# RESPIRATORY INFECTION

# A prospective comparison of severity scores for identifying patients with severe community acquired pneumonia: reconsidering what is meant by severe pneumonia

## K L Buising, K A Thursky, J F Black, L MacGregor, A C Street, M P Kennedy, G V Brown

**Background:** Several severity scores have been proposed to predict patient outcome and to guide initial management of patients with community acquired pneumonia (CAP). Most have been derived as predictors of mortality. A study was undertaken to compare the predictive value of these tools using different clinically meaningful outcomes as constructs for ''severe pneumonia''.

**Methods:** A prospective cohort study was performed of all patients presenting to the emergency department with an admission diagnosis of CAP from March 2003 to March 2004. Clinical and laboratory features at presentation were used to calculate severity scores using the pneumonia severity index (PSI), the revised American Thoracic Society score (rATS), and the British Thoracic Society (BTS) severity scores CURB, modified BTS severity score, and CURB-65. The sensitivity, specificity, positive and negative predictive values were compared for four different outcomes (death, need for ICU admission, and combined outcomes of death and/or need for ventilatory or inotropic support).

**Results:** 392 patients were included in the analysis; 37 (9.4%) died and 26 (6.6%) required ventilatory and/or inotropic support. The modified BTS severity score performed best for all four outcomes. The PSI (classes IV+V) and CURB had a very similar performance as predictive tools for each outcome. The rATS identified the need for ICU admission well but not mortality. The CURB-65 score predicted mortality well but performed less well when requirement for ICU was included in the outcome of interest. When the combined outcome was evaluated (excluding patients aged >90 years and those from nursing homes), the best predictors were the modified BTS severity score (sensitivity 94.3%) and the PSI and CURB score (sensitivity 83.3% for both).

**Conclusions:** Different severity scores have different strengths and weaknesses as prediction tools. Validation should be done in the most relevant clinical setting, using more appropriate constructs of ''severe pneumonia'' to ensure that these potentially useful tools truly deliver what clinicians expect of them.

See end of article for authors' affiliations
.......................

Correspondence to:
Dr K L Buising, Victorian Infectious Diseases Service, 9th Floor, Royal Melbourne Hospital, Grattan Street, Parkville, Victoria 3050, Australia; Kirsty.Buising@mh.org.au

Severity scores have been promoted as useful tools to help clinicians predict the outcome of patients presenting with community acquired pneumonia (CAP).[1–4] For patients identified as likely to have ''severe pneumonia'', management strategies can be appropriately tailored to include admission to hospital, involvement of an experienced clinician in their care, early consideration of intensive care unit (ICU) management, and the use of broad spectrum empirical antibiotics.

Most severity scores for CAP are mortality prediction tools and therefore they identify many elderly patients and patients with complex co-morbidities whose pneumonia may not have been particularly severe, but it serves as the final factor leading to death. For many of these patients, aggressive management strategies would not be clinically appropriate.

Requirement for ICU admission is an alternative definition for ''severe pneumonia'' and some severity scores have been evaluated for their ability to predict this outcome.[5–7] Unfortunately, ICU admission is an imperfect surrogate marker for the construct ''severe pneumonia'' as it involves subjective judgement and the criteria for admission differ between institutions. Using this as a sole outcome of interest risks overlooking patients whose illness may have been underestimated clinically and who died without reaching the ICU.

We suggest that alternative constructs for ''severe pneumonia'' need to be considered. In this study we evaluated the performance of several published severity scores for CAP using different outcomes. The scores included modifications of the British Thoracic Society (BTS) severity score (CURB, CURB-65, and the modified BTS), the revised American Thoracic Society severity score (rATS), and the pneumonia severity index (PSI).[2 3 5–7]

## METHODS

The study was performed at the Royal Melbourne Hospital, an urban adult tertiary teaching hospital with 350 beds including 14 ICU beds. The emergency department assesses 50 000 patients per year, leading to 16 000 admissions.

Consecutive patients presenting to the emergency department between 1 April 2003 and 30 March 2004 with a diagnosis of ''pneumonia'' made by the treating clinician within the first 24 hours of presentation (based on clinical assessment, initial pathology results, and chest radiographic assessment by the clinician) were recruited to the study. All patients prospectively recorded in the emergency department database with symptoms or a diagnosis suggestive of a respiratory or infective illness (including pneumonia, chest

.................................................................

**Abbreviations:** CAP, community acquired pneumonia; PSI, pneumonia severity index

infection, and lower respiratory tract infection) were identified. Those for whom the medical record suggested that the treating doctor made a clinical diagnosis of pneumonia were included in the study. Exclusion criteria included age <18 years; immunosuppression (acquired immune deficiency syndrome with CD4 <200 μl, chemotherapy within the last month, absolute neutrophil count <$0.5 \times 10^9$/l, transplant recipient with ongoing use of immunosuppressants, use of corticosteroids at a dose equivalent to prednisolone >15 mg/day); chronic suppurative lung disease (bronchiectasis, cystic fibrosis); and nosocomial pneumonia (admitted to hospital for >48 hours within 2 weeks prior to presentation).

Data on clinical features and pathology and radiology results available in the first 24 hours after presentation were collected by manual review of the medical record and the pathology computer database. Data collected included age, sex, residency in nursing home, respiratory rate, blood pressure, temperature, heart rate, presence of acute confusion, percutaneous oxygen saturation, co-morbid diseases, initial chest radiographic findings as assessed by the clinician and the radiologist (recorded separately), known antibiotic allergies, prior antibiotic use, tests ordered (and microbiology results obtained), and the site of initial management. The definition of acute confusion was based on the clinician's assessment (that the patient's mental state was altered and that this was a new phenomenon). If pre-existing dementia was known, then deterioration from the preceding usual state was required. A mini-mental state examination was not required. The pathology data collected included arterial blood gas results, serum urea, creatinine, glucose, sodium, haematocrit, and white blood cell count. The clinical and pathology results collected represented the most abnormal result (highest and/or lowest) in the 24 hours from the time of arrival at the emergency department. This time period was deliberately chosen as it most closely resembled the time during which assessments are made in usual clinical practice. Missing values were assumed to be normal, in accordance with methodology in previous studies.

The antibiotics prescribed in the first 48 hours were recorded, as were all antibiotics subsequently prescribed. The progress and outcome of patients were monitored prospectively. This included the length of stay in hospital, requirement for ICU admission at any time during the hospital stay, length of stay in ICU, time to ICU admission, requirement for ventilatory assistance, need for inotropic support, in-hospital mortality, and re-presentation within 2 weeks. If arterial blood gases were not tested, they were assumed to be within the normal range.

Severity scores including the PSI, the modified BTS severity scores, and the revised ATS severity scores were calculated using collected data. The severity scores were defined as follows:

- The PSI developed by Fine *et al*[5] uses 20 clinical variables to determine a score. These scores are then used to define five classes of increasing risk of mortality. We assessed the use of class V alone and classes IV+V to define ''severe pneumonia'' as other guidelines have previously suggested these two definitions.[4 8] This prediction tool has been independently validated and widely endorsed.[2 3 9–11]

- The CURB index[6] was derived from the original BTS study[12] and uses four core clinical features: confusion of new onset (or worsening of existing state for those with background cognitive impairment), serum urea >7 mmol/l, respiratory rate ≥30/min, and blood pressure (systolic blood pressure <90 mm Hg or diastolic blood pressure ≤60 mm Hg). The presence of two or more of these four

criteria led to a ''severe'' classification. This tool has been validated independently.[13–15]

- The CURB-65 index[7] is a further modification of the BTS prediction rules. Age ≥65 years is added as a fifth variable to the four core variables mentioned above. To be classed as severe, a patient needed to meet three or more of the five variables. This tool has been endorsed in some guidelines.[10 16]

- The modified BTS score was suggested in the 2001 BTS guidelines for management of CAP.[2] As a first step the four core CURB variables are assessed and, if a patient has two or more of the four variables, they are classed as severe. If the patient has only one core criteria or is aged ≥50 years or has one co-morbidity, then a second step is required. This step involves assessment for two additional variables: oxygen saturations <92% and the presence of bilateral or multilobar infiltrates on the chest radiograph. If either of these additional criteria is met, then the patient is classed as ''severe''. To our knowledge, no independent validation of this tool has been published.

- The revised ATS (rATS) was proposed by Ewig *et al* and incorporated in the ATS guidelines in 2001.[3 11 17–19] This predictive rule classed a patient as having ''severe pneumonia'' if they met one out of two major criteria (requirement for mechanical ventilation or septic shock) or two out of three minor criteria (systolic blood pressure <90 mm Hg, multilobar chest radiographic changes, or $Pao_2/Fio_2$ <250).

The performance of the severity scores in predicting both death in hospital and the need for ICU admission was evaluated. A variable that included all patients requiring either inotropic support or ventilatory assistance (non-invasive or invasive ventilation) within 48 hours of presentation where no other cause for circulatory or respiratory failure was clinically evident was also evaluated (as it was thought to represent a more objective outcome than ICU admission). Finally, a combined outcome of interest was defined which represented patients who died or required extraordinary interventions to keep them alive—that is, death and/or requirement for ventilatory support or inotropic support. In a subsequent analysis, patients aged ≥90 years, those from nursing homes, and those with advanced illness who were not considered suitable for aggressive treatment (for example, not given antibiotics) were excluded as this patient group was judged unlikely to be the group for whom a prediction tool would need to be applied.

The treating clinicians were unaware of the research being conducted. All decisions regarding diagnostic tests and therapeutics were made by clinicians without intervention by research staff. ICU assessment was based upon usual clinical evaluation. No specific guidelines were promoted at the time. Some clinicians may have been aware of and used severity scores. The current Australian guidelines promote the use of the PSI,[21] and a computerised calculator was available to assist with PSI calculations at the point of care but these were not specifically promoted.

## Statistical analysis

Descriptive analyses were used for patient characteristics. Sensitivity, specificity, positive predictive value, negative predictive value, and 95% confidence intervals were calculated for each severity score, for each outcome of interest. A receiver operator characteristics (ROC) curve was constructed using the performance criteria of each tool, and the area under these curves was reported. For the modified BTS score and the rATS score, the areas under the ROC were approximated as the scores only have binary outcomes.

**Table 1** Patient characteristics, number in each severity score group, management and outcomes

| | |
|---|---|
| Median (range) age (years) | 74 (18–96) |
| Age group (years) | |
|   18–38 | 47 (11.9) |
|   39–58 | 56 (14.3) |
|   59–78 | 137 (34.9) |
|   >78 | 152 (38.7) |
| Female | 158 (40.3) |
| Current smokers | 83 (21.1) |
| Alcohol abuse | 38 (9.6) |
| Nursing home residents | 55 (14.0) |
| Previous presentation within 2 weeks | 19 (4.8) |
| Antibiotics prior to presentation | 99 (25.2) |
| Co-morbidities | |
|   Congestive cardiac failure | 80 (20.4) |
|   Chronic liver disease | 10 (2.5) |
|   Neoplasia | 54 (13.7) |
|   Chronic renal failure | 47 (11.9) |
|   Cerebrovascular disease | 74 (18.8) |
|   COPD | 92 (23.4) |
|   PEG/tracheostomy | 19 (4.8) |
|   Dementia/neurological disease | 52 (13.2) |
|   Diabetes | 87 (22.1) |
|   Aspiration clinically suspected | 39 (9.9) |
| Antibiotic allergy | |
|   No allergy | 350 (89.3) |
|   Mild | 13 (3.3) |
|   Unknown severity | 21 (5.3) |
|   Severe | 8 (2.0) |
| Severity scores | |
|   PSI class | |
|     I | 47 (11.9) |
|     II | 57 (14.5) |
|     III | 68 (17.3) |
|     IV | 131 (33.4) |
|     V | 89 (22.7) |
|     IV + V | 220 (56.1) |
|   rATS severe | 70 (17.8) |
|   CURB severe | 182 (46.4) |
|   CURB-65 severe (group 3) | 161 (41.0) |
|   ModBTS severe | 214 (54.6) |
| Median (range) LOS (days) | 4 (1–76) |
| Site of care | |
|   Ward | 243 (62.0) |
|   ICU direct | 17 (4.3) |
|   Outpatient | 53 (13.5) |
|   Another hospital (not ICU) | 14 (3.6) |
|   Short stay unit | 62 (15.8) |
| Died in emergency department | 2 (0.5) |
| ICU admission | 26 (6.6) |
|   Mechanical ventilation | 23 (5.8) |
|   Inotropes | 26 (6.6) |
| Died in ICU | 8/26 (30.7) |
| Died outside ICU | 29/366 (7.9) |
| Died (all sites) | 37 (9.4) |
|   Admitted to ICU | 8/37 (21.6) |
|   Nursing home residents | 13/37 (35.1) |
|   Aged >90 years | 7/37 (18.9) |
|   Not for aggressive treatment | 4/37 (10.8) |
| Median LOS of those who died (days) | 6.5 |
| Discharge diagnosis of pneumonia | 347 (88.5) |
| Readmitted within 2 weeks | 32 (8.3) |

Values in parentheses are percentages.
PSI, pneumonia severity index; rATS, revised American Thoracic Society; modBTS, modified British Thoracic Society; COPD, chronic obstructive pulmonary disease; LOS, length of stay.

Statistical analysis was performed using Stata 8.0 (Stata Corp, USA).

## RESULTS

### Population

A total of 392 patients with CAP were included in the analysis. Twenty six patients (6.6%) required ICU admission, 17 of whom (65.3%) went directly to the ICU and, of the remainder, eight were admitted to the ICU within 24 hours and one was admitted on day 7 for another medical complication. No patients received non-invasive ventilation and the hospital did not have a separate high dependency unit at the time of the study. Thirty seven patients (9.4%) died while in hospital and, allowing for overlap between groups, 48.4% of these patients were either aged >90 years, or resided in a nursing home, or were considered to be unsuitable for aggressive treatment within 24 hours of presentation due to complex irreversible co-morbidities. The median age of the patients admitted to the ICU was 62.5 years (range 25–85), while the median age of those who died was 82 years (range 43–97) (p<0.001).

Patients were treated with an empirical antimicrobial regimen selected by the treating clinician, usually a beta lactam (either amoxicillin, penicillin or ceftriaxone) in combination with either a macrolide or doxycycline (as per local guidelines).[20] After excluding those patients who were not treated with antibiotics at all or had suspected aspiration pneumonia, 36% of patients did not receive a recommended antibiotic regimen, the most common reason being treatment with a single antibiotic rather than combination therapy. Fifty five patients (14%) received only oral antibiotics and 82.4% received intravenous antibiotics initially. Fourteen patients (3.6%) received no antibiotic treatment and most of these were >90 years of age or from a nursing home. Documentation of the result of a PSI calculation was found in the notes of six patients. Further demographic and clinical data are presented in table 1.

Data were missing for 20 patients who had no blood tests performed (hence serum urea, glucose, creatinine, and white cell counts were unavailable). Only 141 patients (35.9%) underwent arterial blood gas tests. Fourteen patients were transferred directly to a private hospital from the emergency department because the patient requested private care. Forty five patients (11.5%) did not have a discharge diagnosis of pneumonia despite being admitted with this clinical diagnosis. Most of these patients had an upper respiratory tract infection (such as acute bronchitis or an acute exacerbation of chronic obstructive pulmonary disease), and many also had evidence of pulmonary venous congestion on the formal chest radiographic report.

### Sensitivity/specificity of severity scores

Applying the severity scores to our entire population, the predictive value of the PSI for mortality was similar to that described in the original PORT (Pneumonia patient outcomes research team) cohort—that is, class I, 0; class II, 0; class III,

**Table 2** Predictive value of scores for mortality

| Severity score | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) (%) | Area under ROC |
|---|---|---|---|---|---|
| PSI class V | 67.5 (50.2 to 81.9) | 82.1 (77.6 to 85.9) | 28.4 (19.3 to 39.0) | 96.0 (93.1 to 97.9) | 0.82 (0.76 to 0.87) |
| PSI class IV+V | 97.3 (85.8 to 99.9) | 47.9 (42.5 to 53.2) | 16.4 (11.7 to 22.0) | 99.4 (96.7 to 99.9) | 0.82 (0.76 to 0.87) |
| CURB (≥2) | 89.2 (74.5 to 96.9) | 58.1 (52.7 to 63.3) | 18.3 (12.9 to 24.7) | 98.1 (95.1 to 99.4) | 0.82 (0.75 to 0.88) |
| CURB-65 (≥3) | 81.0 (64.8 to 92.0) | 67.9 (62.7 to 72.7) | 20.8 (14.5 to 28.4) | 97.2 (94.2 to 98.8) | 0.82 (0.76 to 0.88) |
| Modified BTS | 91.9 (78.1 to 98.3) | 49.8 (44.5 to 55.2) | 16.2 (11.4 to 21.8) | 98.3 (95.1 to 99.6) | 0.71 (0.65 to 0.75) |
| rATS | 40.5 (24.7 to 57.9) | 84.6 (80.4 to 88.2) | 21.7 (12.7 to 33.3) | 93.1 (89.7 to 95.6) | 0.63 (0.54 to 0.71) |

ROC, receiver operator characteristics curve.

**Table 3** Predictive value of scores for admission to ICU

| Severity score | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) (%) | Area under ROC |
|---|---|---|---|---|---|
| PSI class V | 48.0 (27.8 to 68.7) | 79.0 (74.5 to 83.1) | 13.6 (7.2 to 22.6) | 95.6 (92.7 to 97.6) | 0.69 (0.59 to 0.77) |
| PSI class IV+V | 84.0 (63.9 to 95.4) | 45.4 (40.2 to 50.7) | 9.6 (6.0 to 14.3) | 97.6 (94.0 to 99.3) | 0.69 (0.59 to 0.77) |
| CURB (≥2) | 84.0 (63.9 to 95.4) | 56.2 (50.9 to 61.3) | 11.7 (7.3 to 17.2) | 98.1 (95.1 to 99.4) | 0.76 (0.68 to 0.84) |
| CURB-65 (≥3) | 57.7 (36.9 to 76.6) | 64.7 (59.6 to 69.6) | 10.4 (5.9 to 16.6) | 95.5 (92.2 to 97.7) | 0.66 (0.56 to 0.76) |
| Modified BTS | 96.0 (79.6 to 99.9) | 48.7 (43.5 to 54.0) | 11.4 (7.4 to 16.5) | 99.4 (96.9 to 99.9) | 0.72 (0.67 to 0.76) |
| rATS | 92.0 (73.9 to 99.0) | 87.3 (83.4 to 90.5) | 33.3 (22.4 to 45.7) | 99.3 (97.7 to 99.4) | 0.90 (0.81 to 0.94) |

ROC, receiver operator characteristics curve.

2%; class IV, 8%; class V, 28%.[5] The performance of the tools in identifying patients who died is shown in table 2. If the group of elderly patients (>90 years of age, nursing home residents, and patients identified as not for aggressive treatment at the time of admission) are excluded, the sensitivity of the tools for mortality in the remaining patients was 94.7% (18/19) for both the PSI class IV+V and for CURB; 89.5% (17/19) for CURB-65; 100% for the modified BTS score, and 57.8% (11/19) for the rATS. Twenty nine patients who died were not admitted to the ICU before death, 11 of whom were not in the group aged >90 years, from a nursing home, or identified as not for resuscitation within 24 hours of presentation. The CURB, PSI IV+V, and modified BTS tools all identified 10 of these 11 patients as "severe".

The rates of ICU admission in each of the PSI classes were: class I, 0; class II, 2%; class III, 5%; class IV, 7%; and class V, 14%. The rATS performed well in identifying patients requiring ICU admission, as did the modified BTS, but CURB-65 had a sensitivity of only 57.7% for ICU admission. PSI classes IV+V and CURB had similar predictive values for this outcome of interest (table 3). Eight patients who required ICU admission were not admitted directly from the emergency department; seven of these patients required transfer from the ward to the ICU within 24 hours. In this cohort both the PSI classes IV+V and the CURB definitions of severity correctly identified seven of these eight patients (one patient was misclassified by both tools).

Table 4 gives details on the combined outcome of any patients who died and/or required extraordinary interventions to keep them alive (ventilatory support or inotropes). The modified BTS score performed well for all four outcomes used to define the construct of "severe pneumonia". The PSI classes IV+V and CURB had comparable results for each of the four outcomes; for the combined outcome of death and/or ICU admission excluding the elderly group the sensitivities

were the same with overlapping confidence intervals. The CURB was more specific than the PSI (although confidence intervals overlapped), and this is reflected by a slightly higher area under the ROC curve. When the very elderly patients and nursing home residents were excluded, the sensitivity of the PSI and CURB-65 fell while the other tools remained stable or increased in sensitivity which suggests that this patient group was more often being categorised as severe by these tools than by the other severity scores. The performance of the severity scores was separately assessed only for patients with both an admission and a discharge diagnosis of pneumonia (table 5). In this group the discriminative ability of the tools did not change markedly (area under ROC curves: PSI IV and V, 0.78; CURB, 0.80; CURB-65, 0.74; modified BTS, 0.69; rATS, 0.82).

## DISCUSSION

This is the first study to compare the performance of five published severity scores for CAP for different outcomes of interest. The results show that different severity scores for CAP have different strengths and weaknesses depending on which patients the clinician really wants to identify. CURB-65 predicted mortality well, but not the need for ICU admission or the combined outcomes. With this tool, younger patients were less likely to be identified as "severe" as they needed to qualify for three of the remaining four criteria (after excluding age). The rATS was a sensitive tool for ICU admission but not death; however, the major criteria for this tool are not truly "predictive" in that a requirement for inotropes or ventilatory support needs already to have been appreciated. The modified BTS score showed good performance characteristics for ICU admission, death, and the combined outcomes in this study. The PSI is a widely endorsed and well validated tool and performed well for the different outcomes of interest. CURB is a simple tool which

**Table 4** Predictive value of scores for combined outcomes (death and/or ICU admission*)

| Severity score | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) (%) | Area under ROC |
|---|---|---|---|---|---|
| *All patients* | | | | | |
| PSI class V | 55.5 (41.4 to 69.1) | 82.6 (78.1 to 86.5) | 34.1 (24.3 to 44.9) | 92.0 (88.3 to 94.8) | 0.71 (0.63 to 0.78) |
| PSI class IV+V | 90.7 (79.7 to 96.9) | 49.1 (43.6 to 54.6) | 22.3 (17.0 to 28.4) | 97.0 (93.2 to 99.0) | 0.71 (0.63 to 0.78) |
| CURB (≥2) | 85.2 (72.8 to 93.0) | 59.8 (54.4 to 65.0) | 25.5 (19.3 to 32.5) | 96.1 (92.5 to 98.3) | 0.80 (0.74 to 0.85) |
| CURB-65 (≥3) | 70.9 (57.1 to 82.4) | 68.8 (63.6 to 73.7) | 27.1 (20.0 to 35.1) | 93.5 (89.7 to 96.3) | 0.76 (0.70 to 0.82) |
| Modified BTS | 94.4 (84.6 to 98.8) | 52.4 (46.6 to 59.7) | 24.3 (18.6 to 30.6) | 98.3 (95.1 to 99.6) | 0.73 (0.69 to 0.77) |
| rATS | 57.4 (43.2 to 70.7) | 88.6 (84.7 to 91.8) | 44.9 (32.9 to 57.3) | 92.8 (89.3 to 95.3) | 0.73 (0.65 to 0.79) |
| | | | | | |
| *Excluding elderly group†* | | | | | |
| PSI class V | 47.2 (30.4 to 64.5) | 79.8 (75.2 to 83.8) | 19.1 (11.5 to 28.8) | 93.7 (90.4 to 96.2) | 0.76 (0.70 to 0.81) |
| PSI class IV+V | 83.3 (67.2 to 93.6) | 46.6 (41.3 to 51.9) | 13.6 (9.4 to 18.9) | 96.5 (92.6 to 98.7) | 0.76 (0.70 to 0.81) |
| CURB (≥2) | 83.3 (67.2 to 93.6) | 57.8 (52.5 to 63.0) | 16.7 (11.5 to 22.9) | 97.2 (93.9 to 98.9) | 0.80 (0.73 to 0.86) |
| CURB-65 (≥3) | 66.6 (49.0 to 81.3) | 66.3 (61.1 to 71.2) | 16.7 (10.9 to 23.8) | 95.2 (91.7 to 97.8) | 0.73 (0.65 to 0.81) |
| Modified BTS | 94.3 (81.3 to 99.3) | 49.4 (44.1 to 54.7) | 15.9 (11.2 to 21.5) | 98.8 (96.0 to 99.8) | 0.72 (0.67 to 0.76) |
| rATS | 72.2 (54.8 to 85.8) | 87.9 (84.9 to 91.1) | 37.7 (26.3 to 50.1) | 96.9 (94.4 to 98.5) | 0.80 (0.73 to 0.88) |

*Where ICU is defined as requirement for invasive or non-invasive ventilation and/or inotropic support not judged to be due to another cause.
†Excluding nursing home residents, patients aged >90 years, and patients considered unsuitable for aggressive treatment at presentation (due to advanced irreversible co-morbidities), n = 311.

**Table 5** Predictive value of scores for combined outcomes (death and/or ICU admission*) using only those patients with an admission and discharge diagnosis of pneumonia (n = 347)†

| Severity score | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) (%) | Area under ROC |
|---|---|---|---|---|---|
| PSI class V | 54.5 (40.5 to 68.0) | 82.5 (78.0 to 86.4) | 33.7 (24.0 to 44.5) | 91.8 (88.0 to 94.5) | 0.78 (0.71 to 0.84) |
| PSI class IV+V | 89.0 (77.7 to 95.9) | 49.2 (43.8 to 54.7) | 22.2 (16.9 to 28.4) | 96.5 (92.5 to 98.7) | 0.78 (0.71 to 0.84) |
| CURB (⩾2) | 80.0 (67.0 to 89.6) | 59.6 (54.2 to 64.9) | 24.4 (18.2 to 31.4) | 94.8 (90.9 to 97.4) | 0.80 (0.72 to 0.88) |
| CURB-65 (⩾3) | 70.9 (57.1 to 82.4) | 68.8 (63.6 to 73.8) | 27.1 (20.0 to 35.1) | 93.6 (89.7 to 96.3) | 0.74 (0.64 to 0.83) |
| Modified BTS | 98.2 (90.3 to 99.9) | 51.9 (46.5 to 57.4) | 25.0 (19.3 to 31.3) | 99.4 (96.8 to 99.9) | 0.69 (0.63 to 0.75) |
| rATS | 58.2 (44.4 to 71.3) | 88.7 (84.8 to 91.9) | 45.7 (33.7 to 58.0) | 92.8 (89.5 to 95.4) | 0.82 (0.73 to 0.91) |

*Where ICU is defined as requirement for invasive or non-invasive ventilation and/or inotropic support not judged to be due to another cause.
†Excluding nursing home residents, patients aged >90 years, and patients considered unsuitable for aggressive treatment at presentation (due to advanced irreversible co-morbidities).

showed comparable performance to the PSI. This study aimed to identify patients at the severe end of the spectrum of clinical illness for whom aggressive management strategies might be employed, such as early consultant review, ICU admission, and administration of broad spectrum empirical antibiotics. In this cohort the modified BTS, CURB, and PSI had high sensitivities for the combined outcome of interest chosen to best represent this patient group.

We believe that pneumonia severity scores are likely to be useful for less experienced doctors in order to alert them to a high risk group of patients for whom consultation with more experienced clinicians is required. For this purpose a tool needs a high sensitivity and a good negative predictive value. Patients who are not identified as ''severe'' by the severity score are unlikely to die or to require ICU interventions. The decision about whether or not to admit to hospital patients in this lower risk group is likely then to be influenced by criteria such as social factors, age, general frailty, co-morbidities, etc. A significant proportion of patients who fall into low risk categories of severity scores still do require inpatient care because of factors not assessed by the severity score.[21] [22] The low positive predictive value of these severity scores highlights the need for clinical judgement in guiding the management of those identified in the ''severe pneumonia'' group. Not all patients in the high risk group will require ICU management, but they should receive careful initial and ongoing assessment. No prediction tool is accurate enough to determine appropriate management on its own, and these tools should always only be viewed as augmenting clinical judgement.

The data from this study can be compared with previous studies. The areas under the ROC curve for the CURB and PSI scores were similar to those calculated by Ewig et al[17] for the outcome of ICU admission (0.732 v 0.76 and 0.607 v 0.69). Similarly, the area under the ROC curve for the PSI was very close to that obtained by Aujesky et al[8] when assessing the outcome of mortality (0.81 v 0.82), but the areas under the ROC curve for the CURB and CURB-65 were smaller in our study (0.74 v 0.82 and 0.76 v 0.82). Aujesky et al assessed 30 day mortality while we focused only on death in hospital, and this may account for some of the difference. A change from CURB to CURB-65 (as suggested recently by Macfarlane and Boldy[15]) improved specificity but at the expense of sensitivity for all outcomes with poorer discriminative value. Our data support the suggestion that CURB offers a simple valuable alternative to the PSI, as noted by Ewig et al.[17]

Tools derived to predict mortality are likely to be skewed by elderly patients and patients with complex co-morbidities. For many of these patients aggressive interventions may not be appropriate. For this reason, we chose to evaluate the performance of the tool both including and excluding very elderly patients, those from nursing homes, and those with advanced debilitating co-morbidities. The PSI and CURB-65 have scoring systems that are heavily influenced by patient age. This might explain why they perform better when death is the outcome of interest rather than ICU admission, given the clear difference in age between patients who died and patients admitted to ICU in this cohort. Similarly, in this cohort, most patients requiring admission to the ICU had respiratory or circulatory failure, thus satisfying a major criterion of the rATS tool. However, this tool performed poorly for predicting death in this cohort since most patients who died were not admitted to the ICU.

Our analysis of patients admitted to the ward and then transferred as emergencies to the ICU within 24 hours was intended to identify patients whose severity of illness was possibly underestimated initially. The severity scores correctly identified the majority of these patients as ''severe''. Similarly, in the analysis of patients who died without being admitted to the ICU (excluding the very elderly, nursing home patients, and those identified as not being suitable for aggressive treatment), it is conceivable that the severity of their illness may have been underestimated. These cases illustrate situations in which a severity score may have predicted a poor outcome and identified the need for intensive measures.

A major strength of this study was that patients were assessed using only the data readily available at the point of care when usual management decisions are being made. In some patients the diagnosis of pneumonia was later excluded on the basis of further investigation results, but it is important to include this group in the evaluation of the prediction tools as it best reflects the context in which they will be used. We must be sure that the tools are safe to use in the face of diagnostic uncertainty that accompanies early patient assessments (similar to intention to treat analysis of drug trials). The precision of discharge diagnoses of pneumonia can be poor,[23] [24] hence we have chosen to focus on initial clinical diagnosis to validate these tools. Our data therefore differ from previous studies that used the discharge or final diagnosis to define inclusion. The clinical data collected reflected the most abnormal result in the first 24 hours, which differs from other studies that have used the most abnormal result in the first 48 hours or the first result recorded (at triage).[5–7] We believe the strategy we employed best reflects the decision making process of clinicians at this institution.

In summary, consideration of alternative outcomes to define ''severe pneumonia'' is important when evaluating severity scores for CAP. In addition, validation studies should reflect as closely as possible the context in which the tool is likely to be used. They should use only the data likely to be readily and practically available in real time, avoid retrospective exclusion of cases due to recruitment based on discharge diagnosis, and apply the tool only to patients for whom it would be likely to be used in routine practice. Different severity scores for CAP have different strengths and weaknesses which need to be recognised. The PSI (classes

IV+V), CURB, and modified BTS severity score provide comparable information with regard to identifying high risk patients for whom more aggressive management strategies may be required.

. . . . . . . . . . . . . . . . . . . . .

## Authors' affiliations

**K L Buising, K A Thursky, J F Black, L MacGregor, A C Street, G V Brown,** Victorian Infectious Diseases Service, The Royal Melbourne Hospital, Parkville, Victoria 3050, Australia
**K A Thursky, G V Brown,** Centre for Clinical Research Excellence in Infectious Diseases, Department of Medicine, University of Melbourne, Parkville, Victoria 3050, Australia
**M P Kennedy,** Emergency Department, The Royal Melbourne Hospital, Parkville, Victoria 3050, Australia

## REFERENCES

1 **Lim WS**, Macfarlane JT. Importance of severity of illness assessment in management of lower respiratory infections. *Curr Opin Infect Dis* 2004;**17**:121–5.
2 **British Thoracic Society Standards of Care Committee**. British Thoracic Society guidelines for the management of community acquired pneumonia in adults. *Thorax* 2001;**56**(Suppl IV):iv1–64.
3 **American Thoracic Society**. Guidelines for the management of adults with community acquired pneumonia. *Am J Respir Crit Care Med* 2001;**163**:1730–54.
4 **Johnson PD**, Irving L, Turnidge J. Community acquired pneumonia. *Med J Aust* 2002;**176**:341–7.
5 **Fine MJ**, Auble TE, Yealy DM, *et al*. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997;**336**:243–50.
6 **Neill AM**, Martin IR, Weir R, *et al*. Community acquired pneumonia: aetiology and usefulness of severity criteria on admission. *Thorax* 1996;**51**:1010–6.
7 **Lim WS**, van der Eerden MM, Laing R, *et al*. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003;**58**:377–82.
8 **Aujesky D**, Auble TE, Yealy DM, *et al*. Prospective comparison of three validated prediction rules for prognosis in community-acquired pneumonia. *Am J Med* 2005;**118**:384–92.
9 **Atlas SJ**, Benzer TI, Borowsky LH, *et al*. Safely increasing the proportion of patients with community acquired pneumonia treated as outpatients: and interventional trial. *Arch Intern Med* 1998;**158**:1350–6.
10 **Mandell LA**, Bartlett JG, Dowell SF, *et al*. Update of practice guidelines for the management of community acquired pneumonia in immunocompetent adults. *Clin Infect Dis* 2003;**37**:243–50.
11 **British Thoracic Society and Public Health Laboratory Service**. Community-acquired pneumonia in adults in British hospitals in 1982–1983: a survey of aetiology, mortality, prognostic factors and outcome. *Q J Med* 1987;**62**:195–220.
12 **Lim WS**, Macfarlane JT, Boswell TC, *et al*. Study of community acquired pneumonia aetiology (SCAPA) in adults admitted to hospital: implications for management guidelines. *Thorax* 2001;**56**:296–301.
13 **Lim WS**, Lewis S, Macfarlane JT. Severity prediction rules in community acquired pneumonia: a validation study. *Thorax* 2000;**55**:219–23.
14 **Lim WS**, Macfarlane JT. Defining prognostic factors in the elderly with community acquired pneumonia: a case controlled study of patients aged > or = 75 yrs. *Eur Respir J* 2001;**17**:200–5.
15 **Macfarlane JT**, Boldy D. 2004 update of BTS pneumonia guidelines: what's new? *Thorax* 2004;**59**:364–6.
16 **Ewig S**, Ruiz M, Mensa J, *et al*. Severe community-acquired pneumonia. Assessment of severity criteria. *Am J Respir Crit Care Med* 1998;**158**:1102–8.
17 **Ewig S**, de Roux A, Bauer T, *et al*. Validation of predictive rules and indices of severity for community acquired pneumonia. *Thorax* 2004;**59**:421–7.
18 **Angus DC**, Marrie TJ, Obrosky DS, *et al*. Severe community-acquired pneumonia: use of intensive care services and evaluation of American and British Thoracic Society diagnostic criteria. *Am J Respir Crit Care Med* 2002;**166**:717–23.
19 **Riley PD**, Aronsky D, Dean NC. Validation of the 2001 American Thoracic Society criteria for severe community-acquired pneumonia. *Crit Care Med* 2004;**32**:2398–402.
20 **Therapeutic Guidelines Writing Group**. *Therapeutic guidelines antibiotic*. Version 12. Melbourne, Australia: Therapeutic Guidelines, 2003.
21 **Arnold FW**, Ramirez JA, McDonald LC, *et al*. Hospitalisation for community acquired pneumonia: pneumonia severity index vs clinical judgement. *Chest* 2003;**124**:121–4.
22 **Goss GH**, Rubenfeld GD, Park DR, *et al*. Cost and incidence of social comorbidities in low risk patients with community acquired pneumonia admitted to a public hospital. *Chest* 2003;**126**:2148–55.
23 **Aronsky D**, Haug PJ, Lagor C, *et al*. Accuracy of administrative data for identifying patients with pneumonia. *Am J Med Qual* 2005;**20**:319–28.
24 **Goodpasture H**, Nguyen-Dang CT, Lee TH, *et al*. Miscoding as a cause of elevated simple pneumonia mortality. *Jt Comm J Qual Saf* 2004;**30**:335–41.