Natural Language Processing vs Diagnosis Code-Based Methods for Postherpetic Neuralgia Identification: Development and Validation in Real-World Data

Chengyi Zheng, Bradley Ackerson, Sijia Qiu, Lina S Sy, Leticia I. Vega Daily, Jeannie Song, Lei Qian, Yi Luo, Jennifer H. Ku, Yanjun Cheng, Jun Wu, Hung Fu Tseng

Submitted to: JMIR Medical Informatics
on: March 01, 2024

Table of Contents

Natural Language Processing vs Diagnosis Code-Based Methods for Postherpetic Neuralgia Identification: Development and Validation in Real-World Data

Chengyi Zheng[1] PhD, MS; Bradley Ackerson[2] MD; Sijia Qiu[1] MS; Lina S Sy[1] MPH; Leticia I. Vega Daily[1] MSW; Jeannie Song[1] MPH; Lei Qian[1] PhD; Yi Luo[1] PhD; Jennifer H. Ku[1] PhD, MPH; Yanjun Cheng[1] MS; Jun Wu[1] MD, MS; Hung Fu Tseng[1, 3] PhD, MPH

[1]Department of Research & Evaluation Kaiser Permanente Southern California Pasadena US
[2]South Bay Medical Center, Kaiser Permanente Southern California Harbor City US
[3]Kaiser Permanente Bernard J. Tyson School of Medicine Pasadena US

Corresponding Author:
Chengyi Zheng PhD, MS
Department of Research & Evaluation
Kaiser Permanente Southern California
100 S Los Robles Ave, 2nd Floor
Pasadena
US

Abstract

Background: Diagnosis codes and prescription data are used in algorithms to identify postherpetic neuralgia (PHN), a debilitating complication of herpes zoster (HZ). Because of the questionable accuracy of codes and prescription data, manual chart review is sometimes used to identify PHN in electronic health records (EHR), which can be costly and time-consuming.

Objective: To develop and validate a natural language processing (NLP) algorithm for automatically identifying PHN from unstructured EHR data. To compare its performance with that of code-based methods.

Methods: This retrospective study used EHR data from Kaiser Permanente Southern California, a large integrated healthcare system that serves over 4.8 million members. The source population included members aged ?50 years who received an incident HZ diagnosis and accompanying antiviral prescription between 2018-2020 and had ?1 encounter within 90-180 days of the incident HZ diagnosis. The study team manually reviewed the EHR and identified PHN cases. For NLP development and validation, 500 and 800 random samples from the source population were selected, respectively. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F-score, and Matthews correlation coefficient (MCC) of NLP and the code-based methods were evaluated using chart-reviewed results as the reference standard.

Results: The NLP algorithm identified PHN cases with 90.9% sensitivity, 98.5% specificity, 82.0% PPV, and 99.3% NPV. The composite scores of the NLP algorithm were 0.89 (F-score) and 0.85 (MCC). The prevalences of PHN in the validation data were 6.9% (reference standard), 7.6% (NLP), and 5.4-13.1% (code-based). The code-based methods achieved 52.7-61.8% sensitivity, 89.8-98.4% specificity, 27.6-72.1% PPV, and 96.3-97.1% NPV. The F-scores and MCCs were ranged between 0.45-0.59 and 0.32-0.61, respectively.

Conclusions: The automated NLP-based approach identified PHN cases from the EHR with good accuracy. This method could be useful in population-based PHN research.

(JMIR Preprints 01/03/2024:57949)
DOI: https://doi.org/10.2196/preprints.57949

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Natural Language Processing vs Diagnosis Code-Based Methods for Postherpetic Neuralgia Identification: Development and Validation in Real-World Data

Running title:

Postherpetic Neuralgia Identification in Real-World Data

Chengyi Zheng, PhD, MS[1], Bradley Ackerson, MD[2], Sijia Qiu, MS[1], Lina S. Sy, MPH[1], Leticia I. Vega Daily, MSW[1], Jeannie Song, MPH[1], Lei Qian, PhD[1], Yi Luo, PhD[1], Jennifer H. Ku, PhD, MPH[1], Yanjun Cheng, MS[1], Jun Wu, MD, MS[1], Hung Fu Tseng, PhD, MPH[1,3]

[1] Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, California, USA

[2] South Bay Medical Center**,** Kaiser Permanente Southern California, Harbor City, California, USA

[3] Kaiser Permanente Bernard J. Tyson School of Medicine, Pasadena, California, USA

Corresponding author:

Chengyi Zheng

Department of Research and Evaluation, Kaiser Permanente Southern California

100 S Los Robles Ave, 2nd Floor, Pasadena, CA 91101

Telephone: (626) 986-8665; Fax: (626) 564-7872; Email: Chengyi.X.Zheng@kp.org

Abstract

Background:

Diagnosis codes and prescription data are used in algorithms to identify postherpetic neuralgia (PHN), a debilitating complication of herpes zoster (HZ). Because of the questionable accuracy of codes and prescription data, manual chart review is sometimes used to identify PHN in electronic health records (EHR), which can be costly and time-consuming.

Objective:

To develop and validate a natural language processing (NLP) algorithm for automatically identifying PHN from unstructured EHR data. To compare its performance with that of code-based methods.

Methods:

This retrospective study used EHR data from Kaiser Permanente Southern California, a large integrated healthcare system that serves over 4.8 million members. The source population included members aged ≥50 years who received an incident HZ diagnosis and accompanying antiviral prescription between 2018-2020 and had ≥1 encounter within 90-180 days of the incident HZ diagnosis. The study team manually reviewed the EHR and identified PHN cases. For NLP development and validation, 500 and 800 random samples from the source population were selected, respectively. The sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F-score, and Matthews correlation coefficient (MCC) of NLP and the code-based methods were evaluated using chart-reviewed results as the reference standard.

Results:

The NLP algorithm identified PHN cases with 90.9% sensitivity, 98.5%

specificity, 82.0% PPV, and 99.3% NPV. The composite scores of the NLP algorithm were 0.89 (F-score) and 0.85 (MCC). The prevalences of PHN in the validation data were 6.9% (reference standard), 7.6% (NLP), and 5.4-13.1% (code-based). The code-based methods achieved 52.7-61.8% sensitivity, 89.8-98.4% specificity, 27.6-72.1% PPV, and 96.3-97.1% NPV. The F-scores and MCCs were ranged between 0.45-0.59 and 0.32-0.61, respectively.

Conclusions:

The automated NLP-based approach identified PHN cases from the EHR with good accuracy. This method could be useful in population-based PHN research.

Keywords

postherpetic neuralgia; herpes zoster; natural language processing; electronic health record; real-world data; artificial intelligence; development; validation; diagnosis; EHR; algorithm; EHR data; sensitivity; specificity; validation data; neuralgia; recombinant zoster vaccine

Introduction

Herpes zoster (HZ) or shingles is a painful dermatomal vesicular disease that results from the reactivation of latent varicella-zoster virus in the nerve ganglia. [1] Nearly all adults have the varicella-zoster virus dormant in their nervous system, [2] and the estimated lifetime risk of HZ was approximately 30% prior to the availability of zoster vaccine. [3] HZ usually begins with a prodromal stage of discomfort, followed by a painful, itchy rash on one unilateral dermatome that lasts two to four weeks. [4] HZ patients may develop postherpetic neuralgia (PHN), dermatomal pain persisting at least 90 days after the appearance of the acute HZ rash. [3, 5] PHN is the most common complication of HZ and greatly lowers patients' quality of life. [3]

Population-based studies using real-world data are cost-effective ways to address many questions about PHN. [3] However, accurately identifying PHN is difficult. Clinical trials rely on predetermined follow-up visits, which are difficult to replicate in real-world settings. [6, 7] Due to time and resource constraints, prospective studies have mainly been limited to hundreds of HZ patients and smaller numbers of PHN cases. [3] Retrospective studies of PHN have relied heavily on diagnosis codes, [8-13] which lack accuracy, [3, 14] or manual chart review, [14-16] which is costly and time-consuming. Moreover, despite the widespread use of code-based algorithms, only a few publications included PHN algorithm validation results. [8, 10]

Natural language processing (NLP), a subfield of artificial intelligence, has been

used to identify and extract information from unstructured clinical data. We previously developed NLP methods to identify herpes zoster ophthalmicus (HZO) and HZO with eye involvement, which are also common HZ complications. [17, 18] In this study, we developed and validated an NLP algorithm to identify PHN. Using manual chart-reviewed results as a reference standard, we compared the performance of the NLP algorithm with that of five previously published code-based algorithms.

Methods

Setting

This study was conducted at Kaiser Permanente Southern California (KPSC), an integrated healthcare system with 16 hospitals and 197 medical offices that serves over 4.8 million members. The prepaid health plan incentivizes members to use services at KPSC facilities. The EHR system at KPSC stores all aspects of member care, including sociodemographic characteristics, medical encounters, diagnoses, laboratory tests, pharmacy use, immunization records, membership history, and billing and claims.

PHN Case Definition

PHN was defined as pain/discomfort consistent with the HZ episode ≥90 days after the initial HZ diagnosis; the symptoms were at the location of the initial HZ rash and were not due to other obvious causes. [19-21]

Datasets

This study utilized EHR data of patients aged ≥50 years who each had an incident HZ diagnosis and associated antiviral prescription between 2018 and 2020 at KPSC. All patients had to have at least 1 year of membership prior to the index (incident HZ diagnosis) date so that comorbidities and healthcare utilization could be ascertained. Among patients with ≥1 encounter during the 90-180 days after the incident HZ diagnosis, trained research associates reviewed their EHRs based on the PHN abstraction instructions (Multimedia Appendix 1). An infectious disease physician (BKA) reviewed all possible or unclear cases. From these reviewed cases, we randomly selected 500 cases for NLP development and 800 cases for NLP validation. Because the NLP work was done concurrently with the manual review, the development dataset was collected at an earlier stage, when the reviewed cohort had a greater proportion of Asian and recombinant zoster vaccine (RZV) vaccinated patients.

Reference Standard

Among the 800 cases in the validation dataset, BKA reviewed 37 HZ cases that research associates had identified as unclear PHN cases. Because reviewers sometimes missed positive mentions of PHN, BKA re-reviewed cases in the validation set where NLP results differed from reviewer results. Nine cases were corrected from negative to positive PHN. These manually reviewed results served as the reference standard for assessing the performance of PHN identification algorithms.

NLP Algorithm Development

We developed the NLP algorithm based on our previous work. [17, 18, 22-26] Multimedia Appendix 2 describes the steps for pre-processing text and generating nomenclature. We created the rule-based NLP algorithm using the Linguamatics I2E software (Linguamatics, an IQVIA company, Cambridge, United Kingdom). Each note was searched at different levels: section (e.g., “Physical Exam”, “Assessment/Plan”), cross-sentence, intra-sentence, and phrase. A distance-based relationship algorithm was applied to identify related terms based on the number of words or sentences between them. The relationship search identified the words or phrases (e.g., negated, uncertain, and hypothetical statements) that modified the concepts of interest.

Figure 1 depicts an overview of the NLP algorithm. We separated the extracted clinical texts into three time periods: index (acute HZ) period (-7 to 21 days from incident HZ diagnosis date), transitional (subacute HZ) period (22 to 89 days), and risk (defined PHN) period (90 to 180 days). We developed search queries to identify the HZ anatomic locations in the index episode and PHN-related evidence in the transitional and risk periods. Supporting evidence of PHN included explicit mention of ongoing PHN, symptom location and causality, and PHN listed in the assessment and plan section. Counterevidence of PHN included differential diagnoses, recurrent HZ, and resolved PHN. We excluded sections and statements that may have been copied forward as historical information.

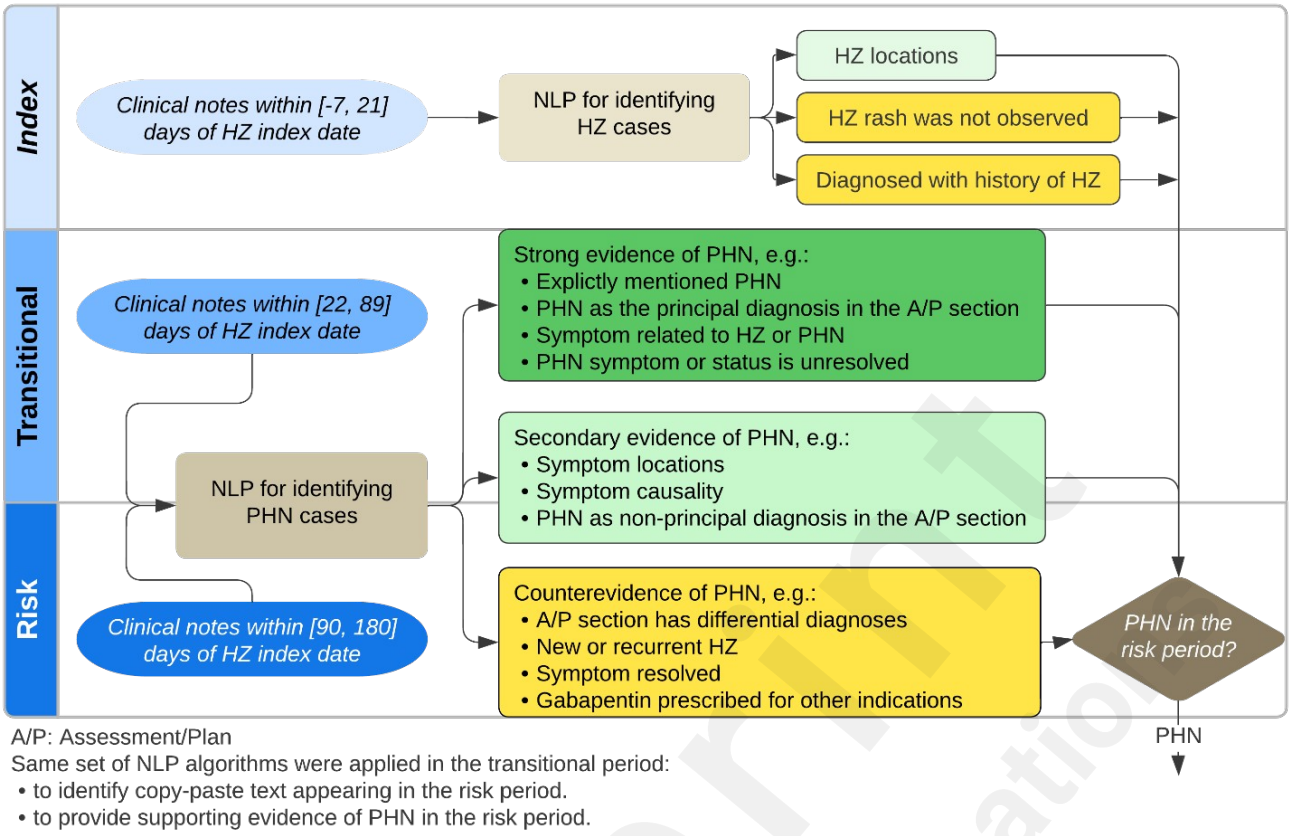


Figure 1. Diagram of natural language processing algorithm.

The PHN decision algorithm was implemented in Python language, which incorporated the evidence from the NLP search queries and classified each case based on decision rules. To exclude the copy-pasted results, the NLP program ran search queries on both the transitional and risk periods and compared the results to locate identical sequences of text. The algorithm considered the time sequences of identified evidence. The symptom location during the risk period was compared to the index HZ location. Because adjacent dermatomes might be difficult to distinguish clinically, symptom location during the index and risk periods had to occur in the same or surrounding dermatomes (e.g., face and neck). Based on the development dataset, we tested and updated the algorithm.

Implementation of Published PHN Identification Algorithms

We selected and implemented five code-based PHN identification algorithms based on the variety of their algorithms, the journal category and impact factor, the publication year, the total citations, and the size of the study (Table 1). The first code-based method (C1: Yanni et al.) exclusively used PHN-related diagnosis codes (Multimedia Appendix 3). [27] The remaining four algorithms (C2: Klompas et. al., [8] C3: Klein et al., [10] C4: Forbes et al., [9] C5: Munoz-Quiles et al. [11]) used additional structured data, such as diagnosis codes for HZ, neuralgia, and chronic pain; prescriptions for analgesics, antidepressants and anticonvulsants; and clinical visit data.

Table 1. List of Sources for Selected Code-Based Methods.

Method	HZ cases (n)	Journal category	Journal (IF[a])	Year	TC[b]
C1 [27]	21,146	General medicine	BMJ Open (2.9)	2018	67
C2 [8]	2,089	General medicine	Mayo Clinic Proceeding (8.9)	2011	125
C3 [10]	62,205	Immunology	Vaccine (5.5)	2019	32
C4 [9]	119,413	Neurology	Neurology (10.1)	2016	155
C5 [11]	87,086	Infectious diseases	Journal of Infection (28.2)	2018	38

[a]IF: impact factor.
[b]TC: total citations based on Google Scholar as of July 1, 2024.

Validation and Analysis

The results generated from the various algorithms were evaluated against the chart-reviewed reference standard validation dataset. We counted the numbers of true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN) cases to calculate the performance metrics: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F-score [28], and Matthews correlation coefficient (MCC) [29].

The F-score is a combination metric in machine learning and NLP research. It is defined as a weighted harmonic mean of sensitivity and PPV, where the

parameter β represents the relative importance of sensitivity vs. PPV.

$$F-score = \frac{(\beta^2+1) * PPV * sensitivity}{\beta^2 * PPV + sensitivity}$$

Since a minority of patients with HZ will develop PHN and false negatives and sensitivity are more important than PPV, we chose β=2 to favor sensitivity over PPV. The F-score's value ranges from 0 to 1, with higher values suggesting better prediction. However, because the F-score does not include TN in its formula, MCC has been proposed as a better overall measurement than the F-score as well as the area under the receiver operating characteristic curve in binary classification. [29, 30] The MCC formula considers all four confusion matrix categories, with values between -1 to 1, where ±1 denotes perfect agreement or disagreement between actuals and predictions, and 0 indicates randomness.

$$MCC = \frac{TP * TN - FP * FN}{\sqrt{(TP+FP) * (TP+FN) * (TN+FP) * (TN+FN)}}$$

The prevalence proportion of PHN was calculated as the number of identified PHN cases per 100 cases of HZ.

Ethical Considerations

The KPSC Institutional Review Board approved this study. A waiver of informed consent was granted for this study because this was a data-only minimal-risk study.

Results

Study Population

The characteristics of the study population are presented in Table 2. The mean (standard deviation) ages of the development and validation datasets were 69.5 (9.1) and 70.0 (9.8) years, respectively, with 65.8% and 67.8% being female. The development dataset had a higher proportion of Asian (35.4% vs. 13.8%) and RZV-vaccinated patients (12.2% vs. 3.5%). There were no significant differences between the development and validation datasets in terms of clinical visits and comorbidities prior to the index date. In the development and validation datasets, approximately one-third of the patients had diabetes (31.8% and 32.8%, respectively), while less than a quarter had chronic pulmonary disease (17.6% and 20.4%) and depression (20.6% and 23.6%). The development data had a higher proportion of patients with cancer as compared to the validation data (10.4% vs. 6.8%).

Table 2. Characteristics of patients in the development and validation datasets.

Characteristic	Patients, n (%)		*P* value[a]
	Development	Validation	
Mean (SD) age, year	69.5 (9.1)	70.0 (9.8)	.47
Age group (years)			.52
50-59	70 (14.0%)	103 (12.9%)	
60-69	172 (34.4%)	282 (35.3%)	
70-79	187 (37.4%)	280 (35.0%)	
≥ 80	71 (14.2%)	135 (16.9%)	
Gender			.47
Female	329 (65.8%)	542 (67.8%)	
Male	171 (34.2%)	258 (32.3%)	
Race/ethnicity			<.01
Non-Hispanic White	198 (39.6%)	424 (53.0%)	
Hispanic	102 (20.4%)	213 (26.6%)	
Asian/Pacific Islanders	177 (35.4%)	110 (13.8%)	
Non-Hispanic Black	18 (3.6%)	41 (5.1%)	
Other/Multiple/Unknown	5 (1.0%)	12 (1.5%)	

Number of outpatient/virtual visits 6 months before HZ diagnosis date			.91
0-1	68 (13.6%)	106 (13.3%)	
2-5	168 (33.6%)	262 (32.8%)	
≥ 6	264 (52.8%)	432 (54.0%)	
Number of emergency department visits 6 months before HZ diagnosis date			.08
0	420 (84.0%)	641 (80.1%)	
≥ 1	80 (16.0%)	159 (19.9%)	
Number of hospitalizations 6 months before HZ diagnosis date			.45
0	475 (95.0%)	752 (94.0%)	
≥ 1	25 (5.0%)	48 (6.0%)	
Comorbidity 1 year before HZ diagnosis date			
Allergic rhinitis	38 (7.6%)	51 (6.4%)	.39
Asthma	41 (8.2%)	82 (10.3%)	.22
Atopic dermatitis	5 (1.0%)	9 (1.1%)	.83
Cancer	52 (10.4%)	54 (6.8%)	.02
Chronic pulmonary disease	88 (17.6%)	163 (20.4%)	.22
Depression	103 (20.6%)	189 (23.6%)	.20
Diabetes	159 (31.8%)	262 (32.8%)	.72
Epilepsy and recurrent	8 (1.6%)	8 (1.0%)	.34
Heart failure	24 (4.8%)	55 (6.9%)	.13
Rheumatoid arthritis	25 (5.0%)	38 (4.8%)	.84
Systemic lupus erythematosus	6 (1.2%)	6 (0.8%)	.41
Recombinant zoster vaccine, n (%)			<.01
Unvaccinated	439 (87.8%)	772 (96.5%)	
1-dose vaccinated	32 (6.4%)	9 (1.1%)	
Fully (2-dose) vaccinated	29 (5.8%)	19 (2.4%)	

[a]Chi-square test was used for categorical variables, and Wilcoxon test was used for continuous variables.

Validation Dataset

In the validation dataset, the numbers of clinical notes in the index, transitional, and risk periods were 12158, 14446, and 18895, respectively. The percentages of HZ or PHN relevant notes were 26.2%, 8.2%, and 3.2%, respectively for the index, transitional, and risk periods. Most of the HZ index visits occurred in primary care, urgent care, emergency departments, and hospital settings (Multimedia Appendix 4). After the index period, HZ-related mentions were much less frequently documented in urgent care visit notes, but more frequently documented in specialist visit notes (n=41 specialties).

Application of NLP on Validation Dataset

Out of the 800 patients in the validation dataset, the NLP algorithm identified 796 HZ patients who had at least one note with HZ or PHN-related terms in the index period. Among the four remaining patients, two patients had their index HZ diagnosed outside KPSC and had no follow-up visits in the index period. For the remaining two patients, HZ-related symptoms were documented, but no mention of HZ or PHN was made in the clinical notes. Among these 796 patients, the NLP algorithm identified the HZ anatomic location for 751 (94.3%) patients, and among them, 611 (81.3%) had laterality information (Multimedia Appendix 5). In the transitional and risk periods, the NLP algorithm identified positive mentions of any pain/discomfort in 370 (46.3%) and 425 (53.1%) patients, respectively.

Validation Results

In the validation dataset, the NLP algorithm achieved 90.9% sensitivity, 98.5% specificity, 82.0% PPV, and 99.3% NPV (Table 3). The composite scores of the NLP algorithm were 0.89 (F-score) and 0.85 (MCC). Of the 800 patients in the validation dataset, 55 (6.9%) were chart-confirmed as PHN. The prevalence proportion of PHN identified by the NLP algorithm was 7.6%.

Table 3. Performance characteristics of natural language processing and code-based methods for identifying PHN as compared to chart-confirmed reference standard.

Method	PHN[a] (%)	TP[b]	TN[c]	FN[d]	FP[e]	Sensitivity (%)	Specificity (%)	PPV[f] (%)	NPV[g] (%)	F-score	MCC[h]
NLP[i]	7.6	50	734	5	11	90.9	98.5	82.0	99.3	0.89	0.85
C1	5.4	31	733	24	12	56.4	98.4	72.1	96.8	0.59	0.61

C2	10.9	34	692	21	53	61.8	92.9	39.1	97.1	0.55	0.44
C3	5.5	31	732	24	13	56.4	98.3	70.5	96.8	0.59	0.61
C4	13.1	29	669	26	76	52.7	89.8	27.6	96.3	0.45	0.32
C5	9.3	31	702	24	43	56.4	94.2	41.9	96.7	0.53	0.44

[a]PHN: postherpetic neuralgia.

[b]TP: true positive.

[c]TN: true negative.

[d]FN: false negative.

[e]FP: false positive.

[f]PPV: positive predictive value.

[g]NPV: negative predictive value.

[h]MCC: Matthews correlation coefficient.

[i]NLP: natural language processing.

Error Analysis of NLP Validation Results

Error analysis of the FN and FP cases is presented in Table 4. Some of the NLP-related errors were caused by the selection of data sources. For two FN cases, NLP incorrectly classified them as PHN negative when statements were found indicating HZ-associated pain had resolved even though additional evidence showed the patients still had other PHN-related symptoms. The FP cases were caused by copied-and-pasted text, incorrect causality attribution of symptoms, misclassified recurrent HZ cases as PHN, and unclear clinical documentation.

Table 4. Error analysis of NLP false negatives and false positives.

Type of NLP[a] Error	# of cases	Description
False negative	5	
EHR[b] data source	2	Case 1: We did not include one free text table (formatted messages) from the Epic EHR.

		Case 2: PHN[c] was mentioned in a clinical note from the hematology department which was excluded from NLP processing.
Unclear documentation	1	HZ[d] and/or PHN were not stated in the clinical note which was required by NLP to reduce false positive hits.
Symptom	2	While the patient stated that HZ-associated pain had resolved, documents also indicated that the patient still had other PHN-related symptoms *(*prickling sensation and itchy).
False positive	11	
EHR data source	2	We included Epic's SmartData elements which lacked specificity for PHN identification.
Unclear documentation	4	In two cases, the text was copied from the clinical notes in the index period. The NLP copy-and-paste detection algorithm was only applied to the clinical notes in the transitional period. In another two cases, PHN and PHN-related medications were listed in the assessment and plan sections. However, it was unclear whether the patient had ongoing symptoms.
Acute HZ	2	NLP misclassified two acute HZ cases that occurred in the risk period as PHN.
Causality	2	Case 1: Pain thought to be due to chalazion based on information in follow-up visits. Case 2: PHN was listed in the assessment section and tramadol and gabapentin were listed in the plan section. However, the medications were likely for lumbosacral radiculopathy.
Symptom	1	The patient reported generalized symptoms (nausea) since HZ, but there was no mention of concomitant sensory changes such as pain, thus the case did not meet our PHN definition.

[a]NLP: natural language processing.

[b]EHR: electronic health record.

[c]PHN: postherpetic neuralgia.

[d]HZ: herpes zoster.

Code-based Methods

The prevalence proportions of PHN identified by code-based methods ranged from 5.4% to 13.1%. The code-based methods achieved 52.7-61.8% sensitivity, 89.8-98.4% specificity, 27.6-72.1% PPV, and 96.3-97.1% NPV. The F-scores and

MCCs were ranged between 0.45-0.59 and 0.32-0.61, respectively. The more sophisticated algorithms were no better than the PHN diagnosis code-only method as measured by the F-score or MCC. Although each component of the code-based methods identified PHN cases, most of them did not contribute to identifying additional true PHN cases beyond those identified by PHN diagnosis codes, and those that did had much lower PPVs (C4.3: 10.3%, C4.2: 26.1% and C2.2: 37.7%) than the PHN diagnosis code only method (C1, PPV 72.1%) (Table 5). We re-reviewed all FP cases from code-based methods C1 and C3 and randomly sampled the remaining FP cases from approaches C2, C4, and C5. Among the 20 reviewed FP cases, we found that none were true PHN cases.

Table 5. PHN cases identified by code-based methods.

Method	PHN[a] diagnosis code used	PHN (%)	TP[b] (PPV[c] %)	Supplementary TP[d] n
C1[e]	✓	**43 (5.4)**	**31 (72.1)**	
C2	✓	**87 (10.9)**	**34 (39.1)**	**3**
C2.1	✓	43 (5.4)	31 (72.1)	
C2.2		69 (8.6)	26 (37.7)	3
C2.3		4 (0.5)	1 (25.0)	
C3	✓	**44 (5.5)**	**31 (70.5)**	
C3.1	✓	24 (3)	18 (75.0)	
C3.2	✓	7 (0.9)	7 (100.0)	
C3.3	✓	41 (5.1)	29 (70.7)	
C3.4	✓	23 (2.9)	17 (73.9)	
C4	✓	**105 (13.1)**	**29 (27.6)**	**2**
C4.1	✓	36 (4.5)	26 (72.2)	
C4.2		69 (8.6)	18 (26.1)	2
C4.2.1		7 (0.9)	6 (85.7)	
C4.2.2		2 (0.3)	1 (50.0)	
C4.2.3		64 (8.0)	16 (25.0)	1
C4.2.4		2 (0.3)	1 (50.0)	1
C4.3		29 (3.6)	3 (10.3)	2
C4.3.1		25 (3.1)	1 (4.0)	1
C4.3.2		2 (0.3)	1 (50.0)	1
C4.3.3		2 (0.3)	1 (50.0)	
C5	✓	**74 (9.3)**	**31 (41.9)**	
C5.1	✓	43 (5.4)	31 (72.1)	
C5.2		18 (2.3)	14 (77.8)	
C5.3		34 (4.3)	2 (5.9)	

[a]PHN: postherpetic neuralgia.

[b]PPV: positive predictive value.

[c]TP: true positive.

[d]Supplementary contributions to the number of correctly identified positive cases, apart from method C1.

[e]Method C1 only used PHN diagnosis codes.

Discussion

Principal Findings

We developed and validated NLP algorithms to identify PHN using various clinical data sources from EHRs. Compared to the chart-reviewed reference standard, the NLP algorithms showed high accuracy. This study demonstrates the feasibility of population-based PHN studies using EHR data with an automated method.

Using manual review to identify PHN cases is often infeasible for population-based research because a large volume of clinical notes would need to be reviewed. In contrast, the size of the study population and length of follow-up have little impact on running the NLP algorithm. Moreover, our NLP algorithm can readily capture PHN at varied time intervals, providing an efficient method to assess the long-term impact of PHN and compare results with studies using different PHN risk windows. Furthermore, studies can use NLP alone or with manual review confirmation. For example, manual review of the NLP-positive cases (n=61) could increase the specificity and PPV to 100% and improve the F-score from 0.89 to 0.93 and MCC from 0.85 to 0.95; this is more efficient than manual review of all 800 HZ cases.

Implementing NLP on EHR data presents challenges. In this study, data sources accounted for a quarter of NLP errors (2 FN and 2 FP). First, clinical data were stored in a variety of locations within our institution's complex EHR system, which contains over 900,000 database tables. It is often difficult to locate the database table storing the data displayed in the EHR user interface. One FN case resulted from not including a previously unknown table. Second, selecting data sources for NLP processing is often a tradeoff. One FN and two FP cases resulted from including or excluding certain data sources. EHRs have also made it easy to create lengthy and bloated notes. [31, 32] According to recent research, over half of clinical note content is duplicated or copied from earlier notes. [32-34] Clinicians may copy from prior visit notes to improve recall and clinical reasoning. [35] However, these replicated contents may lack temporal or contextual information, making them difficult to identify manually and challenging for NLP.

Because PHN-related symptoms such as pain and discomfort are common in a variety of medical conditions with numerous plausible causes, identifying PHN necessitates integrating the NLP-identified PHN symptoms with their associated anatomic location, temporality, and causality. These elements, however, are not always explicitly stated in clinical documents. About half of the NLP FP cases were from incorrectly attributing the complaint or treatment to PHN. These FP cases were partially explained by the NLP algorithm's preference for sensitivity over specificity.

Another popular method of PHN identification is using coded data from

administrative claims and/or EHR, which could include a large sample size at a low cost. However, many of the code-based PHN identification algorithms have not been validated. [3] We implemented and validated five code-based algorithms, including one that solely uses PHN diagnosis codes (C1) and two that had previously been validated (C2 and C3). To maximize their sensitivity, algorithms C2-C5 used the "OR" statement to combine various criteria. The downside of using the “OR” logic is the loss of PPV. Algorithms C2-C5 all had worse PPV than the diagnosis codes-only algorithm (C1). However, in our study, the sensitivity of these algorithms ranged from 53-62%, with only C2 outperforming C1 (62% vs. 56%). Algorithms C2-C5 had lower PPVs (28-71%) than C1 (72%). With such limited sensitivities, these algorithms may miss roughly half of the PHN cases. In our study, aside from the PHN diagnosis codes, the other diagnosis codes and prescription data had little impact on true case identification, instead adding complexity and increasing false positives.

Studies have employed the similarity of the PHN proportions as construct validity of their case-finding algorithms. [8-11] Administrative database studies reported PHN (pain persisting for ≥90 days) prevalences of 3-14% (Multimedia Appendix 6), [3] which are comparable to the 5.4-13.1% prevalences of the code-based approaches in our study. The broad range of prevalences identified in previous code-based studies could be caused by variations in study design, population, and data source. [3] However, the code-based approaches in this study had the same population and data source. Only the variation in algorithms could cause such a wide disparity.

We expanded the validations conducted for the two previously validated

algorithms, which were performed on EHR data. The C2 (Klompas et al.) algorithm was only validated with the 30-day definition in the original study, and it had 86% sensitivity and 78% PPV. [8] In our study, algorithm C2 with the 90-day definition had notably lower sensitivity (62%) and PPV (39%). One main contributor to the variability in performance is the difference in the temporal criteria. According to Yawn, [36] up to 75% of pain present at 30 days disappears at 90 days, and the prevalence of PHN decreased by sixfold when the definition was changed from 30 days to 90 days. As prevalence decreases, so do the sensitivity and PPV. [37, 38] The same trends were also reported in the original C2 paper; the PPVs for different PHN search criteria using the 30-day definition (29-95%) were nearly double that of using the 90-day definition (15-52%). The discrepancy in C2 algorithm performance between the original study and this study could be further explained by the differences in case definition. Our case definition for PHN is based on the persistent PHN-related symptoms and causal attribution, not diagnosis code or medication. Algorithm C2 used ongoing symptoms or renewal of medication for HZ. The use of medications to identify PHN has some drawbacks, as PHN-related medications have a wide range of indications. For example, gabapentin, a first-line therapy for PHN, has over twenty approved and off-label uses. [39] Furthermore, prescriptions can be refilled in the absence of active PHN symptoms for various non-PHN disorders.

The original C3 (Klein et al.) algorithm was only validated on potential PHN cases identified by its four component criteria, rather than randomly selected HZ cases; only PPVs were reported. [10] In this study, the four criteria of the C3 algorithm had PPVs ranging from 71% to 100%, which is consistent with the

previous study's findings (PPVs ranging from 73% to 96%). The C3 algorithm was one of the best-performing code-based algorithms based on F-score and MCC. However, its low sensitivity (56%) and PPV (71%) indicate considerable misclassification. The lower overall PPV is partly due to the “OR” logic of the four criteria. Because Klein et al. did not describe the case definition or chart review rules, we were unable to assess their impact on the performance differences between the original C3 study and this study.

The substantial misclassification of coded methods as observed in this study could have a substantial impact on measuring incidence, identifying risk factors, and assessing vaccine effectiveness. Code-based method studies (C4 and C5) had identified depression, diabetes mellitus, heart failure, and chronic obstructive pulmonary disease as risk factors for PHN. It is conceivable that the link between depression and PHN is caused by using anticonvulsants and tricyclic antidepressants to identify PHN. The inclusion of prescriptions for pain medications and chronic pain codes may contribute to the association of diabetes mellitus, [40] heart failure, [41] and chronic obstructive pulmonary disease [42] with PHN.

Study Strengths and Limitations

This study was conducted within a large integrated healthcare system with comprehensive EHRs. Because the health plan provides strong incentives for members to use its facilities, clinical documentation is expected to be more detailed. We developed NLP algorithms to identify PHN from various

unstructured data sources within EHRs, such as clinical notes, which contain a wealth of information but differ greatly in structure, content, and quality. The algorithms were highly accurate, as evidenced by our validation. Compared to studies based on self-reported pain scores collected through surveys, EHR-based studies measure the healthcare burden of PHN, which is more clinically relevant. This study also has limitations. The reference standard relied on the review of EHRs which could be erroneous and incomplete. [14] Moreover, re-reviewing cases in the validation set where NLP results differed from research associates' results may result in bias in favor of higher performance of the NLP algorithm. On the other hand, reconciling discrepant results improved the quality of the reference standard. Additionally, diagnosis codes, prescriptions, clinical documentation language and style can differ between institutions and physicians. Our NLP method may perform differently in other test datasets.

Conclusions

PHN-related diagnosis codes have low sensitivity for identifying PHN cases. Additional diagnosis codes and prescription data did little to improve sensitivity while significantly lowering PPV. Using clinical text from the EHR, the NLP-based method identified PHN cases with high accuracy. Our NLP method can be used in EHR-based studies to identify PHN risk factors and evaluate the effectiveness of vaccinations and treatments against PHN.

Acknowledgments

A part of this work was previously presented at IDWeek 2023 held in Boston, Massachusetts on October 12, 2023.

Conflicts of Interest

The authors have no conflicts of interest relevant to this article to disclose.

Funding/Support:

This work was supported by Kaiser Permanente Southern California internal research funds.

Multimedia Appendix 1

PHN Abstraction Decision Rules

Multimedia Appendix 2

Additional Details on Natural Language Processing (NLP) Algorithm Development

Multimedia Appendix 3

Code-based Methods

Multimedia Appendix 4

The proportion of HZ- or PHN-related Notes by Department/Specialty

Multimedia Appendix 5

Number and Percentage of HZ Locations as Identified by NLP

Multimedia Appendix 6

Reported PHN Rates in Previous Administrative Database Studies

References

1. Gershon AA, Breuer J, Cohen JI, Cohrs RJ, Gershon MD, Gilden D, et al. Varicella zoster virus infection. Nature reviews Disease primers. 2015;1(1):1-18.
2. Gnann JW, Jr., Whitley RJ. Clinical practice. Herpes zoster. N Engl J Med. 2002 Aug 1;347(5):340-6. PMID: 12151472. doi: 10.1056/NEJMcp013211.
3. Kawai K, Gebremeskel BG, Acosta CJ. Systematic review of incidence and complications of herpes zoster: towards a global perspective. BMJ Open. 2014 Jun 10;4(6):e004833. PMID: 24916088. doi: 10.1136/bmjopen-2014-004833.
4. Johnson RW, Alvarez-Pasquin MJ, Bijl M, Franco E, Gaillat J, Clara JG, et al. Herpes zoster epidemiology, management, and disease and economic burden in Europe: a multidisciplinary perspective. Ther Adv Vaccines. 2015 Jul;3(4):109-20. PMID: 26478818. doi: 10.1177/2051013615599151.
5. Johnson RW, Rice ASC. Postherpetic Neuralgia. New England Journal of Medicine. 2014;371(16):1526-33. PMID: 25317872. doi: 10.1056/NEJMcp1403062.
6. Rowbotham M, Harden N, Stacey B, Bernstein P, Magnus-Miller L. Gabapentin for the treatment of postherpetic neuralgia: a randomized controlled trial. JAMA. 1998 Dec 2;280(21):1837-42. PMID: 9846778. doi: 10.1001/jama.280.21.1837.
7. Lal H, Cunningham AL, Godeaux O, Chlibek R, Diez-Domingo J, Hwang SJ, et al. Efficacy of an adjuvanted herpes zoster subunit vaccine in older adults. N Engl J Med. 2015 May 28;372(22):2087-96. PMID: 25916341. doi: 10.1056/NEJMoa1501184.
8. Klompas M, Kulldorff M, Vilk Y, Bialek SR, Harpaz R. Herpes zoster and postherpetic neuralgia surveillance using structured electronic data. Mayo Clin Proc. 2011 Dec;86(12):1146-53. PMID: 21997577. doi: 10.4065/mcp.2011.0305.
9. Forbes HJ, Bhaskaran K, Thomas SL, Smeeth L, Clayton T, Mansfield K, et al. Quantification of risk factors for postherpetic neuralgia in herpes zoster patients: A cohort study. Neurology. 2016 Jul 5;87(1):94-102. PMID: 27287218. doi: 10.1212/WNL.0000000000002808.
10. Klein NP, Bartlett J, Fireman B, Marks MA, Hansen J, Lewis E, et al. Long-term effectiveness of zoster vaccine live for postherpetic neuralgia prevention. Vaccine. 2019 Aug 23;37(36):5422-7. PMID: 31301920. doi: 10.1016/j.vaccine.2019.07.004.
11. Munoz-Quiles C, Lopez-Lacort M, Orrico-Sanchez A, Diez-Domingo J. Impact of postherpetic neuralgia: A six year population-based analysis on people aged 50 years or older. J Infect. 2018 Aug;77(2):131-6. PMID: 29742472. doi: 10.1016/j.jinf.2018.04.004.
12. Suaya JA, Chen SY, Li Q, Burstin SJ, Levin MJ. Incidence of herpes zoster and persistent post-zoster pain in adults with or without diabetes in the United States. Open Forum Infect Dis. 2014 Sep;1(2):ofu049. PMID: 25734121. doi: 10.1093/ofid/ofu049.
13. Hillebrand K, Bricout H, Schulze-Rath R, Schink T, Garbe E. Incidence of herpes zoster and its complications in Germany, 2005-2009. J Infect. 2015 Feb;70(2):178-86. PMID: 25230396. doi: 10.1016/j.jinf.2014.08.018.
14. Yawn BP, Wollan P, St Sauver J. Comparing shingles incidence and complication rates from medical record review and administrative database estimates: how close are they? Am J Epidemiol. 2011 Nov 01;174(9):1054-61.

PMID: 21920944. doi: 10.1093/aje/kwr206.
15. Tanenbaum HC, Lawless A, Sy LS, Hong V, Ackerson B, Bruxvoort K, et al. Differences in Estimates of Post-Herpetic Neuralgia Between Medical Chart Review and Self-Report. J Pain Res. 2020;13:1757-62. PMID: 32765050. doi: 10.2147/JPR.S255238.
16. Tseng HF, Lewin B, Hales CM, Sy LS, Harpaz R, Bialek S, et al. Zoster Vaccine and the Risk of Postherpetic Neuralgia in Patients Who Developed Herpes Zoster Despite Having Received the Zoster Vaccine. J Infect Dis. 2015 Oct 15;212(8):1222-31. PMID: 26038400. doi: 10.1093/infdis/jiv244.
17. Zheng C, Luo Y, Mercado C, Sy L, Jacobsen SJ, Ackerson B, et al. Using natural language processing for identification of herpes zoster ophthalmicus cases to support population-based study. Clin Exp Ophthalmol. 2019 Jan;47(1):7-14. PMID: 29920898. doi: 10.1111/ceo.13340.
18. Zheng C, Sy LS, Tanenbaum H, Tian Y, Luo Y, Ackerson B, Tseng HF. Text-Based Identification of Herpes Zoster Ophthalmicus With Ocular Involvement in the Electronic Health Record: A Population-Based Study. Open Forum Infectious Diseases. 2021;8(2). doi: 10.1093/ofid/ofaa652.
19. Delaney A, Colvin LA, Fallon MT, Dalziel RG, Mitchell R, Fleetwood-Walker SM. Postherpetic neuralgia: from preclinical models to the clinic. Neurotherapeutics. 2009 Oct;6(4):630-7. PMID: 19789068. doi: 10.1016/j.nurt.2009.07.005.
20. Forbes HJ, Thomas SL, Smeeth L, Clayton T, Farmer R, Bhaskaran K, Langan SM. A systematic review and meta-analysis of risk factors for postherpetic neuralgia. Pain. 2016 Jan;157(1):30-54. PMID: 26218719. doi: 10.1097/j.pain.0000000000000307.
21. Coplan PM, Schmader K, Nikas A, Chan IS, Choo P, Levin MJ, et al. Development of a measure of the burden of pain due to herpes zoster and postherpetic neuralgia for prevention trials: adaptation of the brief pain inventory. J Pain. 2004 Aug;5(6):344-56. PMID: 15336639. doi: 10.1016/j.jpain.2004.06.001.
22. Zheng C, Rashid N, Wu YL, Koblick R, Lin AT, Levy GD, Cheetham TC. Using natural language processing and machine learning to identify gout flares from electronic clinical notes. Arthritis Care Res (Hoboken). 2014 Nov;66(11):1740-8. PMID: 24664671. doi: 10.1002/acr.22324.
23. Zheng C, Sun BC, Wu YL, Lee MS, Shen E, Redberg RF, et al. Automated Identification and Extraction of Exercise Treadmill Test Results. Journal of the American Heart Association. 2020 Mar 3;9(5):e014940. PMID: 32079480. doi: 10.1161/JAHA.119.014940.
24. Zheng C, Yu W, Xie F, Chen W, Mercado C, Sy LS, et al. The use of natural language processing to identify Tdap-related local reactions at five health care systems in the Vaccine Safety Datalink. International Journal of Medical Informatics. 2019 2019/07/01/;127:27-34. doi: 10.1016/j.ijmedinf.2019.04.009.
25. Zheng C, Duffy J, Liu IA, Sy LS, Navarro RA, Kim SS, et al. Identifying Cases of Shoulder Injury Related to Vaccine Administration (SIRVA) in the United States: Development and Validation of a Natural Language Processing Method. JMIR Public Health Surveill. 2022 May 24;8(5):e30426. PMID: 35608886. doi: 10.2196/30426.
26. Zheng C, Rashid N, Koblick R, An J. Medication Extraction from Electronic Clinical Notes in an Integrated Health System: A Study on Aspirin Use in Patients

with Nonvalvular Atrial Fibrillation. Clin Ther. 2015 Sep 1;37(9):2048-58 e2. PMID: 26233471. doi: 10.1016/j.clinthera.2015.07.002.
27. Yanni EA, Ferreira G, Guennec M, El Hahi Y, El Ghachi A, Haguinet F, et al. Burden of herpes zoster in 16 selected immunocompromised populations in England: a cohort study in the Clinical Practice Research Datalink 2000-2012. BMJ Open. 2018 Jun 7;8(6):e020528. PMID: 29880565. doi: 10.1136/bmjopen-2017-020528.
28. Derczynski L, editor. Complementarity, F-score, and NLP Evaluation. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16); 2016; Portorož, Slovenia.
29. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020 Jan 2;21(1):6. PMID: 31898477. doi: 10.1186/s12864-019-6413-7.
30. Chicco D, Jurman G. The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. BioData Min. 2023 Feb 17;16(1):4. PMID: 36800973. doi: 10.1186/s13040-023-00322-4.
31. Downing NL, Bates DW, Longhurst CA. Physician Burnout in the Electronic Health Record Era: Are We Ignoring the Real Cause? Ann Intern Med. 2018 Jul 3;169(1):50-1. PMID: 29801050. doi: 10.7326/M18-0139.
32. Rule A, Bedrick S, Chiang MF, Hribar MR. Length and Redundancy of Outpatient Progress Notes Across a Decade at an Academic Medical Center. JAMA Netw Open. 2021 Jul 1;4(7):e2115334. PMID: 34279650. doi: 10.1001/jamanetworkopen.2021.15334.
33. Steinkamp J, Kantrowitz JJ, Airan-Javia S. Prevalence and Sources of Duplicate Information in the Electronic Medical Record. JAMA Netw Open. 2022 Sep 1;5(9):e2233348. PMID: 36156143. doi: 10.1001/jamanetworkopen.2022.33348.
34. Wang MD, Khanna R, Najafi N. Characterizing the Source of Text in Electronic Health Record Progress Notes. JAMA Intern Med. 2017 Aug 1;177(8):1212-3. PMID: 28558106. doi: 10.1001/jamainternmed.2017.1548.
35. Zheng C, Lee MS, Bansal N, Go AS, Chen C, Harrison TN, et al. Identification of Recurrent Atrial Fibrillation using Natural Language Processing Applied to Electronic Health Records. Eur Heart J Qual Care Clin Outcomes. 2023 Mar 30. PMID: 36997334. doi: 10.1093/ehjqcco/qcad021.
36. Yawn BP. Post-shingles neuralgia by any definition is painful, but is it PHN? Mayo Clin Proc. 2011 Dec;86(12):1141-2. PMID: 22134931. doi: 10.4065/mcp.2011.0724.
37. Murad MH, Lin L, Chu H, Hasan B, Alsibai RA, Abbas AS, et al. The association of sensitivity and specificity with disease prevalence: analysis of 6909 studies of diagnostic test accuracy. CMAJ. 2023 Jul 17;195(27):E925-E31. PMID: 37460126. doi: 10.1503/cmaj.221802.
38. Tenny S, Hoffman MR. Prevalence. StatPearls. Treasure Island (FL)2024.
39. Yasaei R, Katta S, Patel P, Saadabadi A. Gabapentin. StatPearls. Treasure Island (FL)2024.
40. Dyck PJ, Kratz KM, Karnes JL, Litchy WJ, Klein R, Pach JM, et al. The prevalence by staged severity of various types of diabetic neuropathy, retinopathy, and nephropathy in a population-based cohort: the Rochester

Diabetic Neuropathy Study. Neurology. 1993 Apr;43(4):817-24. PMID: 8469345. doi: 10.1212/wnl.43.4.817.
41. Alemzadeh-Ansari MJ, Ansari-Ramandi MM, Naderi N. Chronic Pain in Chronic Heart Failure: A Review Article. J Tehran Heart Cent. 2017 Apr;12(2):49-56. PMID: 28828019.
42. Andenaes R, Momyr A, Brekke I. Reporting of pain by people with chronic obstructive pulmonary disease (COPD): comparative results from the HUNT3 population-based survey. BMC Public Health. 2018 Jan 25;18(1):181. PMID: 29370850. doi: 10.1186/s12889-018-5094-5.

Abbreviations

EHR: electronic health record

FN: false negative

FP: false positive

HZ: herpes zoster

KPSC: Kaiser Permanente Southern California

MCC: Matthews correlation coefficient

NLP: natural language processing

NPV: negative predictive value

PHN: postherpetic neuralgia

PPV: positive predictive value

RZV: recombinant zoster vaccine

TN: true negative

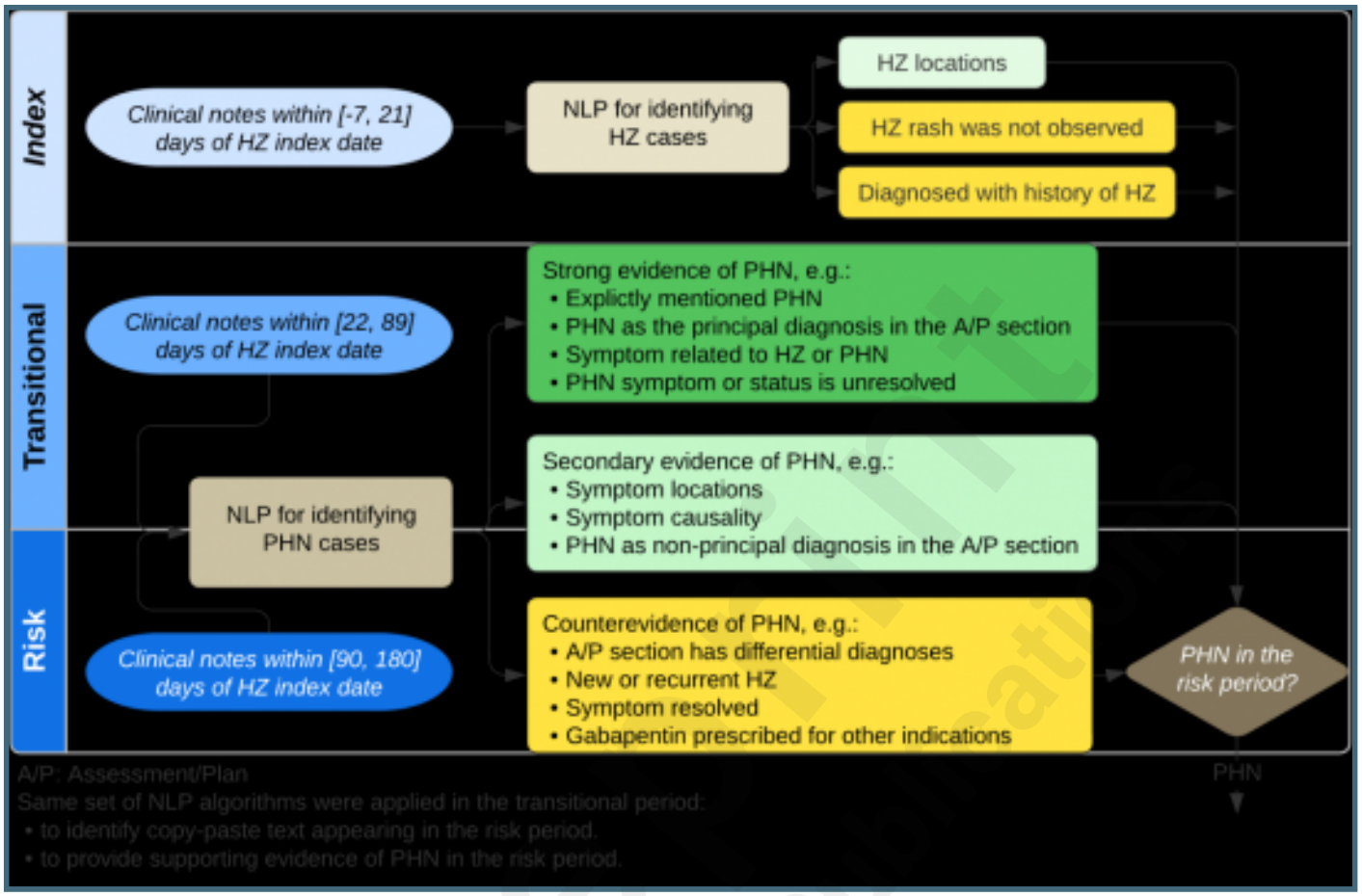
TP: true positive

Supplementary Files

Untitled.
URL: http://asset.jmir.pub/assets/325e61a71fc2798231bc46e7a08a55c7.docx

Figures

Diagram of natural language processing algorithm.



Multimedia Appendixes

PHN abstraction decision rules.
URL: http://asset.jmir.pub/assets/582751de5ead2cc9cd67dbbef345a6af.docx

Additional details on natural language processing (NLP) algorithm development.
URL: http://asset.jmir.pub/assets/f2caaa85a0c1168b22d195f22dfa8f54.docx

Code-based methods.
URL: http://asset.jmir.pub/assets/9ab27fb69f0c0a6182b1162fd41fff83.docx

The proportion of HZ- or PHN-related notes by department/specialty.
URL: http://asset.jmir.pub/assets/30d4e55b78b557ad70ab1901d8b35401.docx

Number and percentage of HZ locations as identified by NLP.
URL: http://asset.jmir.pub/assets/1e4907913c604478f234e068b0ae534d.docx

Reported PHN rates in previous administrative database studies.
URL: http://asset.jmir.pub/assets/b9a3b6c3ebaeb1041393609e218a0133.docx