

## ORIGINAL ARTICLES

# A Systematic Review of the History and Physical Examination to Diagnose Influenza

Mark H. Ebell, MD, MS, Linda L. White, RN, MPH, and Tracy Casault, BS

**Background:** Although influenza is a commonly encountered condition in primary care, and diagnosis is increasingly important given the availability of new treatments, there has been no systematic review of the evidence on clinical diagnosis.

**Methods:** This was a systematic review of the literature with meta-analysis where appropriate. We included cohort studies and randomized trials that compared the history and physical examination with a reference laboratory test for the diagnosis of influenza A and/or B. The primary outcomes were the sensitivity, specificity, likelihood ratios, and area under the receiver-operating characteristic (ROC) curve.

**Results:** Seven studies reported the sensitivity and specificity for a total of 59 variables. We combined studies of influenza A or B alone with those of influenza A and B. Rigors [likelihood ratio (LR) +7.2], the combination of fever and presenting within 3 days of the onset of illness (LR +4.0), and sweating (LR +3.0) were best at ruling-in influenza when present. When absent, the following decreased the likelihood of influenza: any systemic symptoms (LR -0.36), coughing (LR -0.38), not being able to cope with daily activities (LR -0.39), and being confined to bed (LR -0.50). Cough, nasal congestion, and fever (subjective or objective) had the highest calculable areas under the ROC curve.

**Conclusions:** Individual signs and symptoms are of limited value for the diagnosis of influenza. Development of clinical decision rules that systematically combine symptoms may be a more useful strategy. (J Am Board Fam Pract 2004;17:1-5.)

The increasing availability of office and reference laboratory tests to diagnose influenza<sup>1-4</sup> and the development of viable treatments for the disease<sup>5,6</sup> make it more important than ever to make the best possible use of the history and physical examination (HPE) to accurately establish the pretest probability. Patients with a low likelihood of influenza based on the HPE and a negative in-office test have a very low likelihood of influenza, whereas those with a high pretest probability and a negative test may still have a clinically important likelihood of the disease. We have therefore systematically reviewed the literature on diagnosis of influenza using the HPE.

## Methods

### Search Strategy

We searched the Medline database in mid-2000 using the following strategy: "influenza/diagnosis" [MeSH Terms] AND ("sensitivity and specificity" [MeSH Terms] OR "predictive value of tests" [MeSH Terms] OR "medical history taking" [MeSH Terms] OR "physical examination" [MeSH Terms]). We also reviewed the bibliography of every identified study, contacted domain experts, and reviewed the Database of Abstracts of Reviews of Effectiveness. The Medline search was repeated in November 2001 to ensure that there had been no intervening publications.

### Inclusion Criteria

We included articles that reported information about the accuracy of the HPE for the diagnosis of influenza A, B, or both sufficient to calculate both the sensitivity and specificity. We included only independent cohort studies and data from randomized trials that were the functional equivalent of independent cohort studies for the purposes of studying a diagnostic test and that used a reference

Submitted, revised 17 April 2003.

From the Department of Family Practice (ME, LW), Michigan State University (TC), East Lansing. Address correspondence to Mark H. Ebell, MD, MS, 330 Snapfinger Drive, Athens GA 30605 (e-mail: ebelle@msu.edu).

A table containing the complete data used for the meta-analysis is available at <http://www.jabfm.org/cgi/content/full/17/1/1/DC1>.

laboratory test as the reference standard for diagnosis of influenza.

### Study Protocol

Two investigators reviewed all the abstracts of identified studies and by a consensus approach decided on the articles to review in full. Two investigators then read each article, first deciding whether the article met inclusion criteria and then abstracting relevant data to a standard data collection form. The decisions about inclusion and data abstractions were compared, and conflicts were resolved by consensus discussion with a third investigator. We felt that certain variables were similar enough that they could be combined: body aches were included under "myalgias"; feverishness under "fever (subjective)"; pharyngitis under "sore throat"; expectoration of sputum under "sputum"; and purulent nasal discharge with "nasal secretion (purulent)."

### Data Analysis

Where data for a variable came from a single study, we calculated the sensitivity, specificity, and likelihood ratios using standard formulas. The positive likelihood ratio corresponds to how well a positive test includes the diagnosis and a negative likelihood ratio to how well a negative test excludes it. Test results associated with a likelihood ratio between 0.5 and 2.0 have little impact on the likelihood of disease. If more than one study reported data for a variable, we calculated summary estimates of the sensitivity and specificity using a random effects model (MetaTest 0.6; used by permission from Joseph Lau, MD) and we also reported the range. The likelihood ratio was then calculated from the summary estimates. The area under the receiver operating characteristic (ROC) curve<sup>7</sup> [a measure of how well a test discriminates patients with disease from those without disease; scores range from 0.5 (worthless test) to 1.0 (perfect test)] was calculated by the MetaTest software.

### Results

We identified 93 studies in our initial survey Medline search, and 4 additional studies from the bibliographies of these studies. Of this group of 97 studies, 7 met our inclusion criteria.<sup>8-14</sup> Study characteristics are shown in Table 1. All were independent cohort studies in the community or outpatient setting, largely primary care and during times of an influenza epidemic. None was explicit

about blinding, but because of the usual timing of the HPE and reference laboratory tests, we assume that physicians performing the HPE were not aware of the results of the reference standard test.

The 7 studies reported the sensitivity and specificity for a total of 59 variables. Because there was no clear pattern regarding the accuracy of variables for influenza A versus influenza B, because it would reduce statistical power, and because this would introduce another layer of complexity into the analysis and presentation of data, we chose to combine studies of influenza A or B alone with those of influenza A and B.

Signs and symptoms with a positive likelihood ratio (LR+) greater than 2.0 or a negative likelihood ratio (LR-) less than 0.5 are shown in Table 2. Table 2 also summarizes the findings for all variables reported by more than 1 study, including the area under the ROC curve, when it could be calculated. Signs and symptoms with an LR- greater than 0.5 and an LR+ less than 2.0 are generally not useful clinically; those with likelihood ratios falling in this range that were measured only by a single study are not reported and include abdominal pain, antipyretics before consultation, any lower respiratory symptom, any other symptom, asthenia, bronchiolitis, conjunctival injection, dry cough, earache, emesis, face ache, general practitioner consultation, gritty eyes, high risk condition, hoarseness, home visit by physician, lacrimation or conjunctival injection, loss of appetite, lower respiratory tract illness (age >65 years), male gender, moderate or severe fatigue, rhinorrhea, otitis media, pain on respiration, painful cervical adenopathy, received antibiotics, weakness, and wheeze. Monto<sup>8</sup> reported on combinations of variables; only 2 combinations (fever, cough, and nasal congestion; fever, cough, and weakness) had a LR+ greater than 2.0.

The elements of the HPE that best ruled in influenza when *present* were rigors (LR +7.2), the combination of fever and presenting within 3 days of the onset of illness (LR +4.0), and sweating (LR +3.0). The signs and symptoms best able to rule out influenza when *absent* were having any systemic symptoms (LR -0.36), coughing (LR -0.38), not being able to cope with daily activities (LR -0.39), and being confined to bed (LR -0.50).

The area under the ROC curve could not be calculated for signs and symptoms or clusters of signs and symptoms reported by only a single study. The highest calculable areas under the ROC

**Table 1. Design Characteristics of Included Studies (All Were Cohort and Independent, with Blinding Not Explicitly Stated)**

Study	Setting	Influenza	n	Population	Reference Standard	Inclusion Criteria
Nicholson et al, 1997 <sup>10</sup>	Community	A or B	291	Mean age, 73 years; 51.8% female	4× rise in titer for Influenza A	Respiratory infection (over age 60)
Carrat et al, 1997 <sup>12</sup>	Primary care office and home visit	A or B	130	Mean age, 33.2 years; 59% female	Fluorescent monoclonal antibodies	Suspected influenza (influenza-like illness, upper or lower respiratory tract infection, and/or fever >38 without other evidence of infection within 36 hours of onset)
Carrat et al, 1999 <sup>9</sup>	Primary care office and home visits	A	600	Mean age, 38 years; 58% female	DIF and ELISA; 25% culture + PCR	Suspected influenza (all patients over age 1 with ≥1 of the following: influenza-like illness, upper or lower respiratory tract infectious syndrome, and/or temp >38°C without any infectious signs or symptoms.)
Monto et al, 1996 <sup>11</sup>	Primary care office	A or B	897	NA	Viral culture or positive serology	Suspected influenza (influenza-like illness with fever and cough or sore throat presenting between November and April)
Lina et al, 1996 <sup>13</sup>	Primary care office	A or B	340	NA (broad age range)	Culture or ELISA	Suspected influenza; data not available for patients without positive viral nasal swab.
Long et al, 1997 <sup>14</sup>	Primary care office	A or B	788	Mean age, 76 in adults, 5 in children	Tissue culture, with ELISA to type the virus	Adults over age 65 and children with respiratory or febrile illness in community medical and pediatric practices.
Monto et al, 2000 <sup>8</sup>	Outpatient	A or B	3744	Mean age, 34.7 years; 52% female.	Positive culture or ≥4× increase in influenza antibody titer in convalescent vs acute samples	Suspected influenza [fever (≥37.8°C; ≥37.2°C for patients ≥65 years in one study) or feverishness (subjective fever or chills) and 2 of the following: headache, myalgia, cough, or sore throat].

DIF, direct immunofluorescence; ELISA, enzyme-linked immunosorbent assay; NA, not available

**Table 2. Test Characteristics for Selected Signs and Symptoms**

Variable	n	Sens	Spec	LR+	LR–	Weighted AUROCC
Abdominal pain		0.17	0.88	1.46	0.94	
Variables with LR+ >2.0 and/or LR– < 0.5						
Rigors (10)	291	0.16	0.98	7.16	0.86	NA
Fever and ≥3 days ill (age >65) (8)	677	0.40	0.90	4.03	0.67	NA
Sweating (10)	291	0.47	0.83	2.86	0.63	NA
Fever and cough >36 hours (8)	3744	0.50	0.81	2.64	0.61	NA
Confined to bed (10)	291	0.63	0.74	2.45	0.50	NA
Smoker (10)	291	0.32	0.87	2.39	0.79	NA
Unable to cope with daily activities (10)	291	0.74	0.68	2.30	0.39	NA
Fever + cough + nasal congestion (8)	3744	0.59	0.74	2.27	0.55	NA
Fever + cough + weakness (8)	3744	0.60	0.72	2.10	0.56	NA
Any systemic symptom (10)	291	0.84	0.43	1.49	0.36	NA
Cough (8, 9, 11, 13)		0.88 (0.72 to 0.96)	0.32 (0.17 to 0.52)	1.29	0.38	0.679
Commonly measured variables with LR+ <2.0 and LR– >0.5						
Chills (9, 11)	1497	0.83 (0.79 to 0.86)	0.25 (0.22 to 0.27)	1.11	0.68	NA
Subjective temp (8, 10, 11, 14)	5720	0.68 (0.62 to 0.72)	0.60 (0.43 to 0.75)	1.7	0.53	0.672
Objective temp (8, 9, 13)	4684	0.70 (0.46 to 0.86)	0.49 (0.29 to 0.69)	1.37	0.61	0.653
Headache (8, 9, 10, 11, 13)	5872	0.77 (0.65 to 0.86)	0.39 (0.23 to 0.58)	1.26	0.59	0.606
Myalgia (8, 10, 11, 13)	5272	0.68 (0.57 to 0.77)	0.47 (0.18 to 0.78)	1.26	0.60	0.620
Nasal congestion (8, 10, 11)	4932	0.70 (0.56 to 0.81)	0.42 (0.23 to 0.63)	1.21	0.71	0.654
Nasal secretions (purulent) (10, 13)	631	0.17 (0.04 to 0.48)	0.79 (0.63 to 0.89)	0.81	1.05	0.613
No sneezing (9, 10)	891	0.56 (0.39 to 0.72)	0.53 (0.31 to 0.74)	1.19	0.83	0.555
Not vaccinated (12, 14)	912	0.69 (0.19 to 0.96)	0.30 (0.02 to 0.89)	0.98	1.03	0.534
Sore throat (8, 10, 11, 13)	5272	0.68 (0.61 to 0.74)	0.36 (0.26 to 0.47)	1.06	0.89	0.558
Sputum (9, 11)	1497	0.36 (0.25 to 0.48)	0.68 (0.43 to 0.86)	1.13	0.94	0.450

NA, not applicable because there was only one study; n, number of patients; AUROCC, area under the receiver operating characteristic curve.

curve were 0.679 for cough, 0.672 for subjective temperature, 0.654 for nasal congestion, and 0.653 for objectively measured temperature.

## Discussion

Physicians have traditionally used information about the presence or absence of an influenza epidemic in the community and certain signs and symptoms such as sudden onset of fever, cough, myalgias, and chills to diagnose influenza. Our systematic review identifies 3 variables that, when present, help rule in influenza (rigors, sweating, and fever and onset of symptoms less than 3 days before) and 4 additional symptoms that rule it out (no systemic symptoms, not coughing, being able to cope with daily activities, and not being confined to bed). Other commonly used symptoms such as sore throat, chills, and myalgias) were less useful.

A bias toward lower estimates of sensitivity and specificity may have been introduced by the fact that most studies only included patients with suspected influenza.<sup>8,9,11–13</sup> For example, if fever was part of the inclusion criteria for a study, it will make it impossible for this variable to contribute to discriminating between patients with and without influenza. This bias particularly affects the estimates

for fever, headache, myalgias, cough, and sore throat that were part of the inclusion criteria for the large Monto study.<sup>8</sup>

Our study had several limitations. The size of one study,<sup>8</sup> a pooled analysis of the results of several randomized trials, meant that it often dominated the analysis. Any flaws in this study (eg, lack of blinding, an imprecise reference standard, selection bias) would therefore also dominate our analysis. There was considerable heterogeneity between studies, which is why we report the range as well as a summary measure of effect for sensitivity and specificity estimates based on data from more than one study. Finally, several of the variables that had the highest LR+ or lowest LR– came from a single study; again, any flaws in that study's design would have an important impact on our findings.

The literature review was repeated just before publication (February 2004) and identified only one additional article. This article studied patients over age 65 or with underlying cardiopulmonary disease who were admitted to the hospital with a respiratory diagnosis; approximately 20% had influenza.<sup>15</sup> Despite the highly selected nature of the group, these findings were similar to ours. The best predictor of influenza A was the combination of cough,

temperature of 38°C or higher, and illness duration of 7 days or less (LR+ 2.9, LR– 0.3).

Previous studies have shown that individual signs and symptoms rarely include or exclude a disease.<sup>16,17</sup> A more successful strategy is the use of several key symptoms in a clinical decision rule that stratifies patients into low-, moderate-, and high-risk groups. This information can be used in conjunction with the results of office laboratory tests and perhaps imaging studies to make a more accurate diagnosis while also limiting unnecessary testing and overtreatment. This strategy has been successfully implemented for sore throat, deep vein thrombosis, and pulmonary embolism.<sup>16–18</sup> More than anything, this systematic review points out the need for well-designed studies in the primary care setting to develop and validate such a rule.

## References

- Sharma V, Dowd MD, Slaughter AJ, Simon SD. Effect of rapid diagnosis of influenza virus type A on the emergency department management of febrile infants and toddlers. *Arch Pediatr Adolesc Med* 2002;156:41–3.
- Hindiyeh M, Goulding C, Morgan H, et al. Evaluation of BioStar FLU OIA assay for rapid detection of influenza A and B viruses in respiratory specimens. *J Clin Virol* 2000;17:119–26.
- Vabret A, Sapin G, Lezin B, et al. Comparison of three non-nested RT-PCR for the detection of influenza A viruses. *J Clin Virol* 2000;17:167–75.
- Schultze D, Thomas Y, Wunderli W. Evaluation of an optical immunoassay for the rapid detection of influenza A and B viral antigens. *Eur J Clin Microbiol Infect Dis* 2001;20:280–3.
- Jefferson TO, Demicheli V, Deeks JJ, Rivetti D. Amantadine and rimantadine for preventing and treating influenza A in adults. *Cochrane Database Syst Rev* 2002;(3):CD001169.
- Jefferson T, Demicheli V, Deeks J, Rivetti D. Neuraminidase inhibitors for preventing and treating influenza in healthy adults. *Cochrane Database Syst Rev* 2000;(2):CD001265.
- Hanley JA. Alternative approaches to receiver operating characteristic analyses. *Radiology* 1988;168:568–70.
- Monto AS, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. *Arch Intern Med* 2000;160:3243–7.
- Carrat F, Tachet A, Rouzioux, Housset B, Valleron A-J. Evaluation of clinical case definitions of influenza: detailed investigation of patients during 1995–1996 epidemic in France. *Clin Infect Dis* 1999;28:283–90.
- Nicholson KG, Kent J, Hammersley V, Cancio E. Acute viral infections of upper respiratory tract in elderly people living in the community: comparative, prospective, population based study of disease burden. *BMJ* 1997;315:1060–4.
- Monto AS and Ohmit SE. The evolving epidemiology of influenza infection and disease, in Brown LE, Hampson AW, Webster RG, editors. *Options for the control of influenza III: Proceedings of the Third International Conference on Options for the Control of Influenza*; Cairns, Australia; 4–9 May 1996. Amsterdam: Elsevier Science BV; 1996.
- Carrat F, Tachet A, Housset B, Valleron A-J, Rouzioux C. Influenza and influenza-like illness in general practice: drawing lessons for surveillance from a pilot study in Paris, France. *Br J Gen Pract* 1997;47:217–20.
- Lina B, Valette M, Foray S, et al. Surveillance of community-acquired viral infections due to respiratory viruses in Rhone-Alpes (France) during Winter 1994 to 1995. *J Clin Microbiol* 1996;34:3007–11.
- Long CE, Hall CB, Cunningham CK, et al. Influenza surveillance in community-dwelling elderly compared with children. *Arch Fam Med* 1997;6:459–65.
- Walsh EE, Cox C, Falsey AR. Clinical features of influenza A virus infection in older hospitalized persons. *J Am Geriatr Soc* 2002;50:1498–503.
- Ebell MH, Smith MA, Barry HC, Ives K, Carey M. Does this patient have strep throat? *JAMA* 2000;284:2912–8.
- Wells PS, Hirsh J, Anderson DR, et al. Accuracy of clinical assessment of deep-vein thrombosis. *Lancet* 1995;345:1326–30.
- Wells PS, Ginsberg JS, Anderson DR, et al. Use of a clinical model for safe management of patients with suspected pulmonary embolism. *Ann Intern Med* 1998;129:997–1005.