



# Severity assessment tools for predicting mortality in hospitalised patients with community-acquired pneumonia. Systematic review and meta-analysis

James D Chalmers,<sup>1</sup> Aran Singanayagam,<sup>2</sup> Ahsan R Akram,<sup>2</sup> Pallavi Mandal,<sup>2</sup> Philip M Short,<sup>3</sup> Gourab Choudhury,<sup>2</sup> Victoria Wood,<sup>1</sup> Adam T Hill<sup>2</sup>

► Additional tables are published online only. To view these files please visit the journal online (<http://thorax.bmj.com>).

<sup>1</sup>University of Edinburgh, Edinburgh, UK

<sup>2</sup>Department of Respiratory Medicine, Royal Infirmary of Edinburgh, Edinburgh, UK

<sup>3</sup>University of Dundee, Tayside, UK

## Correspondence to

Dr James Chalmers, Department of Respiratory Medicine, Royal Infirmary of Edinburgh, 51 Little France Crescent, Old Dalkeith Road, Edinburgh EH16 4SA, UK; [jamesdchalmers@googlemail.com](mailto:jamesdchalmers@googlemail.com)

Received 13 December 2009

Accepted 28 May 2010

Published Online First

20 August 2010

## ABSTRACT

**Introduction** International guidelines recommend a severity-based approach to management in community-acquired pneumonia. CURB65, CRB65 and the Pneumonia Severity Index (PSI) are the most widely recommended severity scores. The aim of this study was to compare the performance characteristics of these scores for predicting mortality in community-acquired pneumonia.

**Methods** A systematic review and meta-analysis was conducted according to MOOSE (meta-analysis of observational studies in epidemiology) guidelines. PUBMED and EMBASE were searched (1980–2009). 40 studies reporting prognostic information for the PSI, CURB65 and CRB65 severity scores were identified. Performance characteristics were pooled using a random effects model. Relationships between sensitivity and specificity were plotted using summary receiver operator characteristic (sROC) curves.

**Results** All three scores predicted 30 day mortality. The PSI had the highest area under the sROC curve, 0.81 (SE 0.008), compared with CURB65, 0.80 (SE 0.008),  $p=0.1$ , and CRB65, 0.79 (0.01),  $p=0.09$ . These differences were not statistically significant. Performance characteristics were similar across comparable cut-offs for low, intermediate and high risk for each score. In identifying low risk patients, PSI (groups I and II) had the best negative likelihood ratio 0.08 (0.06–0.12) compared with CURB65 (score 0–1) 0.21 (0.15–0.30) and CRB65 (score 0), 0.15 (0.10–0.22).

**Conclusion** There were no significant differences in overall test performance between PSI, CURB65 and CRB65 for predicting mortality from community-acquired pneumonia.

## INTRODUCTION

Community-acquired pneumonia (CAP) is the leading cause of death from infectious disease in western countries and a major burden on healthcare resources.<sup>1</sup>

Although many hospitalised patients have an uncomplicated course, the inpatient mortality for CAP is reported to be between 5.7% and 14%, and CAP accounts for a significant proportion of intensive care unit (ICU) admissions.<sup>2–3</sup> There is evidence that, using clinical judgement, physicians may both overestimate and underestimate the severity of CAP, leading to inappropriate hospitalisation for mild cases that may be treated at home, or insufficiently aggressive interventions for patients at high risk of complications.<sup>4–5</sup>

Severity scores may overcome these difficulties by providing objective classification of patients into low, intermediate and high risk categories based on robust, validated markers of poor outcome.<sup>6</sup> The most extensively studied scoring system, the Pneumonia Severity Index (PSI), was introduced in 1997 following a study of >50 000 patients with CAP.<sup>7</sup> This 20-point score (details are included in the online supplement) classified patients into five risk categories (I–V) based on their percentage risk of death within 30 days. Patients in class I–II were recommended for outpatient therapy based on their low risk of death (0.1–0.7%).<sup>7</sup>

The PSI has been successfully applied in clinical practice to increase the use of outpatient treatment in CAP and is recommended by a number of national and international guidelines.<sup>1–8</sup> The PSI, however, has limitations. It is heavily weighted by age and co-morbid illnesses, and the large number of variables makes it complex to use in a busy emergency department.<sup>9</sup>

Recognising this, the British Thoracic Society (BTS) subsequently derived its own, more simple prediction tool (CURB), also based on the risk of 30 day mortality. In 2003 Lim and colleagues modified this score with the addition of age  $\geq 65$  years as a risk factor, to create CURB65 (details of each severity score are included in the online supplement).<sup>10</sup> This score is significantly easier to remember and use than the PSI, being composed of only five variables with a single point awarded for each. CRB65, without requirement to measure blood urea, is recommended for outpatient use and is widely used in Europe, including for hospitalised patients.<sup>10–11</sup>

CURB65 has been adopted internationally, and is now recommended by a large number of national and international guidelines alongside the PSI.<sup>1–2 12–14</sup>

Studies comparing these scoring systems have given conflicting results, with some suggesting the PSI may be superior,<sup>15–16</sup> while others have found no advantage to the PSI over CURB65.<sup>17–18</sup> Studies of CRB65 have suggested this more simple rule may be equivalent to CURB65 and PSI for prediction of 30 day mortality.<sup>17–20</sup>

This study aimed to systematically review the published literature in relation to these scoring systems and to determine if there is any difference in performance between the PSI, CURB65 and CRB65 rules in predicting mortality from CAP.

## METHODS

The present study was a systematic review and meta-analysis conducted according to MOOSE (meta-analysis of observational studies in epidemiology) guidelines.<sup>21</sup>

### Search criteria

This systematic review and meta-analysis was based on a search of PUBMED using search terms:

1. ('CURB65' OR 'CURB 65' OR 'CURB-65' OR 'C.U.R.B.65' OR 'C-U-R-B-65' OR 'CRB65' OR 'CRB 65' OR 'CRB-65' OR 'C.R.B.65' OR 'C-R-B-65' OR 'Pneumonia severity index' OR 'PSI' OR 'PORT score' OR 'Fine Score') AND ('Pneumonia' OR 'community-acquired pneumonia' OR 'CAP')
2. ('Severity score\*' OR 'predict\*' OR 'prognosis' OR 'mortality score\*') AND ('Pneumonia' OR 'community-acquired pneumonia' OR 'CAP')

The search included articles published between 1980 and August 2009. No language criteria were applied. Full articles of all potentially appropriate abstracts were reviewed. Only peer-reviewed data were included; therefore, conference abstracts were excluded. The search was repeated in EMBASE to obtain any articles missed by the original search. The search strategy was supplemented by reviewing of reference lists, bibliographies and the investigators files. The PUBMED search results are presented in the online supplement.

### Data extraction

Two investigators independently assessed articles to determine study eligibility. Non-relevant studies were excluded based on title and abstract review only. Potentially relevant studies were reviewed by at least two researchers who carried out data extraction and quality assessment in a blinded fashion. Any disagreement between abstractors was resolved independently by a third abstractor. Where appropriate, we contacted the authors to clarify inconsistencies or to obtain missing data.

### Study inclusion and study quality assessment

All studies were considered eligible if they fulfilled the following criteria: original publications; inclusion of consecutive/unselected patients with CAP; radiographic confirmation of CAP and exclusion of non-CAP diagnoses—for example, non-pneumonic exacerbation of chronic obstructive pulmonary disease; and calculation of severity score based on admission data.

A priori we determined to exclude studies that included diagnoses other than CAP and studies reporting selected populations (eg, restricted to a single age group, or restricted to one causative organism or a single site of care; eg, ICU patients). In the case of duplicate publication, the largest study, if applicable, was included. If there was doubt about overlap between studies, the authors were contacted to clarify this.

There are no widely accepted quality criteria for observational studies. In order to assess quality, modified criteria based on the criteria of Hayden *et al* were used<sup>22</sup> (see table E1 in the online supplementary material). Two reviewers independently assessed quality, and the agreement between the two reviewers was measured using the  $\kappa$  statistic.

### Statistical analysis

Analyses were conducted for discrimination and calibration of the scores. For discrimination, the incidence of each outcome in the high, intermediate and low risk groups was calculated and these ORs were weighted by the inverse of their variance and pooled across all studies using the Dersimonian/Laird random effects model. A random effects model was used due to expected heterogeneity between studies. For each severity score, pooled

sensitivity, specificity, positive and negative likelihood ratios and diagnostic ORs are reported. A summary receiver operator characteristic (sROC) curve was constructed describing the relationship between sensitivity and specificity across the included studies. The areas under the ROC curves (AUCs) were compared using the method described by Hanley and MacNeil for comparing curves derived from different sets of cases.<sup>23</sup>

For calibration, the predicted mortalities obtained from the original derivation studies for PSI, CURB65 and CRB65 were compared with observed mortality rates obtained in each study. The observed:predicted RRs are presented with 95% CIs. Risk ratios were calculated using the Mantel–Haenszel method and pooled using a random effects model. An RR >1 indicates that observed mortality is higher than expected from the original derivation study. An RR <1 indicates that fewer deaths than expected occurred in the validation studies than in the original derivation.

Statistical heterogeneity was assessed using the Cochran Q ( $\chi^2$ ) test and the Higgins  $I^2$  tests. For the Cochran Q test,  $p < 0.1$  was considered to represent significant heterogeneity. For the Higgins test,  $I^2 < 25\%$  indicates low heterogeneity, 25–50% moderate and >50% severe heterogeneity.

A priori the authors decided to conduct subgroup analyses to explore sources of heterogeneity in the main analysis. Preplanned analyses included: analyses limited to studies only including hospital inpatients; analysis limited to high quality studies only; and analysis of prospective studies only.

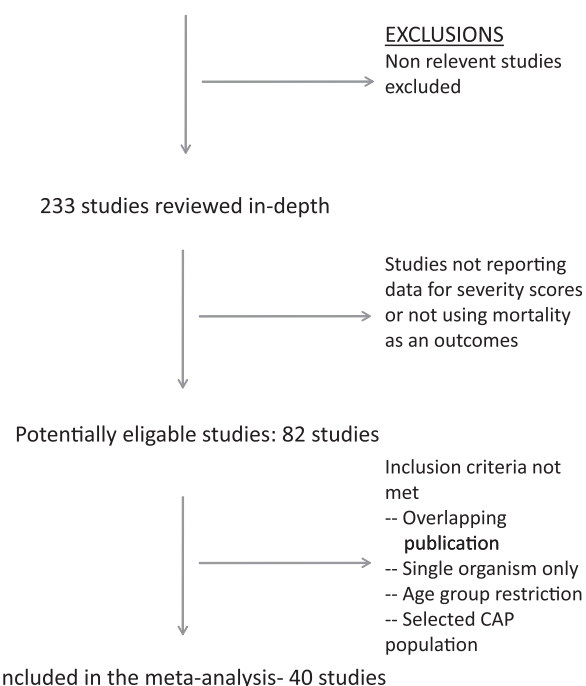
Analyses were conducted using Metadisc software (Barcelona, Spain), SPSS version 13 for windows (SPSS, Chicago, Illinois, USA) and Review manager version 5 (Cochrane Collaboration, Oxford, UK).

## RESULTS

### Literature review and description of included studies

A total of 5102 abstracts were reviewed, and 233 papers were potentially eligible and were reviewed in depth. Figure 1 shows

Search performed: Title and abstract reviewed: 5102



**Figure 1** The process of literature review.

the process of literature review. The majority of studies reviewed in depth were not included because they did not report data for any of the severity scores under investigation.

Forty-two studies containing severity data were excluded (see table E3 in the online supplementary material).

Characteristics of each of the included studies are shown in table E2 in the online supplement. Twenty-one studies were described as prospective observational cohort studies. Thirteen cohorts were described as retrospective. Four studies were clinical trials. One study was described as a population-based observational study and one study was a combination of three prospective observational studies.

Seventeen studies reported data for CURB65,<sup>10 15 18 20 24–36 11</sup> studies reported data for CRB65<sup>10 18 20 25–27 31 33 36–38</sup> and a total of 31 papers reported data for PSI.<sup>7 15 18 20 24 26 27 29–33 35 39–56</sup> comprising 33 individual cohorts. The majority of studies used 30 day mortality as their primary outcome measure, although in-hospital mortality was used in a few studies. The size of studies varied from 137 patients, up to >50 000 patients in the MEDISGROUP retrospective database used to derive the PSI,<sup>7</sup> and 388 406 patients in a retrospective study from Germany assessing CRB65.<sup>38</sup> Reported mortality rates varied from 4.3% to 22.6% (supplementary table E2). The majority of included studies consisted exclusively of hospitalised patients. A proportion of studies included a mix of inpatients and patients managed in the community.<sup>7 15 27 29 31 33 37 49 51</sup> The results of the quality assessment are shown in the online supplement.

## Meta-analysis

### PSI

Thirty-one papers reporting the results for 33 cohorts were included in the meta-analysis for PSI. These studies comprised 81 797 patients with a cumulative mortality rate of 8.3%. The performance characteristics for each cut-off are shown in table 1. The overall test performance across all cut-offs, measured by the area under the sROC curve was good, AUC 0.81 (SE 0.008) (figure 2). Based on the Infectious Disease Society of American/American Thoracic Society (IDSA/ATS) guidelines,<sup>1</sup> patients in risk class I–II should be treated as outpatients, risk class III patients may be treated as outpatients or with short hospitalisation, and patients in risk class IV and V are at higher risk and should be hospitalised. Performance characteristics suggested a good negative likelihood ratio 0.08 (0.06–0.12) for PSI group I–II. Using this cut-off, 34.9% of patients across all studies were identified as low risk. Using PSI I–III to define low risk, the negative likelihood ratio was less good 0.2 (0.17–0.24), but a greater proportion of patients (46.1%) are classified as low risk.

There was significant heterogeneity in all analyses of discrimination. These data and the results of the subanalyses are shown in the online supplement. Calibration results are shown below.

### CURB65

Seventeen studies were including in the analysis, comprising 15 596 patients with CAP. The combined mortality rate was 8.8%. The performance characteristics for each cut-off are shown in table 1. Across all cut-offs, the area under the sROC curve was good, AUC 0.80 (SE 0.008) (figure 2). IDSA/ATS and British Thoracic Society (BTS) guidelines<sup>1 2</sup> suggest that patients with scores of 0–1 are low risk and may be managed as outpatients. Patients with scores of 2 may require brief hospitalisation, and patients with a score  $\geq 3$  are high risk and should be hospitalised. CURB65 0–1 had a negative likelihood ratio of 0.21 (0.15–0.30), and these groups comprised 48.8% of patients in the CURB65 studies. Lowering the threshold for outpatient care to CURB65 0 alone improved the negative likelihood ratio to 0.10 (0.06–0.16), although only 24.2% of patients would be classified as low risk by this approach. Also of note, CURB65  $\geq 4$  had a high specificity of 95.3% (95.0–95.7%) and high positive likelihood ratio 5.4 (4.4–6.6), indicating a strength in identifying high risk patients.

### CRB65

Eleven studies were included for analysis of CRB65. A total of 397 211 patients were included. The analysis was dominated by a large study from Germany of 388 406 patients. IDSA/ATS and BTS guidelines<sup>1 2</sup> suggest that patients with scores of 0 are low risk and may be managed as outpatients. Patients with scores of 1–2 are at moderate risk and may require hospitalisation while patients with scores of  $\geq 3$  are high risk and should be hospitalised. The performance characteristics for each cut-off are shown in table 1. The overall predictive value of CRB65 was good, across all cut-offs, with an area under the curve of 0.79 (SE 0.013) (figure 2). For identifying low risk patients, 33.7% of patients had a CRB65 score of 0, with a negative likelihood ratio of 0.15 (0.10–0.22). As with CURB65, CRB65  $\geq 3$  was specific (specificity 90.9% (90.8–91.0%)) with a high positive likelihood ratio 4.4 (3.6–5.5) suggesting a strength in identifying high risk patients.

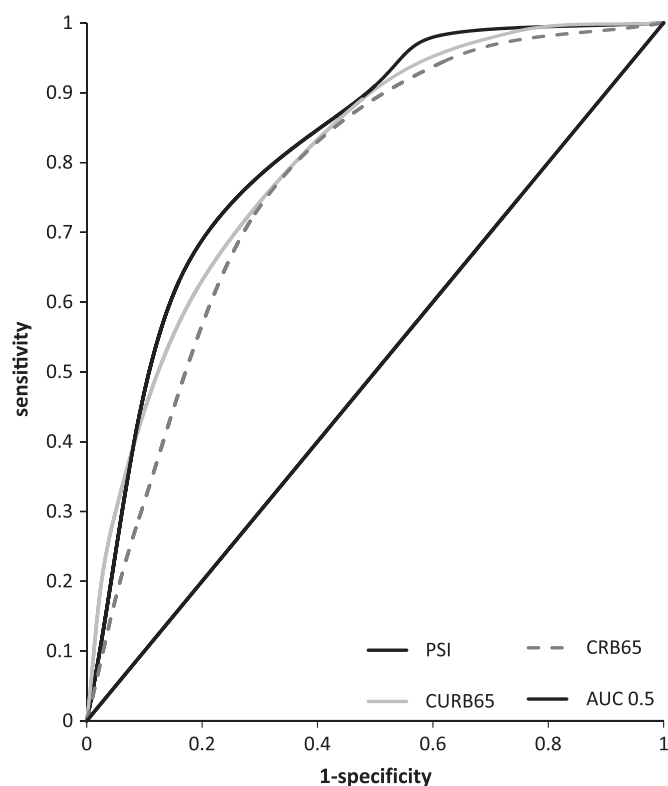
### Calibration

The performance of scores was compared with predicted values obtained from the initial derivation studies. For PSI, Fine *et al*<sup>7</sup> reported mortality rates of 0.3, 0.4, 9.3 and 27% for PSI classes I–II (lowest risk), I–III (low risk), IV and V, respectively. For CURB65, predicted values<sup>10</sup> were 1.2% (CURB65 0–1), 9.0% (CURB65 2) and 22.6% (CURB65 3–5). For CRB65, predicted mortality rates<sup>10</sup>

**Table 1** Pooled performance characteristics of severity scores for predicting mortality in community-acquired pneumonia

|               | Sensitivity        | Specificity        | PLR              | NLR                 | DOR                 | AUROC       |
|---------------|--------------------|--------------------|------------------|---------------------|---------------------|-------------|
| <b>PSI</b>    |                    |                    |                  |                     |                     |             |
| $\geq$ III    | 98.2% (97.8–98.5%) | 38.8% (38.4–39.2%) | 1.4 (1.3 to 1.6) | 0.08 (0.06 to 0.12) | 18.1 (11.5 to 28.5) | 0.88 (0.02) |
| $\geq$ IV     | 91.4% (90.8–92.1%) | 49.5% (49.2–49.9%) | 1.9 (1.8 to 2.0) | 0.2 (0.17 to 0.24)  | 9.6 (8.0 to 11.6)   | 0.82 (0.01) |
| $\geq$ V      | 63.2% (62.1–64.3%) | 83.6% (83.4–83.9%) | 4.1 (3.7 to 4.5) | 0.5 (0.44 to 0.58)  | 8.4 (7.2 to 9.8)    | 0.81 (0.01) |
| <b>CURB65</b> |                    |                    |                  |                     |                     |             |
| $\geq$ 1      | 98.6% (97.6–99.2%) | 26.5% (26.5–27.4%) | 1.2 (1.1 to 1.3) | 0.10 (0.06 to 0.16) | 13.6 (9.3 to 22.3)  | 0.86 (0.02) |
| $\geq$ 2      | 89.1% (87.1–90.8)  | 52.2% (51.3–53.1%) | 1.7 (1.5 to 1.9) | 0.21 (0.15 to 0.30) | 8.3 (5.8 to 11.9)   | 0.81 (0.02) |
| $\geq$ 3      | 62.0% (59.3–64.6%) | 80.8% (80.2–81.4%) | 3.1 (2.7 to 3.5) | 0.46 (0.40 to 0.54) | 7.0 (5.8 to 8.3)    | 0.79 (0.01) |
| $\geq$ 4      | 29.0% (26.3–31.8%) | 95.3% (95.0–95.7%) | 5.4 (4.4 to 6.6) | 0.73 (0.65 to 0.83) | 7.8 (6.3 to 9.6)    | 0.80 (0.01) |
| <b>CRB65</b>  |                    |                    |                  |                     |                     |             |
| $\geq$ 1      | 94.4% (94.2–94.6%) | 38.3% (38.1–38.5%) | 1.3 (1.2 to 1.5) | 0.15 (0.10 to 0.22) | 9.4 (5.8 to 15.3)   | 0.82 (0.03) |
| $\geq$ 2      | 72.7% (69.3–76.0%) | 70.8% (69.8–71.8%) | 2.4 (1.9 to 3.0) | 0.39 (0.28 to 0.54) | 6.3 (4.4 to 9.2)    | 0.78 (0.02) |
| $\geq$ 3      | 29.1% (28.8–29.5%) | 90.9% (90.8–91.0%) | 4.4 (3.6 to 5.5) | 0.72 (0.66 to 0.79) | 6.9 (4.9 to 9.5)    | 0.79 (0.02) |

AUROC, area under the receiver operator characteristic curve; DOR, diagnostic OR; NLR, negative likelihood ratio; PLR, positive likelihood ratio; PSI, Pneumonai Severity Index.



**Figure 2** Summary receiver operator characteristic curves for the Pneumonia Severity Index (PSI), CURB65 and CRB65. AUC, area under the curve.

were 0.9% (CRB65 0), 8.1% (CRB65 1–2) and 31.2% (CRB65 3–4). The results of the calibration analysis are shown in table 2. The observed mortality in the validation studies was significantly higher for both PSI and CURB65 for low risk patients (PSI I–III and CURB65 0–1) than estimated in the original derivation studies. These were consistent findings across all studies, with little heterogeneity (table 2). All three scores were well calibrated ( $p > 0.05$ ) at the higher cut-offs, although significant heterogeneity was observed, suggesting variation in mortality rates between different studies within these groups. Forest plots for each group are displayed in the online supplement.

### Subanalyses

The results of the subanalyses are presented in table E3 for each scoring system in the online supplement. None of the

**Table 2** Calibration of severity scores against predicted mortality from the original derivation studies

|               | Observed mortality | Predicted mortality | RR (95% CI)         | p Value  | I <sup>2</sup> |
|---------------|--------------------|---------------------|---------------------|----------|----------------|
| <b>PSI</b>    |                    |                     |                     |          |                |
| I–II          | 0.75%              | 0.3%                | 2.24 (1.27 to 3.96) | 0.005    | 0%             |
| I–III         | 1.6%               | 0.4%                | 3.64 (2.72 to 4.87) | <0.00001 | 0%             |
| IV            | 8.9%               | 9.3%                | 0.91 (0.79 to 1.06) | 0.2      | 48%            |
| V             | 28.2%              | 27%                 | 0.99 (0.88 to 1.11) | 0.9      | 57%            |
| <b>CURB65</b> |                    |                     |                     |          |                |
| 0–1           | 2.0%               | 1.2%                | 1.63 (1.11 to 2.40) | 0.01     | 27%            |
| 2             | 8.3%               | 9%                  | 0.95 (0.74 to 1.21) | 0.7      | 55%            |
| 3–5           | 22.3%              | 22.6%               | 1.03 (0.84 to 1.25) | 0.8      | 76%            |
| <b>CRB65</b>  |                    |                     |                     |          |                |
| 0             | 2.3%               | 0.9%                | 1.15 (0.45 to 2.94) | 0.8      | 52%            |
| 1–2           | 13.3%              | 8.1%                | 1.13 (0.83 to 1.55) | 0.4      | 90%            |
| 3–4           | 34.4%              | 31.2%               | 1.06 (0.91 to 1.23) | 0.5      | 31%            |

PSI, Pneumonia Severity Index.

subanalyses impacted the main conclusions. The results of the quality assessment are also shown in table E4 (online supplementary material).

### Comparisons between PSI, CURB65 and CRB65

There were no significant differences in the AUC between PSI, CURB65 and CRB65 in the main analysis (PSI vs CURB65,  $p=0.1$ , PSI vs CRB65,  $p=0.09$ , CURB65 vs CRB65,  $p=0.5$ ) or in any of the extensive subanalyses. As noted above and shown in table 1, PSI had a superior negative likelihood ratio and identified a higher proportion of patients as low risk compared with CURB65 and CRB65. The high risk groups of CURB65 and CRB65 had a higher positive likelihood ratio.

### Discussion

This systematic review and meta-analysis found that all three of the major pneumonia severity prediction tools predict 30 day mortality in CAP. The PSI, CURB65 and CRB65 scores all predicted 30 day mortality with moderate to good accuracy.<sup>57</sup> There were no significant differences in overall test performance between these scores, suggesting that clinicians may choose the scoring system best suited to their local needs. Although overall test accuracy was similar, there were some differences in the performance characteristics between the scores. The low negative likelihood ratio suggests that PSI may be superior at identifying low risk patients, while a higher positive predictive value suggests that CURB65/CRB65 may be superior for identifying high risk patients.<sup>58</sup> The clinical importance of these differences is, however, difficult to establish.

National and international guidelines recommend a severity-based approach to diagnosis and treatment of CAP.<sup>1–2</sup> For example, the BTS recommends using the CURB65 score to determine the use of microbiology investigations, admission to hospital, antibiotic choice and admission to high dependency units and ICUs.<sup>2</sup>

Thus the majority of key early decisions in CAP management are based on this severity assessment tool. The IDSA/ATS guidelines, on which many guidelines internationally are modelled, recommend both the PSI and CURB65 scores.<sup>1</sup>

The accuracy of these tools for predicting the outcome of interest is therefore critical. The results of this meta-analysis show that each of the severity scores, PSI, CURB65 and CRB65, predict 30 day mortality from CAP with moderate–good performance characteristics. The PSI is significantly more complex than the CURB65, requiring the measurement of 20 parameters with different points awarded for each.<sup>7</sup> Evidence suggests clinicians find it difficult to use and it is underutilised in clinical practice.<sup>9</sup> This meta-analysis found that PSI had a significant advantage in terms of its negative likelihood ratio, therefore indicating it is superior at identifying low risk patients. PSI was originally designed to identify low risk patients, and the primary objective of clinical studies utilising the PSI has been to increase the proportion of low risk patients managed as outpatients.<sup>7–8, 59</sup> The results of this meta-analysis suggest it is potentially better than alternative scores for this purpose.

CURB65 and CRB65 appear to be superior to PSI for identifying patients at the highest risk, and therefore may be more suitable to guide investigations and the use of broad-spectrum antibiotic therapy (as suggested by the BTS guidelines) where these measures are specifically targeted at high risk patients.<sup>2</sup> Although this may also lead to the conclusion that CURB65 would be superior to identify patients requiring ICU care, this meta-analysis did not consider ICU care, as an outcome and studies suggest CURB65 and PSI are less sensitive to predict ICU care.<sup>3, 6, 26, 29</sup>



CRB65 has been promoted primarily for outpatient use.<sup>2 10</sup> This meta-analysis, however, found only one study that considered its use exclusively in outpatients.<sup>60</sup> The purpose of the current study was to assess its value in hospitalised patients. Further validation studies are required before application of any of these prediction rules in primary care. CRB65 is widely used in European countries for hospitalised patients.<sup>12 38</sup> It is the most simple rule available, composed of the fewest factors and classifying patients accurately into three risk categories. It also has the advantage of not requiring venous blood tests and is therefore immediately available at patient triage, when many management decisions need to be made. In this study, CRB65 was equivalent to CURB65 and PSI in all analyses.

Typically, a positive likelihood ratio of >10 or a negative likelihood ratio of <0.1 is regarded as necessary for a predictive or diagnostic test to be considered robust.<sup>58</sup> It is important to note that only PSI achieved a negative likelihood ratio <0.1 and none achieved a positive likelihood ratio >10, suggesting that all three scores are imperfect. Recent studies have aimed to improve several prediction tools by incorporating information from biomarkers,<sup>41 43</sup> modifying the existing scores<sup>20</sup> or deriving new scores.<sup>26 29 32</sup> Further studies are needed to assess the impact of these modifications.

This meta-analysis was significantly limited by heterogeneity among the studies included. Many studies included both inpatients and outpatients, and different outcome measures were used, such as 30-day mortality or in-hospital mortality. Studies used varying designs, including prospective observational studies, retrospective case note reviews, administrative databases and randomised controlled trials. Despite these variations, extensive subanalyses found similar results to the main analysis. Results were not affected by considering separately those studies using different designs, direct comparisons between the scores, study quality and studies including outpatients. These analyses significantly improved homogeneity while not impacting on the main conclusions. Despite the multiple subanalyses performed, the meta-analysis may still have been affected by biases inherent in the design of the included studies. This meta-analysis sought to compare the three scores recommended by the IDSA/ATS guidelines. A number of other scoring systems have been derived and have a varying degree of validation.<sup>26 29 32</sup> This study only addressed the value of PSI, CURB65 and CRB65, and does not comment on the performance or utility of other scores.

The performance of scores did vary significantly between different studies in different healthcare systems. If the population of patients to which the score is being applied is significantly different from the original derivation it may be necessary to perform local recalibration of the score.

It is widely accepted that prognostic model development is a three-stage process, comprising derivation (creating the rule), validation (applying the rule to new populations of patients to confirm its accuracy) and, perhaps the most important stage, impact analysis (applying the rule and determining if it can improve clinical outcomes for patients).<sup>61</sup> This meta-analysis confirms that CURB65 and CRB65 have similar prognostic performance to PSI, but does not prove that they can improve patient outcome. An impact analysis for CURB65 and CRB65 is needed.

## Conclusion

This systematic review and meta-analysis found no difference in overall test performance between the PSI, CURB65 or CRB65 for predicting mortality in CAP.

**Funding** JDC is supported by a clinical research fellowship from the Medical Research Council (UK).

**Competing interests** None.

**Contributors** All authors had full access to the data in the study. The analysis was conducted by JDC and he takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

- Mandell LA, Wunderink RG, Anzueto A, *et al.* Infectious Disease Society of American/American Thoracic Society consensus guidelines for the management of community acquired pneumonia in adults. *Clin Infect Dis* 2007;**44**:S27–72.
- Lim WS, Baudouin SV, George RC, *et al.* The British Thoracic Society Guidelines for the management of community acquired pneumonia in adults. Update 2009. *Thorax* 2009;**64**(Suppl 3):iii1–55.
- 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.
- Tang CM, Macfarlane JT. Early management of younger adults dying of community-acquired pneumonia. *Respir Med* 1993;**87**:289–94.
- Neill AM, Martin IR, Weir R, *et al.* Community-acquired pneumonia: aetiology and usefulness of severity criteria on admission. *Thorax* 1996;**51**:1010–16.
- Singanayagam A, Chalmers JD, Hill AT. Severity assessment in community-acquired pneumonia. *QJM* 2009;**102**:379–88.
- 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.
- Renaud B, Coma E, Labarere J, *et al.* Routine use of the pneumonia severity index for guiding the site-of-treatment decision of patients with pneumonia in the emergency department: a multicenter, prospective, observational, controlled cohort study. *Clin Infect Dis* 2007;**44**:41–9.
- Lee RW, Lindstrom ST. A teaching hospital's experience applying the Pneumonia Severity Index and antibiotic guidelines in the management of community-acquired pneumonia. *Respirology* 2007;**12**:754–8.
- 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.
- Ewig S, Welte T. CRB-65 for the assessment of pneumonia severity: who could ask for more? *Thorax* 2008;**63**:665–6.
- Alfrageme I, Aspa J, Bello S, *et al.* Guidelines for the diagnosis and management of community-acquired pneumonia. Spanish Society of Pulmonology and Thoracic Surgery (SEPAR). *Arch Bronconeumol* 2005;**41**:272–89.
- Working Group of the South African Thoracic Society. Management of community-acquired pneumonia in adults. *S Afr Med J* 2007;**97**:1296–306.
- Brazilian Society for Infectious Diseases practice guidelines committee. Antimicrobial therapy for community-acquired pneumonia in adults. *Braz J Infect Dis* 2002;**6**:82–7.
- 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.
- Aujesky D, Fine MJ. The pneumonia severity index: a decade after the initial derivation and validation. *Clin Infect Dis* 2008;**47**(Suppl 3):S133–9.
- Buising KL, Thursky KA, Black JF, *et al.* A prospective comparison of severity scores for identifying patients with severe community-acquired pneumonia: reconsidering what is meant by severe pneumonia. *Thorax* 2006;**61**:419–24.
- Man SY, Lee N, Ip M, *et al.* Prospective comparison of three predictive rules for assessing severity of community-acquired pneumonia in Hong Kong. *Thorax* 2007;**62**:348–53.
- Lim WS. Severity assessment in community-acquired pneumonia: moving on. *Thorax* 2007;**62**:287–8.
- Chalmers JD, Singanayagam A, Hill AT. Systolic blood pressure is superior to other haemodynamic predictors of outcome in community acquired pneumonia. *Thorax* 2008;**63**:698–702.
- Stroup DF, Berlin JA, Morton SC, *et al.* Meta-analysis of observational studies in epidemiology. *JAMA* 2000;**283**:2008–12.
- Hayden JA, Cote P, Bombardier C. Evaluation of the quality of prognostic studies in systematic reviews. *Ann Intern Med* 2006;**144**:427–37.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operator characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.
- Ananda-Rajah MR, Charles PG, Melvani S, *et al.* Comparing the pneumonia severity index with CURB-65 in patients admitted with community acquired pneumonia. *Scand J Infect Dis* 2008;**40**:293–300.
- Barlow GD, Nathwani D, Davey PG. The CURB65 pneumonia severity score outperforms generic sepsis scores and early warning in predicting mortality in community acquired pneumonia. *Thorax* 2007;**62**:253–9.
- Buising KL, Thursky KA, Black JF, *et al.* Identifying severe community-acquired pneumonia in the emergency department: a simple clinical prediction tool. *Emerg Med Australas* 2007;**19**:418–26.
- Capelastegui A, Espana PP, Quintana JM, *et al.* Validation of a predictive rule for the management of community-acquired pneumonia. *Eur Respir J* 2006;**27**:151–7.

28. **Challen K**, Bright J, Bentley A, *et al.* Physiological-social score (PMEWS) vs. CURB65 to triage pandemic influenza: a comparative validation study using community-acquired pneumonia as a proxy. *BMC Health Serv Res* 2007;**7**:33.
29. **Charles PG**, Wolfe R, Whitby M, *et al.* SMART-COP: a tool for predicting the need for intensive respiratory or vasopressor support in community-acquired pneumonia. *Clin Infect Dis* 2008;**47**:375–84.
30. **Huang DT**, Weissfeld LA, Kellum JA, *et al.* Risk prediction with procalcitonin and clinical rules in community-acquired pneumonia. *Ann Emerg Med* 2008;**52**:48–58.e2.
31. **Menendez R**, Martinez R, Reyes S, *et al.* Biomarkers improve mortality prediction by prognostic scales in community-acquired pneumonia. *Thorax* 2009;**64**:587–91.
32. **Phua J**, See KC, Chan YH, *et al.* Validation and clinical implications of the IDSA/ATS minor criteria for severe community-acquired pneumonia. *Thorax* 2009;**64**:598–603.
33. **Schuetz P**, Koller M, Christ-Crain M, *et al.* Predicting mortality with pneumonia severity scores: importance of model recalibration to local settings. *Epidemiol Infect* 2008;**136**:1628–37.
34. **Shindo Y**, Sato S, Maruyama E, *et al.* Comparison of severity scoring systems A-DROP and CURB-65 for community-acquired pneumonia. *Respirology* 2008;**13**:731–5.
35. **Tejera A**, Santolaria F, Diez ML, *et al.* Prognosis of community acquired pneumonia (CAP): value of triggering receptor expressed on myeloid cells-1 (TREM-1) and other mediators of the inflammatory response. *Cytokine* 2007;**38**:117–23.
36. **Zuberi FF**, Khan JA. Prospective comparison of prediction rules of mortality risk for CAP in a developing country. *Int J Tuberc Lung Dis* 2008;**12**:447–52.
37. **Bauer TT**, Ewig S, Marre R, *et al.* CRB-65 predicts death from community-acquired pneumonia. *J Intern Med* 2006;**260**:93–101.
38. **Ewig S**, Birkner N, Strauss R, *et al.* New perspectives on community-acquired pneumonia in 388,406 patients. *Thorax* 2009;**64**:1062–9.
39. **Johnstone J**, Eurich DT, Majumdar SR, *et al.* Long-term morbidity and mortality after hospitalization with community-acquired pneumonia: a population-based cohort study. *Medicine (Baltimore)* 2008;**87**:329–34.
40. **Ortega L**, Sierra M, Dominguez J, *et al.* Utility of a pneumonia severity index in the optimization of the diagnostic and therapeutic effort for community-acquired pneumonia. *Scand J Infect Dis* 2005;**37**:657–63.
41. **Chen CZ**, Fan PS, Lin CC, *et al.* Repeated pneumonia severity index measurement after admission increases its predictive value for mortality in severe community-acquired pneumonia. *J Formos Med Assoc* 2009;**108**:219–23.
42. **Davydov L**, Ebert SC, Restino M, *et al.* Prospective evaluation of the treatment and outcomes of community-acquired pneumonia according to the pneumonia severity index in VHA hospitals. *Diagn Microbiol Infect Dis* 2006;**54**:267–75.
43. **Dedier J**, Singer DE, Chang Y, *et al.* Processes of care, illness severity, and outcomes in the management of community-acquired pneumonia at academic hospitals. *Arch Intern Med* 2001;**161**:2099–104.
44. **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.
45. **Feagan BG**, Marrie TJ, Lau CY, *et al.* Treatment and outcomes of community-acquired pneumonia at Canadian hospitals. *CMAJ* 2000;**162**:1415–20.
46. **Flanders WD**, Tucker G, Kirshnadasan A, *et al.* Validation of the pneumonia severity index. Importance of study-specific recalibration. *J Gen Intern Med* 1999;**14**:333–40.
47. **Garau J**, Baquero F, Perez-Trallero E, *et al.* Factors impacting on length of stay and mortality of community-acquired pneumonia. *Clin Microbiol Infect* 2008;**14**:322–9.
48. **Garcia-Vazquez E**, Soto S, Gomez J, *et al.* Simple criteria to assess mortality in patients with community-acquired pneumonia. *Med Clin (Barc)* 2008;**131**:201–4.
49. **Goss CH**, Rubenfeld GD, Park DR, *et al.* Cost and incidence of social co-morbidities in low-risk patients with community-acquired pneumonia admitted to a public hospital. *Chest* 2003;**124**:2148–55.
50. **Migliorati PL**, Boccoli E, Bracci LS, *et al.* A survey on hospitalised community-acquired pneumonia in Italy. *Monaldi Arch Chest Dis* 2006;**65**:82–8.
51. **Renaud B**, Coma E, Hayon J, *et al.* Investigation of the ability of the Pneumonia Severity Index to accurately predict clinically relevant outcomes: a European study. *Clin Microbiol Infect* 2007;**13**:923–31.
52. **Querol-Ribelles JM**, Tenias JM, Querol-Borras JM, *et al.* Validation of the pneumonia severity index for hospitalizing patients with community-acquired pneumonia. *Med Clin (Barc)* 2004;**122**:481–6.
53. **Restrepo MI**, Mortensen EM, Velez JA, *et al.* A comparative study of community-acquired pneumonia patients admitted to the ward and the ICU. *Chest* 2008;**133**:610–17.
54. **Reyes Calzada S**, Martinez Tomas R, Cremades Romero MJ, *et al.* Empiric treatment in hospitalized community-acquired pneumonia: impact on mortality, length of stay and re-admission. *Respir Med* 2007;**101**:1909–15.
55. **Roson B**, Carratala J, Dorca J, *et al.* Etiology, reasons for hospitalization, risk classes and outcomes of community-acquired pneumonia in patients hospitalized on the basis of conventional admission criteria. *Clin Infect Dis* 2001;**33**:158–65.
56. **Van Der Eerden MM**, De Graaff CS, Bronsveld W, *et al.* Prospective evaluation of pneumonia severity index in hospitalised patients with community-acquired pneumonia. *Respir Med* 2004;**98**:872–8.
57. **Swets JA**. Measuring the accuracy of diagnostic systems. *Science* 1998;**240**:1285–93.
58. **Deeks JJ**, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ* 2004;**329**:168–9.
59. **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.
60. **Bont J**, Hak E, Hoes AW, *et al.* Predicting death in elderly patients with community-acquired pneumonia: a prospective validation study re-evaluating the CRB65 severity assessment tool. *Arch Intern Med* 2008;**168**:1465–8.
61. **McGinn TG**, Guyatt GH, Wyer PC. Users' guides to the medical literature:XXII: how to use articles about clinical decision rules. Evidence based medicine working group. *JAMA* 2000;**284**:79–84.