Original Article

Influenza infection screening tools fail to accurately predict influenza status for hospitalized patients during pandemic H1N1 influenza season

Sunita Mulpuru MD FRCP C1,2,3, Virginia R Roth MD FRCP C MBA1,3,4,5, Nadine Lawrence MSc5, Alan J Forster MD FRCP C MSc1,3,4,5

Correspondence: Dr Sunita Mulpuru, The Ottawa Hospital, General Campus, Division of Respirology, 501 Smyth Road, Box 211, Ottawa, Ontario K1H 8L6. Telephone 613-737-8899 ext 73994, fax 613-761-4152, e-mail smulpuru@toh.on.ca

BACKGROUND: Following the severe acute respiratory syndrome outbreak in 2003, hospitals have been mandated to use infection screening questionnaires to determine which patients have infectious respiratory illness and, therefore, require isolation precautions. Despite widespread use of symptom-based screening tools in Ontario, there are no data supporting the accuracy of these screening tools in hospitalized patients.

OBJECTIVE: To measure the performance characteristics of infection screening tools used during the H1N1 influenza season.

METHODS: The present retrospective cohort study was conducted at The Ottawa Hospital (Ottawa, Ontario) between October and December 2009. Consecutive inpatients admitted from the emergency department were included if they were ≥18 years of age, underwent a screening tool assessment at presentation and had a most responsible diagnosis that was cardiac, respiratory or infectious. The gold-standard outcome was laboratory diagnosis of influenza.

RESULTS: The prevalence of laboratory-confirmed influenza was 23.5%. The sensitivity and specificity of the febrile respiratory illness screening tool were 74.5% (95% CI 60.5% to 84.8%) and 32.7% (95% CI 25.8% to 40.5%), respectively. The sensitivity and specificity of the influenza-like illness screening tool were 75.6% (95% CI 61.3% to 85.8%) and 46.3% (95% CI 38.2% to 54.7%), respectively.

CONCLUSIONS: The febrile respiratory illness screening tool missed 26% of active influenza cases, while 67% of noninfluenza patients were unnecessarily placed under respiratory isolation. Results of the present study suggest that infection-control practitioners should re-evaluate their strategy of screening patients at admission for contagious respiratory illness using symptom- and sign-based tests. Future efforts should focus on the derivation and validation of clinical decision rules that combine clinical features with laboratory tests.

Key Words: H1N1; Infection control; Infection screening tools; Influenza; Test characteristics

Contagious respiratory illnesses are important and preventable causes of health care-acquired infections (1). One way of preventing nosocomial transmission is to screen patients for infectious signs and symptoms at the time of admission to hospital. Special precautions can then be applied, such as isolating high-risk patients and using personal protective equipment during care. This approach was one of several infection-control precautions attributed to the control of the 2003 severe acute respiratory syndrome (SARS) outbreak in Canada (2-4) and has become a cornerstone technique in controlling health care-acquired respiratory infections.

Routine screening of patients presenting to emergency departments (EDs) has been recommended by various groups, including the Ministry of Health and Long-Term Care, Ontario Provincial Infectious Diseases Advisory Committee and the Public Health Agency of Canada (5-7). Based on this mandate, many hospital EDs hired dedicated personnel to screen all patients with two symptom-based screening tools during the pandemic H1N1 season. Ontario hospitals used the Febrile Respiratory Illness (FRI) and Influenza-Like Illness (ILI) questionnaires (8) (Appendix 1). The FRI screening tool queried symptoms of new shortness of breath, cough and fever, while the ILI asked more pointed questions about influenza-associated symptoms including sore throat, arthralgias, myalgias, prostration and diarrhea (Appendix 1). If the FRI screen was positive, patients were placed under droplet isolation precautions. The FRI and ILI screening tools were used in conjunction to guide decisions regarding respiratory isolation precautions during the H1N1 influenza pandemic before confirmatory testing was available.

Les outils de dépistage de l’infection par la grippe n’ont pas réussi à prédire avec exactitude le statut d’influenza des patients hospitalisés pendant la saison de pandémie de grippe H1N1

HISTORIQUE : Après l’éclosion de syndrome respiratoire aigu sévère en 2003, les hôpitaux ont été mandatés pour utiliser les questionnaires de dépistage des infections afin de déterminer les patients atteints d’une maladie respiratoire infectieuse et, par conséquent, nécessitant un isolement préventif. Malgré l’utilisation généralisée d’outils de dépistage fondés sur les symptômes en Ontario, aucune donnée n’en établie l’exactitude chez les patients hospitalisés.

OBJECTIF : Mesurer les caractéristiques de rendement des outils de dépistage des infections utilisés pendant la saison de la grippe H1N1.

MÉTHODOLOGIE : La présente étude rétrospective de cohorte a été menée à L’Hôpital d’Ottawa, en Ontario, entre octobre et décembre 2009. Les patients hospitalisés consécutifs admis par le département d’urgence y étaient inclus s’ils avaient 18 ans ou plus, avaient subi une évaluation par l’outil de dépistage à leur présentation et avaient obtenu un diagnostic principal de nature cardiaque, respiratoire ou infectieuse. L’issue de référence était un diagnostic de grippe en laboratoire.

RÉSULTATS : La prévalence de grippes confirmées en laboratoire s’élevait à 23,5 %. La sensibilité et la spécificité de l’outil de dépistage de maladie respiratoire fébrile étaient de 74,5 % (95 % IC 60,5 % à 84,8 %) et de 32,7 % (95 % IC 25,8 % à 40,5 %), respectivement. La sensibilité et la spécificité de l’outil de dépistage de maladie pseudogrippale correspondaient à 75,6 % (95 % IC 61,3 % à 85,8 %) et à 46,3 % (95 % IC 38,2 % à 54,7 %), respectivement.

CONCLUSIONS : L’outil de dépistage de maladie respiratoire fébrile ne distinguait pas 26 % des cas de grippe active, tandis que 67 % des patients n’ayant pas la grippe étaient placés à tort en isolement respiratoire. D’après les résultats de la présente étude, les praticiens de contrôle des infections devraient réévaluer leur stratégie de dépistage des maladies respiratoires contaminantes des patients à leur admission au moyen de tests fondés sur les signes et symptômes. Il faudra porter les futurs efforts sur la dérivation et la validation des règles de décision cliniques qui associent les caractéristiques cliniques aux tests de laboratoire.
While this type of symptom-based screening approach helped to control the SARS outbreak in 2003 (4), there are no data supporting its application in preventing spread of other illnesses, such as influenza, in the hospitalized patient population.

We conducted the present retrospective cohort study to measure the performance test characteristics of the FRI and ILLI screening questionnaires as they were applied during the 2009 H1N1 pandemic influenza season. Given that screening tools were used routinely along with gold-standard laboratory testing, this provided a unique opportunity to assess the accuracy of the current screening approach in a pandemic influenza season.

METHODS

Study design and setting
The present retrospective cohort study was performed at the Ottawa Hospital (TOH), a 1000-bed academic hospital located in Ottawa, Ontario. TOH provides both secondary and tertiary care for patients in the surrounding areas. The present study was approved by TOH’s institutional Research Ethics Board.

Study population
Consecutive inpatients who were at least 18 years of age, admitted from the ED between October 2 and December 29, 2009, underwent a screening tool assessment at presentation, and had a most responsible diagnosis that was cardiac, respiratory or infectious in nature (Appendix 2) were included. Although it was mandated that all patients presenting to the emergency room be screened by dedicated personnel, there were some cases in which the screening tool was incompletely or incorrectly applied. These cases represented <10% of all patients screened for inclusion in the study. Any patient with incomplete screening tool data was excluded from the study.

Application of infection screening tools
All patients were initially screened using the FRI screening tool. The FRI was considered to be positive if the patient exhibited shortness of breath or cough, and the presence of fever or chills (Appendix 1). Droplet isolation precautions were applied if the FRI screen was positive. All patients with positive FRI screens underwent additional screening using the ILLI tool. If the ILLI was negative, droplet precautions were maintained. If the ILLI was positive, patients may have also been placed under airborne isolation. Due to evolving ministry policy and resource limitations, airborne isolation was not used consistently throughout the study period, even with a positive ILLI screen.

Outcomes
The gold-standard outcome in the present study was laboratory-confirmed influenza by polymerase chain reaction (PCR). Influenza-specific PCR was conducted on samples from nasopharyngeal swabs. For most studies, PCR is considered to be the gold standard for influenza diagnosis (9,10). Multiplex PCR was not used routinely at the TOH during this time and data regarding other respiratory viruses were not collected. The secondary outcome was a clinical diagnosis of influenza, as indicated by the most responsible diagnosis on the hospital discharge summary.

Data collection
Using the Ottawa Hospital Data Warehouse, a list of consecutive adult patients admitted to hospital via the ED was obtained. The Ottawa Hospital Data Warehouse is a relational database containing information from several of TOH’s most important operation information systems. Extensive assessments of quality are executed routinely as new data are loaded. All eligible patient encounters underwent chart review using an objective abstraction tool to determine baseline characteristics, the treating physician’s clinical diagnosis, the FRI and ILLI results, and the microbiological test results. Standard approaches for calculating test characteristics and 95% confidence limits were applied (11). The Charlson Index was calculated using a validated method for administrative data (12). Codes used for ‘complications’ were excluded and only ‘comorbid’ diagnoses were used.

Analysis
In the primary analysis, performance test characteristics were calculated using a laboratory-confirmed diagnosis of influenza as the gold standard. The primary analysis group consisted of a subset of patients in the study cohort who underwent laboratory testing (200 of 391 patients). The secondary analysis was conducted on all patients in the study cohort (n=391), and used a clinical diagnosis of influenza as the gold-standard diagnosis based on the hospital discharge summary. This was used as the secondary analysis to assess the possibility of bias introduced through selective laboratory testing for influenza. Agreement between clinical and laboratory diagnoses of influenza was assessed using the kappa statistic (13).

RESULTS
Baseline characteristics are described in Table 1. Forty-seven of 200 (23.5%) patients in the primary analysis group had laboratory-confirmed influenza according to PCR analysis. Fifty-four of 391 (13.8%) patients in the secondary analysis group had a clinical diagnosis of influenza as documented on the hospital discharge summary.

Compared with patients in whom laboratory tests were negative, patients with laboratory-confirmed influenza were younger, had a lower prevalence of comorbid conditions (according to Charlson Index), had shorter length of hospital stay and had lower baseline probability of in-hospital mortality.

Test characteristics of the FRI and ILLI are reported in Table 2. In the primary analysis, the sensitivity and specificity of the FRI screening tool were 74.5% (95% CI 60.5% to 84.8%) and 32.7% (95% CI 25.8% to 40.5%), respectively. The sensitivity and specificity of the ILLI screening tool were 75.6% (95% CI 61.3% to 85.8%) and 46.3% (95% CI 38.2% to 54.7%), respectively. In the secondary analysis (n=391), sensitivity and positive predictive values (PPVs) were similar.

The strength of agreement between the laboratory and clinical diagnosis of influenza was calculated among 200 patients in the present study. These 200 patients had both a clinical diagnosis and microbiological testing performed. The kappa statistic was 0.799 (95% CI 0.70 to 0.90), which is considered to be ‘good agreement’ (13).

DISCUSSION
The FRI and ILLI screening tools implemented in hospitalized patients presenting to the ED during the pandemic H1N1 influenza season were inaccurate reflections of whether a patient had influenza. The FRI screening tool missed 26% of active laboratory-confirmed influenza cases, while 67% of noninfluenza patients were inappropriately isolated. Misclassification of influenza cases in the ED has detrimental consequences. As observed during the SARS epidemic in 2003, infected patients who were not isolated transmitted illness to other patients and health care workers (3,5). The consequences of nosocomial transmission were drastic during the SARS epidemic (2). Over a six-month period in 2003, the Ontario Ministry of Health and Long-Term Care reported 375 cases and 44 SARS-related deaths in Ontario (2). More than 1000 physicians were quarantined during this time, representing 5% of all physicians in Ontario (14). Furthermore, isolating noninfectious patients is associated with adverse clinical events and negative psychological consequences, including higher patient scores for anxiety and depression (15-17). Studies have also demonstrated that health care providers spend less time with and perform physical examinations less frequently on patients under isolation precautions (15,16). Unnecessary isolation procedures are also financially wasteful and can lead to backlogs in the ED.

Studies of symptom prediction rules for influenza have found similar results in the outpatient population (18-25). Combining laboratory information with clinical findings may improve screening tool accuracy; Bewick et al (26) used this approach to identify a more accurate rule. Screening and diagnostic strategies that combine clinical data...
Flu screening tools fail to predict infection

with rapid influenza diagnostic tests (RIDTs) may prove to be the optimal solution in the future. RIDTs or ‘point-of-care’ tests can detect influenza viral proteins and provide results in $\leq 30$ min (27). There are a variety of commercially available tests in Canada but their sensitivity is poor (28). A recent study by Sutter et al (29) demonstrated the poor sensitivity of these tests when compared with PCR (29). If RIDTs become less expensive and more accurate in the future, they could play an important role in screening processes for patients with FRI. They could help inform clinical decisions regarding isolation precautions and antiviral therapy in a relevant time frame to reduce costs and the consequences of inappropriate isolation.

We also noted that patients with laboratory-confirmed influenza in the present study had a shorter length of hospital stay and lower risk-adjusted mortality (Table 1). This trend could be explained by the observation that patients who tested negative for influenza by PCR were actually older on average and had a higher burden of comorbidities (according to Charlson Index) compared with patients without laboratory-confirmed influenza. This trend is also noted in the secondary analysis group. Other confounding factors, such as specific respiratory and cardiac comorbidities and patient functional status, may have contributed to the higher mortality and length of stay in the noninfluenza group.

Our study had several limitations. First, the retrospective methodology made it difficult to control for all residual confounding factors. For example, we did not analyze single comorbidities, such as chronic respiratory conditions, malignancy, cardiac disease or pulmonary thromboembolic disease, in the cohort. These comorbidities could affect the result of the screening tool (because patients are more likely to have a positive screen due to the nature of these disease-related symptoms) as well as the likelihood of contracting influenza.

Second, our primary analysis group represented a subset of the study cohort because only 200 of 391 patients were swabbed for influenza. There may have been factors that led to the decision to test patients for influenza that we could not account for. This introduces a selection bias into the study and represents a limitation of its retrospective design. To account for this, we compared the strength of agreement between influenza diagnosis in the laboratory-confirmed

### Table 1

**Characteristics of the study cohort according to laboratory-confirmed and clinically diagnosed influenza cases**

<table>
<thead>
<tr>
<th>Admitting service</th>
<th>Laboratory-confirmed influenza (n=200)</th>
<th>Clinical diagnosis of influenza (n=391)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes (n=47)</td>
<td>No (n=153)</td>
</tr>
<tr>
<td>Internal medicine</td>
<td>30 (63.8)</td>
<td>103 (67.3)</td>
</tr>
<tr>
<td>Respirology</td>
<td>$\leq 5$ (10.6)</td>
<td>7 (4.6)</td>
</tr>
<tr>
<td>Family medicine</td>
<td>$\leq 5$ (4.3)</td>
<td>13 (8.5)</td>
</tr>
<tr>
<td>Oncology</td>
<td>$\leq 5$ (4.3)</td>
<td>$\leq 5$ (3.3)</td>
</tr>
<tr>
<td>Critical care</td>
<td>$\leq 5$ (4.3)</td>
<td>7 (4.6)</td>
</tr>
<tr>
<td>Hematology</td>
<td>$\leq 5$ (4.3)</td>
<td>$\leq 5$ (3.3)</td>
</tr>
<tr>
<td>Nephrology</td>
<td>$\leq 5$ (2.1)</td>
<td>$\leq 5$ (1.3)</td>
</tr>
<tr>
<td>Obstetrics and gynecology</td>
<td>$\leq 5$ (4.3)</td>
<td>$\leq 5$ (1.3)</td>
</tr>
<tr>
<td>Other</td>
<td>n/a</td>
<td>9 (5.8)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>24 (51.1)</td>
<td>84 (54.9)</td>
</tr>
<tr>
<td>Male</td>
<td>23 (48.9)</td>
<td>69 (45.1)</td>
</tr>
<tr>
<td>Charlson Index score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>23 (48.9)</td>
<td>49 (32.0)</td>
</tr>
<tr>
<td>1–2</td>
<td>18 (38.3)</td>
<td>55 (35.9)</td>
</tr>
<tr>
<td>3–4</td>
<td>$\leq 5$ (6.4)</td>
<td>28 (18.3)</td>
</tr>
<tr>
<td>$\geq 5$</td>
<td>$\leq 5$ (6.4)</td>
<td>21 (13.7)</td>
</tr>
<tr>
<td>Admission month</td>
<td></td>
<td></td>
</tr>
<tr>
<td>October</td>
<td>6 (12.8)</td>
<td>20 (13.1)</td>
</tr>
<tr>
<td>November</td>
<td>41 (87.2)</td>
<td>96 (62.7)</td>
</tr>
<tr>
<td>December</td>
<td>0 (0.0)</td>
<td>37 (24.2)</td>
</tr>
<tr>
<td>Age, years, mean ± SD</td>
<td>49.1±17.36</td>
<td>66.8±18.31</td>
</tr>
<tr>
<td>Length of stay, days, mean ± SD</td>
<td>7.94±9.64</td>
<td>10.74±13.02</td>
</tr>
<tr>
<td>Intensive care unit admission</td>
<td>11 (23.4)</td>
<td>35 (22.9)</td>
</tr>
<tr>
<td>Risk-adjusted mortality, mean ± SD</td>
<td>0.05±0.07</td>
<td>0.15±0.13</td>
</tr>
</tbody>
</table>

Data presented as n (%) unless otherwise indicated. n/a Not applicable

### Table 2

**Performance test characteristics of the febrile respiratory illness (FRI) and influenza-like illness (ILI) infection screening tools in primary and secondary analysis groups**

<table>
<thead>
<tr>
<th>Laboratory-confirmed influenza (n=200)</th>
<th>Positive, n</th>
<th>Negative, n</th>
<th>Sensitivity (% (95% CI))</th>
<th>Specificity (% (95% CI))</th>
<th>Predictive value</th>
<th>Likelihood ratio, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRI</td>
<td>35</td>
<td>103</td>
<td>74.5 (60.5–84.8)</td>
<td>32.7 (25.8–40.5)</td>
<td>25.4</td>
<td>80.6</td>
</tr>
<tr>
<td>ILI*</td>
<td>34</td>
<td>73</td>
<td>75.6 (61.3–85.8)</td>
<td>46.3 (38.2–54.7)</td>
<td>31.8</td>
<td>85.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clinical diagnosis of influenza (n=391)</th>
<th>Positive, n</th>
<th>Negative, n</th>
<th>Sensitivity (% (95% CI))</th>
<th>Specificity (% (95% CI))</th>
<th>Predictive value</th>
<th>Likelihood ratio, % (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FRI</td>
<td>39</td>
<td>131</td>
<td>72.2 (59.1–82.3)</td>
<td>61.1 (55.8–66.1)</td>
<td>22.9</td>
<td>93.2</td>
</tr>
<tr>
<td>ILI*</td>
<td>38</td>
<td>87</td>
<td>73.1 (59.8–83.2)</td>
<td>72.4 (67.3–77.1)</td>
<td>30.4</td>
<td>94.2</td>
</tr>
</tbody>
</table>

*Total numbers in the ILI groups were fewer because all participants with a FRI did not have an ILI screen completed
and clinical-diagnosis groups and found ‘good agreement’ according to the kappa statistic. In addition, sensitivities of the FRI tool were similar between primary analysis and secondary analysis groups, suggesting that the selection bias was not significant. A prospective study would better address residual confounders and bias, and is certainly the next step to assess the performance of the FRI and ILI screening tools in hospitals. We believed that a retrospective study was first warranted given the lack of peer-reviewed data on screening tool performance in hospitals. We believed that a retrospective study was first warranted given the lack of peer-reviewed data on screening tool performance in the hospitalized patient population.

Third, the present study was performed during a unique influenza season (pandemic H1N1) at a single academic hospital, where the prevalence of influenza was unusually high (23.5% in the laboratory-confirmed influenza group). Furthermore, clinician awareness of influenza was heightened and there may have been greater frequency of testing compared with other influenza seasons. Given this, we cannot generalize the performance characteristics of the FRI and ILI screening tools to all influenza seasons. However, the PPV of the FRI and ILI screening tools as calculated in the present study would represent a conservative estimate of the ‘true’ PPV, given the unusually high prevalence of influenza. The PPV of the FRI and ILI questionnaires in low-prevalence influenza seasons is likely to be lower. Therefore, in times of low prevalence, one would be less likely to have influenza with a positive screening tool. We are not aware of and do not have any data from other influenza seasons in which the performance of the FRI and ILI screening tools were evaluated in hospitalized patients. Future prospective research should examine influenza seasons with variable prevalence to better understand the FRI screening tool performance.

Fourth, our institution was not routinely using multiplex PCR during the study period and, therefore, coinfection or infection with other respiratory viruses may not have been detected. Based on viral culture results, we identified another respiratory virus in only two other patients (two of 200). However, because we did not systematically test for other pathogens, we did not include these data. It is possible that excluding other respiratory viruses may have led to reduced specificity and PPV of the FRI and ILI screening tools. Future work using multiplex PCR as the gold-standard viral diagnostic test is required to overcome this limitation.

Finally, some may argue that a sensitivity of 74.5% (95% CI 60.5% to 84.8%) is adequate for this type of infection screening tool. We believe that 75% (three of four) is too low when dealing with a contagious illness that has a significant morbidity and mortality impact within a hospitalized population. To illustrate this, we calculated the pretest and post-test probability of influenza if the screening tool is negative during the pandemic season. In our cohort of patients with laboratory-confirmed influenza, the prevalence of influenza was 23.5% and the negative likelihood ratio was 0.78. Therefore, the post-test probability of having influenza with a negative screening tool is 19.3%. A post-test probability >10% is too high for cases missed by screening. However, if the prevalence of influenza is lower (for example, 5% or 10%) in a given influenza season, then the post-test probability of having influenza with a negative screening tool drops below 10%. This illustrates that while these screening tools maybe more effective during times of low influenza activity, they do not perform adequately in times of greater influenza burden. We liken this example to that of tuberculosis (TB), another contagious respiratory disease. The tuberculin skin test is falsely negative 30% of the time when used in the assessment of active TB (30). This means that 30% of active disease would be missed and would represent a substantial transmission risk. Thus, guidelines have not advocated for use of the tuberculin skin test in detection of active TB (30). Transmission of viral illnesses in hospital can also have significant morbidity and mortality, as seen with SARS; therefore, a sensitivity of 75% is inadequate.

Our study suggests that policy makers and infection-control practitioners should be cautious of the accuracy of symptom-based screening for contagious respiratory illness in the hospitalized patient population. During times of increased prevalence of influenza, use of these screening tools in hospitals may lead to significant misuse of isolation precautions, increased costs and the potential for nosocomial infection. Further prospective research is required to confirm the accuracy of infection screening tools and inform current policies.

ACKNOWLEDGEMENTS/DISCLOSURES: The authors acknowledge and thank the Ontario Thoracic Society and the sponsors of the Cameron C Gray Fellowship Award. Dr Sunita Mulpuru is supported by a 2011-2012 Cameron C Gray Fellowship Award from the Ontario Thoracic Society. SM, VRR and AF contributed to the study design, data analysis and critical review of the manuscript. SM completed the data collection, chart review and initial draft of the manuscript. VRR and AF provided content and methodological expertise. NL completed the chart abstraction from The Ottawa Hospital Data Warehouse. All authors gave final approval of the submitted version.

APPENDIX 1 Sample of the febrile respiratory illness and influenza-like illness screening tools used at The Ottawa Hospital, Ottawa, Ontario

APPENDIX 2 International Classification of Diseases, 10th Revision diagnostic codes used to identify the study cohort

CARDIOVASCULAR CONDITIONS

Respiratory conditions: All J codes
Infections and sepsis: A15-A19, A37, A40, A41.5, A41.8, A49
Cardiac conditions: I26, I28, I50, I51.4, R57

REFERENCES
Flu screening tools fail to predict infection


