey D, Wang N, Yu P. A review of data quality assessment methods for public health information systems. *Int J Environ Res Public Health*. May 14, 2014;11(5):5170-5207. [doi: [10.3390/ijerph110505170](https://doi.org/10.3390/ijerph110505170)] [Medline: [24830450](https://pubmed.ncbi.nlm.nih.gov/24830450/)]
19. Fulcher IR, Boley EJ, Gopaluni A, et al. Syndromic surveillance using monthly aggregate health systems information data: methods with application to COVID-19 in Liberia. *Int J Epidemiol*. Aug 30, 2021;50(4):1091-1102. [doi: [10.1093/ije/dyab094](https://doi.org/10.1093/ije/dyab094)]
20. Macinko J, Harris MJ, Rocha MG. Brazil's National Program for Improving Primary Care Access and Quality (PMAQ): fulfilling the potential of the world's largest payment for performance system in primary care. *J Ambul Care Manage*. 2017;40 Suppl 2 Supplement(2 Suppl):S4-S11. [doi: [10.1097/JAC.000000000000189](https://doi.org/10.1097/JAC.000000000000189)] [Medline: [28252498](https://pubmed.ncbi.nlm.nih.gov/28252498/)]

## Abbreviations

**DQI:** data quality index

**EHR:** electronic health record

**EWS:** early warning system

**ICD-10:** *International Statistical Classification of Diseases, Tenth Revision*

**ICPC-2:** International Classification of Primary Care

**ILI:** influenza-like illness

**MoH:** Ministry of Health

**PHC:** Primary Health Care

**SISAB:** Sistema de Informação em Saúde para a Atenção Básica

**SUS:** Brazilian Unified Health System

**ÆSOP:** Alert-Early System of Outbreaks with Pandemic Potential

*Edited by Amaryllis Mavragani; peer-reviewed by Chinmay Haridas, Oluwatayo Olasunkanmi; submitted 30.09.2024; final revised version received 19.12.2024; accepted 20.12.2024; published 21.02.2025*

*Please cite as:*

*Florentino PTV, Bertoldo Junior J, Barbosa GCG, Cerqueira-Silva T, Oliveira VdA, Garcia MHdO, Penna GO, Boaventura V, Ramos PIP, Barral-Netto M, Marcilio I*

*Impact of Primary Health Care Data Quality on Infectious Disease Surveillance in Brazil: Case Study*

*JMIR Public Health Surveill 2025;11:e67050*

URL: <https://publichealth.jmir.org/2025/1/e67050>

doi: [10.2196/67050](https://doi.org/10.2196/67050)

© Pilar Tavares Veras Florentino, Juracy Bertoldo Junior, George Caique Gouveia Barbosa, Thiago Cerqueira-Silva, Vinicius de Araújo Oliveira, Marcio Henrique de Oliveira Garcia, Gerson Oliveira Penna, Viviane Boaventura, Pablo Ivan Pereira Ramos, Manoel Barral-Netto, Izabel Marcilio. Originally published in *JMIR Public Health and Surveillance* (<https://publichealth.jmir.org>), 21.02.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Public Health and Surveillance*, is properly cited. The complete bibliographic information, a link to the original publication on <https://publichealth.jmir.org>, as well as this copyright and license information must be included.