The CURB65 pneumonia severity score outperforms generic sepsis and early warning scores in predicting mortality in community-acquired pneumonia

Gavin Barlow, Dilip Nathwani, Peter Davey

Background: The performance of CURB65 in predicting mortality in community-acquired pneumonia (CAP) has been tested in two large observational studies. However, it has not been tested against generic sepsis and early warning scores, which are increasingly being advocated for identification of high-risk patients in acute medical wards.

Method: A retrospective analysis was performed of data prospectively collected for a CAP quality improvement study. The ability to stratify mortality and performance characteristics (sensitivity, specificity, positive predictive value, negative predictive value and area under the receiver operating curve) were calculated for stratifications of CURB65, CRB65, the systemic inflammatory response syndrome (SIRS) criteria and the standardised early warning score (SEWS).

Results: 419 patients were included in the main analysis with a median age of 74 years (men = 47%). CURB65 and CRB65 stratified mortality in a more clinically useful way and had more favourable operating characteristics than SIRS or SEWS; for example, mortality in low-risk patients was 2% when defined by CURB65, but 9% when defined by SEWS and 11–17% when defined by variations of the SIRS criteria. The sensitivity, specificity, positive predictive value and negative predictive value of CURB65 was 71%, 69%, 35% and 91%, respectively, compared with 62%, 73%, 35% and 89% for the best performing version of SIRS and 52%, 67%, 27% and 86% for SEWS. CURB65 had the greatest area under the receiver operating curve (0.78 vs 0.73 for CRB65, 0.68 for SIRS and 0.64 for SEWS).

Conclusions: CURB65 should not be supplanted by SIRS or SEWS for initial prognostic assessment in CAP. Further research to identify better generic prognostic tools is required.
that SIRS and SEWS would perform at least as well, or better, than CURB65 and CRB65 in predicting mortality in CAP.

METHODS
A retrospective analysis of prospectively collected data was performed. The data used were collected as part of a controlled before-and-after study over two winter periods (November to April 2001/02 and 2002/03) to evaluate the implementation of a quality improvement programme to improve the delivery and appropriateness of prescribing antibiotics for patients hospitalised with CAP.23 Potential subjects were identified by a review of admission records from two hospitals, one a 1000-bed medical/nursing records (ie, on admission to hospital). Patients were excluded if they had one or more of the following criteria: (1): a non-pneumonia diagnosis; (2) aspiration, hypostatic or hospital-acquired pneumonia; (3) the initial diagnosis of CAP was changed before discharge from the hospital; (4) the patient was HIV-positive, neutropenic (<1.0 x10⁹/l) secondary to chronic illness or treatment, or markedly immunosuppressed (long term (>2 weeks) prednisolone (or equivalent) of >10 mg or immunosuppressive therapy such as methotrexate, azathioprine, mycophenolate, etc); (5) progressive malignancy; (6) the patient had chronic respiratory disease other than asthma or chronic obstructive pulmonary disease; (7) age <16 years. Demographic, clinical and outcomes data were collected using a pre-piloted data collection form. The criteria used to establish the CURB65, CRB65, SIRS and SEWS scores were taken from the earliest recorded reading/result in the patients medical/nursing records (ie, on admission to hospital). Patients were reviewed on alternate days until discharge from the hospital or death. Deaths after discharge, but within 30 days of admission to hospital were established by the hospital’s computer database. Data were subsequently audited and double entered into an Epi-Info database (Centers for Disease Control, Atlanta and World Health Organisation, Geneva). Statistical analyses were performed using SPSS V.10. Descriptive statistics are given as medians or percentages with 95% confidence intervals (CI) where appropriate. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and accuracy were calculated for stratifications of the four tools. A receiver operator curve (ROC) was produced for each tool. The area under the curve (AUC) for each of these associations and standard errors (SE) and 95% CI were also calculated. Table 2 shows the definitions of the above performance characteristics.24 25

Table 1 shows the definitions of each of the four tools studied.

Table 1  Definitions of each of the four prognostic tools studied

<table>
<thead>
<tr>
<th>Severity grade</th>
<th>Score/definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CURB65 score</td>
<td></td>
</tr>
</tbody>
</table>
Based on the presence or absence of the following criteria: new confusion, urea >7 mmol/l, respiratory rate >30/min, systolic blood pressure <90 mm Hg or diastolic blood pressure <60 mm Hg and age >65 years
Severe ≥3
Non-severe, moderate risk 2
Non-severe, low risk 0 or 1
| CRB65 score |
Based on the presence or absence of the following criteria: new confusion, respiratory rate >30/min, systolic blood pressure <90 mm Hg or diastolic blood pressure <60 mm Hg and age >65 years
Severe 3 or 4
Non-severe, moderate risk 1 or 2
Non-severe, low risk 0
| SIRS |
The diagnosis of SIRS is based on the presence of at least two of the following criteria: temperature <36˚C or >38˚C, pulse >90/min, respiratory rate >20/min and white cell count <4 or >12 cells/mm³
Severe sepsis (severe) Severe sepsis plus hypotension and/or organ hypo-perfusion and the patient’s response to adequate fluid resuscitation
Severe sepsis (severe) Severe sepsis plus hypotension and/or organ hypo-perfusion, which has failed to respond to adequate fluid resuscitation
SIRS (non-severe, intermediate risk) ≥2 of the SIRS criteria above
No SIRS (non-severe, low risk) <2 of the SIRS criteria above
| SEWS |
Inform nurse in charge and aim for doctor (at least senior) ≥4
House officer review within 30 min (severe) ≥2
Intermediate risk
Intermediate risk
Intermediate risk
Intermediate risk
Routine (4 hourly) observations (low risk) 0 or 1

SaO₂, oxygen saturation; SEWS, standardised early warning system; SIRS, systemic inflammatory response syndrome.
The CURB65 pneumonia severity score was the only tool that identified a genuinely low risk group of patients (2% in CURB65 = 0 or 1 patients vs 9% in CURB65 = 0 patients vs 9% in the SEWS = 0 or 1 patients and 11% in patients without SIRS or hypotension or hypoperfusion). Figure 1 shows the ROCs. The ROC for CURB65 had the greatest AUC (0.78, SE 0.025, 95% CI 0.73 to 0.83) followed by CRB65 (0.73, SE 0.029, 95% CI 0.67 to 0.79), SIRS (0.68, SE 0.035, 95% CI 0.61 to 0.75) and SEWS (0.64, SE 0.035, 95% CI 0.57 to 0.70). The overall accuracy of the four tools was 70% for CURB65, 79.5% for CRB65 (62% for a cut-off of 2 for severe CAP; see later discussion), 71% for SIRS and 64% for SEWS.

Sub-group analyses were performed on patients who had a chest radiograph reported by a consultant radiologist or seen by a consultant respiratory physician with associated documentation in the patient’s case notes (n = 218). The characteristics of this cohort are shown in Appendix B, available online at http://thorax.bmj.com/supplemental. The ROC for each tool is shown in Appendix D, available online at http://thorax.bmj.com/supplemental. As for the main analyses, CURB65 and CRB65 were the only tools to identify a low risk cohort of patients. In contrast to the main analyses, SIRS (as...
defined above) performed better than CURB65 with respect to sensitivity, specificity, PPV, NPV and accuracy. The ROC for CURB65 still had the greatest AUC (0.79, SE 0.037, 95% CI 0.72 to 0.86), however, followed by CRB65 (0.75, SE 0.043, 95% CI 0.67 to 0.85), SIRS (0.70, SE 0.057, 95% CI 0.59 to 0.81) and SEWS (0.61, SE 0.059, 95% CI 0.49 to 0.72). The overall accuracy of the four tools in this new cohort was 69% for CURB65, 86% for CRB65 (62% for a cut-off of >2 for severe CAP; see later discussion), 76% for SIRS and 61% for SEWS. Table 5 compares the results of this study with two previously reported validation studies.

DISCUSSION
Severity assessment is the key to appropriately managing patients with CAP. The results of this study show that two potential generic prognostic tools, SIRS and SEWS, should not be used in preference to CURB65 or CRB65 for predicting mortality in adult patients who present to hospital with CAP. CURB65 and CRB65 outperform both of these tools in two ways. Firstly, and most importantly, their stratification of mortality is more clinically useful and identifies a genuinely low risk group of patients, whereas SIRS and SEWS do not. This means that CURB65 and CRB65 can be used to identify patients who do not require inpatient care unless they have additional comorbidities or signs of respiratory failure. Secondly, they performed better with regard to most of the other performance criteria (except when compared with our modified definition of SIRS in the chest radiograph defined cohort) and had the greatest AUC in all analyses. It is worth noting, however, that none of the tools performed particularly well and all were well below the standard required of population screening tests. This emphasises the importance of combining predictive tools with clinical judgement.

The ease of using each tool in clinical practice should also be considered. CURB65 requires four bedside and one laboratory criteria.
The definition of sepsis, which incorporates SIRS, was published in a consensus statement in 1992 and has since been widely adopted in research and clinical practice. In our own hospitals, for example, the definitions given in Table 1 have been included in sepsis protocols to guide the intensity of antibiotic treatment. A North American study showed that mortality due to infection increased with the number of SIRS criteria (7% with two criteria, 10% with three criteria and 17% with four criteria) and with severe sepsis (20%) and septic shock (40%).

Jones and Lowes also found a similar relationship in patients with bacteraemia (mortality in patients with no SIRS = 12%, SIRS 2 = 14%, SIRS 3 = 26%, SIRS 4 = 36%, severe sepsis = 38% and septic shock 56%) and it was suggested, on the basis of these studies, that SIRS was “of generalised use in predicting outcome from infection.” Interestingly, as with our study, Jones and Lowes also found a cohort of patients with clinical evidence of hypotension and/or hypoperfusion, but without the classical definition of SIRS. The mortality in this cohort of patients was 29% vs 28% in our study, which explains the higher mortality (17% when these patients were included versus 11% when they were classified as a separate cohort) for patients without SIRS (ie, infection only patients).

Since then, the value of the SIRS criteria and the relationship between an increasing number of SIRS criteria and infection has been questioned. In a study of 300 internal medicine patients with a new onset of fever at a university teaching hospital in The Netherlands, Bossink et al found that although there was a statistically significant association between the number of positive SIRS criteria and mortality (SIRS 1 = 0%, SIRS 2 = 3%, SIRS 3 = 8%, SIRS 4 = 17%), the performance of the definition of sepsis for predicting mortality was not as good as an alternative model proposed in the paper. In our study, sepsis had a sensitivity of 73%, specificity of 30%, PPV of 20% and NPV of 83% for predicting mortality when compared with 63%, 60%, 13% and 94%, respectively, in the study by Bossink et al. A recent, multicentred study using data from 3608 ICU patients who had taken part in the European Sepsis Study found a gradation in mortality from uncomplicated infection or sepsis (25%) to severe sepsis (40%) to septic shock (60%).

We found a similar association depending on how the sepsis definitions were used: from 13% (infection and SIRS) to 38% (severe sepsis/septic shock) with the classical definition and from 11% (infection and SIRS) to 28% (hypotension or hypoperfusion without SIRS) to 38% (severe sepsis/septic shock) with our alternative definition. As with our study, they did not find any difference in mortality between patients with infection without SIRS and sepsis or an association between the number of SIRS criteria and mortality in these groups. The SIRS criteria may also have performed less well in our study because...
two of the criteria, heart rate and white cell count, have not been strongly associated with outcome in CAP.\(^3\)\(^4\)

In contrast to SIRS, there are less data supporting the use of SEWS in infection. This is a concern, given the increasing rate at which it is being implemented in acute medical admissions units in the UK. Indeed, implementation is being encouraged by major organisations interested in clinical effectiveness, such as NHS Quality Improvement Scotland.\(^1\)\(^9\) SEWS is based on an EWS, which was validated for use in acute medical patients in 2001.\(^1\)\(^6\) The AUC in the validation study was 0.67 compared with 0.62 in our study (versus 0.78 for CURB65 and 0.68 for SIRS). A subsequent study of 1695 acute medical patients, who were compared with a cohort of patients admitted to the same unit in the previous year, but before implementation of the EWS, did not show any change in mortality as a result of implementation.\(^2\) A recent study, again from the UK, of 1047 ward patients assessed by an intensive care outreach service, found a strong statistical association between the EWS and the need for intervention or mortality.\(^3\)\(^2\) As with SIRS, the EWS has been tested in different cohorts of patients in different contexts and it is debatable as to whether this evidence can be generalised to all patient populations. In our study, SEWS was better than SIRS at stratifying mortality. It is possible therefore, that it could still be used to identify patients at high risk of needing critical care, once the initial decisions about an appropriate site of care and antibiotic treatment have been made. Overall, however, SEWS performed less well than CURB65, CRB65 and SIRS with regard to other operating characteristics.

There are a number of caveats to the interpretation of the results of our study. Patients were included using a pragmatic, real-life definition of CAP. Sub-group analysis of a chest radiograph defined cohort of patients, however, confirmed the findings of our main analyses. Because the performance of all tests is context dependent, one may not be able to extrapolate our results to healthcare systems dissimilar to the NHS. We also used a limited definition of hypoperfusion to define severe sepsis and septic shock. We feel that this is justified given that the inclusion of acidosis in clinical practice would require an additional blood test, which is not performed in all patients with CAP. Indeed, the BTS guidelines recommend arterial blood gas measurement only when the patient’s oxygen saturation is <92% or other features of severe pneumonia are present.\(^9\) Additionally, acidosis probably affects only a relatively small number of the most severely ill patients. As urine output cannot be measured accurately on admission to hospital and would therefore delay the assessment of severity and reduce the practicality of the tool, oliguria was also excluded as a criterion of hypoperfusion and was not scored in SEWS. When using SEWS, it is recommended that \(\geq 3\) h of urine output be assessed. Given that there are six other SEWS criteria, and that oliguria would be an unusual isolated finding in severe CAP, it is unlikely that the omission of this would have resulted in poorer SEWS performance. Also, oliguria was not included in the validation study by Subbe et al.\(^7\)\(^9\) Nevertheless, it is possible that the omission of these criteria, in particular for SIRS, may have changed performance characteristics.

In summary, CURB65 and CRB65 were better at stratifying mortality and outperformed SIRS and SEWS in predicting 30-day mortality in CAP. For the time being, other prognostic tools should not supplant CURB65 in the initial assessment of patients with CAP. There is clearly a need for corroboration of our results and the development of better generic predictive tools for use in acute medicine and sepsis.

CONTRIBUTORS
GB had the initial study idea, collected and analysed the data and wrote the initial draft of the paper. PD and DN were involved in developing the initial idea and edited the initial and subsequent drafts of the paper. PD is the guarantor.

Authors’ affiliations
Gavin Barlow, Castle Hill Hospital, Hull and East Yorkshire Hospitals NHS Trust, Cottingham, East Yorkshire, UK
Dilip Nathwani, Ninewells Hospital and Medical School, Tayside University Hospitals NHS Trust, Dundee, UK
Peter Davey, Health Informatics Centre, University of Dundee, Dundee, UK

Funding: The original quality improvement project was funded by NHS Education Scotland and The Chief Scientist Office, Scotland.

Competing interests: None.

Ethical approval: Collection of data was approved by both Tayside Universities NHS Trust’s medical ethics committee and Caldecott guardian.

REFERENCES
21. Subbe CP, Davies RG, Williams E, et al. Effect of introducing the Modified Early Warning Score on clinical outcomes, cardio-pulmonary arrests and...


**LUNG ALERT**

**New combination therapy improves survival in non-small cell lung cancer**


T he survival rate of patients with metastatic non-small-cell lung cancer remains poor even with chemotherapy. This randomised-controlled trial looks at the effect of monoclonal antibody against vascular endothelial growth factor (bevacizumab) in the treatment of patients with metastatic non-squamous-cell, non-small-cell lung cancer.

Eight hundred and seventy-eight patients with recurrent or advanced non-small-cell lung cancer (stage IIIB or IV) were randomised to receive chemotherapy with paclitaxel and carboplatin alone (444) or paclitaxel–carboplatin and bevacizumab (434). Patients with squamous-cell cancers were excluded from the study because a phase 2 trial had shown serious haemorrhagic events in this sub-group. The primary endpoint of the study was overall survival. The authors report an improvement in median survival by 2 months on addition of bevacizumab to chemotherapy (12.3 months vs 10.3 months). The median progression-free survival in the two groups was 6.2 and 4.5 months, respectively. Surprisingly, the addition of bevacizumab also improved the response to chemotherapy, probably by improving drug delivery to the tumour. The baseline vascular endothelial growth factor (VEGF) levels did not correlate with overall survival.

The addition of bevacizumab resulted in increased treatment related deaths (15 vs 2; p = 0.001), mainly due to haemorrhage and neutropenia.

This study does show significant survival benefits of addition of bevacizumab to a chemotherapeutic regimen, for treatment of metastatic non-squamous-cell, non-small-cell lung cancer in a select group of patients, at the risk of increased treatment-related deaths. The exact mechanism by which the survival is improved is not clear as the VEGF levels did not correlate with overall survival.

Vinod Aiyappan

Specialist Registrar (LAT) in Respiratory Medicine, Queen Elizabeth Hospital, Kings Lynn, Norfolk, UK;

drivinodaiyappan@doctors.org.uk

The CURB65 pneumonia severity score

259


