

The use of classification and regression trees to predict the likelihood of seasonal influenza

Anna M Afonso^a, Mark H Ebell^{a,*}, Ralph Gonzales^b, John Stein^c,
Blaise Genton^d and Nicolas Senn^e

^aDepartment of Epidemiology and Biostatistics, University of Georgia, Athens, Georgia, USA, ^bDepartment of Internal Medicine and ^cDepartment of Emergency Medicine, University of California, San Francisco, CA, USA, ^dInfectious Disease Service, Department of Ambulatory Care and Community Medicine, University Hospital, Lausanne, Switzerland and ^eDepartment of Ambulatory Care and Community Medicine, University of Lausanne, Lausanne, Switzerland.

*Correspondence to Mark H Ebell, Department of Epidemiology and Biostatistics, University of Georgia, N129A Coverdell Building, Athens, GA 30602, USA; E-mail: ebell@uga.edu

Received 9 October 2011; Revised 5 February 2012; Accepted 6 February 2012.

Background. Individual signs and symptoms are of limited value for the diagnosis of influenza.

Objective. To develop a decision tree for the diagnosis of influenza based on a classification and regression tree (CART) analysis.

Methods. Data from two previous similar cohort studies were assembled into a single dataset. The data were randomly divided into a development set (70%) and a validation set (30%). We used CART analysis to develop three models that maximize the number of patients who do not require diagnostic testing prior to treatment decisions. The validation set was used to evaluate overfitting of the model to the training set.

Results. Model 1 has seven terminal nodes based on temperature, the onset of symptoms and the presence of chills, cough and myalgia. Model 2 was a simpler tree with only two splits based on temperature and the presence of chills. Model 3 was developed with temperature as a dichotomous variable ($\geq 38^{\circ}\text{C}$) and had only two splits based on the presence of fever and myalgia. The area under the receiver operating characteristic curves (AUROCC) for the development and validation sets, respectively, were 0.82 and 0.80 for Model 1, 0.75 and 0.76 for Model 2 and 0.76 and 0.77 for Model 3. Model 2 classified 67% of patients in the validation group into a high- or low-risk group compared with only 38% for Model 1 and 54% for Model 3.

Conclusions. A simple decision tree (Model 2) classified two-thirds of patients as low or high risk and had an AUROCC of 0.76. After further validation in an independent population, this CART model could support clinical decision making regarding influenza, with low-risk patients requiring no further evaluation for influenza and high-risk patients being candidates for empiric symptomatic or drug therapy.

Keywords. Clinical decision rules, common illnesses, infectious disease, influenza, primary care, public health, respiratory infections.

Introduction

In a typical year, ~5% to 20% of the population is affected by seasonal influenza.¹ The public health impact includes a loss in workforce productivity, ~200 000 hospitalizations each year and up to 49 000 flu-related deaths per year in the USA alone.¹ While vaccination is the cornerstone of influenza prevention, the vaccine has incomplete uptake and limited effectiveness.²

A timely diagnosis of influenza has the potential to improve the allocation of medical resources, prevent the inappropriate use of antibiotics and permit the

prompt initiation of antiviral therapy if patients are diagnosed within 36 hours of symptom onset.³ The latter has been demonstrated to reduce the duration of symptoms by ~24 hours.²

Patients with seasonal influenza typically present with fever, chills, cough, myalgias, sore throat, headache and fatigue.⁴ Because other respiratory infections may have a similar clinical presentation,⁵ information from individual signs and symptoms is of limited value for diagnosing influenza. Clinical decision rule that integrates several signs and symptoms could be used in conjunction with information about the baseline prevalence of influenza in the community to classify

patients as low risk (below the test threshold, requiring no further evaluation), above the treatment threshold (candidates for empiric therapy) or between the test and treatment threshold (candidates for further testing).⁶ While previous studies^{7–13} have identified simple heuristics to identify patients with influenza (i.e. a ‘fever and cough’ rule), their accuracy is variable and limited, with a sensitivity and specificity ranging from 30% to 80% and 55% to 95%¹⁴, respectively. Previous attempts to create multivariate models have been limited by poor reporting or failure to validate the resulting models using either a ‘split sample’ approach or a new population.^{10,15} In a recent study, we developed and validated a clinical decision rule based on a logistic regression model that was able to classify ~50% of subjects as being high risk or low risk.¹⁶

Classification and regression trees (CARTs)¹⁷ are an alternative to logistic regression for the creation of clinical decision rules. This approach has the advantage of not making any assumptions about the underlying statistical model (i.e. it is a model-free estimator) and the resulting decision trees have good face validity and are easily applied at the bedside. The aim of this investigation is to use CART analysis to develop and internally validate a clinical decision rule to stratify the risk of influenza using the history and physical examination.

Methods

Dataset

We identified two studies, one from California and one from Switzerland, that evaluated the accuracy of the history and physical examination in consecutive adults with suspected influenza or acute respiratory tract infection in the outpatient setting during flu

season. Characteristics of the combined dataset are reported in Table 1.^{7,18} Because the study populations were so similar, we chose to combine the datasets in order to be able to create a development set with 70% of the data ($n = 322$) and a separate validation set with 30% of the data ($n = 137$). This assured an adequate number of patients for model development and increased generalizability. Simply using one community’s data for the development set would not have resulted in enough cases in that group to reliably create a CART model. The final prevalence of influenza is typical of that during peak flu season in US surveillance studies. Thus, rather than using two independent populations for the development and validation sets, each set was randomly selected from the combined dataset. The validation set was not used during model development; it was reserved to perform the final validation of the candidate models.

Variables that were considered to have nearly the same clinical meaning were combined under a single variable (i.e. ‘rhinitis’ and ‘runny nose’, ‘myalgia’ and ‘muscle pain’ and ‘chills and sweating’ and ‘sweats’). Acute onset was defined as the presentation of symptoms to a physician within 48 hours of symptom onset. Variables reported in one study but not another were eliminated from the combined dataset.

The prevalence of influenza for the study conducted in a primary care clinic (53%) was higher than the prevalence of influenza in the study conducted in an urgent care setting (20%), yielding a prevalence of 33% in the final dataset. The latter pretest probability is typical of the peak of influenza season in the USA, which ranged from 25% to 40% in 2010–11, 25% to 45% in 2009–10 and 20% to 30% in 2008–09.¹⁹ These are rough generalizations based on the Centers for Disease Control and Prevention’s weekly surveillance data and do not represent the total prevalence of influenza A and B, but the proportion of influenza

TABLE 1 Characteristics of included studies

	Swiss population ¹⁸	US population
Number of patients	201	258
Setting	University primary care clinic that serves an urban population of 150 000 in Lausanne, Switzerland	Emergency Department or urgent care ambulatory patients in a large tertiary care University Hospital in San Francisco, CA
Date	December 1999 to February 2000	January 2002 to March 2002
Inclusion criteria	Adult outpatients with influenza-like illness as determined by the primary care physician	Consecutive adults with symptoms of an acute respiratory tract infection (cough, sinus pain, congestion/rhinorrhoea, sore throat or fever developing in past 3 weeks)
Mean age (range)	34.3 (17–86)	38.8 (18–90)
Prevalence of influenza (%)	104/201 (52.8)	53/258 (20.5)
Reference standard	Culture	Polymerase chain reaction
Independent predictors of influenza odds ratio (95% confidence interval)	Fever; 4.24 (2.33–7.71) Myalgia; 2.76 (1.01–7.49)	Myalgia; 4.22 (1.96–9.1) Fever; 3.84 (1.98–7.45) Chills; 3.37 (1.6–7.06) Rhinitis; 2.22 (1.02–4.82)

diagnoses out of the total outpatient visits for influenza-like illness (ILI). It is possible that the higher prevalence in the Swiss study could be due to differing selection criteria that identified patients more likely to have the flu. Culture was the reference standard test in the Swiss study, while polymerase chain reaction was the reference standard test in the American study.

Software

The univariate analysis was performed with Stata version 11.0 (College Station, TX). The creation of a CART was performed with JMP 8.0.2 (SAS Institute).

CART analysis

Unlike logistic regression, CART analysis does not require postulation of an underlying model.²⁰ Therefore, CART is able to discover complex interactions between variables, an advantage for data in which the relationships between predictors and outcome is unclear. It is a non-parametric technique, not requiring assumptions about the distribution of data.

In our analysis, the outcome variable was binary: identification of flu by the reference standard. All symptoms recorded in both studies were included in the determination of the model (excluding gender and age), regardless of whether they were significant predictors of flu (Table 3). Gender ($P = 0.60$) and age were omitted from the analysis because they did not show clinical or statistical significance. The only variables that were included in the model but were not significant predictors of flu were headache ($P = 0.05$) and sore throat ($P = 0.75$). Temperature was the only continuous predictor, while binary predictor variables included the presence of myalgia, cough, rhinitis, sore throat, headache, fatigue, chills/sweating, acute onset of symptoms and fever. In Models 1 and 2, temperature was a continuous variable. In Model 3, temperature was dichotomized at a temperature of 38°C to create a binary fever variable.

CART involves binary recursive partitioning, which splits a single parent node into two daughter nodes based on the predictor variable that best stratifies the population into groups with and without the outcome of interest. Each daughter node can be split further into two nodes and so on. The partition algorithm obtains all possible splits and calculates a likelihood ratio (LR) chi-square statistic for a test of independence for each possible split. The P value for the chi-square test represents the probability of getting a chi-square value greater than the one found by chance alone. The criterion used for selecting splits of the nodes was set to 'maximize significance'. This means that splits were chosen based on significance values for each split candidate rather than the raw test statistic. Each candidate variable is ranked by its logworth statistic to identify the optimal split for each node (the logworth statistic is

the negative log of adjusted P values for the chi-square statistic).²¹

For continuous variables, splits are constructed around a cutting value that maximizes the separation of the groups.²² The separation of groups is measured by the sum of squares for the differences between the means of the two groups.²⁰ Missing values were assigned to 'closes which assigns the missing value based on patterns identified from the non-missing data.'²¹ There were only 11 cases in the derivation set that were missing a value for temperature and 3 in the validation set missing temperature.

In this analysis, the minimum split size was set to JMP's default setting: five patients in a single group (1% of the total sample). We set a fairly low minimum split size in order to be able to grow out and then manually prune the tree back. We 'manually' pruned the tree by examining how many additional patients could be classified as high or low risk, which we considered the most desirable groups for classification since they would not require further diagnostic testing according to the test and treatment thresholds defined below. Other goals for our CART models were that they maintain good face validity for patients and clinicians and that they are simple to use.

Models 1 and 3 were built by first growing out the tree by manually splitting according to the optimal logworth statistic. We then manually pruned back the tree, omitting variables that did not further classify a substantial percentage of patients into a high- or low-risk group. Model 2 is a reduced version of Model 1. Another method used to develop CART models involved automatic repeated splitting according to the logworth statistic until the R -square was better than what the next 10 splits would obtain.²¹ Automatic splitting obtained the same result as Model 1 except that myalgia and temperature $>37.85^{\circ}\text{C}$ are omitted from the model. Automatic splitting obtained the same result for Model 3. Results for the manually modified trees are shown in Tables 3 and 4.

Test and treatment thresholds

In a previous study,¹⁶ we surveyed a group of 20 generalist physicians about their testing and treatment preferences regarding influenza. Based on these data, we established a test threshold of 10%, below which neither testing nor treatment was indicated, and a treatment threshold of 50%, above which empiric antiviral therapy should be considered. These thresholds are also consistent with those reported in a previous decision threshold analysis.²³

Analysis

As noted earlier, 30% of the dataset was not included in this analysis as it was reserved for validation. These data were not used during model development. We calculated receiver operating characteristic curves to

assess the overall diagnostic accuracy of the CART models.²⁴ CARTs were used to stratify patients into low-, moderate- and high-risk groups. The post-test probability of influenza and LR was calculated for each of the terminal nodes in the CART tree. We used this approach (rather than a bootstrap evaluation) because it allowed us to specify cut-offs for classification into low-, moderate- and high-risk groups prior to evaluation in the validation set.

Results

An analysis of patient symptoms according to influenza status for the combined dataset (Table 2) shows that patients with influenza had a much higher percentage of fever (63% versus 24%) and myalgia (90% versus 63%) as well as a moderately higher percentage of chills/sweating (85% versus 63%). However, there is significant similarity in the prevalence of other symptoms between cases and non-cases.

We developed three candidate CART models, shown in Figures 1–3 (the full models are shown in Fig. A1–A3). Model 1 (Fig. A1) initially had six splits and seven terminal nodes but was simplified (Fig. 1) by combining terminal nodes into a single ‘moderate-risk’ group and by omitting ‘temperature >37.85°C’ and ‘myalgia’ splits that added only 16 or 7 patients, respectively, into high- or low-risk categories. Models 2 and 3 had only two splits and three terminal nodes. In Models 1 and 2, temperature $\geq 37.4^\circ\text{C}$ (99.3 F) was the most important predictor variable. This cut-off for

temperature was chosen by the CART algorithm as the optimal split but may be problematic since it is at the upper bound of the normal range. In Model 3, we included temperature as a prespecified dichotomous variable with a cut-off $\geq 38^\circ\text{C}$ (100.4 F), which is more typical of an abnormal temperature in usual clinical practice. Although gender and age were considered optimal candidates for splits by the algorithm, we did not include these splits in the model as previous multivariate and univariate analyses did not identify age or gender as important variables.

Because our modelling goals were to maximize identification of low- and high-risk patients who would not require further diagnostic testing and to favour simplicity, we combined the terminal nodes in Model 1 that had a probability of influenza between 10% and 50% into a single moderate probability node.

Model 2 only had three terminal nodes, which we designated as low, moderate and high risk. The post-test probability and LR of flu for each node and for the combined low-, moderate- and high-risk groups of each model are shown in Table 3.

The most clinically useful information gained by using the prediction model is the percentage of patients classified as low and high risk since that information has the potential to change clinical practice. We report separately the accuracy of Models 1, 2 and 3 for the development, validation and combined datasets. This percentage decreased less between derivation and validation groups for Model 2 (71%–67%) than it did for Model 1 (41%–32%) or Model 3 (57%–48%). The key design and performance characteristics of the

TABLE 2 Symptom prevalence by influenza status

Symptom/characteristic	% With symptom	Prevalence of symptom in influenza positive (N = 157)	Prevalence of symptom in influenza negative (N = 302)	P value
Mean age (SD)	37 (14)	36 (13)	37 (15)	
Temperature ($^\circ\text{C}$) ^a (SD)	37.51 (0.88)	38.03 (0.80)	37.23 (0.79)	
Male ^b	45	67/156 (43%)	137/301 (46%)	0.60
Fever (>37.8°C) ^a	37	99/157 (63%)	71/302 (24%)	<0.0001 ^a
Chills	71	134/157 (85%)	190/302 (63%)	<0.0001 ^a
Myalgia	72	142/157 (90%)	193/302 (64%)	<0.0001 ^a
Duration of symptoms <2 days before presentation	33	82/157 (52%)	69/302 (23%)	<0.0001 ^a
Fatigue	83	141/157 (90%)	240/302 (79%)	0.005 ^a
Cough	92	150/157 (96%)	271/302 (90%)	0.03 ^a
Rhinitis	76	128/157 (82%)	220/302 (73%)	0.04 ^a
Headache	78	131/157 (83%)	228/302 (76%)	0.05
Sore throat	72	115/157 (73%)	217/302 (72%)	0.75

Note: The following variables that were reported in one study but not another were eliminated from the combined dataset: race, whether the visit was the first for the given illness, days missed from work, duration of illness in days, presence of a co-morbidity, sinus pain, whether discharge was present, colour of discharge, throat swelling, difficult swallowing, whether phlegm was dry or scant, whether blood was present in phlegm, wheezing, shortness of breath, painful breathing, chest pain, abdominal pain, vomiting, diarrhoea, pulse, blood pressure, respiratory rate, oxygen saturation, toxic appearance, TM abnormality, purulent sinus drainage, sinus tenderness, tonsillar swelling, tonsillar exudates, cervical lymphadenopathy, prolonged expiration, decreased breath sounds, rales, rinchi, wheezes, abdominal tenderness, a clinician’s diagnosis of influenza, the results of a rapid flu test and the week that the patient presented with symptoms.

^aTemperature missing for 14 patients.

^bGender missing for two patients.

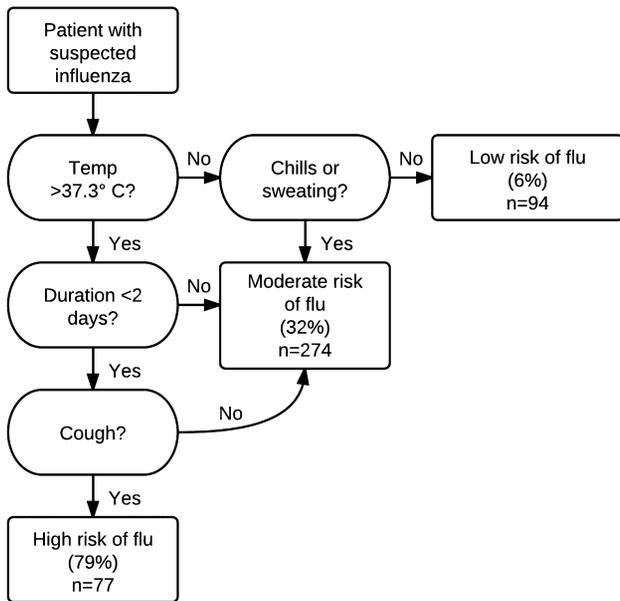


FIGURE 1 Simplified CART for Model 1 (see Fig. A1 for full model)

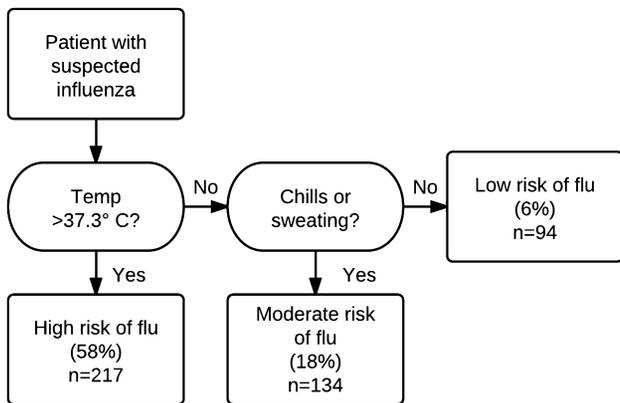


FIGURE 2 CART for Model 2 (see Fig. A2 for full model)

three candidate CART models are summarized in Table 4, including the area under the receiver operating characteristic curve (AUROCC), LR's and percentage of patients classified into useful risk categories for the development and validation set for each model.

Discussion

Model 1 has a higher AUROCC than Model 2, may be more acceptable or believable because it uses more clinical variables and has a higher LR for the high-risk category. Model 2, on the other hand, is simpler and easier to use and classifies more patients below the test threshold or above the treatment threshold than Model 1 (67% versus 32% for the validation set).

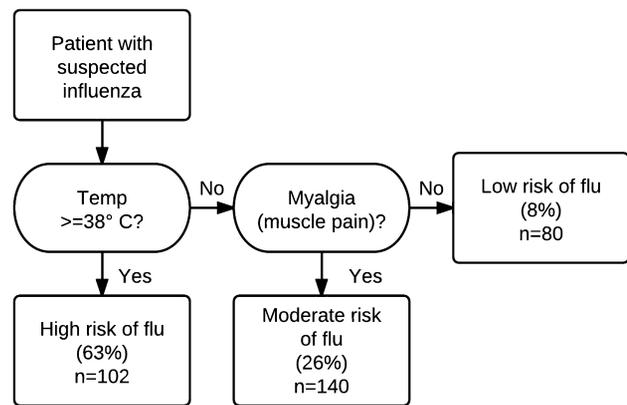


FIGURE 3 CART for Model 3 (see Fig. A3 for full model)

These are the most clinically useful classifications because they help a physician rule out or rule in influenza. Model 3 has a slightly higher AUROCC than Model 2 but classifies only 49% of validation set patients into high- or low-risk categories. An advantage of Model 3, however, is that it utilizes a temperature cut-off for fever that may be more acceptable to physicians.

While on balance we feel that Model 2 provides the most clinically useful model, all the models are reasonable options depending on patient and physician preferences. Models 2 and 3 in particular could be easily memorized. For example, during flu season, patients with fever and an onset of symptoms <48 hours should be empirically treated (unless there is some other obvious aetiology for their fever such as exudative pharyngitis, otitis media or sinus pain and tenderness). Those without fever but with chills or sweats should be tested for influenza, while those without either symptom are unlikely to have influenza. Note that this simplified model is only an aid to clinician judgement; it is important to remain vigilant for pneumonia and other uncommon but serious causes of similar symptoms. While these models were accurate in our population, they should be further validated in an independent population that also assesses the impact on cost, test ordering, prescribing and clinical outcomes.

CART has several advantages as a tool for developing clinical decision rules. While traditional statistical techniques require the postulation of a model, CART does not. When complex interactions and patterns exist in data, they can be difficult or virtually impossible to model.²² Furthermore, multivariate models are complex, and even point scores may generate non-intuitive results (for example when different combinations of two or three variables give different results). A decision tree, on the other hand, is easily understood by both patients and physicians.

TABLE 3 Classification accuracy of the CART risk models in the derivation, validation and full datasets

	Derivation group			Validation group			All patients		
	Flu/total (%)	LR	% In group	Flu/total (%)	LR	% In group	Flu/total (%)	LR	% In group
Model 1									
High risk	47/60 (78)	7.2	19	14/17 (82)	7.8	13	61/77 (79)	7.1	17
Moderate risk	56/191 (29)	0.83	59	34/92 (37)	0.98	68	89/274 (32)	0.89	62
Low risk	4/71 (5.6)	0.12	22	2/25 (8)	0.15	19	6/94 (6)	0.13	21
Model 2									
High risk	86/157 (55)	2.4	49	40/65 (62)	2.7	48	126/217 (58)	2.6	49
Moderate risk	17/94 (18)	0.44	29	8/44 (18)	0.37	33	24/134 (18)	0.40	30
Low risk	4/71 (6)	0.12	22	2/25 (8)	0.15	19	6/94 (6)	0.13	21
Model 3									
High risk	64/102 (63)	3.4	32	25/36 (70)	3.9	26	89/139 (64)	3.4	30
Moderate risk	37/140 (26)	0.72	43	24/71 (34)	0.89	52	61/211 (29)	0.78	46
Low risk	6/80 (7.5)	0.16	25	1/30 (3)	0.06	22	7/109 (6)	0.13	24

Note: 11 cases in the derivation set and 3 in the validation set had missing temperature or gender data. Random values were used for these variables in these cases during development and validation of the CART model.

TABLE 4 Characteristics of the three candidate CART models for predicting the likelihood of influenza

	Model 1	Model 2	Model 3
Number of nodes	8	6	6
Number of terminal nodes	3	3	3
AUROC (derivation/validation)	0.82/0.80	0.75/0.76	0.76/0.77
Percentage of patients classified as low or high risk (derivation/validation)	41%/32%	71%/67%	57%/48%
Likelihood ratio for influenza in high-risk group (derivation/validation)	7.2/7.8	2.4/2.7	3.4/3.9
Likelihood ratio for influenza in low-risk group (derivation/validation)	0.12/0.15	0.12/0.15	0.16/0.06

A strength of the current study is that it is one of the first to use separate derivation and validation subgroups to develop and evaluate a clinical decision tool for influenza.¹³ In addition, the models can be easily memorized for use at the point of care, even without a computer, and successfully classify a clinically meaningful percentage of patients into low- or high-risk groups that do not require further evaluation. A limitation of the current study is the use of the same population for derivation and validation, known as a 'split sample' approach. The next step in our programme of research is to prospectively validate these CART models in a completely different population. The impact of use of this and other clinical scores on the rates of ordering office-based diagnostic tests, on prescriptions for anti-influenza drugs, on cost and on clinical outcomes requires further study, as demonstration of accuracy alone is not sufficient.

Ultimately, a clinical rule or decision tree for influenza should be linked to prevalence data such as that reported regularly by the Center for Disease Control and Prevention²⁵ or that reported by Euroflu,²⁶ a World Health Organization surveillance network for European regions. A clinical rule should also be evaluated in populations with different prevalences of

disease (for example, outside of flu season or during the shoulder of flu season). The same clinical decision rule may give different recommendations depending on the underlying pretest probability. For example, during shoulder season before and after the peak of influenza, a high-risk patient may have a post-test probability <50% and require diagnostic testing to confirm the diagnosis. On the other hand, during peak influenza season, they would be candidates for empiric therapy. Thus, use of prevalence data integrated with a clinical rule will help a clinician make the best use of the history, physical examination and optionally diagnostic testing and will help them make the best decision about use of diagnostic tests and antiviral therapy.

A limitation of our study is the merging of two distinct datasets gathered from separate studies. Although we believe that this decision increased the generalizability of our model, differences in prevalence could reflect differences in selection criteria between the two studies. Overfitting the CART model to our dataset is another possible limitation, although pruning the tree reduces that possibility. Our criterion for assessing the benefit of an additional split was based on percentage of patients correctly classified. Idiosyncrasies in our own data may have been built into

the model and may not validate as well in an independent population. The test and treatment thresholds that we used were based on a small unrepresentative survey of primary care physicians: further work is needed to refine these estimates.

In conclusion, we consider this to be an internal validation study since the validation group was randomly drawn from the same population as the derivation group. Further validation in a completely independent population, in different flu seasons and at different times within the flu season when the prevalence is <34% is needed. We would also like to explore use of this model by patients, as part of a guide to self-care that is integrated with prevalence data and treatment recommendations.

Declaration

Funding: this study was unfunded.

Ethical approval: this study was approved by the Human Subjects Committee of the University of Georgia and classified as 'exempt' from full board review.

Conflict of interest: none.

References

- 1 Centers for Disease Control and Prevention. Seasonal Influenza (Flu): Questions & Answers. <http://www.cdc.gov/flu/about/qa/disease.htm> (accessed on 16 April 2011).
- 2 Jefferson T, Di Pietrantonj C, Rivetti A *et al*. Vaccines for preventing influenza in healthy adults. *Cochrane Database of Systematic Reviews* 2010. issue 7. Art No.: CD001269.
- 3 Treanor J, Hayden F, Vrooman P *et al*. Efficacy and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: a randomized controlled trial. US Oral Neuraminidase Study Group. *JAMA* 2000; **283**: 1016.
- 4 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 Montalto NJ. An office-based approach to influenza: clinical diagnosis and laboratory testing. *Am Fam Physician* 2003; **67**: 111.
- 6 Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980; **302**: 1109.
- 7 Stein J, Louie J, Flanders S *et al*. Performance characteristics of clinical diagnosis, a clinical decision rule, and a rapid influenza test in the detection of influenza infection in a community sample of adults. *Ann Emerg Med* 2005; **46**: 412–9.
- 8 Govaert T, Dinant G, Aretz K, Knottnerus J. The predictive value of influenza symptomatology in elderly people. *Fam Pract* 1998; **15**: 16.
- 9 Boivin G, Hardy I, Tellier G, Maziade J. Predicting influenza infections during epidemics with use of a clinical case definition. *Clin Infect Dis* 2000; **31**: 1166–9.
- 10 Monto A, Gravenstein S, Elliott M, Colopy M, Schweinle J. Clinical signs and symptoms predicting influenza infection. *Arch Intern Med* 2000; **160**: 3243.
- 11 Van den Dool C, Hak E, Wallinga J *et al*. Symptoms of influenza virus infection in hospitalized patients. *Infect Control Hosp Epidemiol* 2008; **29**: 314–9.
- 12 Walsh E, Cox C, Falsey A. Clinical features of influenza A virus infection in older hospitalized persons. *J Am Geriatr Soc* 2002; **50**: 1498–503.
- 13 Friedman M, Attia M. Clinical predictors of influenza in children. *Arch Pediatr Adolesc Med* 2004; **158**: 391.
- 14 Ebell MH, Afonso A. A systematic review of clinical decision rules for the diagnosis of influenza. *Ann Fam Med* 2011; **9**: 69.
- 15 Carrat F, Tachet A, Rouzioux C, Housset B, Valleron A. Evaluation of clinical case definitions of influenza: detailed investigation of patients during the 1995-96 epidemic in France. *Clin Infect Dis* 1999; **28**: 283–90.
- 16 Ebell MH, Afonso AM, Gonzalez R *et al*. Development and validation of a clinical decision rule for the diagnosis of influenza. *J Am Board Fam Med* 2012; **25**: 55–62.
- 17 Breiman L. Classification and Regression Trees. London, UK: Chapman & Hall/CRC, 1984.
- 18 Senn N, Favrat B, D'Acremont V, Ruffieux C, Genton B. How critical is timing for the diagnosis of influenza in general practice? *Swiss Med Wkly* 2005; **135**: 614.
- 19 Centers for Disease Control and Prevention. United States Surveillance Data: 1997-1998 through 2009-2010 Seasons. <http://www.cdc.gov/flu/weekly/ussurvdata.htm> (accessed on 21 January 2012).
- 20 Gaudard M, Ramsey P, Stephens M. Interactive Data Mining and Design of Experiments: The JMP Partition and Custom Design Platforms. Cary, NC: North Haven Group, 2006.
- 21 Recursive Partitioning: The Partition Platform. JMP Statistical Discovery from SAS: Statistics and Graphics Guide, 2nd edn. Cary, NC, 2007, pp. 657–679.
- 22 Ramsey P, Stephens M, Gaudard M. Six Sigma, Data Mining, and the JMP Partition Platform. Cary, NC: North Haven group, LLC, 2005.
- 23 Sintchenko V, Gilbert G, Coiera E, Dwyer D. Treat or test first? Decision analysis of empirical antiviral treatment of influenza virus infection versus treatment based on rapid test results. *J Clin Virol* 2002; **25**: 15–21.
- 24 Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001; **323**: 157.
- 25 Seasonal Influenza (Flu): Flu Activity & Surveillance. <http://www.cdc.gov/flu/weekly/fluactivitysurv.htm> (accessed on 15 June 2011).
- 26 WHO/Europe Influenza Surveillance (EuroFlu.org). <http://www.euroflu.org/> (accessed on 29 July 2011).