Meta-analysis of Calibration, Discrimination, and Stratum-Specific Likelihood Ratios for the CRB-65 Score

Mark H. Ebell, MD, MS[1], Mary E. Walsh, PhD[2], Tom Fahey, MD[2], Maggie Kearney, MPH[1], and Christian Marchello, PhD[1]

[1]Department of Epidemiology and Biostatistics, College of Public Health , University of Georgia, Athens, GA, USA; [2]HRB Centre for Primary Care Research, Department of General Practice, Royal College of Surgeons in Ireland, Dublin, Republic of Ireland.

BACKGROUND: The CRB-65 score is recommended as a decision support tool to help identify patients with community-acquired pneumonia (CAP) who can safely be treated as outpatients.

OBJECTIVE: To perform an updated meta-analysis of the accuracy, discrimination, and calibration of the CRB-65 score using a novel approach to calculation of stratum-specific likelihood ratios.

DESIGN: Meta-analysis of accuracy, discrimination, and calibration.

METHODS: We searched PubMed, Google, previous systematic reviews, and reference lists of included studies. Data was abstracted and quality assessed in parallel by two investigators. The quality assessment used an adaptation of the TRIPOD and PROBAST criteria. Measures of discrimination, calibration, and stratum-specific likelihood ratios are reported.

KEY RESULTS: Twenty-nine studies met our inclusion criteria and provided usable data. Most studies were set in Europe, none in North America, and 12 were judged to be at low risk of bias. The pooled estimate of area under the receiver operating characteristic curve was 0.74 (95% CI 0.71–0.77) for all studies. Calibration was good although there was significant heterogeneity; the pooled estimate of the ratio of observed to expected mortality for all studies was 1.04 (95% CI 0.91–1.19). The corresponding values for studies at low risk of bias where patients could be treated as outpatients or inpatients were 0.76 (0.70–0.81) and 0.88 (0.69–1.13). Summary estimates of stratum-specific likelihood ratios for all studies were 0.19 for the low-risk group, 1.1 for the moderate-risk group, and 4.5 for the high-risk group, and 0.13, 1.3, and 5.6 for studies at low risk of bias where patients could be treated as outpatients or inpatients.

CONCLUSIONS: The CRB-65 is useful for identifying low-risk patients for outpatient therapy. Given a 4% overall mortality risk, patients classified as low risk by the CRB-65 had an outpatient mortality risk of no more than 0.5%.

Electronic supplementary material *The online version of this article (https://doi.org/10.1007/s11606-019-04869-z) contains supplementary material, which is available to authorized users.*

Received September 6, 2018
Revised November 19, 2018
Accepted January 24, 2019
Published online April 16, 2019

KEY WORDS: community-acquired pneumonia; risk prediction models; clinical decision rules; CRB-65; adults; meta-analysis.
J Gen Intern Med 34(7):1304–13
DOI: 10.1007/s11606-019-04869-z

Physicians in the ambulatory and emergency department setting must decide whether patients with community-acquired pneumonia (CAP) can safely be treated as outpatients or whether they should be hospitalized. Current guidelines from the American Thoracic Society and Infectious Disease Society of America recommend the use of the Pneumonia Severity Index (PSI) or CURB-65 clinical decision rules to assist in this decision.[1] However, the PSI is cumbersome, with 20 questions including several laboratory tests.[2] The CURB-65 is simpler, but also requires a blood test (blood urea nitrogen) that may not be readily available in primary care settings and could delay decision-making.[3]

The CRB-65 places patients into low-, moderate-, or high-risk groups for mortality based on four easily obtained clinical indicators: confusion (new onset); respiratory rate $\geq$ 30/min; blood pressure (systolic $<$ 90 mmHg or diastolic $\leq$ 60 mmHg); and 65 years or older. A previous meta-analysis found that it had good calibration between observed and expected results.[4] The CRB-65 is recommended by British Thoracic Society guidelines for evaluation of outpatients with CAP; patients with a score of 0 may be safely treated as outpatients, whereas those with a score of 1 or higher should be considered for hospitalization.[5]

A previous meta-analysis of the CRB-65 reported the calibration of the CRB-65, but not the diagnostic accuracy.[4] It concluded that while the CRB-65 was well calibrated for inpatients, it overestimated mortality in community settings. Also, a number of potentially relevant studies evaluating the CRB-65 have been published since that meta-analysis was published in 2010. In the current study, we perform an updated meta-analysis of the accuracy of the CRB-65 for mortality prediction, and in addition to evaluating calibration and discrimination, we apply a novel approach for performing meta-analysis of stratum-specific likelihood ratios (SSLRs).

METHODS

Search Strategy

This meta-analysis was performed in collaboration with the leader of a team of researchers that did a previous meta-analysis of the accuracy of the CRB-65 score, published in 2010.[4] Their search was complete through June 2009; we therefore searched PubMed beginning in January 2009 using the following strategy: "CRB-65" OR "CRB65" OR "CURB65" OR "CURB-65." We also searched the reference lists of included articles, and the first 100 results on Google Scholar using the search term "CRB-65." The abstracts were reviewed in parallel by two investigators, and any article identified as potentially meeting inclusion criteria by either investigator was reviewed in full, again by both investigators in parallel.

Inclusion and Exclusion Criteria

We included studies reporting the accuracy of the CRB-65 score among patients with community-acquired pneumonia (CAP). Studies had to provide sufficient data to calculate mortality for the following risk groups: low risk = 0 points, moderate risk = 1 to 2 points, and high risk = 3 to 4 points. We included both prospective and retrospective cohort studies, with treatment in the inpatient or outpatient settings.

We excluded studies in children; studies in special populations such as immunocompromised patients or those characterized by a comorbidity such as asthma, cancer, or diabetes; studies of patients with sepsis; studies of patients with hospital-acquired or ventilator-acquired pneumonia; abstracts from a meeting without a full publication; and case control studies. Studies performed in countries classified as low income or lower-middle income by the World Bank (https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups) were excluded, as the case mix and resources for treating pneumonia are likely different from those in better resourced settings, and therefore may not be generalizable.

Data Abstraction

Data regarding study characteristics, study quality, and test accuracy were abstracted in parallel by two investigators. Any discrepancies were resolved by review and discussion between the investigators. Where data had been communicated to the authors as part of the previous meta-analysis, we used those data.[4] The final data abstraction has been approved by all authors.

Assessment of Study Quality

The TRIPOD statement describes a set of 22 quality criteria for studies describing prediction models (PMID 25560730). However, many of these criteria are focused on presentation of results and process rather than study quality and how likely the study is to avoid bias. We identified a subset of TRIPOD criteria for our quality assessment, focusing on description of the study design and data source, the study population, handling of missing data, and patient flow. We also reviewed the PROBAST statement from the Cochrane group (https://abstracts.cochrane.org/2017-global-evidence-summit/probast-%E2%80%93-risk-bias-tool-prediction-modelling-studies) and added additional items to capture study characteristics such as the spectrum of included patients, whether predictors and outcomes were defined the same way for all participants, whether the clinical decision rule was determined prospectively, and whether a relevant measure of accuracy was reported. The final quality assessment framework is summarized in Appendix Figure 3 and Table 2.

Meta-analysis Method

We conducted meta-analysis for the following groups of studies: (1) all studies, (2) only studies with low risk of bias, (3) only studies where participants could be treated in either the inpatient or outpatient settings, (4) only studies at low risk of bias where patients could be treated in either the outpatient or inpatient setting.

Measures of discrimination (AUC) and calibration (*O*:*E* ratio, calibration slope or plot) were extracted from included studies, where reported. Measures of uncertainty were also extracted. Where the *O*:*E* ratio was not reported, the expected number of deaths in each validation study was calculated by applying the probabilities reported in the derivation study[3] to the numbers of patients in each risk category. Where measures of uncertainty were not reported, the standard error of the total *O*:*E* ratio and AUC was estimated using equations proposed by Debray et al. (2017) if possible.[6] In two studies, AUC was provided for the validation cohort in the study.[7, 8]

To evaluate the case mix of each study, the mean and SD of the total CRB-65 score among patients in the study were calculated where sufficient information was provided; a higher mean CRB-65 score would indicate a sicker population. Where the number of patients was reported only for 3 categories of risk rather than each of the 5 scores of the CRB-65, the mid-score of the risk category was used to calculate mean and SD. The mean and standard deviation of demographic variables including age were also extracted.[6]

To summarize the performance of the CRB-65 score, a random effects meta-analysis of *O*:*E* and AUC values was conducted with REML estimation using the metaan procedure in Stata 14 (Stata Corp, College Station, TX).[6, 9] This was conducted as recommended on the log scale for the *O*:*E* ratio and logit scale for the AUC.[6, 10] The proportion of heterogeneity due to between study variation was estimated using the I^2 statistic.

We also calculated stratum-specific likelihood ratios. The likelihood ratio (LR) for a test or risk score with 3 or more risk groups, with mortality as the outcome of interest, is calculated as LR = [(deaths in risk group) (total deaths)] [(survivors in risk group) (total survivors)]. This

is similar conceptually to a risk ratio (RR) for a treatment trial, which is the ratio of the risk of an outcome in the treatment group to the risk of that outcome in the control group. We calculated stratum-specific likelihood ratios (SSLRs) by treating the likelihood ratios as risk ratios. We used the metan procedure (version 9) in Stata 15.1 (StataCorp, College Station, TX) to perform a random effects meta-analysis of likelihood/risk ratios. The proportion of heterogeneity due to between study variation is estimated using the I^2 statistic.[11]

RESULTS

Search Results

The original meta-analysis[4] identified 14 studies.[7, 12–24] The current PubMed search, including a bridge search in January of 2018, yielded 348 articles, of which 154 required a full review, and 16 met our inclusion criteria.[8, 18, 25–38] Of the 100 articles identified with the Google Scholar search, 13 were reviewed in full, 5 were abstracts only, and 2 met our inclusion criteria and had

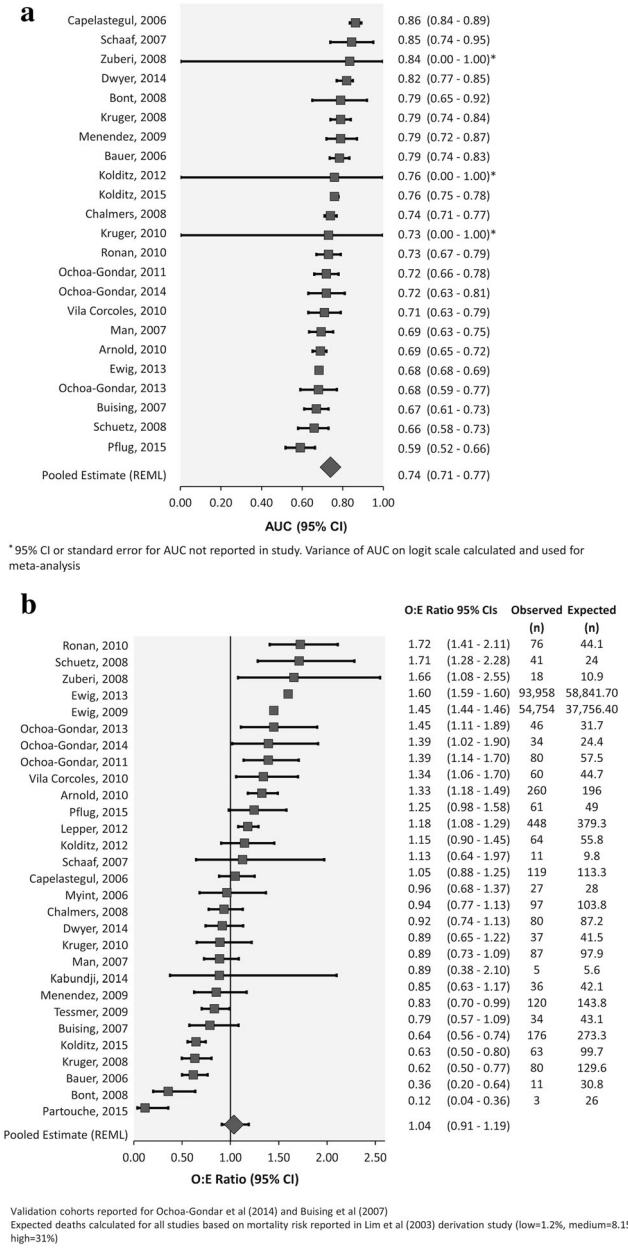

Figure 1 Forest plots for all studies with available data of (a) the area under the receiver operating characteristic curve (AUC) and (b) the ratio of observed to expected mortality.

Table 1 Characteristics of Included Studies

Study	Treatment Setting (% ambulatory if reported)	Data collection	Number of participants	Mean or median age (range)	Outcome measured	Mortality rate	Mean CRB-65 score (SE)
Arnold, 2010	Hospitalized	Retrospective	3085	Mean 65.5	28-day mortality	8.4%	0.87 (0.83)
Bauer, 2006	Both (25% ambulatory)	Prospective	1959	Mean 62.5 (18–99)	30-day mortality	4.1% (0.6% ambulatory, 5.2% hospitalized)	1.03 (0.85)
Bont, 2008	Ambulatory	Prospective	314	Mean 77.3	30-day mortality	3.5%	1.34 (0.61)
Buising, 2007	Both (16% ambulatory)	Prospective	740	Median 74 (18–98)	In-hospital mortality	9.7%	1.82 (1.12)
Capelastegul, 2006	Both (38% ambulatory)	Prospective	1776	Mean 61.8 (18–96)	30-day mortality	6.7% (0.2% ambulatory, 11% inpatient)	0.86 (0.87)
Chalmers, 2008	Hospitalized	Prospective	1007	Mean 66 (50–78)	30-day mortality	9.6%	1.50** (1.06**)
Dwyer, 2014	Both (29% ambulatory)	Retrospective	1172	Median 65 (18–100)	30-day mortality	6.8%	1.03 (0.9)
Ewig, 2009	Hospitalized	Prospective	388,406	Median 76	In-hospital mortality	14.1%	1.49** (0.92**)
Ewig, 2013	Hospitalized	Prospective	669,594	Mean 72.8	In-hospital mortality	14.0%	1.41** (0.82**)
Kabundji, 2014	Both (55% ambulatory	Prospective	152	Median 36.5 (20–87)	14-day mortality	3.3%	0.48** (0.80**)
Kolditz, 2012	Hospitalized	Prospective	984	Mean 58.7	30-day mortality	6.5%	0.75 (0.76)
Kolditz, 2015	Both	Prospective	4432	Median 64 survivors and 78 non-survivors (18–101)	28-day mortality	4.0%	0.84 (0.8)
Kruger, 2008	Both (31% ambulatory)	Prospective	1508	Mean 61 (18–98)	28-day mortality	4.2% (0.4% ambulatory, 5.9% inpatient)	0.86 (0.83)
Kruger, 2010	Both	Prospective	728	Mean 59.0 (18–96)	6-month mortality	5.1%	0.74 (0.75)
Lepper, 2012	Both	Prospective	6142	Mean 59.8	90-day mortality	7.3%	0.83 (0.83)
Man, 2007	Hospitalized	Prospective	1016	Mean 72 (17–103)	30-day mortality	8.6%	1.38 (0.86)
Menendez, 2009	Hospitalized	Prospective	447	Mean 67.3	30-day mortality	8.1%	1.34 (0.93)
Myint, 2006	Hospitalized	Prospective	192	Median 75 (17–96)	6-week mortality	14.1%	2.05** (0.94**)
Ochoa-Gondar, 2011	Both (13% ambulatory)	Prospective	590	Mean 77.4	30-day mortality	13.6% (1.4% ambulatory, 15.3% inpatient)	1.38 (0.63)
Ochoa-Gondar, 2013	Both (19% ambulatory)*	Retrospective	350	Mean 78.3	30-day mortality	13.1% (4.4% ambulatory, 15.2% inpatient)	1.33 (0.55)
Ochoa-Gondar, 2014	Both (22% ambulatory)	Retrospective	260	Mean 78.1	30-day mortality	13.1%	1.33 (0.57)
Partouche, 2015	Ambulatory	Retrospective	642	Median 55 (18–102)	30-day mortality	0.5%	0.45 (0.64)
Pflug, 2015	Hospitalized	Retrospective	559	Mean 74.1 (18–104)	30-day mortality	10.9%	1.28 (0.83)
Ronan, 2010	Hospitalized	Retrospective	422	69% age >65	30-day mortality	18.0%	1.46 (0.98)
Schaef, 2007	Hospitalized	Prospective	105	Mean 64.9	In-hospital mortality	10.5%	1.33** (1.11**)
Schuetz, 2008	Hospitalized	Prospective	373	Median 73	30-day mortality	11.0%	0.81 (0.61)
Tessmer, 2009	Hospitalized	Prospective	1854	Mean 66.3	30-day mortality	6.5%	1.24** (0.85**)
Vila Corcoles, 2010	Both (24% ambulatory)	Retrospective	473	Mean 76.8	30-day mortality	12.7% (1.7% ambulatory and 16.2% inpatient)	1.32 (0.6)
Zuberi, 2008	Hospitalized	Retrospective	137	Mean 60.4	30-day mortality	13.1%	1.18 (0.9)

Only the "validation cohort" for Ochoa-Gondor (2014) was used to avoid duplication with data already published in Ochoa-Gondar (2013)
***Estimated based on frequency tables per risk category (midpoint score of category used)*

Table 2 Instrument for Evaluating the Quality of a Study Describing a Clinical Decision Rule for Prognosis

Study, year	Study design and reporting			Patient selection			Predictors and outcomes				Flow		Risk of bias (low or high)
	1	2	3	4	5	6	7	8	9	10	11	12	
	Described design and data source	Study setting described	Relevant performance measures reported	Described eligibility criteria for participants	Characteristics of participants described	Appropriate spectrum of patients	CRB-65 determined prospectively	Predictors defined similarly for all	Outcome and its assessment defined	Outcome defined say way for all	Described flow of participants	Missing data handling described	
Kruger, 2008	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	U	L
Man, 2007	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	U	L
Schuetz, 2008	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	Y	U	L
Bauer, 2006	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	L
Capelastegul, 2006	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	N	Y	L
Kolditz, 2012	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	N	Y	L
Kolditz, 2015	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	N	Y	L
Kruger, 2010	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	N	Y	L
Menendez, 2009	Y	Y	Y	Y	Y	Y	Y	Y	U	Y	Y	N	L
Pflug, 2015	Y	Y	Y	Y	Y	Y	N	U	Y	Y	Y	Y	L
Buising, 2007	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	L
Lepper, 2012	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	L
Chalmers, 2008	Y	Y	Y	Y	Y	Y	Y	U	U	U	N	U	H
Ewig, 2013	Y	Y	Y	Y	Y	Y	U	U	U	U	N	Y	H
Ewig, 2009	Y	Y	N	Y	Y	Y	U	U	U	U	N	Y	H
Partouche, 2015	Y	Y	N	Y	Y	Y	Y	U	U	U	Y	N	H
Schaaf, 2007	Y	Y	Y	Y	Y	Y	Y	Y	U	U	N	N	H
Tessmer, 2009	Y	Y	N	Y	Y	N	Y	U	Y	Y	Y	U	H
Arnold, 2010	Y	Y	Y	Y	Y	Y	N	Y	U	Y	N	Y	H
Zuberi, 2008	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	U	H
Ochoa-Gondar, 2013	Y	Y	Y	Y	Y	N	N	U	U	U	N	U	H
Ochoa-Gondar, 2011	Y	Y	Y	Y	Y	N	N	U	U	U	N	Y	H
Vila Corcoles, 2010	Y	Y	Y	Y	Y	N	N	U	Y	U	N	U	H
Myint, 2006	Y	Y	N	Y	Y	N	Y	Y	U	U	N	Y	H
Ochoa-Gondar, 2014	Y	Y	Y	Y	Y	N	N	U	Y	Y	N	U	H
Bont, 2008	Y	Y	Y	Y	Y	N	Y	U	N	Y	N	Y	H
Ronan, 2010	Y	Y	Y	Y	N	Y	N	U	Y	N	Y	Y	H
Dwyer, 2014	Y	Y	Y	Y	N	U	U	U	N	U	Y	Y	H
Kabundji, 2014	Y	Y	Y	Y	Y	Y	U	U	N	U	N	Y	H

Table 3 Summary Estimates of Observed/Expected (*O*:*E*) Ratios and Area Under the Receiver Characteristic Curve (AUC) for Subgroups of Included Studies

Analysis description	*O*:*E* ratio			AUC		
	Pooled *O*:*E* ratio (95% CI)	No. of studies (no. of patients)	I^2	Pooled AUC (95% CI)	No. of studies (no. of patients)	I^2
Including all studies with available information	1.04 (0.91–1.19)	29 (1,089,009)	99.7%	0.74 (0.71–0.77)	23 (691,621)	92.6%
Including only studies with low risk of bias	0.93 (0.78–1.11)	12 (20,254)	89.1%	0.74 (0.69–0.79)	11 (14,112)	89.9%
Including only studies where the rule was applied in ED or primary care settings and patients could be treated as either outpatients or inpatients	1.05 (0.87–1.27)	15 (20,667)	91.3%	0.75 (0.71–0.78)	13 (14,373)	85.1%
Including only studies at low risk of bias where the rule was applied in ED or primary care settings and patients could be treated as either outpatients or inpatients	0.88 (0.69–1.13)	8 (17,248)	92.7%	0.76 (0.70–0.81)	7 (11,106)	91.0%

not been previously identified,[26, 29] Of these 32 studies, one study did not report mortality data, and correspondence with the authors revealed that there were no deaths in this population, so it was excluded.[39] In addition, two pairs of studies used the same dataset, so we excluded one study in each pair.[12, 40] This resulted in a final total of 29 included studies (see Fig. 1).

Study Characteristics

Study characteristics are summarized in Table 1. All but 3 studies were set in Europe, including 10 in Germany and 6 in Spain; none were set in the USA or Canada. Most studies reported 30-day mortality, with an overall mortality rate ranging from 0.5% (in a study recruiting only outpatients) to 18.0%. Nine studies gathered data retrospectively to determine the CRB-65 score, while the remainder gathered data prospectively, often as part of the CAPNETZ disease registry. The mean or median age in most studies was over 60 years, although one South African study reported a median age of 36 years.[29]

Assessment of Study Quality

Study quality is summarized in Table 2 and Appendix Table 5. Twelve of 29 studies were judged to be at low risk of bias. Of the remaining studies, common deficiencies included missing or unclear descriptions of how the CRB-65 was measured, how confusion was defined and determined, and how the outcome of mortality was determined for all patients. The two largest studies by far were judged to be at high risk of bias due to no description of patient flow, unclear descriptions of how CRB-65 was determined, and unclear definition of predictors and the outcome.[17, 30]

Discrimination

In total, 23 of 29 studies reported an AUC and 20 reported 95% confidence intervals. Standard error of the logit (AUC) was calculated based on 95% CIs for 20 studies and based on the number of events and non-events for 3 studies based on equations derived by Debray et al.[6] Table 3 summarizes the pooled AUC for all studies and for specified subgroups. The pooled AUC for all studies was 0.74 (95% CI 0.71–0.77), and the AUC for subgroups was similar, suggesting that the CRB-65 performs moderately well for predicting mortality within

Table 4 Summary Estimates of Stratum-Specific Likelihood Ratios an Estimate of Heterogeneity (I^2). Low Risk = 0 Points, Moderate Risk = 1 to 2 Points, High Risk = 3 to 4 Points

Analysis description	Risk group	Studies	LR (95% CI)	I^2
Including all studies with available information	Low	24	0.19 (0.16–0.24)	85.8%
	Moderate	29	1.08 (1.04–1.12)	96.2%
	High	29	4.49 (4.07–4.95)	91.0%
Including only studies with low risk of bias	Low	12	0.15 (0.10–0.21)	24.6%
	Moderate	12	1.19 (1.06–1.34)	91.6%
	High	12	4.98 (3.66–6.76)	80.2%
Including only studies where the rule was applied in ED or primary care settings and patients could be treated as either outpatients or inpatients	Low	11	0.12 (0.07–0.19)	34.6%
	Moderate	15	1.10 (0.96–1.25)	93.8%
	High	15	5.59 (4.25–7.34)	75.6%
Including only studies at low risk of bias where the rule was applied in ED or primary care settings and patients could be treated as either outpatients or inpatients	Low	8	0.13 (0.08–0.21)	40.0%
	Moderate	8	1.30 (1.17–1.44)	84.7%
	High	8	5.61 (3.71–8.47)	85.6%

30 days. A forest plot displaying the study specific results and pooled results for the AUC is presented in Figure 1a.

Calibration

Though the total *O*:*E* ratio was typically not reported, it could be calculated for all included validation populations from the reported number of observed deaths per risk category. The standard error of the log (*O*:*E*) was therefore calculated for all 29 studies as described by Debray et al.[6] Only one study provided a calibration plot[23] and only two presented calibration tables.[8, 23] The calibration slope was not reported for any validation study and could not be derived using other information. Table 3 shows the results of each meta-analysis. Calibration was good although there was significant heterogeneity between studies; the pooled estimate of the ratio of observed to expected mortality for all studies was 1.04 (95% CI 0.91–1.19) for all studies, and in the specified subgroups, the range of *O*:*E* ratio was 0.88 to 1.05. A forest plot displaying the study specific results and pooled results for the *O*:*E* ratio is presented in Figure 1b.

Stratum-Specific Likelihood Ratios for Prediction of Mortality

The summary estimates of the stratum-specific likelihood ratios, overall and stratified by subgroups, are shown in Table 4. The overall estimates of the likelihood ratios for low-, moderate-, and high-risk groups were 0.19, 1.1, and 4.5 respectively. For studies judged to be at low risk of bias, the likelihood ratios were 0.15, 1.2, and 5.0 respectively, and for studies where patients could be treated in either the outpatient or inpatient setting, 0.12, 1.1, and 5.6 respectively. When the latter group was limited to studies at low risk of bias, results were similar (0.13, 1.3, 5.6).

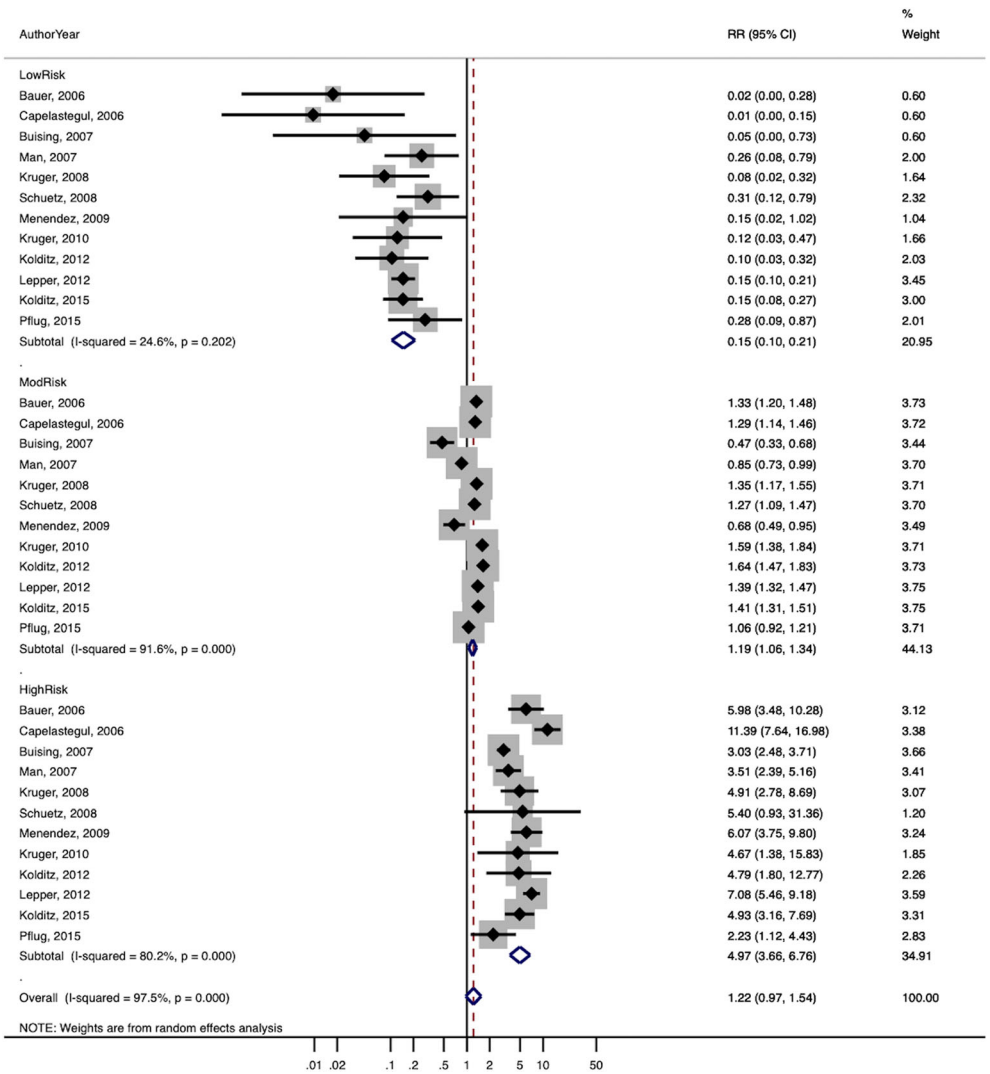

Figure 2 Forest plots of stratum-specific likelihood ratios for (a) studies at low risk of bias and (b) studies at low risk of bias where patients could be treated as either inpatient or outpatients.

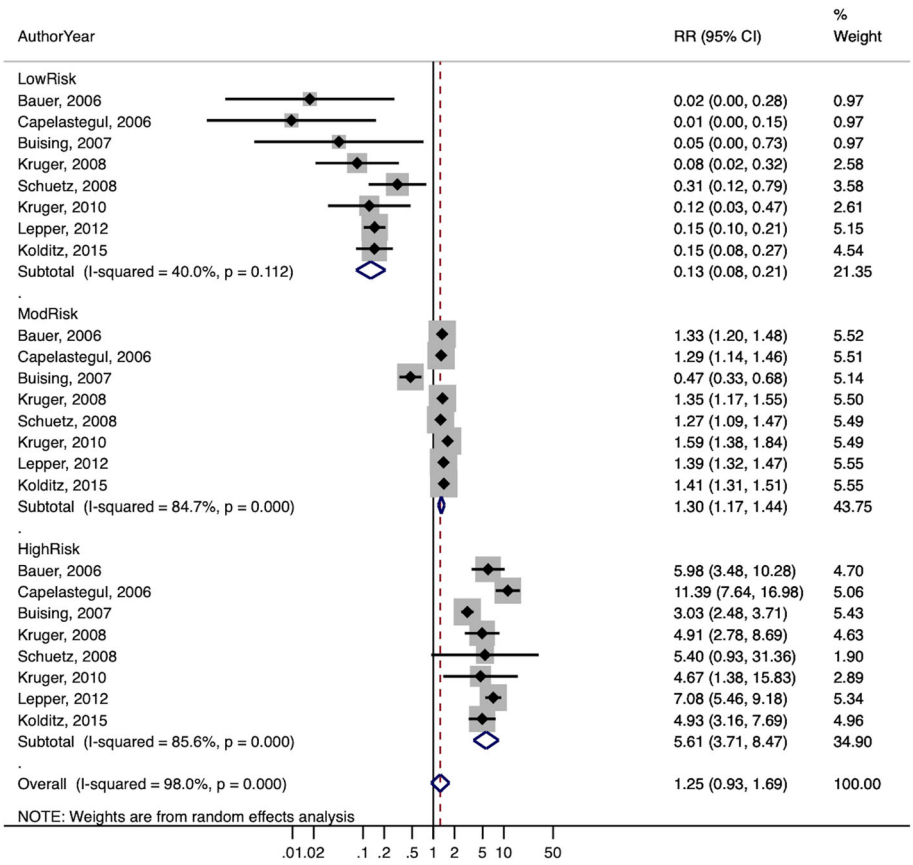

Fig. 2 (continued)

A forest plot of the likelihood ratios for studies at low risk of bias is shown in Figure 2a, and for studies at low risk of bias where patients could be treated as inpatients or outpatients in Figure 2b. Heterogeneity was generally high based on the I^2 statistic, although this is of somewhat limited value with a small number of studies.[41]

DISCUSSION

We have updated a meta-analysis of the accuracy of the CRB-65 score, identifying 15 additional studies and providing a full evaluation of the score's calibration as measured by the ratio of observed to expected deaths, discrimination as measured by the AUC, and prognostic accuracy as measured by stratum-specific likelihood ratios. The latter analysis uses a novel approach to calculating stratum-specific likelihood ratios, by organizing the data so they can be treated as risk ratios.

The summary estimate of calibration is near 1.0 for all studies and for the subset of studies at low risk of bias. The overall discrimination is good, with a pooled AUC of 0.74, and the AUC was consistent across subgroups. Studies at low risk of bias where patients could be treated in either the outpatient or inpatient setting avoid the selection bias inherent in studies where patients were treated only as outpatients and inpatients. We believe that this is the most relevant group for guiding clinical decision-making in the ambulatory setting. Using the LR estimates from these studies (0.13, 1.3, 5.6) and an overall mortality risk of 4%, the probability of death is 0.5% in the low-risk group, 5.1% in the moderate-risk group, and 18.9% in the high-risk group.

Two German studies reported data for patients treated as both inpatients and outpatients from the CAPNETZ registry, and found an overall mortality rate of 4.1%. The decision to admit was at the discretion of the clinician, and mortality was only 0.3% when the clinician chose to treat them as outpatients, and 5.5% when the patient was admitted for treatment.[13, 18] Thus, clinicians were able to use their overall clinical impression ("clinical gestalt") to accurately identify patients at low risk for mortality who could be treated as outpatients, with a mortality rate similar to that in the low-risk group identified by the CRB-65 score. This suggests that clinical rules such as the CRB-65 should serve as

a decision support tool, but should not necessarily replace clinician decisions.

It is notable that there have been no validation studies in the USA, or in fact in all of North and South America. Prospective validation studies of patients presenting in primary care, urgent care, and the emergency department are needed outside of Europe, as there are differences in how care is organized, who gets hospitalized, and for how long they are hospitalized.

A limitation of the current study is the heterogeneity of the estimates of calibration, discrimination, and prognostic accuracy. However, there is consistency across the summary estimates of AUC (0.74 to 0.76), *O*:*E* (0.88 to 1.05), and LRs across subgroups (0.12–0.19, 1.1–1.3, and 4.5–5.6). Also, for the low-risk group, which is the clinically most important because it identifies patients who do not require hospitalization, heterogeneity is lower when limited to studies at low risk of bias (I^2 = 24.6% vs 85.8% for all studies). It is also lower for the low-risk group when limited to studies where patients could be treated in either the inpatient or outpatient setting, for both all studies in this group (34.6%) and for studies in this group at low risk of bias (40.0%).

Another limitation is the modest overall quality of included studies. In part, this is due to poor reporting. Future studies should clearly state how and when the CRB-65 score (or any clinical decision rule) is assessed and by whom. Studies should also provide complete data regarding calibration, discrimination, and accuracy to facilitate future meta-analyses. They should also ideally include all patients and determine their outcome whether treated as inpatients or outpatients. Studies that directly compare the CRB-65 with the physician's overall clinical impression are also needed.

In conclusion, the CRB-65 can be used by physicians to estimate mortality risk, and can serve as a useful check on physician judgment. Patients in the low-risk group with a score of 0 have a very low mortality risk (0.5% given a typical mortality rate of 4% for CAP) and can in most cases safely be treated as outpatients. Most patients in the moderate- and high-risk groups should be hospitalized, although other considerations may alter these decisions regarding treatment setting.

Funding Source: *Dr. Ebell's collaboration with Drs. Fahey and Walsh was supported by a 2018/2019 Fulbright Teaching/Research award.*

Role of Investigators: *Mark Ebell and Tom Fahey were responsible for overall design and conduct of the study. Mark Ebell participated in data abstraction, analysis, and primary authorship of the manuscript. Maggie Kearney, Mary Walsh, and Christian Marchello participated in data abstraction. Mary Walsh performed the random effects meta-analysis of observed/expected ratio and AUC, while Mark Ebell performed meta-analysis of stratum-specific likelihood ratios. Tom Fahey also participated in writing the manuscript and in the analysis, and all authors reviewed and approved the final manuscript.*

Corresponding Author: *Mark H. Ebell, MD, MS; Department of Epidemiology and Biostatistics, College of Public Health University of Georgia, Athens, GA, USA (e-mail: ebell@uga.edu).*

Compliance with Ethical Standards:

Conflict of Interest: *The authors declare that they do not have a conflict of interest.*

Publisher's Note: *Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.*

REFERENCES

1. **Mandell LA, et al.** Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44(Suppl 2):S27–72.
2. **Fine MJ, et al.** A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med. 1997;336(4):243–50.
3. **Lim WS, et al.** Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377–82.
4. **McNally M, et al.** Validity of British Thoracic Society guidance (the CRB-65 rule) for predicting the severity of pneumonia in general practice: systematic review and meta-analysis. Br J Gen Pract. 2010;60(579):e423–33.
5. **Lim WS, et al.** BTS guidelines for the management of community acquired pneumonia in adults: update 2009. Thorax. 2009;64(Suppl 3): iii1–55.
6. **Debray TPA, et al.** A guide to systematic review and meta-analysis of prediction model performance. BMJ. 2017;356.
7. **Buising KL, et al.** Identifying severe community-acquired pneumonia in the emergency department: a simple clinical prediction tool. Emerg Med Australas. 2007;19(5):418–26.
8. **Ochoa-Gondar O, et al.** Validation of the CORB75 (confusion, oxygen saturation, respiratory rate, blood pressure, and age >/= 75 years) as a simpler pneumonia severity rule. Infection. 2014;42(2):371–8.
9. **Kontopantelis E, Reeves D.** metaan: Random-effects meta-analysis. Stata J, 2010. 10(3): p. 395–407.
10. **Snell KI, et al.** Meta-analysis of prediction model performance across multiple studies: Which scale helps ensure between-study normality for the C-statistic and calibration measures? Stat Methods Med Res. 2017;962280217705678.
11. **Higgins JP, Thompson SG.** Quantifying heterogeneity in a meta-analysis. Stat Med. 2002;21(11):1539–58.
12. **Barlow G, Nathwani D, Davey P.** The CURB65 pneumonia severity score outperforms generic sepsis and early warning scores in predicting mortality in community-acquired pneumonia. Thorax. 2007;62(3):253–9.
13. **Bauer TT, et al.** CRB-65 predicts death from community-acquired pneumonia. J Intern Med. 2006;260(1):93–101.
14. **Bont J, et al.** Predicting death in elderly patients with community-acquired pneumonia: a prospective validation study reevaluating the CRB-65 severity assessment tool. Arch Intern Med. 2008;168(13):1465–8.
15. **Capelastegui A, et al.** Validation of a predictive rule for the management of community-acquired pneumonia. Eur Respir J. 2006;27(1):151–7.
16. **Chalmers JD, Singanayagam A, Hill AT.** Systolic blood pressure is superior to other haemodynamic predictors of outcome in community acquired pneumonia. Thorax. 2008;63(8):698–702.
17. **Ewig S, et al.** New perspectives on community-acquired pneumonia in 388 406 patients. Results from a nationwide mandatory performance measurement programme in healthcare quality. Thorax. 2009;64(12):1062–9.
18. **Kruger S, et al.** Procalcitonin predicts patients at low risk of death from community-acquired pneumonia across all CRB-65 classes. Eur Respir J. 2008;31(2):349–55.
19. **Man SY, et al.** Prospective comparison of three predictive rules for assessing severity of community-acquired pneumonia in Hong Kong. Thorax. 2007;62(4):348–53.

20. **Menendez R, et al.** Biomarkers improve mortality prediction by prognostic scales in community-acquired pneumonia. Thorax. 2009;64(7):587–91.
21. **British Thoracic S, et al.** Severity assessment criteria recommended by the British Thoracic Society (BTS) for community-acquired pneumonia (CAP) and older patients. Should SOAR (systolic blood pressure, oxygenation, age and respiratory rate) criteria be used in older people? A compilation study of two prospective cohorts. Age Ageing. 2006;35(3):286–91.
22. **Schaaf B, et al.** Sepsis severity predicts outcome in community-acquired pneumococcal pneumonia. Eur Respir J. 2007;30(3):517–24.
23. **Schuetz P, et al.** Predicting mortality with pneumonia severity scores: importance of model recalibration to local settings. Epidemiol Infect. 2008;136(12):1628–37.
24. **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(4):447–52.
25. **Kolditz M, et al.** Assessment of oxygenation and comorbidities improves outcome prediction in patients with community-acquired pneumonia with a low CRB-65 score. J Intern Med. 2015;278(2):193–202.
26. **Dwyer R, et al.** Improvement of CRB-65 as a prognostic tool in adult patients with community-acquired pneumonia. BMJ Open Respir Res. 2014;1(1):e000038.
27. **Partouche H, et al.** Suspected community-acquired pneumonia in an ambulatory setting (CAPA): a French prospective observational cohort study in general practice. NPJ Prim Care Respir Med. 2015;25:15010.
28. **Pflug MA, et al.** Short-term mortality of adult inpatients with community-acquired pneumonia: external validation of a modified CURB-65 score. Postgrad Med J. 2015;91(1072):77–82.
29. **Kabundji DM, et al.** Determining need for hospitalisation: Evaluation of the utility of the CRB-65 score in patients with community-acquired pneumonia presenting to an emergency department. S Afr Med J. 2014;104(11):769–772.
30. **Ewig S, et al.** Prediction of in-hospital death from community-acquired pneumonia by varying CRB-age groups. Eur Respir J. 2013;41(4):917–22.
31. **Ochoa Gondar O, et al.** Ability of the modified CRB75 severity scale in assessing elderly patients with community acquired pneumonia. Aten Primaria. 2013;45(4):208–15.
32. **Kolditz M, et al.** Serum cortisol predicts death and critical disease independently of CRB-65 score in community-acquired pneumonia: a prospective observational cohort study. BMC Infect Dis. 2012;12:90.
33. **Ochoa-Gondar O, et al.** Comparison of three predictive rules for assessing severity in elderly patients with CAP. Int J Clin Pract. 2011;65(11):1165–72.
34. **Lepper PM, et al.** Serum glucose levels for predicting death in patients admitted to hospital for community acquired pneumonia: prospective cohort study. BMJ. 2012;344:e3397.
35. **Arnold FW, et al.** Predictive accuracy of the pneumonia severity index vs CRB-65 for time to clinical stability: results from the Community-Acquired Pneumonia Organization (CAPO) International Cohort Study. Respir Med. 2010;104(11):1736–43.
36. **Ronan D, et al.** Predicting mortality in patients with community-acquired pneumonia and low CURB-65 scores. Eur J Clin Microbiol Infect Dis. 2010;29(9):1117–24.
37. **Vila Corcoles A, Ochoa Gondar O, Blanco TR.** Usefulness of the CRB-65 scale for prognosis assessment of patients 65 years or older with community-acquired pneumonia. Med Clin (Barc). 2010;135(3):97–102.
38. **Tessmer A, et al.** Impact of intravenous {beta}-lactam/macrolide versus {beta}-lactam monotherapy on mortality in hospitalized patients with community-acquired pneumonia. J Antimicrob Chemother. 2009;63(5):1025–33.
39. **Francis NA, et al.** Severity assessment for lower respiratory tract infections: potential use and validity of the CRB-65 in primary care. Prim Care Respir J. 2012;21(1):65–70.
40. **Nullmann H, et al.** External validation of the CURSI criteria (confusion, urea, respiratory rate and shock index) in adults hospitalised for community-acquired pneumonia. BMC Infect Dis. 2014;14:39.
41. **von Hippel PT.** The heterogeneity statistic I(2) can be biased in small meta-analyses. BMC Med Res Methodol. 2015;15:35.