ORIGINAL RESEARCH

A Systematic Review of Clinical Prediction Rules for the Diagnosis of Influenza

Mark H. Ebell, MD, MS, Ivan Rahmatullah, MD, MPH, Xinyan Cai, MPH, Michelle Bentivegna, MPH, Cassie Hulme, MPH, Matthew Thompson, MBChB, MPH, DPhil, and Barry Lutz, PhD

Background: Clinical prediction rules (CPRs) can assist clinicians by focusing their clinical evaluation on the most important signs and symptoms, and if used properly can reduce the need for diagnostic testing. This study aims to perform an updated systematic review of clinical prediction rules and classification and regression tree (CART) models for the diagnosis of influenza.

Methods: We searched PubMed, CINAHL, and EMBASE databases. We identified prospective studies of patients presenting with suspected influenza or respiratory infection and that reported a CPR in the form of a risk score or CART-based algorithm. Studies had to report at a minimum the percentage of patients in each risk group with influenza. Studies were evaluated for inclusion and data were extracted by reviewers working in parallel. Accuracy was summarized descriptively; where not reported by the authors the area under the receiver operating characteristic curve (AUROCC), predictive values, and likelihood ratios were calculated.

Results: We identified 10 studies that presented 14 CPRs. The most commonly included predictor variables were cough, fever, chills and/or sweats, myalgias, and acute onset, all which can be ascertained by phone or telehealth visit. Most CPRs had an AUROCC between 0.7 and 0.8, indicating good discrimination. However, only 1 rule has undergone prospective external validation, with limited success. Data reporting by the original studies was in some cases inadequate to determine measures of accuracy.

Conclusions: Well-designed validation studies, studies of interrater reliability between telehealth an in-person assessment, and studies using novel data mining and artificial intelligence strategies are needed to improve diagnosis of this common and important infection. (J Am Board Fam Med 2021;34:1123–1140.)

Keywords: Clinical Decision Rules, Clinical Medicine, Influenza, Physical Examination, Prospective Studies, Respiratory Diseases, Systematic Reviews

Introduction

Seasonal influenza accounts for losses in workforce productivity, a strain on health services, and an average of approximately 40,000 deaths[1] and a range of 140,000 to 710,000 hospitalizations[2] in the United States every year. Neuraminidase inhibitors such as zanamivir and oseltamivir modestly reduce the duration of symptoms by about 30 hours if initiated within 24 hours, and by 20 hours if initiated within 36 hours of symptoms onset for both influenza virus type A and B.[3,4] Therefore, if prescribed prompt diagnosis of influenza is needed for maximal benefit and to institute infection control measures such as mask wearing and social distancing.[5]

Clinical prediction rules (CPRs) use a combination of signs, symptoms, and sometimes simple

point-of-care tests to assist diagnostic and therapeutic decisions. They can take the form of a risk score (point score) that is typically derived from the β coefficients of a logistic regression, or as a nonparametric classification and regression tree (CART) model that has as its output a decision tree. As has been demonstrated with other conditions such as sore throat,[6] pulmonary embolism,[7,8,9] and ankle injury,[10] clinical prediction rules can help establish a patient's probability of a given condition and inform the interpretation of subsequent diagnostic tests. They have the potential to reduce the need for diagnostic tests, as patients initially classified as low probability for the condition in question may not require further testing or evaluation, and patients with high probability for the condition may be directed to proceed with therapy or other intervention.

Systematic reviews have been performed of individual signs and symptoms of influenza, but were limited by the search strategy, by not using modern analytic methods for diagnostic meta-analysis, and by their age.[11,12] A systematic review of CPRs for the diagnosis of influenza found that the positive predictive value of simple clinical heuristics such as "cough and fever" (26% to 87%) and "cough, fever, and acute onset" (30% to 77%) varied widely between studies.[13] However, that systematic review had similar limitations to the meta-analyses of individual signs and symptoms. It also failed to identify more sophisticated risk scores or algorithms. Our goal is therefore to perform an updated, comprehensive systematic review and meta-analysis of clinical prediction rules for the diagnosis of influenza.

Methods
This study is a systematic review of previously published studies of the accuracy of CPRs for the diagnosis of influenza. A CPR was defined as a point score or algorithm (such as those that result from a classification and regression tree analysis) for the diagnosis of laboratory-confirmed influenza. The study was registered with the PROSPERO database (#CRD42020161801) and followed PRISMA guidance regarding conduct and reporting of a diagnostic meta-analysis.

Data Sources and Searches
We searched PubMed, CINAHL and EMBASE databases in August 2020, with a bridge search in June 2021. The primary search strategy is shown in Appendix 1. It included individual signs and symptoms, whose accuracy is reported in a separate systematic review (manuscript in preparation). We also used a recently developed search strategy designed to be highly sensitive for detection of CPRs, shown in Appendix 2.[14] The limits "has abstract" and "human" were applied to both searches. In addition, the reference lists of included studies were reviewed for additional articles, as were 3 older systematic reviews identified by our search.[10–12]

Study Selection
Studies were included if they recruited a prospective cohort of children, adolescents, or adults presenting with symptoms of respiratory tract infection (RTI) or clinically suspected influenza in the inpatient or outpatient setting. Studies had to report sufficient information to calculate the accuracy of a risk score or classification tree for the diagnosis of laboratory-confirmed influenza; studies presenting a multivariate model not easily used by clinicians were excluded. Simple heuristics that studied the presence or absence of 2 or 3 symptoms for the diagnosis of influenza, such as "cough and fever indicates influenza," were excluded from this study but are addressed in a companion systematic review (manuscript in preparation). No limits were set for country or year; language was limited to publications in English. The reference standard had to be polymerase chain reaction (PCR), culture or serology and had to have been performed in all participants to avoid verification bias. Studies also had to report at a minimum the percentage of patients in each risk group with influenza.

Studies were excluded if they did not enroll patients with acute RTI or suspected influenza, or if they were in a specialized population such as only patients in skilled nursing facilities, immunosuppressed patients, or patients with chronic lung disease. Studies were also excluded if they used a case-control design (ie, recruited patients with known influenza and healthy controls) as this has been shown to inflate the apparent accuracy of diagnostic tests.[15] Finally, studies that reported a multivariate model but did not convert it to a point score that is easily usable by clinicians were excluded.

Data Extraction and Quality Assessment
All abstracts were reviewed for inclusion in parallel by 2 authors, 1 of whom was a physician. For any

abstract deemed potentially of interest by any reviewer, the full article was obtained and reviewed by the corresponding author and 1 other reviewer. Studies meeting inclusion and exclusion criteria were reviewed in parallel by 2 authors who each abstracted study characteristics including the type of validation (if any), study quality, and test accuracy data (true positive, false positive, false negative, and true positive in comparison with a valid reference standard). Discrepancies were resolved through consensus discussion. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool was adapted for our study and definitions for low, unclear and high risk of bias prespecified for each domain as shown in Appendix 3.[14]

Data Synthesis and Analysis

For published CPRs, the probability of influenza in each risk group, likelihood ratios, and the proportion of patients in that risk group are reported. If a CPR reported more than 2 risk groups, the likelihood ratio was calculated for each risk group ("stratum specific likelihood ratios"). If a point score was proposed without identification of low, moderate and high-risk groups, we used clinical judgment to propose

such groups post hoc based on test and treatment thresholds from a previous study and reported those findings separately from the original study findings. For published studies of a point score but reporting the area under the receiver operating characteristic curve (AUROCC), AUROCC was calculated using the pROC package in R.[16]

Results

The search strategy described in Appendix 1 yielded 1209 unique studies from 3 databases, and the initial and bridge searches using the RCSI filter for clinical prediction rule searches in Appendix 2 identified 404 studies, of which 240 were unique for a total of 1449 unique studies. After screening abstracts, the full text of 234 studies was reviewed to determine whether they met the inclusion and exclusion criteria for the systematic review. A total of 10 met those criteria and reported the accuracy of a CPR (either a point score or a classification tree) for influenza. Our review of previous systematic reviews identified no additional studies. Therefore, a total of 10 studies met our inclusion criteria and were included in the analysis. See Figure 1 for the PRISMA flow diagram.

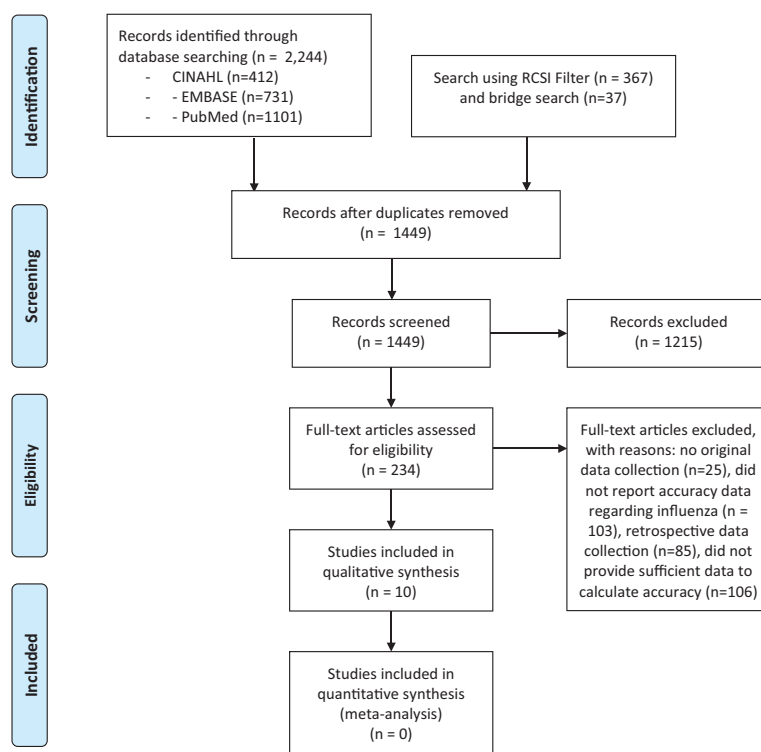**Figure 1. PRISMA diagram describing the search.**

Table 1. Characteristics of Included Studies That Develop or Validate Clinical Prediction Rules for Influenza

Study	No. of Patients	Mean Age (Years)	Population Studied	Flu Strain	Reference Std	Country (Year)
Afonso, 2012*	456	Swiss: 34.3 US: 38.8	Swiss: Adult outpatients with ILI US: adult outpatients with acute RTI symptoms	A + B	Swiss: Culture US: PCR	Switzerland (1999 to 2000), US (2002)
Ebell, 2012*,†	456	Swiss: 34.3 US: 38.8	Swiss: Adult outpatients with ILI US: adult outpatients with acute RTI symptoms	A + B	Swiss: Culture US: PCR	Switzerland (1999 to 2000), US (2002)
Anderson, 2018	4572	10.7	Patients > 6 months of age with fever and cough or sore throat, presenting within 3 days if outpatient and within 5 days if in-patient	A + B	PCR	Thailand (2009 to 2014)
Dugas, 2019	1941	48.6	An ED patient who reported 1 or more of the following: measured fever >100.4°F, subjective fever, cough, nasal congestion, sinus congestion, rhinorrhea, sore throat, or shortness of breath	NR	PCR	US (2013 to 2014)
Govaert, 1998	1791	Range: 60–91	Adults aged 60 or older presenting to GP with ILI Patients with heart, lung, kidney disease, or diabetes excluded	NR	Serology	Netherlands (1991 to 1992)
Ranjan, 2012	638	46.4	Travelers quarantined in an airport during swine flu outbreak with ILI	A	PCR	India (2009)
van Vugt, 2015†	1801	48	Adult outpatients and older with clinical presentation of LRTI, symptom onset < 7 days	A + B	PCR	12 European countries (2007 to 2010)
Vuichard-Gysin, 2019	2191	NR	Children and adults of Canadian Hutterite communities with ARTI defined as at least 2 of: chills, cough, earache, fatigue, fever, headache, myalgias, coryza, or sore throat	A + B	PCR	Canada (2008 to 2011)
Woolpert, 2012	789	31.2	Outpatients with subjective fever and/or temperature >38.0°C, plus cough or sore throat	A + B	PCR	US (2007 to 2008)
Zimmerman, 2016	4173	34.1	Outpatients > 6 months of age, seeking care for ARTI with cough or fever	A + B	PCR	US (2011 to 2012)

ARTI, acute respiratory tract infection; ED, emergency department; GP, general practice; ILI, influenza-like illness; LRTI, lower respiratory tract infection; NR, not reported; PCR, polymerase chain reaction; RTI, respiratory tract infection; Std, standard; US, United States.

*These studies used the same dataset but different methods for generating a clinical prediction rule.

†The FluScore was developed in the study by Ebell and colleagues and validated in the study by van Vugt and colleagues.

Study Characteristics

Characteristics of the included studies are summarized in Table 1. Published stiudies ranged in size from 456 to 4572 patients, and all were based on outpatient populations. Four studies were set in North America, 2 in Europe, 2 in Asia, and 2 publications used a combined dataset from Switzerland and the United States. The reference standard test was PCR for 7 studies, serology for 1, and culture or PCR for 2. Most tested for both influenza A and B and the prevalence of laboratory-confirmed influenza ranged from 6.8% to 34.2%.

These 10 studies reported a total of 7 CART algorithms and 7 risk scores, summarized in Appendix 4. Two studies used the same dataset.[17,18] For 2 studies, we proposed a new "post hoc" risk score for each 1 using their data based on our assessment of what we considered to be clinically useful low, moderate and high-risk groups.[19,20]

Study Quality

Assessment of study quality is summarized in Table 2. Six studies were judged to be at low risk of bias, while 4 were judged at moderate risk of bias. For the latter, in 1 dataset (reported by 2 studies) the potential bias was because 2 different reference standards were used (culture and PCR).[17,18] A second study did not account for all patients at the end of the study,[21] while a third study did not use PCR for any patients.[19]

Validation Procedures

With regards to validation of proposed CPRs, 2 studies performed no validation and only presented data from the derivation population.[20,22] Five studies performed internal validation using a split sample;[17,18,21,23,24] this approach randomly divides a single dataset into 2 groups, 1 to derive the CPR and 1 to validate or test it. This approach is considered less robust than true external validation in a completely new population, since the characteristics

Table 2. Evaluation of Study Quality Using the QUADAS-2 Framework for Studies of Influenza Clinical Prediction Rules

Study, Year	Patient Selection					Index Test				Reference Std				Flow and Timing				Overall
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
	Consecutive	Not Case Control or Retrospective-Cohort	Exclusion Criteria	Risk of Bias	Applicability	Index Blinded	Threshold Pre-Specified	Risk of Bias	Applicability	PCR Used for Some or All Patients	Reference Blinded	Risk of Bias	Applicability	All Got Reference Std	All Got Same Ref Std	All Accounted	Risk of Bias	Overall
Anderson, 2018	Y	Y	Y	L	N	Y	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Ranjan, 2012	Y	Y	Y	L	N	Y	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
van Vugt, 2015	Y	Y	Y	L	N	Y	Y	L	L	Y	Y	L	L	Y	Y	Y	L	L
Vuichard-Gysin, 2019	Y	Y	Y	L	N	Y	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Woolpert, 2012	Y	Y	Y	L	N	Y	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Zimmerman, 2016	Y	Y	Y	L	N	Y	Y	L	L	Y	U	L	L	Y	Y	Y	L	L
Afonso, 2012	Y	Y	Y	L	N	Y	Y	L	L	Y	U	L	L	Y	N	Y	H	M
Ebell, 2012	Y	Y	Y	L	N	Y	Y	L	L	Y	U	L	L	Y	N	Y	H	M
Dugas, 2019	Y	Y	Y	L	N	Y	Y	L	L	Y	U	L	L	Y	Y	N	H	M
Govaert, 2013	Y	Y	Y	L	N	Y	Y	L	L	N	U	H	L	Y	Y	Y	L	M

H, High; L, Low; M, Moderate; N, No; PCR, polymerase chain reaction; Std, standard; U, Uncertain; Y, Yes.

of the patients in the 2 groups are very similar.[25] One study performed prospective temporal internal validation, developing the score with data from 2009 to 2013 and validating it with data from the same clinical center in 2014.[26] The preferred approach of external, prospective validation in a completely new population was performed for a single risk score, the FluScore, in 1 published study.[27]

Predictor Variables

The predictor variables in each risk score or CART model are summarized in Table 3. Cough, fever, chills and/or sweats, and myalgias, were present in between 7 and 12 models. Acute onset, coryza, sore throat, age, headache, and fatigue or malaise were present in between 2 and 4 models, while sneezing, whether influenza was circulating in the community, close contact with influenza, male sex, and sinus problems were each present in a single model.

Accuracy of Clinical Prediction Rules

The accuracy of CPRs varied widely; it is summarized in Table 4. The AUROCC ranged from 0.64 to 0.79 in studies of risk scores, and was not reported by 4 studies. The AUROCC ranged from 0.68 to 0.82 for CART models. The proportion of patients classified as low risk by a CPR ranged from 16% to 89%, with a wide range of probabilities of influenza in these low-risk groups depending on the risk score and the overall prevalence of influenza in the population.

Validation of Clinical Prediction Rules

Two CPRs were prospectively validated. Van Vugt and colleagues used data from the European multicenter GRACE study to validate the FluScore.[27] The AUROCC was 0.79 in the original study (combining both derivation and validation) and 0.71 in the prospective validation. The stratum specific likelihood ratios were 0.10/1.02/2.47 and 0.51/2.53/3.24 for the original and prospective validation groups respectively. More patients were classified as low risk in the study by van Vugt and colleagues,[27] which is clinically useful for telephone or other triage, but this group had a 13.6% prevalence of influenza compared with 5.0% in the original validation.

Anderson and colleagues developed a CART model using data from 3782 patients identified in 2009 to 2013, and validated it in the same population using 2014 data for 790 patients.[26] They report

only an overall AUROCC of 0.69 for derivation and validation groups combined, and positive likelihood ratio (LR+) of 2.73 and negative likelilhood ratio (LR-) of 0.52 for the derivation group (data to calculate likelihood ratios not reported for the validation group). The prevalence of influenza in their low- and high-risk groups was consistent, 21.9% vs 20.2% and 59.7% vs 63.8% in derivation and validation groups respectively.

Discussion

Our systematic review identified 7 published risk scores and 7 published CART algorithms for diagnosis of influenza. Only 2 of the CPRs have had any kind of prospective validation. One CPR did not validate particularly well,[27] but the validation study did not directly ask patients about the presence of chills or sweats, and apparently assumed that they were present in all patients with fever. This assumption could adversely affect validation.[27] The other prospective validation study was prospective temporally but studied patients at the same clinical site,[24] which is considered inferior to prospective validation in an entirely new population.[25] The overall accuracy of most rules was moderately good based on areas under the ROC curve that were typically between 0.7 and 0.8. Despite being a common condition, the development and especially validation of CPRs for influenza remains poorly studied.[25] While the use of anti-influenza drugs is not universally endorsed, making an accurate, early diagnosis could help implement mask wearing and distancing to reduce spread of the disease in the community, as was seen during the COVID-19 pandemic.[5]

Only 4 CPRs divided patients into low, moderate and high-risk groups,[17,18] with the remainder generally identifying only low- and high-risk groups. We would argue that it is most useful to clinicians to identify 3 risk groups, which can potentially correspond to the the 3 groups identified by Pauker and Kassirer's threshold model of diagnosis: below the test threshold (rule out), between the test and treatment thresholds (gather more information), and above the treatment threshold (treat empirically).[28] These thresholds have been studied, with 1 study finding a test threshold of 5% for US physicians and 32% for Swiss physicians, and treatment thresholds of 55% and 68% respectively.[29] A second study in older adults

Table 3. Predictor Variables Included in Each Influenza Risk Score or Classification and Regression Tree (CART) Model

Study, Year	Cough	Fever	Chills or Sweats	Myalgia	Acute Onset	Coryza	Sore Throat	Age	Headache	Fatigue or malaise	Sneezing	Flu in community	Close Contact	Sex	Sinus Problems
Afonso, 2012	×		×		×										
Afonso, 2012	×	×	×												
Afonso, 2012	×	×		×											
Anderson, 2018	×		×	×		×		×							
Dugas, 2019	×	×				×			×						
Ebell, 2012 (FluScore)	×	×	×	×	×										
Govaert, 1998	×	×	×	×	×		×		×	×					
Ranjan, 2012	×	×				×	×				×	×	×		
Vuichard-Gysin, 2019 (children)	×	×	×			×	×					×	×	×	
Vuichard-Gysin, 2019 (children)	×	×	×					×							
Vuichard-Gysin, 2019 (adults)	×		×	×			×								×
Vuichard-Gysin, 2019 (adults)	×		×	×											
Woolpert, 2012	×			×	×										
Zimmerman, 2016	×	×								×					
Number of models:	12	11	9	7	4	3	3	2	2	2	1	1	1	1	1

Table 4. Accuracy of Clinical Prediction Rules for Influenza

Study or Dataset	Type of Validation and Number of Patients	AUROCC	Overall Prevalence of Flu	% Flu and LR by Risk Group (Derivation)	% Flu and LR by Risk Group (Validation)	% of Patients in Each Risk Group
FluScore risk score						
Ebell, 2012	Split sample internal: derivation n = 326, validation n = 133	0.79 *	34.2%	Low: 9.2%, LR 0.20	Low: 5.0%, LR 0.10	Low: 32.5% *
				Mod: 27.8%, LR 0.76	Mod: 35.9%, LR 1.02	Mod: 28.1%
				High: 59.1%, LR 2.83	High: 57.4%, LR 2.47	High: 39.4%
van Vugt, 2015 (flu season)	Prospective external: n = 505	0.71 †	23.6%	NA	Low: 13.6%, LR 0.51	Low: 60% ‡
					Mod: 32.1%, LR 1.53	Mod: 27%
					High: 50.0%, LR 3.24	High: 14%
Classification and Regression Tree (CART) models						
Afonso, 2012: Model 1	Split sample internal: derivation n = 322, validation n = 134	0.82/0.80 *	34.2%	Low: 5.6% flu, LR 0.12	Low: 8%, LR 0.15	Low: 21% *
				Mod: 29% flu, LR 0.83	Mod: 37%, LR 0.98	Mod: 62%
				High: 78% flu, LR 7.2	High: 82%, LR 7.8	High: 17%
Afonso, 2012: Model 2	Split sample internal: derivation n = 322, validation n = 134	0.75/0.76 *	34.2%	Low: 5.6% flu, LR 0.12	Low: 8%, LR 0.15	Low: 21% *
				Mod: 18% flu, LR 0.44	Mod: 18%, LR 0.37	Mod: 30%
				High: 55% flu, LR 2.4	High: 62%, LR 2.7	High: 49%
Afonso, 2012: Model 3	Split sample internal: derivation n = 322, validation n = 137	0.76/0.77 *	34.2%	Low: 7.5% flu, LR 0.16	Low: 3%, LR 0.06	Low: 24% *
				Mod: 26% flu, LR 0.72	Mod: 34%, LR 0.89	Mod: 46%
				High: 63% flu, LR 3.4	High: 70%, LR 3.9	High: 30%
Anderson, 2018	Prospective internal: derivation n = 3782 (2009 to 2013), validation n = 790 (2014)	0.69 *	32.7%	Low: 21.9% flu, LR 0.52	Low: 20.2% flu, LR NR	Low: 71% †
				High: 59.7% flu, LR 2.73	High: 63.8%, LR NR	High: 29%
Vuichard-Gysin, 2019 (children)	Split sample internal: derivation n = 819, validation n = 422	0.77/0.74 $	12.3%	Low: 8% flu, LR 4.5	Low: 8%, LR 0.72	Low: 88.6% †
				High: 36% flu, LR 4.5	High: 34%, LR 4.3	High: 11.4%
Vuichard-Gysin, 2019 (adults)	Split sample internal: derivation n = 627, validation n = 323	0.80/0.75 $	7.1%	Low 4% flu, LR 0.59	Low: 5%, LR 0.68	Low: 89.3% †
				High 30% flu, LR 5.8	High: 26%, LR 4.6	High: 10.7%
Zimmerman, 2016	Split sample internal: derivation n = 2087, validation n = 2086	0.68/0.69 $	15.4%	Low: 6% flu, LR 0.33	Low 5%	43% *
				High: 23% flu, LR 1.63	High: 23%	57%
Other risk scores						
Dugas, 2019	Split sample internal: derivation n = 1553, validation n = 388	NR	9.4%	Low (0 to 2): 1.5%, LR 0.16	Low (0 to 2): 3.3%, LR 0.25	Low: 33.4% *
				High (3+): 12.5%, LR 1.48	High (3+): 16.2%, LR 1.40	High: 66.6%
Govaert, 1998 (post hoc)	None, derivation only, n = 1791	NR	6.8%	Low (0): 3.2%, LR 0.46;	NA	Low: 64.5% †
				Mod (1 to 2): 6.2%, LR 0.91		Mod: 8.1%
				High (≥ 3): 15%, LR 2.49		High: 27.4%
Ranjan, 2012	None, derivation only, n = 638	NR	19.9%	Low (0 to 6): 4%	NA	NR
				High (7+): 64%		
Vuichard-Gysin, 2019 (children)	Split sample internal: derivation n = 819, validation n = 422	0.76/0.70 $	12.3%	Low (0 to 4): 6% flu, LR 0.53	Low (0 to 4): 7%, LR 0.59	Low: 80.1% *
				High (5+): 31% flu, LR 3.7	High (5+): 28%, LR 3.1	High: 19.9%
Vuichard-Gysin, 2019 (adults)	Split sample internal: derivation n = 627, validation n = 323	0.78/0.79 $	7.1%	Low (0 to 3): 6% flu, LR 0.59;	Low (0 to 3): 5%, LR 0.67	Low: 94.2% *
				High (4+): 34% flu, LR 5.8	High (4+): 26%, LR 4.8	High: 5.8%

Continued

Table 4. Continued

Study or Dataset	Type of Validation and Number of Patients	AUROCC	Overall Prevalence of Flu	% Flu and LR by Risk Group (Derivation)	% Flu and LR by Risk Group (Validation)	% of Patients in Each Risk Group
Woolpert, 2012 (original)	None, derivation only, n = 523	NR	30.0%	0 pts: 0.0%, LR 0.0	NA	0 pts: 0.6% *
				1 pts: 10.1%, LR 0.26		1 pt: 15.1%
				2 pts: 20.9%, LR 0.62		2 pts: 53.0%
				3 pts: 50.0%, LR 2.33		3 pts: 26.0%
				4 pts: 82.1%, LR 10.7		4 pts: 5.4%
Woolpert, 2012 (post hoc)	None, derivation only, n = 523	NR	30.0%	Low (0 to 1): 9.8%, LR 0.25	NA	Low: 15.7% *
				Mod (2): 20.9%, LR 0.61		Mod: 53.0%
				High (3+): 55.5%, LR 2.87		High: 31.4%

AUROCC, area under the receiver operating characteristic curve; EAST-PC, Enhancing Antibiotic Stewardship in Primary Care dataset; Mod, moderate; LR, likelihood ratio; NA, not applicable; NR, not reported.
*All subjects.
†Derivation group only.
‡ Validation group only.
$ Derivation/Validation.

estimated test and treament thresholds of 10% and 41% respectively.[30] The test threshold in particular will be influenced by the availability and adoption of diagnostic testing versus relying on clinical diagnosis alone. A CPR in the US would be most useful if the low-risk group had a very low risk of influenza, so testing could be avoided, while the high-risk group had a probability of influenza of at least 55% so treatment could be started empirically. This is especially relevant for telehealth settings, and when advising patients whether they should be evaluated.

An often overlooked metric for CPRs is the proportion of patients classified as low or high risk. If the CPR is designed properly with test and treatment thresholds in mind, the probability of disease in the low-risk group is low enough that the disease can be ruled out without testing, and the probability of disease in the high-risk group is high enough that empiric therapy may be appropriate. In this case, it is most useful if a larger proportion of patients fall into those low or high-risk groups. The moderate risk group requires further evaluation and potentially testing (for example using a point-of-care influenza test) which might require an in-person visit and is therefore more expensive and burdensome. How low a risk is "low enough" will vary with the test threshold, which as we have previously demonstrated can vary considerably between countries.[29] It also likely varies at the individual level, as

providers may vary in their perception of risk of influenza complications and/or benefits of treatment or other measures such as quarantine.

The most commonly used predictor variables were cough, fever, chills or sweats, myalgias, acute onset, coryza, sore throat, headache and fatigue. Future studies of influenza diagnosis should consider this a minimum dataset, but should also consider combinations of symptoms as interaction terms in multivariate models, differential weighting by symptom severity, and novel symptoms such as the time to maximum symptoms severity. Different models are likely needed for different age groups, and vaccine status should be included more often in models.[31,32]

Limitations of our meta-analysis include the failure of many studies to fully report diagnostic accuracy, for example the area under the receiver operating characteristic curve. We were also unable to perform meta-analysis, since with only 1 exception[17,27] each study reported a different proposed CPR.

In the COVID-19 pandemic, more and more care is being provided via telehealth.[33] This is likely to persist for some time, and if high quality care of respiratory infections such as COVID-19 or influenza can be delivered remotely, there is the potential for telehealth to reduce transmission to others in the outpatient setting. The most commonly used

predictor variables (Table 3) can all be ascertained remotely, and could potentially be supplemented by home testing for influenza.[34]

Future studies should follow reporting recommendations of the "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis" (TRIPOD) statement including clear definitions of predictors and outcomes, full reporting of models, and prospective validation in new populations.[25] In addition, there is a need for prospective validation of existing rules, as well as the development and prospective validation of more CPRs developed using novel machine learning strategies.[31,32] Studies of interrater reliability of patient and physician assessment of symptoms of influenza are also needed. In addition to prospectively gathering new data, existing datasets could be pooled for individual patient meta-analysis of diagnosis as well as to develop and internally validate new CPRs. Rules designed to use symptoms only to rule out influenza could be especially useful for telehealth or triage, reducing the need for in-person visits.

To see this article online, please go to: http://jabfm.org/content/34/6/1123.full.

References

1. Dushoff J, Plotkin JB, Viboud C, Earn DJD, Simonsen L. Mortality due to influenza in the United States - An annualized regression approach using multiple-cause mortality data. Am J Epidemiol 2006;163:181–7.
2. Rolfes MA, Foppa IM, Garg S, et al. Annual estimates of the burden of seasonal influenza in the United States: a tool for strengthening influenza surveillance and preparedness. Influenza Other Respir Viruses 2018;12:132–7.
3. Ebell MH, Call M, Shinholser J. Effectiveness of oseltamivir in adults: a meta-analysis of published and unpublished clinical trials. Fam Pract 2013;30:125–33.
4. Jefferson T, Jones M, Doshi P, Del Mar C. Neuraminidase inhibitors for preventing and treating influenza in healthy adults: systematic review and meta-analysis. BMJ 2009;339:b5106.
5. Chiu NC, Chi H, Tai YL, et al. Impact of wearing masks, hand hygiene, and social distancing on influenza, enterovirus, and all-cause pneumonia during the coronavirus pandemic: Retrospective national epidemiological surveillance study. J Med Internet Res 2020;22:e21257.
6. Centor RM, Witherspoon JM, Dalton HP, Brody CE, Link K. The diagnosis of strep throat in adults in the emergency room. Med Decis Making 1981;1:239–46.
7. van Belle A, Büller HR, Huisman MV, Christopher Study Investigators, , et al. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. JAMA 2006;295:172–9.
8. Lucassen W, Geersing GJ, Erkens PM, et al. Clinical decision rules for excluding pulmonary embolism: a meta-analysis. Ann Intern Med 2011;155:448–60.
9. Kline JA, Mitchell AM, Kabrhel C, Richman PB, Courtney DM. Clinical criteria to prevent unnecessary diagnostic testing in emergency department patients with suspected pulmonary embolism. J Thromb Haemost 2004;2:1247–55.
10. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Worthington JR. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. Ann Emerg Med 1992;21:384–90.
11. Ebell MH, White LL, Casault T. A systematic review of the history and physical examination to diagnose influenza. J Am Board Fam Pract 2004;17:1–5.
12. Call SA, Vollenweider MA, Hornung CA, Simel DL, McKinney WP. Does this patient have influenza? JAMA 2005;293:987.
13. Ebell MH, Afonso A. A systematic review of clinical decision rules for the diagnosis of influenza. Ann Fam Med 2011;9:69–77.
14. Ebell MH, Fahey T, Murphy ME, Barry A, Barry H, Hickner J. An updated and more efficient search strategy to identify primary care–relevant clinical prediction rules. J Clin Epidemiol 2020;125:26–9.
15. Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. Stat Med 2002;21:1525–37.
16. R Development Core Team R. R: A Language and Environment for Statistical Computing; 2011.
17. Ebell MH, Afonso AM, Gonzales R, Stein J, Genton B, Senn N. Development and validation of a clinical decision rule for the diagnosis of influenza. J Am Board Fam Med 2012;25.
18. Afonso AM, Ebell MH, Gonzales R, Stein J, Genton B, Senn N. The use of classification and regression trees to predict the likelihood of seasonal influenza. Fam Pract 2012;29:671–7.
19. Govaert TME, Dinant GJ, Aretz K, Knottnerus JA. The predictive value of influenza symptomatology in elderly people. Fam Pract 1998;15:16–22.
20. Woolpert T, Brodine S, Lemus H, Waalen J, Blair P, Faix D. Determination of clinical and demographic predictors of laboratory-confirmed influenza with subtype analysis. BMC Infect Dis 2012;12:129.
21. Dugas AF, Hsieh YH, Lovecchio F, Emergency Department National Influenza Network Investigators, et al. Derivation and validation of a clinical decision guideline for influenza testing in 4 US emergency departments. Clin Infect Dis 2020;70:49–58.

22. Ranjan P, Kumari A, Das R, Gupta L, Singh S, Yadav M. Evaluation of clinical features scoring system as screening tool for influenza A (H1N1) in epidemic situations. J Postgrad Med 2012;58: 265–9.

23. Vuichard-Gysin D, Mertz D, Pullenayegum E, Singh P, Smieja M, Loeb M. Development and validation of clinical prediction models to distinguish influenza from other viruses causing acute respiratory infections in children and adults. PLoS ONE 2019;14:e0212050.

24. Zimmerman RK, Balasubramani GK, Nowalk MP, et al. Classification and regression tree (CART) analysis to predict influenza in primary care patients. BMC Infect Dis 2016;16:503.

25. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–73.

26. Anderson KB, Simasathien S, Watanaveeradej V, et al. Clinical and laboratory predictors of influenza infection among individuals with influenza-like illness presenting to an urban Thai hospital over a five-year period. PLoS ONE 2018;13:e0193050.

27. van Vugt SF, Broekhuizen BDL, Zuithoff NPA, on Behalf of the GRACE Consortium, , et al. Validity of a clinical model to predict influenza in patients presenting with symptoms of lower respiratory tract infection in primary care. FAMPRJ 2015;32: cmv039.

28. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med 1980;302: 1109–17.

29. Ebell MH, Locatelli I, Senn N. A novel approach to the determination of clinical decision thresholds. Evidence-Based Med 2015;20.

30. Rothberg MB, Bellantonio S, Rose DN. Management of influenza in adults older than 65 years of age: cost-effectiveness of rapid testing and antiviral therapy. Ann Intern Med 2003;139:321–9.

31. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 1996;49: 1225–31.

32. Phillips ND, Neth H, Woike JK, Gaissmaier W. FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. Judgment and Decision Making 2017;12(4):344–68.

33. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020;20:1193.

34. Lyon V, Zigman Suchsland M, Chilver M, et al. Diagnostic accuracy of an app-guided, self-administered test for influenza among individuals presenting to general practice with influenza-like illness: study protocol. BMJ Open 2020;10:e036298.

Appendix 1. Search for Diagnosis of Influenza Using the History and Physical Examination

Databases Used
1. PubMed
2. CINAHL
3. EMBASE

Keyword Development
    We developed keyword combinations to retrieve articles from the databases that contain 4 following components:

1. Influenza.
2. Methods related to clinical decision rule/predictive modelling to diagnose Influenza
3. Laboratory tests as the reference standards to diagnose Influenza.
4. Sign and symptoms related to Influenza.

Summary of Search Activities
    We generated generic terms and subject headings that are specific to each database based on the 4 components of keyword combinations. We applied the combinations of the generic terms and the subject headings differently in each database using their specific algorithm. The date of our last search was on October 14, 2019.

PubMed

- Keyword component #1:
    ((influenza OR flu [tiab]) OR "Influenza, Human"[Mesh])

- Keyword component #2:
    ("predictive"[tiab] OR "predictor*"[tiab] OR "prediction*"[tiab] OR "clinical model"[tiab] OR "clinical score"[tiab] OR "clinical scoring"[tiab] OR "clinical rule"[tiab] OR "decision guideline"[tiab] OR "decision model"[tiab] OR "validation study"[tiab] OR "validation studies"[tiab] OR "derivation study"[tiab] OR "screening score"[tiab] OR "decision rule"[tiab] OR "diagnostic rule"[tiab] OR "diagnostic score"[tiab] OR "predictive outcome"[tiab] OR "predictive rule"[tiab] OR "scoring"[tiab] OR "risk score"[tiab] OR "risk scoring"[tiab] OR "prognostic score"[tiab] OR "prognostic index"[tiab] OR "prognostic rule"[tiab] OR "prospective validation"[tiab] OR "diagnostic model"[tiab] OR "clinical definition"[tiab] OR "case definition*"[tiab] OR "Bayes Theo-rem"[Mesh] OR "Decision Support Tech-niques"[Mesh] OR "sensitivity and specificity" [MeSH Terms] OR "predictive value of tests" [MeSH Terms] OR ("risk"[tiab] AND "tool"[tiab]) OR (("validate"[tiab] OR "validation"[tiab] OR "validating"[tiab] OR "develop"[tiab] OR "development"[tiab] OR "derivation"[tiab] OR "derive"[tiab] OR "deriving"[tiab] OR "performance"[tiab]) AND ("decision"[tiab] OR "predictive"[tiab] OR "prediction"[tiab] OR "rule"[tiab] OR "score"[tiab] OR "scoring"[tiab] OR "index"[tiab] OR "model"[tiab] OR "scale"[tiab] OR "tool"[tiab] OR "algorithm"[tiab])) OR ("development"[tiab] AND "validation"[tiab]) OR ("derivation"[tiab] AND "validation"[tiab]))

- Keyword component #3:
    ("gold standard"[tiab] OR "diagnostic testing"[tiab] OR "laboratory"[tiab] OR culture[tiab] OR "virus positive"[tiab] OR immunofluorescence[tiab] OR "viral culture"[tiab] OR "Polymerase Chain Reaction"[Mesh] OR "Clinical Laboratory Techniques"[Mesh] OR "Reverse Transcriptase Polymerase Chain Reaction"[MeSH Terms] OR "Influenza A virus/isolation and purification"[MeSH Terms] OR "Predictive Value of Tests"[MeSH Terms] OR PCR[All Fields])

- Keyword component #4:
    (symptom*[tiab] OR sign*[tiab] OR temperature[tiab] OR stiffness[tiab] OR myalgia[tiab] OR rhinorrhea [tiab] OR cough[tiab] OR fever[tiab] OR "sore throat"[tiab] OR headache[tiab] OR "Fever"[Mesh] OR "Cough"[Mesh] OR "Pharyngitis"[Mesh] OR "Headache"[Mesh] OR"Myalgia"[Mesh])

- <u>Filter</u>: English as language and the presence of abstract
- Combined keywords
Search: #1 AND #2 AND #3 AND #4
Filters: Abstract, English

- Results: 1101

CINAHL

- Keyword component (#1):

(MH "Influenza+") OR TI ((Influenza OR flu)) OR AB ((Influenza OR flu))

- Keyword component (#2):

(MH "Predictive Validity") OR (MH "Predictive Value of Tests") OR (MH "Predictive Research") OR (MH "Cox Proportional Hazards Model") OR (MH "Multiple Logistic Regression") OR (MH "Loglinear Models") OR (MH "Decision Trees") OR (MH "Vali-dation Studies") OR TI ("predictive" OR "predic-tor*" OR "prediction*" OR "clinical model" OR "clinical score" OR "clinical scoring" OR "clinical rule" OR "decision guideline" OR "decision model" OR "validation study" OR "validation studies" OR "deriva-tion study" OR "screening score" OR "decision rule" OR "diagnostic rule" OR "diagnostic score" OR "pre-dictive outcome" OR "predictive rule" OR "scoring" OR "risk score" OR "risk scoring" OR "prognostic score" OR "prognostic index" OR "prognostic rule" OR "prospective validation" OR "diagnostic model" OR "clinical definition" OR "case definition*") OR AB ("pre-dictive" OR "predictor*" OR "prediction*" OR "clinical model" OR "clinical score" OR "clinical scoring" OR "clinical rule" OR "decision guideline" OR "decision model" OR "validation study" OR "validation studies" OR "derivation study" OR "screening score" OR "dec-ision rule" OR "diagnostic rule" OR "diagnostic score" OR "predictive outcome" OR "predic-tive rule" OR "scoring" OR "risk score" OR "risk scoring" OR "prognostic score" OR "prognostic index" OR "prognostic rule" OR "prospective validation" OR "diagnostic model" OR "clinical definition" OR "case defini-tion*") OR TI (("risk" AND "tool")) OR AB (("risk" AND "tool")) OR TI ((("validate" OR "validation" OR "validating" OR "develop" OR "development" OR "derivation" OR "derive" OR "deriving" OR "perform-ance") AND ("decision" OR "predictive" OR "prediction" OR "rule" OR "score" OR "scoring" OR "index" OR "model" OR "scale" OR "tool" OR "algorithm"))) OR AB ((("validate" OR "validation" OR "validat-ing" OR "develop" OR "development" OR "derivation" OR "derive" OR "deriving" OR "performance") AND ("decision" OR "predictive" OR "prediction" OR "rule" OR "score" OR "scoring" OR "index" OR "model" OR "scale" OR "tool" OR "algorithm"))) OR TI (("development" AND "validation") OR ("deri-vation" AND "val-idation")) OR AB (("development" AND "validation") OR ("derivation" AND "validation"))

- Keyword component (#3):

(MH "Signs and Symptoms+") OR (MH "Body Temperature Changes+") OR (MH "Fever+") OR (MH "Cough") OR (MH "Muscle Pain") OR (MH "Pharyngitis") OR (MH "Headache+") OR TI (symptom OR symptoms OR sign OR signs OR temperature OR stiffness OR myalgia OR rhinorrhea OR cough OR fever OR "sore throat" OR headache) OR AB (symptom OR symptoms OR sign OR signs OR temperature OR stiffness OR myalgia OR rhinorrhea OR cough OR fever OR "sore throat" OR headache)

- Keyword component (#4):

(MH "Diagnosis, Laboratory+") OR (MH "Polymerase Chain Reaction+") OR (MH "Reverse Transcriptase Polymerase Chain Reaction") OR (MH "Microbial Culture and Sensitivity Tests") OR (MH "Cell Culture Techniques") OR (MH "Fluorescent Antibody Technique+") OR TI ("gold standard" OR "diagnostic testing" OR "laboratory" OR culture OR "virus positive" OR immunofluorescence OR "viral culture" OR PCR) OR AB ("gold standard" OR "diagnostic testing" OR "laboratory" OR culture OR "vi-rus positive" OR immunofluorescence OR "viral culture" OR PCR)

- Filter: None
- Combined keywords:

#1 AND #2 AND #3 AND #4

- Results: 412

EMBASE

- Keyword component (#1):

influenza:ab,ti OR 'influenza'/exp

- Keyword component (#2):

'prediction'/exp OR 'predictive value'/exp OR 'predictor variable'/exp OR 'predictive validity'/exp OR 'bayes theorem'/exp OR 'sensitivity and specificity'/exp OR 'clinical model':ab,ti OR 'clinical score':ab,ti 'clinical scoring':ab,ti OR 'clinical rule':ab,ti OR 'decision guideline':ab,ti OR 'decision model':ab,ti OR 'validation study':ab,ti OR 'validation studies':ab,ti OR 'derivation study':ab,ti OR 'screening score':ab,ti OR 'decision rule':ab,ti OR 'diagnostic rule':ab,ti OR 'diagnostic score':ab,ti OR 'predictive outcome':ab,ti OR 'predictive

rule':ab,ti OR 'scoring':ab,ti OR 'risk score':ab,ti OR 'risk scoring':ab,ti OR 'prognostic score':ab,ti OR 'prognostic index':ab,ti OR 'prognostic rule':ab,ti OR 'prospective validation':ab,ti OR 'diagnostic model':ab,ti OR 'clinical definition':ab,ti OR 'case definition*':ab,ti

- Keyword component (#3):

'gold standard'/exp OR 'dyes, reagents, indicators, markers and buffers'/exp OR 'polymerase chain reaction'/exp OR 'reverse transcription polymerase chain reaction'/exp OR 'immunofluorescence'/exp OR 'virus culture'/exp OR 'gold standard':ab,ti OR 'diagnostic testing':ab,ti OR 'laboratory':ab,ti OR culture:ab,ti OR 'virus positive':ab,ti OR immunofluorescence:ab,ti OR 'viral culture':ab,ti

- Keyword component (#4):

'fever'/exp OR 'coughing'/exp OR 'pharyngitis'/exp OR 'headache'/exp OR 'myalgia'/exp OR 'rhinorrhea'/exp OR 'physical disease by body function'/exp OR symptom:ab,ti OR symptoms:ab,ti OR sign:ab,ti OR signs:ab,ti OR temperature:ab,ti OR stiffness:ab,ti OR myalgia:ab,ti OR rhinorrhea:ab,ti OR cough:ab,ti OR fever:ab,ti OR 'sore throat':ab,ti OR headache:ab,ti

- Filter: None
- Combined keywords:

#1 AND #2 AND #3 AND #4

- Results: 731

Appendix 2. Search for clinical prediction rules using RCSI filter[1]

"Influenza"[ti] and
 ("diagnosis"[tiab] OR "diagnostic") AND (("clinical prediction"[tiab] OR "clinical model"[tiab] OR "clinical score"[tiab] OR "clinical scoring"[tiab] OR "validation of a clinical"[tiab] OR "decision guide-line"[tiab] OR "validation study"[tiab] OR "validation studies"[tiab] OR "derivation study"[tiab] OR "screening score"[tiab] OR "decision rule"[tiab] OR "diagnostic rule"[tiab] OR "diagnostic score"[tiab] OR "predictive outcome"[tiab] OR "predictive rule"[tiab] OR "predictive score"[tiab] OR "predictive value"[tiab] OR "predictive risk"[tiab] OR "prediction outcome"[tiab] OR "prediction rule"[tiab] OR "prediction score"[tiab] OR "scoring"[tiab] OR "prediction value"[tiab] OR "prediction risk"[tiab] OR "risk assessment"[tiab] OR "risk score"[tiab] OR "risk scoring"[tiab] OR "prognostic score"[tiab] OR "prognostic index"[tiab] OR "prognostic rule"[tiab] OR "prospective validation"[tiab] OR ("risk"[tiab] AND "tool"[tiab]) OR (("validate"[tiab] OR "validation"[tiab] OR "validating"[tiab] OR "develop"[tiab] OR "development"[tiab] OR "derivation"[tiab] OR "derive"[tiab] OR "deriving"[tiab] OR "performance"[tiab]) AND ("decision"[tiab] OR "predictive"[tiab] OR "prediction"[tiab] OR "rule"[tiab] OR "score"[tiab] OR "scoring"[tiab] OR "index"[tiab] OR "model"[tiab] OR "scale"[tiab] OR "tool"[tiab] OR "algorithm"[tiab])) OR ("development"[tiab] AND "validation"[tiab]) OR ("derivation"[tiab] AND "validation"[tiab]) OR "signs and symptoms"[tiab]))

NOT "Vaccine"
 This search identified 4 potential studies not identified elsewhere for full text review.
 A bridge search in June, 2021 identified 37 additional abstracts but none met our inclusion criteria.

Appendix 3. Definitions used for QUADAS-2 framework for quality assessment

Patient Selection

1. Was a consecutive or random sample of patients enrolled?
Y: Study enrolled consecutive patients or a random sample of consecutive patients with influenza symptoms from a primary care, urgent care, hospital, or ED setting.
N: A convenience sample or other nonconsecutive or nonrandom sample was used, or it only included patients referred for diagnostic imaging or to an ENT clinic (this does not address exclusion criteria, handled in #3 below).
U: Uncertain
2. Was the study designed to avoid a case-control design (Y/N/U)?
Y: The study population was drawn from a cohort that included patients with a spectrum of disease.
N: The study population consisted of patients with known disease and healthy controls
U: Uncertain
3. Did the study design avoid inappropriate exclusion criteria (Y/N/U)?
Y: There were no inappropriate exclusion criteria, such as excluding those with uncertain findings.
N: The study used inappropriate exclusion criteria.
U: Uncertain
4. **Patient Selection Risk of Bias**: What is likelihood that patient selection could have introduced bias (L/H/U)?
L: Low likelihood of bias due to patient selection or enrollment ("Yes" to 1, 2 and 3)
H: High likelihood of bias due to patient selection ("No" to 1, 2 or 3)
U: Unable to judge degree of bias.
5. **Concerns About Patient Selection Applicability**: Are there concerns that included patients and setting do not match the review question?
N: Patients are likely to be typical of patients with influenza symptoms.
Y: Patients are unlikely to represent typical patients with influenza symptoms, for example come from a specialized population, are immunosuppressed, etc.
U: Uncertain

Index Test: Influenza Symptoms

6. Were index test results interpreted without knowledge of reference standard?
Y: Yes (including when sign or symptom performed by clinician and reference standard performed using reference laboratory PCR that could not be known by clinician)
N: No
U: Uncertain
7. If a threshold was used for the index test, was it pre-specified?
Y: The threshold was prespecified or there was no threshold mentioned
N: The threshold was established post hoc
U: A threshold was used but it is not clear when it was specified
8. **Index Test Risk of Bias**: What is the likelihood that conduct of the index test could have introduced bias (L, H, U)?
L: Low likelihood of bias due to ("Yes" to 6 and "Yes" or "Uncertain" to 7)
H: High likelihood of bias due to failure to mask to reference standard ("No" to 6 or 7)
U: Uncertain
9. **Concerns Regarding Index Test Applicability**: Are there concerns that the index test differs from those specified in the review question?
L: Low likelihood (the index test is a commonly reported sign or symptom)
H: High likelihood (the index test is not a sign or symptom from history and physical examination, or comorbidity)
U: Uncertain
Reference Standard Test
10. Is the reference standard likely to correctly classify patients as having Influenza?
Y: Yes, used PCR.
N: No, used another reference standard
U: Uncertain
11. Was the reference standard interpreted without knowledge of the index test?
Y: Yes, reference standard interpretation masked to index test results
N: No, reference standard interpretation not masked to index test results
U: Uncertain

12. **Reference Standard Risk of Bias**: Could conduct or interpretation of the reference standard could have introduced bias?

L: Low likelihood of bias due to the reference standard ("Yes" to 9, "Yes" or "Uncertain" to 10)

H: High likelihood of bias due to inadequate reference standard ("No" to 9 or 10)

U: Uncertain

13. **Concerns Regarding Applicability of the Reference Standard**: are there concerns that the target conditions defined by the reference standard do not match the review question?

L: Low likelihood of bias, that is, the reference standard was intended to detect Influenza

H: High likelihood of bias, that is, the reference standard was not intended to detect Influenza

U: Uncertain

Patient Flow and Timing

14. Did all patients receive a reference standard?

Y: Yes, all patients received some sort of reference standard (no partial verification bias)

N: No, some patients did not receive any reference standard (partial verification bias)

U: Uncertain

15. Did all patients receive the same reference standard?

Y: Yes, all used the same reference standard (no differential verification bias)

N: No, the reference standard varied depending on the results of the index test (differential verification bias)

U: Uncertain

16. Were all patients included in the analysis?

Y: All patients were properly accounted for in the analysis

N: Some patients were not accounted for or dropped for unclear reasons

U: Uncertain

17. **Patient Flow Risk of Bias**: Could patient flow have introduced bias?

L: Low likelihood of bias based on absence of partial verification bias and good follow-up ("Y" on 14 and15, "Yes" or "Uncertain" for 16)

H: High likelihood of bias based on partial verification bias or poor follow-up ("No" to 14 or 15, or significant number of patients lost to follow-up in 16)

U: Uncertain

Appendix 4. Clinical Prediction Rules Included in the Systematic Review

Study	Risk Score or CART model
Afonso, 2012	**CART model 1** Fever, chills or sweats, duration < 2 days, cough
Afonso, 2012	**CART model 2** Fever, chills, sweating
Afonso, 2012	**CART model 3** Fever, myalgia
Anderson, 2018	**CART model** Age ≥ 5 years, cough, coryza, chills, myalgia
Dugas, 2019	**Risk score** 2 points for new or increased cough, 1 point for headache, 1 point for subjective fever, and 1 point for measured temperature > 100.4 F. Low-risk group: 0 to 2 points High-risk group: ≥ 3 points
Ebell, 2012	**Risk score ("FluScore")** 2 points for fever and cough, 2 points for myalgias, 1 point for chills or sweats, and 1 point for duration < 2 days Low-risk group: 0 to 2 points Moderate risk group: 3 points High-risk group: 4 to 6 points
Govaert, 1998	**Risk score (original)** 1 point for each of the following 8 symptoms: cough, fever, acute onset, malaise, rigor or chills, myalgia, headache, sore throat. Probability of influenza is then reported for each point score.
Govaert, 1998	**Risk score (post hoc)** 1 point for each of the following 8 symptoms: cough, fever, acute onset, malaise, rigor or chills, myalgia, headache, sore throat Low-risk group: 0 points Moderate risk group: 1 to 2 points High-risk group: 3 + points
Ranjan, 2012	**Risk score** 1 to 3 points for different levels of fever, 1 point each for sneezing, coryza, sore throat, cough or wheeze, 1 point for flu circulating in community, and 2 points for close contact with confirmed flu. Low-risk group: 0 to 6 points High-risk group: ≥ 7 points
Vuichard-Gysin, 2019 (children)	**CART model** Fever, chills, cough, coryza, male sex
Vuichard-Gysin, 2019 (children)	**Risk score (children)** 1 point for age 6 to 17 years, 2 points for chills, 2 points for cough, and 3 points for fever. Low-risk group: 0 to 4 points High-risk group: ≥ 5 points
Vuichard-Gysin, 2019 (adults)	**CART model** Chills, cough, myalgia, sinus problem, sore throat
Vuichard-Gysin, 2019 (adults)	**Risk score (adults)** 2 points for chills, 2 points for cough, and 1 point for myalgias Low-risk group: 0 to 3 points High-risk group: ≥ 4 points
Woolpert, 2012	**Risk score (original)** 1 point for each of the following 4 symptoms: acute onset, fever, cough, myalgia. Probability of influenza is then reported for each point score.
Woolpert, 2012	**Risk score (post hoc)** 1 point for each of the following 4 symptoms: acute onset, fever, cough, myalgia Low-risk group: 0 to 1 points Moderate risk group: 2 points High-risk group: 3 + points
Zimmerman, 2016	**CART model** Fever, cough, fatigue

CART, Classification and Regression Tree.