Outbreak, Surveillance, Investigation & Response (OSIR) Journal
Field Epidemiology Training Program, Division of Epidemiology
Department of Disease Control, Ministry of Public Health, Thailand
Tel: +6625903894, Fax: +6625903845, Email: osireditor@osirjournal.net, https://he02.tci-thaijo.org/index.php/OSIR

A Novel Surveillance Evaluation Approach Using Clinical Text Extraction from the Hospital Information System: Case Study of Somdejpraboromrachineenart Natawee Hospital, Songkhla Province, Southern Thailand for Influenza in 2024

Suralai Jongrakwong[1]*, Suphanat Wongsanuphat[2]

1 Somdejpraboromrachineenart Natawee Hospital, Ministry of Public Health, Thailand

2 Division of Epidemiology, Department of Disease Control, Ministry of Public Health, Thailand

*Corresponding author, email address: suralai.putto@gmail.com

Received: 16 Mar 2025; Revised: 15 Jun 2025; Accepted: 18 Jun 2025
https://doi.org/10.59096/osir.v18i2.274302

Abstract

In 2024, Na Thawi District, Songkhla Province, reported the highest influenza cases in Southern Thailand. This cross-sectional study evaluated the influenza surveillance system (R506) at Somdejpraboromrachineenart Natawee Hospital in 2024 using both quantitative and qualitative methods. Stakeholders involved in the epidemiological surveillance system were interviewed to describe the system qualitatively. Quantitatively, 8,758 medical records from the hospital information system (HIS) and 358 R506 reports were reviewed to assess sensitivity, positive predictive value (PPV), completeness, accuracy, and timeliness. The female-to-male ratios were 1.17:1 in HIS meeting the R506 definition and 1.08:1 in R506. Most cases were in the 25–60-year age group in HIS and 5–9-year group in R506. The lowest proportions were among those aged 60 years or more. Cases peaked in July; HIS showed a gradual rise from May, while R506 surged from June. Most cases occurred in Na Thawi and Chana districts. Subdistrict-level data showed consistent hotspots in Na Thawi, Sathon, and surrounding areas. The overall incidence in the area was higher than in the reporting system. The sensitivity was 8.52% and the PPV was 84.92% with R506 showing 100% completeness and accuracy, except for onset date (21.79%). Timeliness was high: 98.88% within 3 days and 99.72% within 7 days. From the qualitative study, the stakeholders accepted the surveillance system, describing it as simple, flexible, stable, and useful for planning and resource allocation. Clinical text extraction enabled full review without the need for sampling.

Keywords: influenza, surveillance system, hospital information system, electronic medical record, clinical text extraction, report 506

Introduction

Influenza is an acute viral respiratory infection causing fever, headache, muscle aches, and fatigue.[1] Global outbreaks occur frequently and contribute significantly to morbidity and mortality.[2] The virus is classified into types A, B, and C, with type A being the most common.[3] Transmission occurs through respiratory droplets and contact with contaminated surfaces. The incubation period is 1–3 days and individuals are infectious from one day before symptom onset to 3–5 days afterward in adults and up to seven days in children.[1,4]

In 2024, Thailand reported over 645,000 influenza cases and 49 deaths, primarily caused by type A.[5] The Southern Health Region 12 reported 42,697 cases, with Songkhla Province reporting the highest incidence. In Na Thawi District, which has a population of about 60,000, 358 cases were reported through its district hospital, with frequent outbreaks occurring in prisons, schools, and hospitals.[6,7]

Thailand's traditional passive surveillance system faces several challenges such as limited coverage, delays, and resource constraints, which hinder timely outbreak detection.[8,9] Recent advancements in information technology and data science have enabled improved data extraction from hospital information systems (HIS).[10–14] Manual record reviews remain time-consuming, while clinical text extraction enables faster and more accurate symptom identification from unstructured data.[15–17] Although this method has demonstrated high precision and recall (0.7–1.0), its use in surveillance evaluation is still limited—likely due to reliance on structured data, the complexity of natural language processing (NLP), and limited collaboration between informatics and public health professionals.[8,18–20] This study aimed to: 1) evaluate clinical text extraction for surveillance evaluation; 2) describe the influenza surveillance system qualitatively; and 3) assess quantitative characteristics using this method at Somdejpraboromrachineenart Natawee Hospital.

Methods

Study Overview

This mixed-methods study evaluated the influenza surveillance system at Somdejpraboromrachineenart Natawee Hospital using both quantitative and qualitative approaches. The quantitative component assessed system attributes—sensitivity, positive predictive value, data quality, timeliness, and representativeness—using data from the hospital information system (HIS) and Report 506 (R506) between 1 Jan 2024 and 31 Dec 2024. The qualitative component included semi-structured interviews with key stakeholders to explore system characteristics: acceptability, simplicity, flexibility, stability, and usefulness.[21]

Operational Definitions

We used two influenza case definitions to evaluate the surveillance system.

Influenza cases based on the R506 definition[22]

Any individual with fever (or a body temperature of 38°C or higher), cough, and one of the following symptoms: sore throat, runny nose, body aches, headache, or fatigue, who visited Somdejpraboromrachineenart Natawee Hospital, during 1 Jan to 31 Dec 2024.

Influenza cases based on Somdejpraboromrachineenart Natawee Hospital (hospital definition)

Any individual who met the R506 definition or had a positive result on a rapid influenza diagnostic test (presumptive diagnosis based on a nasopharyngeal, throat, or nasal swab), who visited Somdejpraboromrachineenart Natawee Hospital from 1 Jan to 31 Dec 2024.

Data Collection

For system description and qualitative evaluation, 11 stakeholders across three levels—executives, department heads, and operators—were interviewed. These included the hospital director, public health officers, physicians, nurses, epidemiologists, and information technologists. The interview focused on disease reporting, case screening, diagnosis, coding, reporting frequency, data analysis, and system feedback.

For quantitative characteristics, an overview of the data analysis process is illustrated in Figure 1. An initial manual review of 30 medical records in the electronic medical record system was conducted to understand the traditional review process and identify relevant variables for HIS extraction, including text-based fields. Additional records were reviewed as needed when new variables of interest were identified. Next, data were extracted from the surveillance system and the HIS. The current capabilities of these systems allow reports to be exported in Excel format, facilitating subsequent data utilization. Cases included in the study were those diagnosed with influenza-related conditions based on the international classification of diseases, 10th revision (ICD-10) (Table 1). Primary ICD-10 (influenza diagnosis group) codes for influenza diagnoses include J09, J100, J101, J108, J110, J111, and J118, while secondary codes (related to influenza) include J00, J029, and J069. Variables collected included hospital number (HN), demographic characteristics, diagnosis codes, patient type, visit and diagnosis dates, laboratory results, report date, and clinical texts such as chief complaint, present illness, past history, and physical examination.

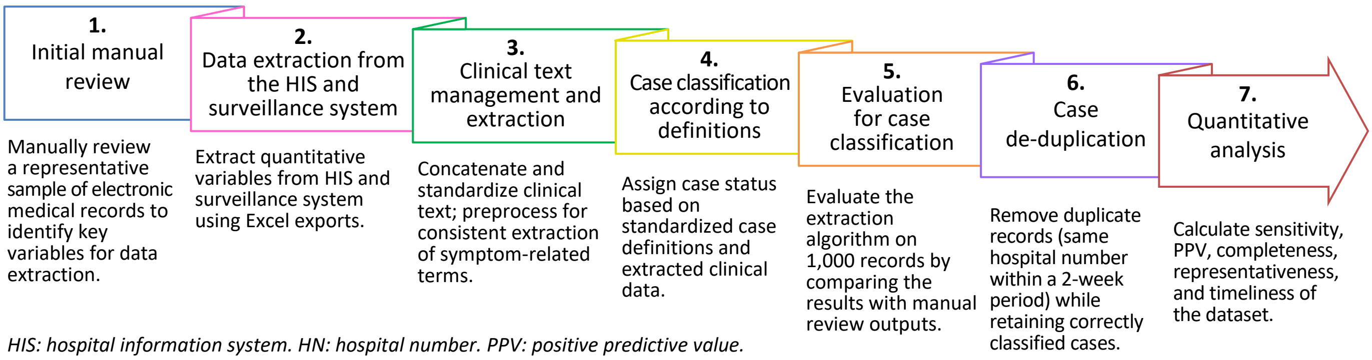

Figure 1. Quantitative data analysis process at Somdejpraboromrachineenart Natawee Hospital, 2024

Table 1. Characteristics of influenza cases by ICD-10 diagnosis at Somdejpraboromrachineenart Natawee Hospital, 2024

Diagnosis	ICD-10 code	Number of medical records	Medical records after deduplication	Met hospital definition n (%)	Met R506 definition n (%)	Reported in R506 (n)
Acute nasopharyngitis/ common cold	J00	6,366	4,374	2,432 (55.60)	2,281 (52.15)	25
Acute pharyngitis, unspecified	J029	391	309	137 (44.34)	127 (41.10)	-
Acute upper respiratory infection, unspecified	J069	297	724	499 (68.92)	484 (66.85)	-
Influenza due to identify avian flu virus	J09	8	6	2 (33.33)	2 (33.33)	-
Influenza due to other identified influenza virus with unspecified type of pneumonia	J100	57	45	40 (88.89)	39 (86.67)	-
Influenza with other respiratory manifestations, seasonal influenza virus identified	J101	194	143	124 (86.71)	124 (86.71)	87
Influenza with other manifestations, seasonal influenza virus identified	J108	272	203	171 (84.24)	169 (83.25)	123
Influenza with pneumonia, virus not identified	J110	18	13	12 (92.31)	12 (92.31)	4
Influenza with other respiratory manifestations, virus not identified	J111	113	91	74 (81.32)	73 (80.22)	27
Influenza with other manifestations, virus not identified	J118	413	358	266 (74.30)	256 (71.51)	92
Total		**8,758**	**6,266**	**3,757 (59.96)**	**3,567 (56.93)**	**358**

ICD-10: the international classification of diseases, 10th revision. R506: Report 506.

Data Management and Analysis

Stakeholder information was summarized to describe the system and create patient flow diagrams. Qualitative data were analyzed using thematic analysis based on acceptability, simplicity, flexibility, stability, and usefulness.

For quantitative analysis, clinical texts—including chief complaint, present illness, and physical exam—were concatenated using Excel's CONCAT() function. Text preprocessing addressed capitalization, spelling errors, and word choice inconsistencies (in Thai and English) for uniformity (Supplementary Table 1). Negation patterns (e.g., "no fever," "denied cough") were identified and standardized. Two doctors and one public health officer iteratively reviewed and refined the process on 500 records over five rounds to eliminate errors. Symptoms such as fever, cough, sore throat, runny nose, body aches, and fatigue were extracted using Excel's SEARCH() and ISNUMBER() functions and then compared against case definitions. For evaluation phase, we randomly selected 1,000 records (excluding 500 records in text pre-processing) to manually assess the accuracy of the clinical text extraction algorithm in identifying whether each record met or did not meet the case definition, compared against manual review of the raw clinical text.

After determining consistency with case definition, records with duplicate HN and visit date (i.e., the same diagnosis within a 2-week period) were identified. If the duplicates met the case definition, only the first occurrence was retained. Otherwise, the records were manually reviewed to determine the correct case classification, and the first matched case definition was retained. To calculate date of onset, duration of symptom was extracted from chief complaint and date of visit. The entire data set was managed and analyzed using Microsoft Excel version 16.78.3.

Sensitivity was defined as the percentage of cases who met the definition of influenza disease surveillance and were reported in the R506 system. It was calculated using the formula:

$$\text{Sensitivity} = \left(\frac{\text{Number of patients meeting the influenza surveillance definition and reported in the system}}{\text{Total number of patients meeting the influenza surveillance definition}} \right) \times 100$$

Positive predictive value (PPV) was defined as the percentage of cases reported in the R506 system who met the definition of influenza and was calculated using the formula:

$$PPV = \left(\frac{\text{Number of patients meeting the influenza definition in Report 506}}{\text{Total number of patients reported in the surveillance definition}} \right) \times 100$$

Completeness and accuracy of recorded variables in R506, including factors such as age (with an acceptable error margin of ±1 year), gender (male/female), and date of hospital admission (which should be on or before the reporting date) were evaluated. Timeliness was categorized into three levels: timely (within 0–3 days), moderately timely (within 4–7 days) and not timely (more than 7 days).

For representativeness, the characteristics of cases reported to R506 were compared with the characteristics of cases who met the R506 definition. The distribution were compared by person, time, and place, using variables such as gender, age, date of onset, place of onset including (province, district, and subdistrict).

Results

System Description

All cases were screened in the outpatient department during office hours and in the emergency room after hours. Nurses recorded symptoms and vital signs, and entered data into the HIS. Patients with influenza-like symptoms were directed to a separate area, instructed to wear masks, and sometimes underwent nasal swab testing. Physicians documented medical history, conducted physical examinations, and recorded diagnoses and ICD-10 codes in the HIS. Laboratory confirmation was not routinely performed. Most cases were treated as outpatients. Epidemiologists reviewed data and reported confirmed cases to the R506 system on working days, with retrospective reporting done after holiday periods. In the emergency room, similar procedures were followed. Laboratory-confirmed positive cases triggered alerts through the hospital's epidemic warning system, prompting verification and R506 reporting.

Hospitalized patients were isolated. Epidemiologists coordinated with local subdistrict hospitals for community control. ICD-10 codes were entered into the HIS within three days of discharge. Monthly data analysis was presented in hospital health meetings. The reporting process is shown in Figure 2.

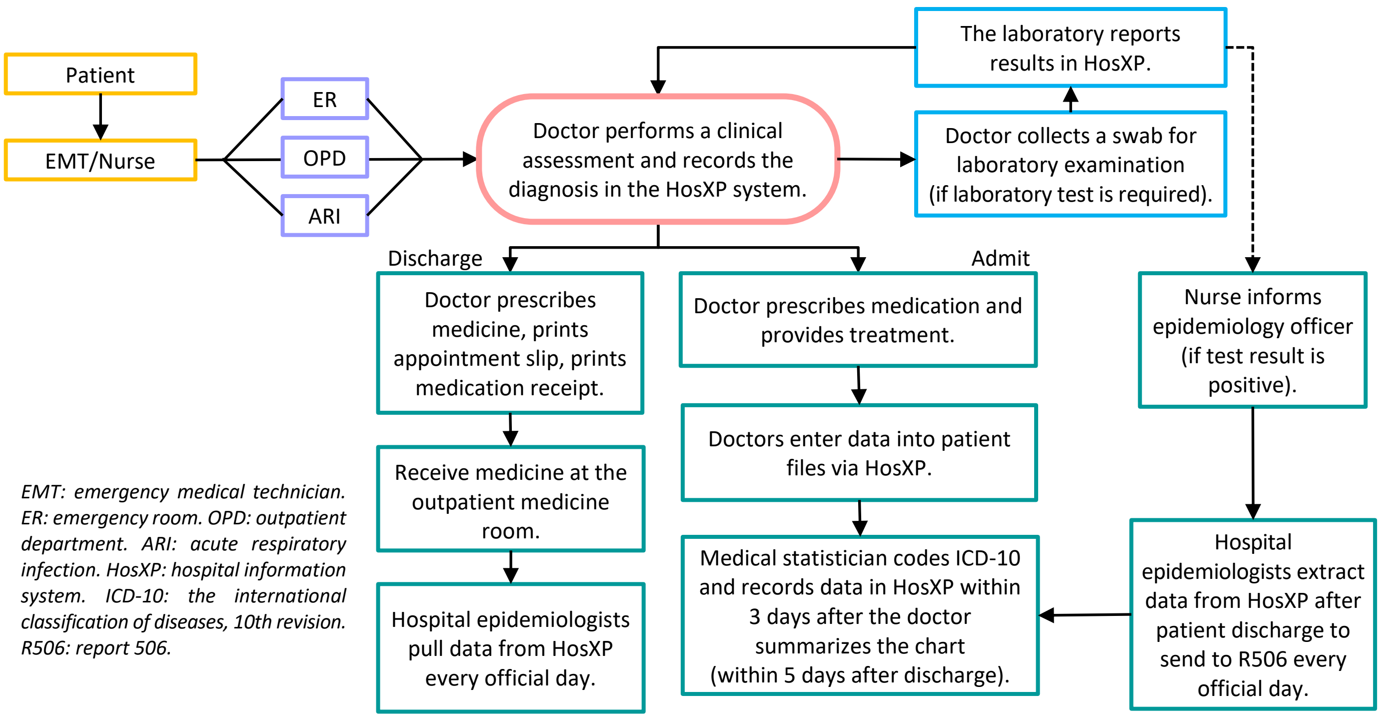

Figure 2. The process of reporting influenza disease at Somdejpraboromrachineenart Natawee Hospital, 2024

Clinical Text Management and Evaluation

We conducted five iterative rounds over three hours to refine negation handling and symptom search (Supplementary Table 1). From 8,258 HIS records based on ICD-10 codes (excluding 500 records in text pre-processing), 1,000 records were randomly selected for evaluation. Of these, 999 were classified as meeting (n=395) or not meeting (n=604) the R506 case definition consistently between using the rule-based extraction algorithm and manual review, yielding an accuracy of 99.9%. The sole misclassification occurred because the keyword "fever" referred to a patient's relative, not the patient, highlighting challenges in contextual interpretation.

Quantitative Characteristics

From 8,758 HIS records, we excluded 2,492 duplicates, leaving 6,266 for analysis, of which 358 were reported to R506. As shown in Table 1, secondary ICD-10 disease groups had more diagnoses than primary disease groups.

Sensitivity

A total of 3,567 cases met the R506 definition with 304 reported to R506 (sensitivity 8.52%). Among unreported cases, 616 (18.89%) had COVID-19 infection, 220 (6.74%) had a negative rapid influenza diagnostic test, and 2,878 (88.21%) were diagnosed using an ICD-10 secondary disease group. After excluding these three groups of cases, the sensitivity values were 10.30%, 9.08%, and 11.13%, respectively. Similarly, 3,757 cases met the hospital definition, with 325 reported to R506 (sensitivity 8.65%). Among unreported cases, 365 (10.64%) were diagnosed with the ICD-10 primary disease group, of which 145 had COVID-19 infection and 220 had a negative rapid influenza diagnostic test. After excluding these two groups of cases, the sensitivity values were 9.00% and 9.19%, respectively and the overall sensitivity was 9.58% (Table 2).

Table 2. Number of influenza cases identified by medical record review according to R506 and hospital case definitions at Somdejpraboromrachineenart Natawee Hospital, 2024

	Number of cases by medical record review			
	R506 definition		Hospital definition	
	Met	Not met	Met	Not met
Reported in R506 system				
Yes	304	54	325	33
No	3,263		3,432	

R506: Report 506.

Positive predictive value

For the R506 definition, 304 cases met the criteria (PPV 84.92%). Among the 54 individuals who did not meet the criteria (21 cases had a positive laboratory test result), 10 had no symptoms recorded (one due to a relative collecting the medication, nine were prisoners brought by prison officers), 10 had no fever, and 25 had no cough (eight had a positive laboratory result). Nine cases tested positive for influenza despite not having influenza symptoms. For hospital criteria, 325 cases met the criteria (PPV 90.78%). Among the 33 individuals who did not meet the criteria, 18 had no cough, nine had no fever, five had no associated symptoms, and one had no symptoms recorded (Table 2).

Completeness, accuracy and timeliness

All variables were recorded completely (100% completeness). For accuracy, all variables except the date of onset were recorded correctly (100% accuracy). The accuracy of "date of onset" was 21.79%. The distribution of reporting timeliness is shown in Figure 3. Of the 354 cases (98.88%) were reported within 3 days and 357 (99.72%) within 7 days. One case was reported 14 days after the date of diagnosis.

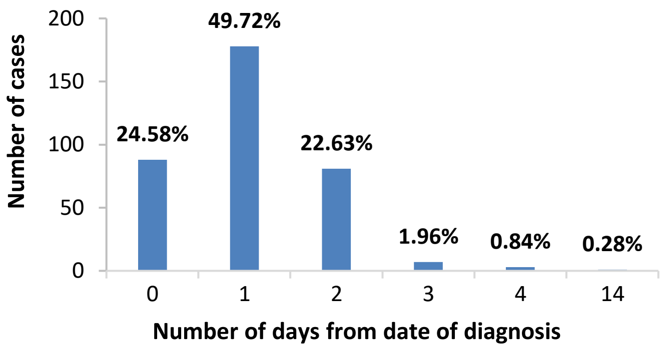

Figure 3. Distribution of time (days) between diagnosis and report among influenza cases reported to R506 at Somdejpraboromrachineenart Natawee Hospital, 2024

Representativeness

For R506 definition, the female to male ratios in HIS and R506 were 1.17:1 and 1.08:1, respectively. The 25–60-year age group had the highest proportion of cases (29.2% in HIS and 26.8% in R506). The age group of 60 years and older had the lowest proportion of cases (6.28% in HIS and 2.79% in R506) (Table 3).

Table 3. Characteristic of cases by R506 case definition and surveillance system at Somdejpraboromrachineenart Natawee Hospital, 2024

Characteristics	Met R506 definition (n=3,567) n (%)	R506 surveillance system (n=358) n (%)
Gender		
Female	1,915 (54.00)	185 (52.00)
Male	1,652 (46.00)	173 (48.00)
Age (years)		
0–4	727 (20.38)	70 (19.55)
5–9	678 (19.01)	89 (24.86)
10–14	447 (12.53)	56 (15.64)
15–24	449 (12.59)	37 (10.34)
25–60	1,042 (29.21)	96 (26.82)
60+	224 (6.29)	10 (2.79)

R506: Report 506.

Influenza cases peaked in July in both groups, with a high in January followed by a steady decline and then an increase from May to July. The number of cases reported to R506 increased sharply from June to July. Reporting was lower in May, June, and February, with no R506 cases in April, although some cases were recorded in HIS during that month (Figure 4).

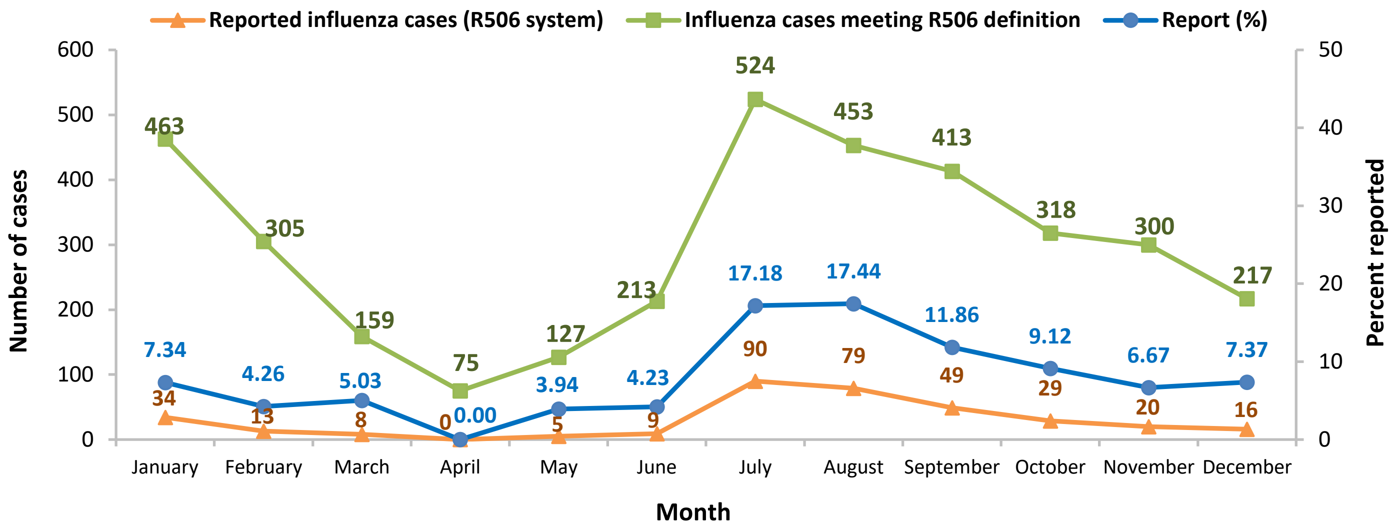

Figure 4. Number of cases per month and the proportion reported to the surveillance system at Somdejpraboromrachineenart Natawee Hospital, 2024

Figure 5 shows the distribution of reported influenza cases and cases meeting the R506 definition by subdistrict. Most cases were reported from Songkhla Province (99.72% for R506 and 98.07% for cases meeting the R506 definition). At the district level, reported R506 cases were concentrated in Na Thawi (72.1%), Chana (19.6%), and Thepha (3.91%), while cases meeting the R506 definition were concentrated in Na Thawi (69.5%), Chana (20.2%), and Saba Yoi (3.03%). Within Na Thawi District, the highest number of cases meeting the R506 definition were from Na Thawi (21.8%), Sathon (11.7%), and Khlong Sai (10.85%) subdistricts. The highest number of reported cases were from Na Thawi (22.1%), Thap Chang (13.9%), and Sathon (12.8%) subdistricts.

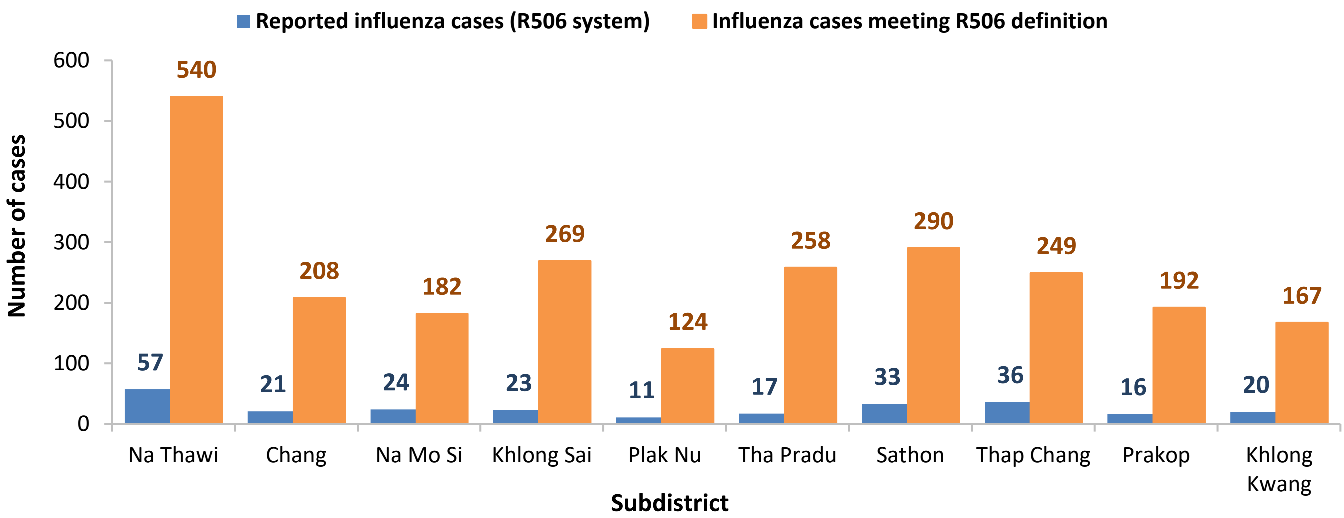

Figure 5. Distribution of reported influenza cases and cases meeting the R506 definition by subdistrict, Somdejpraboromrachineenart Natawee Hospital, 2024

Qualitative Characteristics

Simplicity

Stakeholders can efficiently use the HIS to retrieve influenza data quickly via ICD-10 codes, which are imported into the R506 surveillance system. Multiple reporting channels such as LINE and phone calls exist, though epidemiologists require time to screen and confirm cases.

Flexibility

Stakeholders can promptly update the system for ICD-10 changes or new diseases without requiring extra budget, allowing the system to adapt to similar diseases.

Acceptability

Executives and staff valued the system for outbreak detection and can cooperate in reporting. However, influenza reporting is only triggered by physician diagnosis, laboratory confirmation, or identified clusters, as influenza is not a high-priority policy disease.

Stability

The system benefits from strong leadership support and trained staff. Epidemiologists follow guidelines; however, backup coverage by subordinate staff is limited.

Usefulness

Data informs policy and budgeting and is shared externally for surveillance and outbreak alerts. However, feedback to hospital staff on outbreak status is lacking.

Discussion

A rule-based, iterative text extraction process, combining concatenation, spelling/negation handling, and symptom searches, achieved 99.9% accuracy in classifying records against R506 definitions. In other studies, this method enabled rapid review of over 8,000 records in one day, automating extraction of clinical symptoms and onset dates while improving processing speed and accuracy.[18,19] Despite its effectiveness, challenges remain, including variations in capitalization, spelling, word choice, time expressions, and negation in both Thai and English.[23] Addressing these issues requires advanced natural language processing, spelling correction, and support for medical terminology.[24,8]

The sensitivity of R506 and hospital definitions was 8.65% and 8.52%, respectively, which was lower than reported in prior studies (12.39–27.5%).[15–17] This may be due to the following reasons. First, physician diagnosis relies on clinical judgment, which may miss cases without positive results or correct ICD-10 coding. Second, epidemiologists determine which cases to report, often selecting the hospital definition when case numbers are high. This flexible interpretation can deviate from standardized surveillance criteria meant to guide outbreak detection and public health response.

The positive predictive values for R506 and hospital definitions were 84.92% and 90.87%, respectively, higher than reported in other studies (17.25–30.68%).[15–17] Our high values resulted from epidemiologists selecting cases based on ICD-10 codes and confirming symptoms and laboratory results prior to reporting. Among 54 reported cases that did not meet the R506 definition, 21 had positive laboratory results but lacked clinical symptoms. The hospital definition allows inclusion of asymptomatic, laboratory-confirmed cases. Laboratory confirmation was performed using the CiTEST influenza A+B rapid test, which has high sensitivity and specificity (99.0%/98.9% for influenza A and 98.8%/99.0% for influenza B).[25]

Data completeness was 100% for mandatory fields. Most variables were recorded accurately, except for the onset date, which was accurate in only 21.79% of records, although this value was within the reported range of 10.9–79.1% from other studies.[15,16] The inaccuracy was often due to staff recording the visit date instead. Previous research indicates higher accuracy in larger hospitals with more support from epidemiologists and information technologists.[15]

Timeliness was high. 98.88% and 99.72% of cases were reported within 3 and 7 days, respectively, due to daily reporting by epidemiologists. Delays occurred mainly during holidays, with only one case reported after 14 days.

For representativeness, the male-to-female ratio, date of onset, and place of onset aligned with their respective variables in the subset who met the definition. Although the magnitude does not reflect the real situation due to low sensitivity, the data remained representative in terms of gender, time trend, and location. The reporting proportion was lower in May, June, and February, as epidemic officers could choose to report based on the hospital or R506 definition. Case numbers were also low during this period, and no cases were reported in April, likely due to public holidays affecting reporting. The trend in case numbers did not show significant changes.

Our qualitative study findings showed strong system acceptability among executives and staff. The system was perceived as simple, flexible, and adaptable—ICD-10 codes could be updated without additional funding or external approval. Stability was supported by backup epidemiologists. However, some staff lacked awareness of reporting criteria. Surveillance data were actively used for policy planning and shared with external partners for outbreak detection.

Limitations

While the iterative rule-based process improved consistency in extracting clinical data, internal accuracy estimates do not substitute for external validation. No independent validation study comparing our algorithm to full traditional methods was conducted. Secondary data sources (HIS, medical records) may contain missing or misclassified information. During interviews, stakeholder responses might have been biased by the presence of department heads; however, confidentiality assurances and separate interviews partly mitigated this potential bias. Random sampling across departments and levels was conducted to represent a wide range of perspectives but we did not include all staff.

Recommendations

Epidemiologists can effectively evaluate influenza surveillance using clinical text extraction with predefined search terms, adapted to local documentation styles. Training healthcare staff on surveillance definitions and improving symptom onset documentation are essential. Continued disease reporting with timely feedback and random strain testing will strengthen preparedness. Automating laboratory-based reporting ensures accuracy and reduces discrepancies. Collaboration among national and local health agencies is critical to establish a rapid test–based alert system, enhancing verification, reducing workload, and ensuring consistent reporting to R506.

The Division of Epidemiology should develop tools to extract symptoms and onset dates from clinical texts, including chief complaints and physical exams. Clinical criteria should be refined—such as reassessing the requirement for cough alongside fever—since 6% of RIDTs-positive cases lacked cough.[26] Reporting asymptomatic but laboratory-confirmed influenza cases is also critical, given the potential for transmission before and after symptoms, especially in children where viral shedding may last over seven days.[27,28]

Conclusion

The 2024 evaluation of the influenza surveillance system at Somdejpraboromrachineenart Natawee Hospital showed that actual influenza incidence exceeded reported figures. Case identification by epidemiologists using selective criteria may have introduced bias. Despite low sensitivity, the system demonstrated high positive predictive value, accuracy, and data completeness. It was adequately representative across person, place, and time, with timely reporting. Stakeholders described the system as simple, flexible and stable. Clinical text extraction via Microsoft Excel allowed complete medical record review without sampling, supporting effective surveillance evaluation and health planning.

Acknowledgements

We sincerely thank Dr. Suwat Wiriyapongsukit, Director of Somdejpraboromrachineenart Natawee Hospital, for his support. Thanks to Seree Lamoh for assistance with interviews. The guidance from Dr. Kittiphan Chalom, Dr. Ratchadapon Phabwijitsilp, and the Field Epidemiology and Management Training is much appreciated.

Author Contributions

Suralai Jongrakwong: Conceptualization, methodology, data curation, formal analysis, writing—original draft, writing—review, & editing. **Suphanat Wongsanupat:** Methodology, writing—review & editing, validation, supervision.

Ethical Approval

This study was approved by the Provincial Human Research Ethics Committee of Songkhla Province (approval no. 57/2568), dated 24 Mar 2025.

Informed Consent

Data from the HIS were used with strict confidentiality and without disclosure of personal details. For interviews, the study objectives were explained to participants, and informed consent was obtained prior to participation.

Data Availability

The data used in this study were obtained from the Hospital Information System (HIS) and Report 506 (R506) database at Somdejpraboromrachineenart Natawee Hospital. These data contain sensitive patient information and are not publicly available to protect patient confidentiality. Requests for access to the data may be considered on a case-by-case basis and require approval from the hospital's ethics committee and relevant authorities.

Conflicts of Interest

The authors declare no conflict of interest.

Funding Support

The authors received no specific funding for this study.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used ChatGPT to enhance clarity and correct grammatical errors, and used Napkin to create graphics. The content produced by this tool was reviewed and edited by the authors, who accept full responsibility for the final text.

Suggested Citation

Jongrakwong S, Wongsanuphat S. A novel surveillance evaluation approach using clinical text extraction from the hospital information system: case study of Somdejpraboromrachineenart Natawee Hospital, Songkhla Province, Southern Thailand for influenza in 2024. OSIR. 2025 Jun;18(2):96–105. doi:10.59096/osir.v18i2.274302.

References

1. World Health Organization. Influenza (seasonal) [Internet]. Geneva: World Health Organization; 2023 [cited 2025 Mar 8]. <https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)>

2. Thompson WW, Weintraub E, Dhankhar P, Cheng PY, Brammer L, Meltzer MI, et al. Estimates of US influenza-associated deaths made using four different methods. Influenza Other Respir Viruses. 2009 Jan;3(1):37–49. doi:10.1111/j.1750-2659.2009.00073.x.

3. Department of Disease Control, Ministry of Public Health. Influenza [Internet]. Nonthaburi: Department of Disease Control; [cited 2024 Dec 3]. <https://ddc.moph.go.th/disease_detail.php?d=13>. Thai.

4. Centers for Disease Control and Prevention (US). How flu spreads [Internet]. Atlanta: Centers for Disease Control and Prevention; 2021 [cited 2024 Dec 3]. <https://www.cdc.gov/flu/about/disease/spread.htm>

5. Department of Disease Control, Ministry of Public Health. Influenza situation [Internet]. Nonthaburi: Department of Disease Control; [cited 2024 Dec 16]. <https://ddc.moph.go.th/doe/pagecontent.php?page=607&dept=doe>. Thai.

6. Department of Disease Control, Ministry of Public Health. Influenza situation report [Internet]. Nonthaburi: Department of Disease Control; [cited 2025 Jan 3] <https://dvis3.ddc.moph.go.th/t/DDC_CENTER_DOE/views/DDS2/sheet33?:embed=y&:isGuestRedirectFromVizportal=y>. Thai.

7. Na Thawi District Public Health Office. Influenza situation 2024: Program R506. Songkhla: Na Thawi District Public Health Office; 2024.

8. Weeks HL, Beck C, McNeer E, Williams ML, Bejan CA, Denny JC, et al. medExtractR: a targeted, customizable approach to medication extraction from electronic health records. J Am Med Inform Assoc. 2020 Mar 8;27(3):407–18. doi:10.1093/jamia/ocz207.

9. Rotejanaprasert C, Armatrmontree P, Chienwichai P, Maude RJ. Perspectives and challenges in developing and implementing integrated dengue surveillance tools and technology in Thailand: a qualitative study. PLoS Negl Trop Dis. 2024 Aug 14;18(8):e0012387. doi:10.1371/journal.pntd.0012387.

10. Neira RAQ, de Vries GJ, Caffarel J, Stretton E. Extraction of Data from a Hospital Information System to Perform Process Mining. Stud Health Technol Inform. 2017;245:554–8.

11. Ong TC, Kahn MG, Kwan BM, Brandt P, Folck B, Marquard J. Dynamic-ETL: a hybrid approach for health data extraction, transformation and loading. BMC Med Inform Decis Mak. 2017;17:134. doi:10.1186/s12911-017-0532-3.

12. Wongsanuphat S, Sangwanloy S, Sopha P, Buangsuang W, Denduang S, Thongplean A, et al. Enhancing coronavirus disease (COVID-19) surveillance system through information technology, Thailand, 2020. OSIR. 2020 Sep;13(3):90–100. doi: 10.59096/osir.v13i3.262804.

13. Wongsanuphat S, Sangwanloy S, Sopha P, Buangsuang W, Denduang S, Thongplean A, et al. Evaluation of enhancing the information system for COVID-19 surveillance in Thailand, 2020: A pre-post intervention comparison. OSIR. 2020 Sep;13(3):101–9. doi:10.59096/osir.v13i3.262806.

14. Wongsanuphat S, Malaikham J, Suriya S, Prommongkhol J, Khampha N, Khempetch T. Enhancing digital disease surveillance in Thailand using information technology, data engineering, data science, and artificial intelligence. OSIR. 2025 Mar;18(1):1–10. doi:10.59096/osir.v18i1.272056.

15. Watcharaporn N, Nutthien T, Dis Wises W, Kaewprasit P, Sirithavorn SA. Study of influenza surveillance system and policy proposals for developing the response to emerging respiratory infectious diseases in Bangkok area in 2016. Public Health Academic Journal. 2017;26:63–72.

16. Chobkatanyoo A, Sangsawang C, Thepparat T, Kaesaman S, Malae S. Influenza surveillance system evaluation in Narathiwat Province, Thailand, 2019. Weekly Epidemiological Surveillance Report [Internet]. 2021 [cited 2025 Mar 8];52:625–33. <https://he05.tci-thaijo.org/index.php/WESR/article/view/1762>

17. Thongkham J. Evaluation of the influenza surveillance system health region 8, fiscal year 2020. The Office of Disease Prevention and Control 8 Udon Thani Jounal [Internet]. 2023 [cited 2025 Mar 9];1(3):20–32. <https://he03.t

ci-thaijo.org/index.php/JODPC8/article/view/497/833>

18. Barrett N, Weber-Jahnke JH, Thai V. Engineering natural language processing solutions for structured information from clinical text: extracting sentinel events from palliative care consult letters. Stud Health Technol Inform. 2013;192:594–8. doi:10.3233/978-1-61499-289-9-594.
19. Mincheva Z, Vasilev N, Nikolov V, Antonov A. Extracting structured data from text in natural language. International Journal of Intelligent Information Systems. 2021;10(4):74–80. doi:10.11648/j.ijiis.20211004.16.
20. Montgomery MP, Morris SE, Rolfes MA, Kittikraisak W, Samuels AM, Biggerstaff M, et al. The role of asymptomatic infections in influenza transmission: what do we really know. Lancet Infect Dis. 2024 Jun;24(6):e394–404. doi:10.1016/S1473-3099(23)00619-9.
21. German RR, Lee LM, Horan JM, Milstein RL, Pertowski CA, Waller MN, et al. Updated guidelines for evaluating public health surveillance systems: recommendations from the Guidelines Working Group. MMWR Recomm Rep. 2001 Jul 27;50(RR-13):1–35.
22. Division of Epidemiology, Department of Disease Control. Case definitions and reporting guidelines for dangerous communicable diseases and diseases under surveillance in Thailand [Internet]. Nonthaburi: Department of Disease Control; 2020 [cited 2025 May 22]. 500 p. <https://klb.ddc.moph.go.th/dataentry/handbook/form/113>. Thai.
23. Miller TA, Bethard S, Dligach D, Lin C, Savova GK. Extracting time expressions from clinical text. Proceedings of the 2015 Workshop on Biomedical Natural Language Processing; 2015 Jul 30; Beijing, China. Kerrville (TX): Association for Computational Linguistics; 2015; 11 p. doi:10.18653/v1/w15-3809.
24. Lai KH, Topaz M, Goss FR, Zhou L. Automated misspelling detection and correction in clinical free-text records. J Biomed Inform. 2015 Jun:55:188–95. doi:10.1016/j.jbi.2015.04.008.
25. Citest Diagnostics Inc. Influenza A+B rapid test cassette [Internet]. Vancouver: Diagnostics Inc; [cited 2025 Mar 9] <https://www.citestdiagnostics.com/Home/ProductInfo/232>
26. Muller AM, Schmohl KA, Knoop K, Schug C, Urnauer S, Hagenhoff A, et al. Hypoxia-targeted 131I therapy of hepatocellular cancer after systemic mesenchymal stem cell-mediated sodium iodide symporter gene delivery. Oncotarget. 2016 Aug 23;7(34):54795-54810. doi:10.18632/oncotarget.10758.
27. Freitas FTM, Cabral APS, Barros ENC, Burigo MJO, Prochnow RD, Silva LA, et al. Pre-symptomatic transmission of pandemic influenza H1N1 2009: investigation of a family cluster, Brazil. Epidemiol Infect. 2013 Apr;141(4):763–6. doi:10.1017/S0950268812001501.
28. Kreimeyer K, Foster M, Pandey A, Arya N, Halford G, Jones SF, et al. Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review. J Biomed Inform. 2017 Sep;73:14–29. doi:10.1016/j.jbi.2017.07.012.