Development of a machine learning algorithm based on administrative claims data for identification of ED anaphylaxis patient visits

Ronna L. Campbell, MD, PhD,[a] Mollie L. Alpern, MD,[b] James T. Li, MD, PhD,[b] John B. Hagan, MD,[b] Megan Motosue, MD,[c] Aidan F. Mullan, MA,[d] Lauren S. Harper, MD,[a] Christine M. Lohse, MS,[d] and Molly M. Jeffery, PhD[a,e] *Rochester, Minn; and Honolulu, Hawaii*

Background: Epidemiologic studies of anaphylaxis commonly rely on *International Classification of Diseases* (*ICD*) codes to identify anaphylaxis cases, which may lead to suboptimal epidemiologic classification.

Objective: We sought to develop and assess the accuracy of a machine learning algorithm using *ICD* codes and other administrative data compared with *ICD* code–only algorithms to identify emergency department (ED) anaphylaxis visits.

Methods: We conducted a retrospective review of ED visits from January 2013 to September 2017. Potential ED anaphylaxis visits were identified using 3 methods: anaphylaxis *ICD* diagnostic codes (method 1), *ICD* symptom-based codes with or without a code indicating an allergic trigger (method 2), and *ICD* codes indicating a potential allergic reaction only (method 3). A machine learning algorithm was developed from administrative data, and test characteristics were compared with *ICD* code–only algorithms.

Results: A total of 699 of 2191 (31.9%) potential ED anaphylaxis visits were classified as anaphylaxis. The sensitivity and specificity of method 1 were 49.1% and 87.5%, respectively. Method 1 used in combination with method 2 resulted in a sensitivity of 53.9% and a specificity of 68.7%. Method 1 used in combination with method 3 resulted in a sensitivity of 98.4% and a specificity of 15.1%. The sensitivity and specificity of the machine learning algorithm were 87.3% and 79.1%, respectively.

Conclusions: *ICD* coding alone demonstrated poor sensitivity in identifying cases of anaphylaxis, with venom-related anaphylaxis missing 96% of cases. The machine learning algorithm resulted in a better balance of sensitivity and specificity and improves upon previous strategies to identify ED anaphylaxis visits. (J Allergy Clin Immunol Global 2023;2:61-8)

Key words: *Anaphylaxis, emergency department, epidemiology, machine learning*

Anaphylaxis is an acute systemic hypersensitivity reaction that can be life-threatening.[1] Among health care settings, it is most commonly treated in the emergency department[2,3] (ED). The diagnosis of anaphylaxis can be challenging due to the diverse presentations involving multiple organ systems and lack of timely sensitive and specific confirmatory laboratory testing.[4] Anaphylaxis can present with varying levels of severity, ranging from a relatively mild self-limited to a rapidly fatal condition.

The diagnosis of anaphylaxis is also challenging due to the lack of universal agreement with regard to clinical diagnostic criteria. The National Institutes of Health[5]/Food Allergy and Anaphylaxis Network (NIAID/FAAN)[6] have been the most widely studied anaphylaxis diagnostic criteria to date. Furthermore, they have been prospectively validated and found to have 95% sensitivity and 71% specificity. The World Allergy Organization[1] recently proposed revised diagnostic criteria that have not yet been widely studied, validated, or universally adopted. The Brighton criteria[7] were developed for the diagnosis of anaphylaxis in the setting of immunizations, not for other potential triggers of anaphylaxis, and were shown to result in a discordant diagnosis compared with the NIAID/FAAN criteria in 28.1% of cases in a cohort of ED patients.[8] Furthermore, the Brighton criteria include lip swelling as a major respiratory criterion, which could lead to overdiagnosis of anaphylaxis in some cases. Even if universally adopted diagnostic criteria were available, the subjective nature of many of the symptoms of anaphylaxis (eg, dyspnea, abdominal pain, and odynophagia) would limit specificity. Although diagnostic criteria are helpful at the bedside when the diagnosis of anaphylaxis is suspected, they do not replace clinical judgment. Likewise, the reference standard for the assessment of anaphylaxis diagnostic criteria in the context of anaphylaxis research has also been based on clinical judgment because this allows for the integration of information in addition to diagnostic criteria such as the patient's past medical history, laboratory data, information obtained at follow-up visits, and results of allergy testing.

Most studies of anaphylaxis epidemiology rely on *International Classification of Diseases* (*ICD*) codes to identify cases of anaphylaxis. However, anaphylaxis is frequently underdiagnosed.[9,10] To account for anaphylaxis underdiagnosis, some authors[11] have proposed a symptom-based coding algorithm that includes diagnoses indicating multiple organ-system dysfunction (eg, *ICD* codes for wheezing, urticaria, and hypotension) in

From [a]the Department of Emergency Medicine and [b]the Division of Allergic Diseases, Mayo Clinic, Rochester; [c]the Division of Allergy and Immunology, Kaiser Permanente Honolulu Clinic, Honolulu; and [d]the Division of Clinical Trials and Biostatistics and [e]the Department of Health Sciences Research, Mayo Clinic, Rochester.

Disclosure of potential conflict of interest: R. Campbell is an author for UpToDate and a consultant for Bryn Pharma. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication June 7, 2022; revised August 18, 2022; accepted for publication September 6, 2022.

Available online October 17, 2022.

Corresponding author: Ronna L. Campbell, MD, PhD, Department of Emergency Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905. E-mail: Campbell.ronna@mayo.edu.

The CrossMark symbol notifies online readers when updates have been made to the article such as errata or minor corrections

2772-8293

https://doi.org/10.1016/j.jacig.2022.09.002

Abbreviations used
ED: Emergency department
ICD: *International Classification of Diseases*
ICD-9-CM: *International Classification of Diseases, Ninth Revision, Clinical Modification*
ICD-10-CM: *International Classification of Diseases, Tenth Revision, Clinical Modification*
NIAID/FAAN: National Institutes of Health/Food Allergy and Anaphylaxis Network
OLDW: OptumLabs Data Warehouse

association with a code suggestive of an allergic trigger to identify cases of anaphylaxis that were not coded with *ICD* anaphylaxis-specific codes. However, this algorithm was not validated using patient-level data, and therefore it is unknown whether the algorithm improves the accuracy of anaphylaxis recognition when compared with anaphylaxis-only *ICD* codes.

Manual review of large databases is time consuming and costly. Machine learning, algorithms that harness statistical methods to identify and learn patterns from clinical data, offers a potentially efficient and automated means to process large data sets. The goal of this study was to develop a machine learning algorithm based on *ICD* codes and other administrative data compared with *ICD* code–only algorithms to improve identification of ED anaphylaxis visits using expert review of ED anaphylaxis cases as the reference standard.

METHODS

Study design and setting

We conducted a retrospective case review study of ED patients of the Mayo Clinic Hospital, Saint Marys Campus in Rochester, Minn, a quaternary academic ED with approximately 78,000 annual patient visits. Pa*ient*s during *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9- CM*) era were included from January 1, 2013, to September 30, 2015. Patients during the *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-10-CM)* cohort were included from October 1, 2015, to September 27, 2017. The institutional review board approved our study. Additional methodologic details are provided in this article's Methods section in the Online Repository at www.jaci-global.org.

Participants

ED patients of all ages, including both children and adults, with diagnostic codes of interest were identified on the basis of primary diagnosis and 8 secondary diagnoses. Fig 1 shows a flow diagram of patient identification. To be included in the initial cohort of ED visits screened, patients were required to have provided research authorization and the associated ED visit was required to have *ICD* diagnostic codes shown in Table E1 in this article's Online Repository at www.jaci-global.org, meeting 1 of the following 3 criteria: (1) an anaphylaxis-specific diagnostic code, a code suggestive of an allergic reaction or presence of an allergic trigger (trigger-specific codes), or a code for angioedema; (2) a code indicating respiratory compromise in addition to a code indicating skin and mucosal involvement; or (3) a code indicating reduced blood pressure in addition to a code indicating skin and mucosal involvement.

From the cohort of ED visits screened, ED visits were further stratified for identification on the basis of 3 methods: (1) ED visits with diagnostic codes specifically for anaphylaxis (method 1); (2) ED visits that were not identified by method 1 but that had a combination of symptom codes alone ([respiratory plus skin and mucosal] or [reduced blood pressure plus skin and mucosal]) or a combination of symptom codes from 2 of 4 organ systems (respiratory, cardiovascular, gastrointestinal, or skin and mucosal) in addition to a trigger-specific code (method 2); or (3) ED visits with codes indicating a potential allergic reaction that were not identified by method 1 or method 2 (method 3). Coding combinations are presented in Table E2 in this article's Online Repository at www.jaci-global.org. For more details, see this article's Methods section in the Online Repository.

Data extraction and variables

Standardized extraction forms were developed and pilot tested. The NIAID/FAAN criteria were implemented for diagnosis, but the reviewers were allowed to exercise clinical judgment and to take into consideration any additional information available in the electronic medical record including patient past medical history, laboratory data, follow-up visits, and results of allergy testing. To verify the presence or absence of anaphylaxis, ED visits were reviewed by 2 attending allergists-immunologists (J.H. and J.T.L.) and 1 second-year allergy-immunology fellow (M.A.). Forty-two percent of ED visits were reviewed by 2 attending allergists-immunologists and allergy-immunology fellow, 47% were reviewed by 1 attending allergist-immunologist, and the remaining 11% were reviewed by only an allergy-immunology fellow after completing the double reviews with the attending allergist-immunologists. The 2 attending allergist-immunologists (J.H. and J.T.L.) had 24 and 33 years of postfellowship clinical practice, respectively, at the time the chart reviews were conducted for this study. In addition, both participated in the retrospective and prospective validation of the NIAID/FAAN criteria and had extensive prior training on chart review for the diagnosis of anaphylaxis among ED patients with concordance rates that were greater than the concordance rates of the NIAID/FAAN criteria for the consensus diagnosis of anaphylaxis.[6,12] Any diagnostic differences were resolved through careful joint review. Electronically extracted data are presented in Table E1. For more details, see this article's Methods section in the Online Repository.

Test performance

Using the final diagnosis made by expert review as the reference standard, test characteristics were determined for the *ICD* codes and coding combinations as well as the machine learning algorithm for identification of ED anaphylaxis visits. Two-by-two tables were created to calculate measures of diagnostic test accuracy.

Statistical analysis

The occurrence of each diagnosis code is summarized with frequency counts and percentages. The ability of each set of diagnostic codes to predict anaphylaxis as diagnosed by expert review is described using sensitivity and specificity. CIs for each predictive measure were calculated using Wilson's score interval.

Machine learning algorithm and prediction performance evaluation

To assess the predictive ability of the extracted diagnostic codes for predicting anaphylaxis, a stochastic gradient boosting algorithm was fit to the data. For more details, see this article's Methods section in the Online Repository.

Comparison of Mayo Clinic data to OptumLabs data

The gradient boosting algorithm that was developed in the Mayo ED cohort was applied to predict anaphylaxis in the OptumLabs Data Warehouse (OLDW) population from January 1, 2013, to September 27, 2017. The OLDW is a health care services claims database[13] that includes 20% of the commercially insured population in the United States and 24% of Medicare Advantage beneficiaries. Age, sex, and race or ethnicity distributions among the enrolled population in the database are similar to those among US commercial and Medicare Advantage populations. All US census divisions are represented in the OLDW. For more details, see this article's Methods section in the Online Repository.

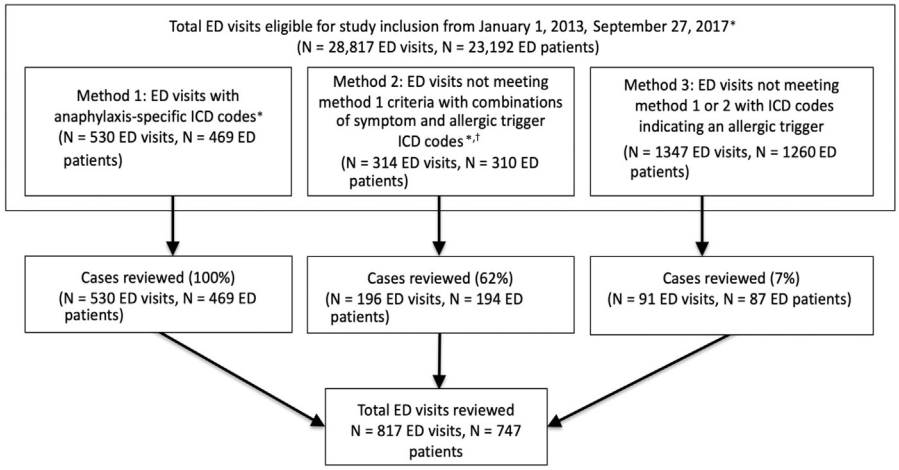

FIG 1. Flow diagram of identification of potential anaphylaxis ED visits. **ICD* codes are presented in Table E1. †Coding combinations are presented in Table E2.

RESULTS

ED anaphylaxis visits

During the study period, the medical records from 817 ED visits with diagnostic codes of interest were reviewed. A flow diagram of ED visit identification is shown in Fig 1. A summary of the diagnostic codes and coding combinations reviewed and the proportion that were classified as anaphylaxis are presented in Tables E2 and E3 in this article's Online Repository at www.jaci-global.org. With appropriate statistical weighting, this represented a study cohort of 2191 patients. Of the weighted patient cohort, most were female (56.9%) and most were White (83.5%). The median age was 42.8 years (interquartile range, 22-61.1 years) (Table I).

After statistical weighting based on a stratified random sample, a total of 699 of 2191 (31.9%) ED visits were classified as anaphylaxis, 577 (26.3%) were classified as allergic reactions, and 914 (41.7%) were classified as other etiologies (Table I). Of the 699 ED anaphylaxis visits, 235 (33.6%) had a food trigger, 196 (28.0%) had a venom trigger, 62 (8.9%) had a medication trigger, 72 (10.3%) had another identified trigger, 134 (19.2%) had an unknown trigger, and 590 (84.4%) received epinephrine for management of their reaction either before or after ED arrival.

Of the total 699 anaphylaxis cases, 343 (49.1%) were identified with anaphylaxis-specific codes (method 1), 34 (4.9%) additional cases were identified using the symptom and trigger-based coding algorithm (method 2), and an additional 322 (46.1%) were identified by individual codes suggestive of a possible allergic reaction (method 3). Among the total 302 (23.7%) cases of food-related anaphylaxis, 159 (52.6%) were identified by method 1, an additional 6 (2.0%) cases were identified by method 2, and 76 (25.3%) were identified by method 3. Among the total 164 cases of medication-related anaphylaxis, 80 (48.8%) were identified by method 1, an additional 22 (13.4%) cases were identified by method 2, and 61.9 (37.8%) were identified by method 3. Among the total 460 cases of venom-related anaphylaxis, 16 (3.5%) were identified by method 1, an additional 14 (3.0%) cases were identified by method 2, and 430 (93.5%) were identified by method 3. Thus, the use of anaphylaxis-specific *ICD* codes alone missed 96% of venom-related anaphylaxis cases.

Test characteristics of methods 1, 2, and 3 for identification of ED anaphylaxis visits

The sensitivity and specificity of the anaphylaxis-specific codes (method 1) were 49.1% (95% CI, 45.3-52.8) and 87.5% (95% CI, 85.7-89.1), respectively (Table II). The addition of the symptom and trigger-based algorithm (method 2) to the anaphylaxis-specific codes increased the sensitivity to 53.9% (95% CI, 50.2-57.7) but reduced the specificity to 68.7% (95% CI, 66.3-71.0). The addition of the individual codes suggestive of an allergic reaction (method 3) to the anaphylaxis-specific codes resulted in a sensitivity of 98.4% (95% CI, 97.1-99.2) but reduced the specificity to 15.1% (95% CI, 13.3-17.0).

Test characteristics of gradient boosting machine learning algorithm

The full gradient boosting algorithm considered all method 1, method 2, and method 3 codes that were present in at least 5 visits from the data set as well as the candidate predictors presented in Table III. Internal cross-validation found an optimal forest size of 1500 trees, which provided a sensitivity of 87.3% (95% CI, 83.6-90.4) and a specificity of 79.1% (95% CI, 74.8-82.8) (Table II).

A secondary algorithm repeated the gradient boosting process, this time starting with only those variables that had a relative importance value greater than 1 from the full algorithm. This smaller algorithm found an optimal forest size of 1500 trees, providing a sensitivity of 81.5% (95% CI, 77.3-85.2) and a specificity of 80.3% (95% CI, 76.0-84.0). Although we observe a slight drop in predictive accuracy, this algorithm reduced the number of candidate predictors by nearly half, from 39 to 19.

As a further reduction in algorithm complexity, the gradient boosting algorithm was fit to only candidate predictors with a relative importance greater than 3 from the initial algorithm. This algorithm found an optimal forest size of 1500 and achieved a sensitivity of 82.0% (95% CI, 77.8-85.6) and a specificity of 76.4% (95% CI, 71.9-80.4). The number of candidate predictors considered was only 9.

TABLE I. ED patient characteristics

ED patient characteristic	Unweighted all ED visits (N = 817)	Weighted all ED visits (N = 2191)	Weighted anaphylaxis ED visits (N = 699.2)
Age (y)			
Mean ± SD	36.3 ± 24.6	41.7 ± 24.4	33.4 ± 22.7
Median (Q1, Q3)	34.4 (16.1, 55.8)	42.8 (22.0, 61.1)	31.6 (16.5, 50.6)
Sex			
Female	457 (55.9)	1246.1 (56.9)	342.8 (49.0)
Male	360 (44.1)	944.9 (43.1)	356.4 (51.0)
Race			
Native American/Alaskan Native	3 (0.4)	17.0 (0.8)	1.0 (0.1)
Asian	17 (2.1)	45.5 (2.1)	4.0 (0.6)
Black	47 (5.8)	126.6 (5.8)	26.0 (3.7)
Native Hawaiian/Pacific Islander	2 (0.2)	2.0 (0.1)	1.0 (0.1)
White	682 (83.5)	1829.2 (83.5)	619.7 (88.6)
Other/did not disclose	66 (8.1)	171.7 (7.8)	47.5 (6.8)
Ethnicity			
Not Hispanic or Latino	786 (96.2)	2063.0 (94.2)	674.7 (96.5)
Hispanic or Latino	22 (2.7)	90.5 (4.1)	21.5 (3.1)
Unknown/did not disclose	9 (1.1)	37.5 (1.7)	3.0 (0.4)
Anaphylaxis	398 (48.7)	699.2 (31.9)	699.2 (100)
Allergic reaction	158 (19.3)	577.4 (26.3)	—
Other etiology (n = 261, 914.4)			—
Urticaria	25 (3.1)	58.7 (2.7)	—
Angioedema	39 (4.8)	280.6 (12.8)	—
Urticaria with angioedema	12 (1.5)	40.9 (1.9)	—
Asthma exacerbation	4 (0.5)	4.0 (0.2)	—
Anxiety	2 (0.2)	16.5 (0.7)	—
Nonallergic drug reaction	3 (0.4)	21.3 (1.0)	—
Nonallergic dermatitis	8 (1.0)	25.8 (1.2)	—
Indeterminate	12 (1.5)	12.0 (0.5)	—
Other	156 (19.1)	454.7 (20.8)	—
Epinephrine administered	498 (61.0)	970.6 (44.3)	590.2 (84.4)
anaphylactic or allergic reaction trigger (n = 556, 1276.6, 699.2)			
Food	174 (31.3)	302.4 (23.7)	234.9 (33.6)
Medication	106 (19.1)	163.9 (12.8)	62.0 (8.9)
IV contrast	30 (5.4)	58.9 (4.6)	33.5 (4.8)
Venom	58 (10.4)	459.5 (36.0)	195.8 (28.0)
Vaccine	5 (0.9)	5.0 (0.4)	3.0 (0.4)
Immunotherapy	7 (1.3)	7.0 (0.5)	6.0 (0.9)
Latex	1 (0.2)	1.0 (0.1)	0.0 (0.0)
Environmental	9 (1.6)	27.3 (2.1)	2.0 (0.3)
Other	45 (8.1)	59.5 (4.7)	28.0 (4.0)
Unknown	121 (21.8)	192.0 (15.0)	134.0 (19.2)
Food trigger (n = 174, 302.4, 234.9)			
Nuts	80 (46.0)	94.5 (31.2)	83.5 (35.5)
Seafood	18 (10.3)	32.4 (10.7)	30.5 (13.0)
Milk	13 (78.5)	27.4 (9.1)	26.5 (11.3)
Eggs	15 (8.6)	43.1 (14.2)	11.0 (4.7)
Fruits/vegetables	8 (4.6)	22.5 (7.4)	6.0 (2.6)
Other	8 (4.6)	22.5 (7.4)	22.5 (9.6)
Unknown	32 (18.4)	60.1 (19.9)	55.1 (23.4)
Medication trigger (n = 106, 163.9, 62)			
NSAID	19 (17.9)	33.5 (20.4)	13.0 (21.0)
Antibiotics	42 (39.6)	42.0 (25.6)	24.0 (38.7)
Antineoplastics	5 (4.7)	5.0 (3.1)	5.0 (8.1)
Opioid	4 (3.8)	4.0 (2.4)	4.0 (6.5)
Unclear class	5 (4.7)	5.0 (3.1)	2.0 (3.2)
Other	31 (29.2)	74.4 (45.4)	14.0 (22.6)

(Continued)

TABLE I. *(Continued)*

ED patient characteristic	Unweighted all ED visits (N = 817)	Weighted all ED visits (N = 2191)	Weighted anaphylaxis ED visits (N = 699.2)
ED disposition for patients with anaphylactic or allergic reactions (n = 556, 1276.6, 699.2)			
Home	276 (49.6)	680.9 (53.3)	271.5 (38.8)
ED observation unit	170 (30.6)	457.6 (35.8)	315.6 (45.1)
Inpatient, non-ICU	64 (11.5)	78.0 (6.1)	65.0 (9.3)
ICU	45 (8.1)	59.0 (4.6)	47.0 (6.7)
Died	1 (0.2)	1.0 (0.1)	0.0 (0.0)
Epinephrine autoinjector prescribed for patients with anaphylactic or allergic reactions (n = 556, 1276.6, 699.2)			
Yes	442 (79.5)	943.5 (73.9)	627.7 (89.8)
No	109 (19.6)	328.0 (25.7)	66.5 (9.5)
Unknown	5 (0.9)	5.0 (0.4)	5.0 (0.7)

Values are n (%) unless otherwise stated.
ICU, Intensive care unit; *IV*, intravenous; *NSAID*, nonsteroidal anti-inflammatory drug; *Q*, quartile.

TABLE II. Test characteristics of *ICD* code–based and machine learning algorithms to identify ED anaphylaxis visits

Algorithm	Sensitivity (%)	Specificity (%)	Positive predictive value* (%)	Negative predictive value* (%)	Positive likelihood ratio	Negative likelihood ratio
Method 1†,‡	49.1	87.5	64.7	78.6	3.93	0.58
Method 1 + method 2†,‡	53.9	68.7	44.7	76.1	1.72	0.67
Method 1 + method 3†,‡	98.4	15.1	35.2	95.3	1.16	0.11
Machine learning§	87.3	79.1	80.1	86.7	4.17	0.16

*Prevalence of anaphylaxis was 31.9%.
†*ICD* codes are presented in Table E1.
‡Coding combinations for methods 1, 2, and 3 are presented in Table E2.
§Gradient boosting algorithm elements are presented in Table III.

Comparison of Mayo cohort to a large national database population

The full gradient boosting algorithm was used to predict anaphylaxis using the OLDW, a large national health care services claims database of commercially insured and Medicare Advantage beneficiaries in the United States. Characteristics of the ED visits identified by the algorithm in the OLDW were compared with characteristics of patients identified by the algorithm among Mayo patients with commercial or Medicare insurance. With regard to demographic characteristics, the OLDW cohort had a higher proportion of patients in the 0- to 4-year-old and 5- to 14-year-old age groups compared with the Mayo cohort (19% vs 12.4% and 15.1% vs 10.2%, respectively) and had a lower proportion of patients in the 45- to 64-year-old and 65 years and older age groups (13.8% vs 17.8% and 5.9% vs 15.6%, respectively) (see Table E4 in this article's Online Repository at www.jaci-global.org). In addition, the Mayo cohort had an overall higher proportion of females compared with the OLDW cohort (47.2% vs 36.1%, respectively). There were no significant differences in any age group with regard to proportions of patients who arrived via EMS or had critical care billing. The overall proportions of patients who had epinephrine administered was not significantly different between the OLDW and Mayo cohorts (46.9% vs 41.6%, respectively). The overall proportion of patients who were dismissed from the ED or ED observation unit was slightly greater in the Mayo cohort compared with the proportion of patients who were dismissed from the ED in the OLDW cohort (93.5% vs 89.5%, respectively). In addition, the proportion of patients who were provided a prescription for self-injectable epinephrine was slightly higher than the proportion of patients who had self-injectable epinephrine dispensed in the OLDW cohort (96.5% vs 93.1%, respectively).

In addition, the percentages of algorithm-identified ED visits in the OLDW and Mayo cohort were assessed by month of the year to compare seasonal trends (Fig E1). The monthly predicted anaphylaxis cases were greater between June and September compared with all other months for both the OLDW and Mayo cohorts (monthly median visits, 2277 for June-September vs 1480 for all other months among the OLDW cohort, $P = .004$; monthly median visits, 48 for June-September vs 30 for all other months among the Mayo cohort, $P = .008$). There was no difference in seasonal pattern between the 2 cohorts ($P = .677$).

DISCUSSION

To our knowledge, this is the first study to develop and assess a machine learning algorithm for the identification of ED anaphylaxis visits. Incorporation of Current Procedural Terminology codes, patient characteristics, ED arrival mode, self-injectable epinephrine prescriptions, and ED disposition in addition to *ICD*

TABLE III. Candidate predictors for machine learning algorithm to identify ED anaphylaxis visits

ICD-9-CM and *ICD-10-CM* codes*	Total cases (N = 2191)†	Relative importance score (GB algorithm)	Anaphylaxis predicted cases (N = 431)†
Anaphylaxis-specific codes			
Anaphylaxis caused by food	164	3.52	143
Other anaphylactic reaction	317	3.43	207
Anaphylaxis due to vaccination	1	—	1
Anaphylaxis due to blood products	1	—	0
Anaphylaxis due to other serum	1	—	1
Anaphylaxis due to drug	38	3.37	30
Trigger-specific codes			
Unspecified adverse drug	14	0.00	1
Dermatitis due to drug	28	0.00	2
Adverse effect of drug E-codes	252	0.28	68
Other drug allergy	8	0.00	1
Toxic venom reaction E-codes	454	2.28	158
Toxic effect of venom	304	4.85	132
E-codes hornet, wasp, bee sting	423	0.84	143
Dermatitis due to food	78	1.32	35
Adverse food reaction	3	—	0
Allergy unspecified	484	4.03	110
Symptom-specific codes			
Respiratory compromise			
Acute respiratory failure	14	0.00	5
Acute respiratory distress	3	—	2
Dyspnea or respiratory distress	169	1.11	57
Stridor	11	0.00	4
Asthma	172	3.10	76
Edema of larynx	2	—	0
Any above respiratory code	344	1.33	139
Reduced blood pressure			
Hypotension	86	0.00	12
Syncope or collapse	68	0.03	11
Any reduced blood pressure code	144	0.03	21
Gastrointestinal			
Allergic gastroenteritis or colitis	3	—	3
Nausea with vomiting	39	0.05	11
Nausea alone	68	0.28	31
Vomiting alone	21	0.04	16
Abdominal pain	64	0.01	22
Any above gastrointestinal code	173	0.23	72
Skin-mucosal			
Edema	12	0.00	0
Flushing	39	0.00	3
Urticaria	176	0.20	61
Pruritus	109	0.04	5
Any above skin-mucosal code	327	1.04	67
Angioedema	442	2.72	52
Demographics and arrival method			
Age	2191	14.96	431
Sex: female	1249	3.91	352
EMS arrival	503	4.41	235
Disposition			
Not dismissed directly from ED‡	839	4.88	344
Management			
IM epinephrine§	419	8.62	317
Nebulizer treatment	246	1.68	123
Intravenous fluids	433	1.34	177
Endotracheal intubation	23	0.02	10
Critical care billing	247	1.25	141
Epinephrine prescription	1484	14.79	706

BP, Blood pressure; *E-codes*, external cause of injury codes; *GI*, gastrointestinal; *ICU*, intensive care unit; *IM*, intramuscular.

*Includes codes with nonzero counts from 2013 to 2017. However, only those codes present in at least 5 visits were included in the recursive partitioning and machine learning algorithms. See Table E1 for specific ICD codes.

†Numbers account for the sampling weights.

‡Includes patients admitted to the ED observation unit, hospital, or ICU.

§As indicated by 96372 Current Procedural Terminology coding.

anaphylaxis-, allergic-, and symptom-based diagnostic codes resulted in a substantial improvement in prediction of ED anaphylaxis visits.

The inaccuracy and complexities of using *ICD* codes for the identification of ED visits for anaphylaxis and allergic reactions are well recognized.[14,15] Based on the known low sensitivity of anaphylaxis-specific *ICD* codes for identifying anaphylaxis visits, an *ICD*-based algorithm incorporating codes for allergic reactions along with symptom-based codes was previously developed.[11] However, this algorithm has not yet been validated on the basis of patient-level data. Because our goal was to explore a more comprehensive range of coding combinations that might identify anaphylaxis visits, the symptom-based coding algorithm used in this study (method 2) is more expansive than the previously published algorithm. However, despite the addition of more coding combinations, we found a minimally increased sensitivity along with a reduced specificity for identification of anaphylaxis visits, suggesting that the addition of symptom-based coding algorithms may not improve the accuracy of anaphylaxis visit identification. As expected, the addition of individual codes suggestive of an allergic reaction, not in combination with symptom codes (method 3), led to a much higher sensitivity and a much lower specificity. The gradient boosting machine learning algorithm resulted in a better balance of sensitivity and specificity, giving it promise as a tool to identify ED anaphylaxis cases.

Although the test characteristics of the machine learning algorithm were substantially better than the use of *ICD* diagnosis codes with or without the addition of symptom-based *ICD* codes, there is still need for further optimization. With the increasing accessibility of data within the electronic medical record, future algorithms could incorporate additional data such as vital signs, elements of the patient review of systems or physical examination, treatments beyond epinephrine including systemic antihistamines (including H1 and H2 antihistamines) and systemic corticosteroids, or features of clinical notes extracted through natural language processing. In addition, prehospital interventions as well as subsequent outpatient follow-up visits, procedures, and diagnoses could be included. Furthermore, more sophisticated machine learning algorithms may enable more accurate identification of anaphylaxis patients.

In this study, anaphylaxis-specific codes alone (method 1) missed 51% of the anaphylaxis cases and 35% of cases with an anaphylaxis-specific code were not categorized as anaphylaxis on the basis of expert opinion. This supports previous studies highlighting the miscoding and misdiagnosis of anaphylaxis in the ED.[10,16,17] It is likely that some misdiagnosis and miscoding is due to the lack of specific *ICD-9* and *ICD-10* codes for anaphylaxis due to many nonfood triggers and the fact that nonspecific anaphylaxis codes (eg, T78.2 Anaphylactic shock, unspecified) include the word shock, which is infrequent in anaphylaxis.[6] The low sensitivity of anaphylaxis-specific codes was most evident in the case of venom-related anaphylaxis in which the use of anaphylaxis-specific *ICD* codes alone missed 96% of cases. This is consistent with findings published by Harduar-Morano et al[11] showing that Hymenoptera sting reactions were more likely to be coded with symptom-based codes without the use of an anaphylaxis *ICD* code. These findings suggest that epidemiological studies relying on anaphylaxis-specific *ICD* codes alone would not only underestimate the number of anaphylaxis cases but also include cases that had been misdiagnosed. Fortunately, *International Classification of Diseases, Eleventh Revision* includes a more comprehensive categorization for anaphylaxis including specific codes for anaphylaxis due to insect venom (ie, anaphylaxis due to insect venom [4A84.2] and systemic allergic reaction due to Hymenoptera venom [4A85.30]). Thus, the sensitivity and specificity of *ICD* coding over time are likely to improve.[18]

Although the improved anaphylaxis classification included in *International Classification of Diseases, Eleventh Revision* is an important step forward in anaphylaxis epidemiology, underrecognition and subsequent underdiagnosis will likely continue to contribute to potential underestimations of anaphylaxis incidence and prevalence. Underrecognition and underdiagnosis of anaphylaxis have been well documented and are likely due to multiple factors including inadequate *ICD* classification in *ICD-9* and *ICD-10* versions, lack of a universally accepted diagnostic criteria and definitions of anaphylaxis, and highly variable patient presentations.[19,20] Thus, the use of machine learning algorithms may enable more accurate identification of anaphylaxis patients for cohort and epidemiologic studies. Moreover, machine learning algorithms applied clinically may be used to alert ED providers to patients with anaphylaxis similar to currently available sepsis alert systems.[21]

To assess the potential generalizability of our findings, we compared the machine learning algorithm-predicted anaphylaxis cases in our cohort with those in a national database. Although the patients in our cohort tended to be older and had a higher proportion of females, we found that the other characteristics assessed were largely similar. The cohorts overall were not significantly different with regard to proportions of patients who arrived by EMS, had critical care billing, or epinephrine administered. We found that 93.5% of patients predicted by the algorithm in our cohort were dismissed from the ED or ED observation unit compared with 89.5% of patients in the OLDW cohort who were dismissed directly from the ED. When comparing dispositions, we combined patients dismissed from our ED observation unit with those dismissed from the ED directly because it is not possible to differentiate patients in the OLDW cohort who were dismissed from an ED observation unit from those who were dismissed after admission to the hospital under observation status. We estimated that the patients dismissed from our observation unit were more similar to those dismissed from the ED than those dismissed after admission to the hospital under observation status because ED observation unit utilization for anaphylaxis patients in our ED is robust (the largest proportion of our anaphylaxis patients are dismissed from the ED observation unit).[22] Furthermore, only about one-third of EDs in the United States have an observation unit[23] so most anaphylaxis patients are dismissed home after observation in the ED proper rather than an observation unit. We also found that 96.5% of predicted patients in our cohort were prescribed self-injectable epinephrine. We were not able to assess the number of prescriptions dispensed. In comparison, 93.1% of patients in the OLDW cohort had self-injectable epinephrine dispensed. This difference can likely be accounted for by the fact that not all prescriptions were filled among our cohort. This would be consistent with a study by Pourang et al[24] that found that 96% of prescriptions for self-injectable epinephrine were dispensed. Finally, the finding that the seasonality of the anaphylaxis cases was similar supports that the algorithm may be appropriately predicting anaphylaxis ED visits in the national cohort.

Our study has multiple limitations. The retrospective design limits the determination of the diagnosis of anaphylaxis to the data documented in the electronic health record. The experience of the coders may have been another limiting factor. Ascertainment bias is another potential limitation. However, this was mitigated by reviewing cases identified by an extensive array of diagnostic codes and combinations of codes, including anaphylaxis-related codes, allergy-related codes, and symptom-based codes. Despite this, cases of unrecognized anaphylaxis could still have been missed.[20,25] Furthermore, because there are no objective tests for the diagnosis of anaphylaxis, expert opinion was used as the reference standard. Expert opinion allows the incorporation of diagnostic criteria as well as the patient's past medical history, laboratory data, information obtained at follow-up visits, and results of allergy testing. Reliability in the determination of the diagnosis of anaphylaxis was enhanced by having 89% of cases reviewed by board-certified allergist-immunologists with extensive clinical experience and prior participation in 2 studies of anaphylaxis diagnosis and a second-year allergy fellow who participated in duplicate extraction of 42% of charts before single extraction for the remaining 11% of cases. Until sensitive and specific serum markers are identified for the diagnosis of anaphylaxis, the subjective nature of many of the potential symptoms of anaphylaxis will make the definitive diagnosis of anaphylaxis challenging. This was a single-center study with limited patient diversity, which could limit external validity. Furthermore, this study was conducted at an academic medical center and the anaphylaxis management and diagnosis could differ from those in other settings. However, the similarities between epinephrine administration and seasonality between the predicted anaphylaxis patients in the OLDW national cohort and our cohort may suggest that our findings could be generalized to a national population. Finally, *ICD* and Current Procedural Terminology coding changes will impact the algorithm as well.

In conclusion, to our knowledge, this study is the first to develop and assess a machine learning algorithm to identify ED cases of anaphylaxis. Future studies are needed to build upon our work to further refine the machine learning algorithm including other potential candidate predictors to improve its accuracy. Such an algorithm would facilitate progress in the field of anaphylaxis research, which will enable improvements in anaphylaxis prevention, diagnosis, and management.

Clinical implications: This study developed and assessed a machine learning algorithm identifying ED anaphylaxis cases laying a foundation for future algorithms that will more accurately estimate the epidemiologic burden of anaphylaxis.

REFERENCES

1. Cardona V, Ansotegui IJ, Ebisawa M, El-Gamal Y, Fernandez Rivas M, Fineman S, et al. World Allergy Organization anaphylaxis guidance 2020. World Allergy Organ J 2020;13:100472.
2. Decker WW, Campbell RL, Manivannan V, Luke A, St Sauver JL, Weaver A, et al. The etiology and incidence of anaphylaxis in Rochester, Minnesota: a report from the Rochester Epidemiology Project. J Allergy Clin Immunol 2008;122:1161-5.
3. Bohlke K, Davis RL, DeStefano F, Marcy SM, Braun MM, Thompson RS, et al. Epidemiology of anaphylaxis among children and adolescents enrolled in a health maintenance organization. J Allergy Clin Immunol 2004;113:536-42.
4. Shaker MS, Wallace DV, Golden DBK, Oppenheimer J, Bernstein JA, Campbell RL, et al. Anaphylaxis—a 2020 practice parameter update, systematic review, and Grading of Recommendations, Assessment, Development and Evaluation (GRADE) analysis. J Allergy Clin Immunol 2020;145:1082-123.
5. Sampson HA, Munoz-Furlong A, Campbell RL, Adkinson NF Jr, Bock SA, Branum A, et al. Second symposium on the definition and management of anaphylaxis: summary report–second National Institute of Allergy and Infectious Disease/Food Allergy and Anaphylaxis Network symposium. Ann Emerg Med 2006;47:373-80.
6. Loprinzi Brauer CE, Motosue MS, Li JT, Hagan JB, Bellolio MF, Lee S, et al. Prospective validation of the NIAID/FAAN criteria for emergency department diagnosis of anaphylaxis. J Allergy Clin Immunol Pract 2016;4:1220-6.
7. Ruggeberg JU, Gold MS, Bayas JM, Blum MD, Bonhoeffer J, Friedlander S, et al. Anaphylaxis: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine 2007;25:5675-84.
8. Erlewyn-Lajeunesse M, Dymond S, Slade I, Mansfield HL, Fish R, Jones O, et al. Diagnostic utility of two case definitions for anaphylaxis: a comparison using a retrospective case notes analysis in the UK. Drug Saf 2010;33:57-64.
9. Ross MP, Ferguson M, Street D, Klontz K, Schroeder T, Luccioli S. Analysis of food-allergic and anaphylactic events in the National Electronic Injury Surveillance System. J Allergy Clin Immunol 2008;121:166-71.
10. Choi B, Kim SH, Lee H. Missed registration of disease codes for pediatric anaphylaxis at the emergency department. Emerg Med Int 2019;2019:4198630.
11. Harduar-Morano L, Simon MR, Watkins S, Blackmore C. Algorithm for the diagnosis of anaphylaxis and its validation using population-based data on emergency department visits for anaphylaxis in Florida. J Allergy Clin Immunol 2010;126:98-104.e4.
12. Campbell RL, Hagan JB, Manivannan V, Decker WW, Kanthala AR, Bellolio MF, et al. Evaluation of National Institute of Allergy and Infectious Diseases/Food Allergy and Anaphylaxis Network criteria for the diagnosis of anaphylaxis in emergency department patients. J Allergy Clin Immunol 2012;129:748-52.
13. Wallace PJ, Shah ND, Dennen T, Bleicher PA, Crown WH. Optum Labs: building a novel node in the learning health care system. Health Aff (Millwood) 2014;33:1187-94.
14. Clark S, Gaeta TJ, Kamarthi GS, Camargo CA. ICD-9-CM coding of emergency department visits for food and insect sting allergy. Ann Epidemiol 2006;16:696-700.
15. Saff RR, Li Y, Santhanakrishnan N, Camargo CA Jr, Blumenthal KG, Zhou L, et al. Identification of inpatient allergic drug reactions using ICD-9-CM codes. J Allergy Clin Immunol Pract 2019;7:259-64.e1.
16. Hocagil H, Karakilic E, Hocagil C, Senlikci H, Buyukcam F. Underdiagnosis of anaphylaxis in the emergency department: misdiagnosed or miscoded? Hong Kong Med J 2013;19:429-33.
17. Choi B, Kim SH, Lee H. Are registration of disease codes for adult anaphylaxis accurate in the emergency department? Allergy Asthma Immunol Res 2018;10:137-43.
18. Tanno LK, Demoly P. Joint Allergy Academies. How can the World Health Organization's International Classification of Diseases (ICD)-11 change the clinical management of anaphylaxis? Expert Rev Clin Immunol 2018;14:783-6.
19. Aun MV, Kalil J, Giavina-Bianchi P. Adults and children with anaphylaxis in the emergency room: why it is not recognized? Curr Opin Allergy Clin Immunol 2018;18:377-81.
20. Sclar DA, Lieberman PL. Anaphylaxis: underdiagnosed, underreported, and undertreated. Am J Med 2014;127:S1-5.
21. Alsolamy S, Al Salamah M, Al Thagafi M, Al-Dorzi HM, Marini AM, Aljerian N, et al. Diagnostic accuracy of a screening electronic alert tool for severe sepsis and septic shock in the emergency department. BMC Med Inform Decis Mak 2014;14:105.
22. Manivannan V, Hess EP, Bellamkonda VR, Nestler DM, Bellolio MF, Hagan JB, et al. A multifaceted intervention for patients with anaphylaxis increases epinephrine use in adult emergency department. J Allergy Clin Immunol Pract 2014;2:294-9.e1.
23. Wiler JL, Ross MA, Ginde AA. National study of emergency department observation services. Acad Emerg Med 2011;18:959-65.
24. Pourang D, Batech M, Sheikh J, Samant S, Kaplan M. Anaphylaxis in a health maintenance organization: International Classification of Diseases coding and epinephrine auto-injector prescribing. Ann Allergy Asthma Immunol 2017;118:186-90.e1.
25. Tuttle KL, Wickner P. Capturing anaphylaxis through medical records: are ICD and CPT codes sufficient? Ann Allergy Asthma Immunol 2020;124:150-5.