PNEUMONIA

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

....

Correspondence to:
Dr G Barlow, Department of Infection and Tropical Medicine, Castle Hill Hospital, Hull and East Yorkshire Hospitals NHS Trust, Cottingham, East Yorkshire HU16 5JQ, UK; gavin.barlow@hey.nhs.uk

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 v 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.

Community-acquired pneumonia (CAP) is an important quality improvement target in acute medicine.[1] Recent, major national and specialist society CAP guidelines suggest the use of prognostic (severity) assessment in guiding clinical decisions about the level of intervention required.[2–5] The British Thoracic Society (BTS)[2], Infectious Diseases Society of America[3] and the Canadian Thoracic Society[4] guidelines recommend the use of validated prognostic tools[6 7] on admission to hospital as adjuncts to clinical judgement in guiding the management of patients with CAP. The pneumonia severity index (PSI) has been used to identify low-risk patients who can be managed equally effectively at home or as inpatients.[8–10]

In the United Kingdom, the BTS guidelines promote the use of CURB65,[7] which is based on four bedside and one laboratory based prognostic marker (table 1). This tool, which is an evolution of two previously validated prognostic rules,[11 12] was shown to have 75% sensitivity and 75% specificity for predicting death at 30 days in CAP in the validation set of a large prospective multicentre, multinational derivation/validation study.[7] CRB65, which does not require a blood urea level, was shown to stratify mortality similarly, but had inferior performance characteristics. There is evidence that junior doctors, however, have poor awareness of the BTS recommendations. In a survey of 83 junior and middle grade doctors, only 4% could correctly state all four prognostic markers of the BTS CURB tool.[13] Woodhead suggested that the use of a generic rather than a pneumonia-specific predictive tool might be easier for doctors to remember and use in the clinical management of patients.[14]

Ewig *et al*[15] have previously compared the performance of CURB, the precursor of CURB65, with that of the PSI, the modified American Thoracic Society (ATS) rule for predicting the need for intensive care unit (ICU) admission, and the American College of Chest Physicians-Society of Critical Care Medicine's definition of sepsis. The ATS rule performed the best with the PSI and CURB having similar performance characteristics. More recently, when compared with the PSI, CURB65 was shown to have equivalent performance.[16] CURB65 has not been compared, however, with generic sepsis or early warning scores. The study reported in this paper compared the performance of CURB65 and CRB65 in predicting death with that of two commonly used generic scores, the systemic inflammatory response syndrome (SIRS) and the standardised early warning score (SEWS) (table 1). SIRS is recognised worldwide as a component of the definition of sepsis.[17] It is often used to define and stratify sepsis in research,[18] and has been incorporated into our and other hospitals' sepsis guidelines. SEWS[19] is a modified version of an early warning score (EWS),[20] which has increasingly been advocated for use in the acute medical environment to guide the intensity of nursing observation and medical management.[20–22] Our hypothesis was

....

Table 1 Definitions of each of the four prognostic tools studied

Severity grade	Score/definition
CURB65 score	
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/mm^3. The diagnosis of severe sepsis and septic shock is based on the presence of SIRS in infection with the presence or absence of hypotension and/or organ hypo-perfusion and the patient's response to adequate fluid resuscitation	
Septic shock (severe)	Severe sepsis plus hypotension and/or organ hypo-perfusion, which has failed to respond to adequate fluid resuscitation
Severe sepsis (severe)	SIRS plus hypotension and/or organ hypo-perfusion
SIRS (non-severe, intermediate risk)	≥2 of the SIRS criteria above
No SIRS (non-severe, low risk)	<2 of the SIRS criteria above
SEWS	
A complex scoring system based on the patient's respiratory rate, SaO$_2$, temperature, blood pressure, heart rate, neurological response and urine output. The SEWS chart is available at: http://www.nhshealthquality.org/nhsqis/files/Supplement%20to%20Final%20Report%20(SEWS%20Chart).pdf	
Inform nurse in charge and aim for doctor (at least senior house officer) review within 30 min (severe)	≥4
Hourly observations and inform nurse in charge (non-severe, intermediate risk)	2 or 3
Routine (4 hourly) observations (low risk)	0 or 1

SaO$_2$, oxygen saturation; SEWS, standardised early warning system; SIRS, systemic inflammatory response syndrome.

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 teaching hospital and the other a 500-bed district general hospital. Patients were included if they were receiving antibiotics for a suspected lower respiratory tract infection and had either a new infiltrate on the chest radiograph or had been clinically diagnosed as having CAP by a specialist registrar or consultant doctor. 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 (<1.0×10^9/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, mycophenalate, 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 and associated 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. Severity was defined according to how one would expect to use the tools in clinical practice. For CURB65, therefore, severe CAP was defined as a score of ≥3.[7] For CRB65, severe CAP was defined as a score of 3 or 4 as suggested by the findings of Lim *et al*.[7] For SEWS, severe CAP was defined as a score of ≥4 as it is at this level of score that the SEWS chart recommends early intervention by a doctor.[19] We did not include a urine output score in the calculation of SEWS. SIRS was analysed in four different ways to establish the best performing variation of this tool and the most appropriate cut-offs to define severe CAP. For SIRS, hypotension was defined as a systolic blood pressure <90 mm Hg. Organ hypoperfusion was defined as new confusion (MSQ ≤8/10 or a 2 point drop in MSQ). We did not attempt to separate patients with severe sepsis from those with septic shock as, by definition, septic shock cannot be diagnosed on admission to hospital until the patient has received adequate intravenous fluid resuscitation. We have

Table 2 Definitions of performance characteristics

Performance characteristics	Definition equation
Sensitivity	How good is the test at identifying patients who die $$\frac{\text{True positives (ie, tested positive and died)}}{\text{All deaths}}$$
Specificity	How good is the test at identifying patients who do not die $$\frac{\text{True negatives (ie, tested negative and lived)}}{\text{All alive}}$$
PPV	In the event of a positive test, the probability that the patient will die $$\frac{\text{True positives}}{\text{All positives}}$$
NPV	In the event of a negative test, the probability that the patient will not die $$\frac{\text{True negatives}}{\text{All negatives}}$$
Accuracy	The proportion of all tests that have given a correct result $$\frac{\text{True positives and negatives}}{\text{All positives and negatives}}$$
ROC	A curve created by mapping sensitivity against $1 - $ specificity. The AUC is a marker of performance with higher values indicating better diagnostic ability. An AUC = 1 indicates perfect performance whereas an AUC = 0.5 indicates that there is a 50:50 chance that the test correctly identifies those who die

Comment
Sensitivity and specificity are useful when considering the performance of a test at the population level. In contrast, PPV, NPV and accuracy are better when considering the performance of a test at the patient level (ie, in the event of a positive test result, what is the probability that the patient in front of me will or will not die?). It is also important to consider how easy a test is to apply in clinical practice (see discussion).

AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; ROC, receiver operating curve.

Table 3 Demographic and clinical characteristics of the patients included

Numbers included	
Number included in the quality improvement project (Appendix A available online at http://thorax.bmj.com/supplemental gives data on full cohort)	503
Number of patients with full data for all four tools	419
Demographics	
Sex	47% male (n = 197)
Age	Median (range) = 74 (16–98) years
	Age >65 years = 70% (n = 292)
Living in own home	84% (n = 351)
Living alone	36% (n = 152)
Clinical characteristics	
Antibiotic from GP before admission	42% (n = 149/356)
CRP >50	76.5% (n = 286/374)
Chest radiograph consistent with pneumonia	95% (218/230)
Severity assessment on admission	
CURB65 score	
0 or 1	33.5% (n = 140)
2	28.5% (n = 119)
≥3	38% (n = 160)
Respiratory rate ≥30/min	22% (n = 91)
Systolic blood pressure <90 mm Hg	6% (n = 24)
Diastolic blood pressure ≤60 mm Hg	25% (n = 103)
Pulse ≥125/min	10% (n = 43)
Pulse oximetry <92% (any FiO$_2$)	33% (n = 140)
Blood urea >7 mmol/l	58% (n = 242)
New confusion	31% (n = 129)
% without comorbidity	35% (n = 147)
% with asthma/COPD	35% (n = 146)
Initial antibiotic regimen	
Broad spectrum betalactam + macrolide	48% (n = 201)
Narrow spectrum betalactam + macrolide	31% (n = 130)
Betalactam monotherapy	10% (n = 42)
Macrolide monotherapy	3% (n = 13)
Levofloxacin	2% (n = 8)
Others	6% (n = 25)
Outcomes	
% of patients transferred to the intensive care unit	3% (n = 13)
30-day post admission mortality	19% (n = 79)
Length of hospital stay (excludes deaths)	Median (range) = 5 (0–116) days

COPD, chronic obstructive pulmonary disease; CRP, C reactive protein; FiO$_2$, fraction of inspired oxygen; GP, general practitioner.

therefore assumed that patients who went on to be diagnosed with septic shock shortly after admission are embedded in the cohort of patients with severe sepsis.

RESULTS
Of the 503 patients (a description of the whole cohort is provided in Appendix A available online at http://thorax.bmj.com/supplemental) included in the quality improvement study, full data for all four tools were available for 419 (83%) patients. Table 3 shows the descriptive statistics for this cohort of patients. Reasons for exclusion from the original quality improvement study have previously been published.[23] Most deaths occurred within the first week of admission (72%) with 14% occurring in the second week and the remainder (14%) between 15 and 30 days. Table 4 shows the mortality for severity stratifications of the four tools and associated sensitivity, specificity, PPV and NPV. Based on the results of these analyses, for SIRS, we defined severe CAP as the presence of hypotension and/or organ hypoperfusion without SIRS or severe sepsis/septic shock (table 4).

CURB65 and CRB65 were the only tools that identified a genuinely low risk group of patients (2% in CURB65 = 0 or 1 patients v 0% in CRB65 = 0 patients v 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 operating characteristics of the four tools in this cohort of patients are shown in Appendix C, 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

Table 4 Operating characteristics of CURB65, CRB65, systemic inflammatory response syndrome criteria used in four different ways and standardised early warning score

Severity	Mortality	Sensitivity %	Specificity %	PPV %	NPV %
CURB65					
CURB65 = 0	0/58 (0%)	100	0	19	NC
CURB65 = 1	3/82 (4%)	100	17	22	100
CURB65 = 2	20/119 (17%)	96	40	27	98
CURB65 = 3	27/105 (26%)	71	69	35	91
CURB65 = 4	27/49 (55%)	37	92	53	86
CURB65 = 5	2/6 (33%)	2.5	99	33	81
CRB65					
CRB65 = 0	0/71 (0%)	100	0	19	NC
CRB65 = 1	21/150 (14%)	100	21	23	100
CRB65 = 2	28/131 (21%)	73	59	29	90.5
CRB65 = 3	28/61 (46%)	38	89	45	86
CRB65 = 4	2/6 (33%)	2.5	99	33	81
SIRS used in four different ways					
No SIRS*	21/124 (17%)	100	0	19	NC
SIRS	21/197 (11%)	73	30	20	83
Severe sepsis/septic shock	37/98 (38%)	47	82	38	87
SIRS = 0†	5/35 (14%)	100	0	19	NC
SIRS = 1	16/89 (18%)	94	9	19	86
SIRS = 2	28/136 (21%)	73	30	20	83
SIRS = 3	27/113 (24%)	38	62	19	81
SIRS = 4	3/46 (6%)	4	87	6	80
SIRS = 0‡	5/35 (14%)	100	0	19	NC
SIRS = 1	16/89 (18%)	94	9	19	86
SIRS = 2	12/89 (13%)	73	30	20	83
SIRS = 3	8/75 (11%)	58	53	22	84
SIRS = 4	1/33 (3%)	48	73	29	86
Severe sepsis/septic shock	37/98 (38%)	47	82	38	87
No SIRS or hypotension/organ hypo-perfusion§	9/81 (11%)	100	0	19	NC
SIRS	21/197 (11%)	89	21	21	89
Hypotension and/or organ hypo-perfusion, but no SIRS	12/43 (28%)	62	73	35	89
Severe sepsis/septic shock	37/98 (38%)	47	82	38	87
SEWS					
SEWS = 0	4/41 (10%)	100	0	19	NC
SEWS = 1	5/58 (9%)	95	11	20	90
SEWS = 2	17/87 (19.5%)	89	26	22	91
SEWS = 3	12/80 (15%)	67	47	23	86
SEWS = 4	10/57 (17.5%)	52	67	27	86
SEWS = 5	8/34 (23.5%)	39	81	32	85
SEWS ≥6	23/62 (37%)	29	88	38	84

NC, not calculable; NPV, negative predictive value; PPV, positive predictive value; SEWS, standardised early warning score; SIRS, systemic inflammatory response syndrome.
*Defined as in table 1.
†Defined by the presence or absence of: temperature <36°C or >38°C, pulse >90/min, respiratory rate >20/min and white cell count <4 or >12 cells/mm³.
‡Defined by above and definition of severe sepsis/septic shock given in table 1.
§New definition (see Results section).

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 0.83), 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.[8–10] 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. Although the laboratory criteria may delay a full
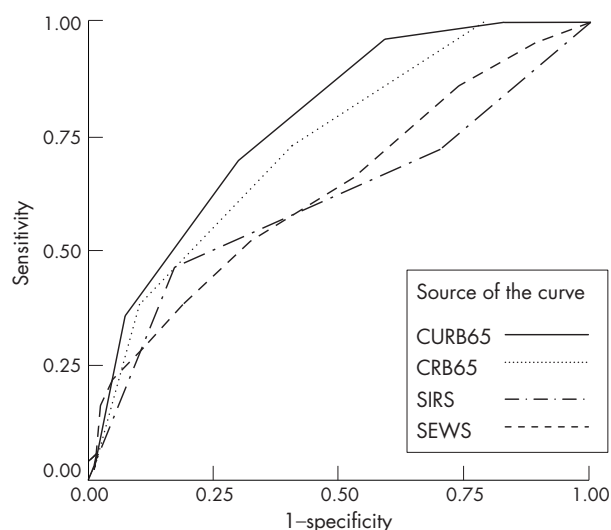
Figure 1 Receiver operating curve for each prognostic tool. SEWS, standardised early warning score; SIRS, systemic inflammatory response syndrome.

Table 5 Comparison of the performance characteristics of CURB65 in three different validation studies

CURB65 score	Study reported here n=419(%)	Study reported here, CXR cohort* n=218(%)	Lim et al[7] n=932(%)	Capelastegui et al[16] n=1100 inpatients, 676 outpatients(%)
0	0	0	0.6	0
1	3.7	2.1	1.7	1.1
2	16.8	14.8	9.0	7.6
3	25.7	19.6	16.1	21
4	55.1	37.5	36.9	41.9
5	33.3	50	20	60

CXR, chest x ray.
*Patients with a chest radiograph that had been reported by a consultant radiologist or seen by a consultant respiratory physician with associated documentation in the case-notes.

assessment on admission to hospital, Lim et al[7] have shown earlier that CRB65 stratifies mortality similarly and can therefore be used while awaiting the urea result. In contrast to the conclusions of Lim et al, the results of our study suggest that a CRB65 score of ⩾2 would be a safer cut-off for defining severe CAP. SIRS requires three bed-side and one laboratory criteria as well as an assessment of hypotension and hypoperfusion. Although SEWS requires the most data, all criteria can be measured at the bedside. It does require the availability of a pulse oximeter, however, and accurate measurement of urine output over a 3-h period (see discussion later).

This is now the third published study to assess the predictive performance of CURB65 in CAP.[7][16] CURB65 stratified mortality similarly across all three studies (table 5). More importantly, mortality in the two least severe stratifications (ie, 0 and 1) was similar (2%, 1.2% and 0.4%, respectively) thereby confirming that CURB65 can safely identify a low risk cohort of patients. The overall higher mortality seen in our study is likely to reflect the characteristics of the study cohort, in particular, the older median age (74 years v 69 years in the derivation/validation study and the greater proportion of patients with severe CAP; 38% v 29%[7]). The sensitivity, specificity, PPV and NPV were also similar. In our study these values were 71%, 69%, 35% and 91% when compared with 68%, 75%, 22% and 96% in the validation set and 75%, 75%, 23% and 97% in the derivation set of the validation/derivation study.[7]

CURB65 has recently been found to have similar performance to the PSI (AUC 0.87 v 0.89) in predicting a 30-day mortality in CAP,[16] but is yet to be compared with the modified ATS criteria.[26] Ewig et al[15] recently compared both of these with the old BTS tool (CURB) and found that the modified ATS criteria performed the best in predicting mortality and the need for admission to an ICU. The performances of CURB and the PSI were comparable. Interestingly, when severe sepsis was used as the cut-off for severity in this study, it had better sensitivity (89% v 51%), PPV (20% v 16%) and NPV (99% v 96%), but worse specificity (70% v 80%) and overall accuracy (71% v 78%) than CURB in predicting mortality. One major caveat of this study is that 17% of patients (versus 3% in our study) were admitted to an ICU, which means that one may not be able to generalise these results to the NHS (National Health Service) setting.

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.[17] 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%).[27] Jones and Lowes[28] 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% v 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[29] 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%).[30] 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.[7] [31]

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.[19] SEWS is based on an EWS, which was validated for use in acute medical patients in 2001.[20] 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.[21] 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.[32] 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.[2] 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 ⩾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.[20] 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

REFERENCES

1 **Jha AK,** Li Z, Orav EJ, et al. Care in US hospitals—the hospital quality alliance program. N Engl J Med 2005;**353**:265–74.
2 **Macfarlane JT,** Boswell T, Douglas G, et al. BTS guidelines for the management of community-acquired pneumonia in adults. Thorax 2001;**56**(Suppl 4):1–64.
3 **Bartlett JG,** Dowell SF, Mandell LA, et al. Practice guidelines for the management of community-acquired pneumonia in adults. Clin Infect Dis 2000;**31**:347–82.
4 **Mandell LA,** Marrie TJ, Grossman RF, et al. Canadian guidelines for the initial management of community-acquired pneumonia: an evidence-based update by the Canadian Infectious Diseases Society and the Canadian Thoracic Society. The Canadian Community-Acquired Pneumonia Working Group. Clin Infect Dis 2000;**31**:383–421.
5 **American Thoracic Society.** Guidelines for the management of adults with community-acquired pneumonia. Am J Respir Crit Care Med 2001;**163**:1730–54.
6 **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.
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 **Carratala J,** Fernandez-Sabe N, Ortega L, et al. Outpatient care compared with hospitalization for community-acquired pneumonia: a randomized trial in low-risk patients. Ann Intern Med 2005;**142**:165–72.
9 **Marrie TJ,** Lau CY, Wheeler SL, et al. A controlled trial of a critical pathway for treatment of community-acquired pneumonia. CAPITAL Study Investigators. Community-Acquired Pneumonia Intervention Trial Assessing Levofloxacin. JAMA 2000;**283**:749–55.
10 **Yealy DM,** Auble TE, Stone RA, et al. Effect of increasing the intensity of implementing pneumonia guidelines: a randomized, controlled trial. Ann Intern Med 2005;**143**:881–94.
11 **British Thoracic Society and the 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;**239**:195–220.
12 **Neill AM,** Martin IR, Weir R, et al. Community-acquired pneumonia: aetiology and usefulness of severity criteria on admission. Thorax 1996;**51**:1010–16.
13 **Barlow GD,** Nathwani D, Davey P. Improving the management of community-acquired pneumonia: barriers to high-quality care in a United Kingdom teaching hospital. Programme and Abstracts of 13th European Congress of Clinical Microbiology and Infectious Diseases. Glasgow, UK, May, 2003.
14 **Woodhead M.** Assessment of illness severity in community acquired pneumonia: a useful new prediction tool? Thorax 2003;**58**:371–2.
15 **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.
16 **Capelastegui A,** España PP, Quintana JM, et al. Validation of a predictive rule for the management of community-acquired pneumonia. Eur Respir J 2006;**27**:151–7.
17 **American College of Chest Physicians-Society of Critical Care Medicine Consensus Conference.** Definitions for sepsis and organ failure and guidelines for use of innovative therapies in sepsis. Crit Care Med 1992;**20**:864–75.
18 **Bernard GR,** Vincent J-L, Laterre P-R, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. N Engl J Med 2001;**344**:699–709.
19 **Emergency medical admissions scoping group final report.** NHS Quality Improvement Scotland, Edinburgh, UK, July 2004. SEWS chart available at http://www.nhshealthquality.org/nhsqis/files/Supplement%20to%20Final%20Report%20(SEWS%20Chart).pdf (accessed 21 December 2006).
20 **Subbe CP,** Kruger M, Rutherford P, et al. Validation of a modified Early Warning Score in medical admissions. Q J Med 2001;**94**:521–6.
21 **Subbe CP,** Davies RG, Williams E, et al. Effect of introducing the Modified Early Warning Score on clinical outcomes, cardio-pulmonary arrests and

intensive care utilisation in acute medical admissions. *Anaesthesia* 2003;**58**:797–802.

22 **McBride J**, Knight D, Piper J, *et al*. Long-term effect of introducing an early warning score on respiratory rate charting on general wards. *Resuscitation* 2005;**65**:41–4.

23 **Barlow G**, Nathwani D, Williams F, *et al*. Reducing door to antibiotic time in community-acquired pneumonia: controlled before and after evaluation and cost-effectiveness analysis. *Thorax* 2006;**62**:67–74.

24 **Greenhalgh T**. Papers that report diagnostic or screening tests. In: Greenhalgh T, eds. *How to read a paper. The basics of evidence based medicine*, 2nd edn. London, UK: BMJ Books, BMJ Publishing Group, 2001.

25 **Giesecke J**. Clinical epidemiology: sensitivity, specificity and misclassification. In: Giesecke T, eds. *Modern infectious disease epidemiology,* 2nd edn. London, UK: Arnold, Hodder Headline Group, 2002.

26 **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.

27 **Rangel-Frausto MS**, Pittet D, Costigan M, *et al*. The natural history of the systemic inflammatory response syndrome (SIRS). *JAMA* 1995;**273**:117–23.

28 **Jones GR**, Lowes JA. The systemic inflammatory response syndrome as a predictor of bacteraemia and outcome from sepsis. *Q J Med* 1996;**89**:515–22.

29 **Bossink AW**, Groeneveld J, Hack CE, *et al*. Prediction of mortality in febrile medical patients: How useful are systemic inflammatory response syndrome and sepsis criteria. *Chest* 1998;**113**:1533–41.

30 **Albert C**, Brun-Buisson C, Goodman SV et al. Influence of systemic inflammatory response syndrome and sepsis on outcomes of critically ill infected patients. *Am J Respir Crit Care Med* 2003;**168**:77–84.

31 **Lim WS**, Lewis S, Macfarlane JT. Severity prediction rules in community acquired pneumonia: a validation study. *Thorax* 2000;**55**:219–23.

32 **Goldhill DR**, McNarry AF, Mandersloot G, *et al*. A physiologically-based early warning score for ward patients: the association between score and outcome. *Anaesthesia* 2005;**60**:547–53.

LUNG ALERT ....

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

▲ Sandler A, Gray R, Perry MC, *et al*. Paclitaxel–carboplatin alone or with bevacizumab for non–small-cell lung cancer. *N Engl J Med* 2006;**355**:2542–50.

The 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; drvinodaiyappan@doctors.org.uk