BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email editorial.bmjopen@bmj.com

BMJ Open

Appetite loss in addition to existing prediction models for the diagnosis of community-acquired pneumonia in the elderly: external validation and extension of diagnostic models

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019155
Article Type:	Research
Date Submitted by the Author:	14-Aug-2017
Complete List of Authors:	Takada, Toshihiko; Fukushima Kenritsu Ika Daigaku, Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR) Yamamoto, Yosuke; Kyoto University, Department of Healthcare Epidemiology Terada, Kazuhiko; Kimitsu Chuo Byoin, Department of General Medicine Ohta, Mitsuyasu; Kanagawa Kenritsu Ashigarakami Byoin, Department of Internal Medicine Mikami, Wakako; Kanagawa-ken Keiyukai Keiyu Byoin, Department of Radiology Yokota, Hajime; Chiba Daigaku Igakubu Fuzoku Byoin, Department of Radiology Hayashi, Michio; Fukushima Kenritsu Ika Daigaku, Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR) Miyashita, Jun ; Fukushima Kenritsu Ika Daigaku, Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR); Kyoto University, Department of Healthcare Epidemiology Azuma, Teruhisa; Fukushima Kenritsu Ika Daigaku, Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR) Fukuma, Shingo; Kyoto University, Department of Healthcare Epidemiology Fukuhara, Shunichi; Kyoto University, Department of Healthcare Epidemiology
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Infectious diseases, Geriatric medicine, Respiratory medicine
Keywords:	GENERAL MEDICINE (see Internal Medicine), INFECTIOUS DISEASES, PRIMARY CARE, Respiratory infections < THORACIC MEDICINE

SCHOLARONE™
Manuscripts

consciousness disturbance. Appetite loss was based on self-assessment of 24-h food intake in proportion to the usual intake.

Outcome measurement

The reference standard for CAP diagnosis was CXR in the posteroanterior and lateral views or in the anteroposterior view for patients who could not stand up. Two certified radiologists (WK and HY) who were blinded to the information about the patients' history and physical examination data independently assessed the CXRs. Any disagreement was resolved by discussion.

Existing prediction models for CAP

Existing models were identified based on a previous external validation study of prediction models for CAP in the primary care setting.[12] Inclusion criteria for these existing models were 1) use of logistic regression for the diagnosis of CAP in adult patients; 2) predictors were items about history and physical examination; 3) the diagnosis of CAP was made by CXR or computed tomography (CT).

Statistical analysis

The existing models in their original form and without adjustment of intercept and coefficients were externally validated for diagnostic performance by calibration plot and discrimination. Calibration plot was tested with the Hosmer–Lemeshow (HL) test and a value of less than 0.05 indicated a lack of good fit.[13] Discrimination was assessed by area under the curve (AUC).

The diagnostic performance of each item in the existing models, as well as the nonspecific symptoms (appetite loss, falls and consciousness disturbance), were assessed in terms of sensitivity, specificity, positive likelihood ratio (LR+), and negative likelihood ratio (LR−), all of which were reported as point estimates of diagnostic odds

For peer review only

ratio (DOR) with 95% confidence interval (CI). DOR ranged from zero to infinity and was calculated as the ratio of LR+ to LR−, with higher values indicative of better discriminative performance; a value of 1 indicated that the test did not discriminate between people with and without the disease.[14] The cut-off points for continuous variables were determined according to existing prediction models[15] or the values with the best sensitivity and specificity in the receiver operating characteristic (ROC) curve. A calibration plot of the logistic regression model with appetite loss as the sole predictor was drawn to evaluate its diagnostic performance. To evaluate the additional value of these nonspecific symptoms to existing prediction models, we developed an extended logistic regression model without changing the coefficients of the existing models. Calibration, discrimination, category-free net reclassification improvement (NRI), and decision curve analysis (DCA) were investigated in the extended model. The NRI was the cumulative net proportion of events reclassified correctly plus the net proportion of non-events reclassified correctly.[16] The DCA was a graphical approach to evaluate the prediction models based on the principle that the relative harm of false positives and false negatives can be expressed in terms of a probability threshold. Net benefit was obtained by subtracting the proportion of false positives from the proportion of true positives, weighed by the relative harm of false positive and false negative results.[17] Net benefit was assessed in the probability threshold, which was decided by 10 physicians who worked in the participating hospitals and who were blinded to the results of our research.

To assess the agreement in the CXR interpretations of the two radiologists, Cohen's Kappa coefficient (κ) was calculated.[18] Imputation method was not applied for missing values because of the possibility of perfect prediction of categorical data. We did not estimate sample size because there was no available consensus on the adequate sample size among external validation studies of prediction models or evaluation studies of diagnostic performance of the added predictors.[13] Statistical

For peer review only

analyses were performed with a commercial software program (STATA, version 14.2 SE; StataCorp LP, College Station, Texas).

Results

Patient characteristics

A total of 109 patients were included in the study population. The baseline characteristics of the patients are listed in Table 1; mean age was 76.7 years [standard deviation (SD) 7.8] and 58.7% were men. CAP was diagnosed by CXR in 22% of patients. There was moderate agreement (κ = 0.58) between the 2 radiologists who read the CXRs.

External validation of existing models

Six existing prediction models for CAP diagnosis were externally validated in our cohort. The formula for each model was as following:

Diehr[6]

−2 for rhinorrhoea, −1 for sore throat, +1 for night sweats, +1 for myalgia, +1 for sputum, +2 for respiratory rate >25, +2 for temperature >37.7°C

Singal[7]

–3.539, +0.884 for cough, +0.681 for fever, +0.464 for crackles, +0.030 for 5.0 (pretest probability of pneumonia)

Heckerling[8]

–1.705, +0.494 for temperature >37.7°C, +0.428 for pulse >100 beats/min, +0.658 for rales, +0.638 for decreased breath sounds, +0.691 for absence of asthma

Melbye[9]

+ 4.7 for fever (reported by patient) with duration of illness of >1 week,

–4.5 for coryza, –2.1 for sore throat, +5.0 for dyspnea, +8.2 for lateral chest pain, +0.9 for crackles

For peer review only

Hopstaken[10]

–2.74, +1.02 for dry cough, +1.78 for diarrhea, +1.13 for temperature ≥38°C

van Vugt[11]

−3.984, +0.446 for dyspnea, +0.698×absence of runny nose, +0.596×diminished vesicular breathing, +1.404×crackles, +0.961×pulse>100 beats/min, +0.980×temperature >37.8°C

The diagnostic performance of each history and physical examination predictor item is presented in Table 2. For each model, a calibration plot was drawn by comparing the predicted probability with the observed probability in our dataset (Fig. 1). All existing models demonstrated significant values in the HL test, indicating poor fit (Table 3). On the other hand, the calibration plot for the model by van Vugt[11] visually demonstrated good fit between the predicted probability and the observed prevalence of CAP. Discrimination was highest for the model by Singal[7] (AUC = 0.76; 95% CI 0.65–0.87), followed by the models by van Vugt[11] (AUC = 0.75; 95% CI 0.63–0.88); Diehr[6] (AUC = 0.75; 95% CI 0.64–0.86); and Heckerling[8] (AUC = 0.75; 95% CI 0.64–0.86) (Table 3).

Diagnostic performance of the nonspecific symptoms

The sensitivity, specificity, LR+, and LR− for appetite loss, falls, and consciousness disturbance are listed in Table 2. The cut-off point for appetite loss was determined at 50% by the ROC curve. The AUC of appetite loss for CAP diagnosis was 0.76 (95% CI 0.65–0.87). The calibration plot of the model for appetite loss showed better fit than the model by van Vugt,[11] which showed the best calibration among the existing models (Fig. 2). Non-significant HL test suggested good fit of the model for appetite loss (Table 3).

Added value of the nonspecific symptoms

For peer review only

In the evaluation of the diagnostic performance of each nonspecific symptom, falls and consciousness disturbance seemed less useful; therefore, we investigated the added value of appetite loss to the model by van Vugt,[11] which showed the best performance among the existing models.

The formula for the extended model included the following: −5.258, −0.135 for appetite, +0.446 for dyspnea, +0.698 for absence of runny nose, +0.596 for diminished vesicular breathing, +1.404 for crackles, +0.961 for tachycardia, and +0.980 for temperature >37.8 °C. The extended model showed improved performance on the calibration plot, especially in the subgroup with high probability of CAP (Fig. 2), and good fit by the HL test (= 0.48; χ^2 = 8.6). Compared with the original model, the extended model tended to have an improved AUC, but this change was not significant [0.77 (95% CI 0.65–0.89) vs. 0.75 (95% CI 0.63–0.88); = 0.48]. NRI of the extended model was 0.53 (95% CI 0.08–0.97) (= 0.019), which meant that 53% of patients obtained correctly increased probability of CAP.

Fig. 3 illustrates the decision curves for the model by van Vugt[11] and the extended model. Among the 10 physicians, the mean threshold for ordering CXR for the diagnosis of CAP in the elderly was 22.5% (SD, 7.1%; range 10–30%). At a threshold of 10% to 30%, higher net benefit was more frequently observed in the extended model than in the original model. At a threshold of 20%, the net benefit of the extended model increased by 0.02 compared with the original model.

Discussion

Our study assessed the external validity of existing models for the diagnosis of CAP in an elderly population at acute care hospitals in Japan. These existing models had been validated in a meta-analysis of individual data of patients who were at least 18 years old.[12] Although the existing models were supposed have poor external validity in the elderly population, the model by van Vugt[11] showed relatively good calibration and

For peer review only

discrimination.[12] In the evaluation of the diagnostic performance of the nonspecific symptoms of CAP in the elderly, falls and consciousness disturbance showed high specificity but low sensitivity, leading to clinically irrelevant LR+ and LR−. On the other hand, information about appetite showed good performance for the diagnosis of CAP in the elderly. It has been reported that nurse-assessed food consumption predicted bacteremia with sensitivity of 0.92 and negative predictive value of 0.98.[19] Therefore, appetite loss could be an important predictor of infection. In the assessment of the added value of appetite loss to the model by van Vugt,[11] calibration was improved, especially in the population with high probability; however, the improvement in discrimination was not significant. NRI was significant because more than 50% of patients were correctly assessed to have increased probabilities. Results of the DCA showed higher net benefit of the extended model at a threshold appetite loss of 10% to 30%. For example, applying the extended model at a threshold appetite loss of 20% could correctly detect CAP in 2 additional patients per 100 patients.

The strength of our study was that this was the first study that evaluated the external validity of existing models for CAP in the elderly. Based on our results, the model by van Vugt[11] can be applied in the elderly population despite its relatively poor calibration. To the best of our knowledge, this was the first study to investigate and identify the value of information on appetite loss, independently and in addition to the model by van Vugt,[11] for the diagnosis of CAP in the elderly.

This study had several limitations. First, our sample size was relatively small. Because this study was conducted in acute care hospitals, many patients were ineligible because they were already prescribed antibiotics in the primary care setting. Therefore, similar investigation in a primary care setting should be further explored. Second, the frequency of CAP was relatively higher than that previously reported,[12] probably because the number of patients with CAP was higher in the acute care setting than in the primary care settings. Therefore, the external validity of our study should be applied

For peer review only

carefully in the primary care setting. Third, although CT has been reported to be superior in terms of diagnostic accuracy,[20] CXR was used as the reference standard of CAP in this study because performing CT on all patients was not feasible and ethical due to difficult access, higher radiation exposure, and cost.

In conclusion, the model by van Vugt[11] demonstrated relatively high performance among the existing models for the diagnosis of CAP in the elderly. Information about appetite loss independently demonstrated diagnostic usefulness for CAP in the elderly and improved the performance of the model by van Vugt.[11]

For peer review only

Contributors:

TT had full access to all of the data in the study; he takes responsibility for integrity of the data and accuracy of the data analysis and wrote the first draft. YY designed the study, interpreted the data, and drafted the paper. KT, MO, WM, and HY collected and interpreted the data. MH, JM, and TA designed the study, collected and interpreted the data. S Fukuma and S Fukuhara supervised the research and revised the work critically for important intellectual content.

Funding:

This work was supported by a Grant-in-Aid for Epidemiological Research of St. Lukes Life Science Institute.

Competing interests:

None declared.

Ethics approval:

Approval was granted by the ethics committees of Kyoto University Graduate School and Faculty of Medicine, Shirakawa Kosei General Hospital, Kimitsu Chuo Hospital, and Ashigarakami Hospital.

Data sharing statement:

No additional data are available.

For peer review only

References

1. Metlay JP, Schulz R, Li YH, et al. Influence of age on symptoms at presentation in patients with community-acquired pneumonia. 1997;**157**:1453–59.

2. Marrie TJ. Community-acquired pneumonia in the elderly. 2000;**31**:1066–78.

3. Berman P, Hogan DB, Fox RA. The atypical presentation of infection in old age. 1987;**16**:201–7.

4. Thiem U, Heppner HJ, Pientka L. Elderly patients with community-acquired pneumonia: optimal treatment strategies. 2011;**28**:519–37.

5. Kataoka H, Matsuno O. Age-related pulmonary crackles (rales) in asymptomatic cardiovascular patients. 2008;**6**:239–45.

6. Diehr P, Wood RW, Bushyhead J, Krueger L, Wolcott B, Tompkins RK. Prediction of pneumonia in outpatients with acute cough--a statistical approach. 1984;**37**:215–25.

7. Singal BM, Hedges JR, Radack KL. Decision rules and clinical prediction of pneumonia: evaluation of low-yield criteria. 1989;**18**:13–20.

8. Heckerling PS, Tape TG, Wigton RS, et al. Clinical prediction rule for pulmonary infiltrates. 1990;**113**:664–70.

9. Melbye H, Straume B, Aasebo U, Dale K. Diagnosis of pneumonia in adults in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. 1992;**10**:226–33.

10. Hopstaken RM, Muris JW, Knottnerus JA, Kester AD, Rinkens PE, Dinant GJ. Contributions of symptoms, signs, erythrocyte sedimentation rate, and C-reactive protein to a diagnosis of pneumonia in acute lower respiratory tract infection. 2003;**53**:358–64.

For peer review only

11. van Vugt SF, Broekhuizen BD, Lammens C, et al. Use of serum C reactive protein and procalcitonin concentrations in addition to symptoms and signs to predict pneumonia in patients presenting to primary care with acute cough: diagnostic study. 2013;**346**:f2450.

12. Schierenberg A, Minnaard MC, Hopstaken RM, et al. External Validation of Prediction Models for Pneumonia in Primary Care Patients with Lower Respiratory Tract Infection: An Individual Patient Data Meta-Analysis. ! 2016;**11**:e0149895.

13. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. 2015;**162**:W1–73.

14. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. 2003;**56**:1129–35.

15. 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. " # 2003;**58**:377–82.

16. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. 2011;**30**:11–21.

17. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. $ 2006;**26**:565–74.

18. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. 1986;**1**:307–10.

19. Komatsu T, Onda T, Murayama G, et al. Predicting bacteremia based on nurse-assessed food consumption at the time of blood culture. 2012;**7**:702–5.

For peer review only

20. Syrjala H, Broas M, Suramo I, Ojala A, Lahde S. High-resolution computed tomography for the diagnosis of community-acquired pneumonia. 1998;**27**:358–63.

For peer review only

Tables

Table 1. Patients' characteristics

Age (Mean ± SD)	76.7±7.8
Male (%)	58.7
Past history of BA/COPD (%)	5.5
Past history of IHD/CHF (%)	12.9
Clinical diagnosis of CAP (%)	29.4
CAP diagnosed by CXR (%)	22.0

SD, standard deviation; BA, bronchial asthma; COPD, chronic obstructive pulmonary disease; IHD, ischemic heart disease; CHF, chronic heart failure; CAP, community-acquired pneumonia; CXR, chest x ray

For peer review only

Table 2. Diagnostic performance of the predictors of CAP in the elderly

	CAP+ (N=27)	CAP- (N=82)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)	Cut-off point
cough (%)	100	98.8	100 (87.2-100)	1.2 (0.3-6.6)	1.0 (1.0-1.0)	0	-	-
sputum (%)	88.9	74.1	88.9 (70.8-97.6)	25.9 (16.8-36.9)	1.2 (1.0-1.4)	0.4 (0.1-1.3)	2.8 (0.8-9.6)	-
sore throat (%)	22.2	40.2	22.2 (8.6-42.3)	59.8 (48.3-70.4)	0.6 (0.3-1.2)	1.3 (1.0-1.7)	0.4 (0.2-1.1)	-
rhinorrhea (%)	34.6	60.5	34.6 (17.2-55.7)	39.5 (28.8-51.0)	0.6 (0.3-1.0)	1.7 (1.1-2.4)	0.3 (0.1-0.9)	-
dyspnea (%)	40.7	34.2	40.1 (22.4-61.2)	65.9 (54.6-76.0)	1.2 (0.7-2.1)	0.9 (0.6-1.3)	1.3 (0.6-3.2)	-
fever (%)	74.1	35.4	74.1 (53.7-88.9)	64.6 (53.3-74.9)	2.1 (1.5-3.0)	0.4 (0.2-0.8)	5.2 (2.0-13.5)	-
chill (%)	29.6	27.2	29.6 (13.8-50.2)	71.6 (60.5-81.1)	1.0 (0.5-2.1)	0.98 (0.7-1.3)	1.1 (0.4-2.7)	-
night sweats (%)	33.3	24.4	33.3 (16.5-54.0)	75.6 (64.9-84.4)	1.4 (0.7-2.6)	0.9 (0.7-1.2)	1.5 (0.6-3.9)	-
myalgia (%)	7.4	18.3	7.4 (0.9-24.3)	81.7 (71.6-89.4)	0.4 (0.1-1.7)	1.1 (1.0-1.3)	0.4 (0-1.5)	-
pleural pain (%)	7.4	7.4	7.4 (0.9-24.3)	92.6 (84.6-97.2)	1.0 (0.2-4.7)	1.0 (0.9-1.1)	1.0 (0-4.7)	-
diarrhea (%)	7.4	7.4	7.4 (0.9-24.3)	92.6 (84.6-97.2)	1.0 (0.2-4.7)	1.0 (0.9-1.1)	1.0 (0-4.7)	-
duration of symptoms (days)	6.1	7.2	42.3 (23.4-63.1)	62.5 (51.0-73.1)	1.1 (0.7-1.9)	0.9 (0.6-1.3)	1.2 (0.5-3.0)	7 days from the model by Melbye[9]
coma (%)	15.4	6.1	15.4 (4.4-34.9)	93.9 (86.3-98.0)	2.5 (0.7-8.7)	0.9 (0.8-1.1)	2.8 (0.7-10.6)	-
fall (%)	11.1	1.2	11.1 (2.4-29.2)	98.8 (93.4-100.0)	9.1 (1.0-84.0)	0.9 (0.8-1.0)	10.1 (1.4-∞)	-
appetite loss (%)	48.8	80.1	66.7 (46.0-83.5)	79.3 (68.9-87.4)	3.2 (2.0-5.3)	0.4 (0.2-0.7)	7.7 (3.0-19.7)	50% determined by the ROC curve
SBP(mmHg)	129.9	130.1	3.7 (0.1-19.0)	100 (95.6-100)	-	1.0 (0.9-1.0)	-	90mmHg from CURB65[15]
DBP (mmHg)	72.5	73.9	22.2 (8.6-42.3)	84.1 (74.4-91.3)	1.4 (0.6-3.3)	0.9 (0.7-1.2)	1.4 (0.5-4.4)	60mmHg from CURB65[15]
pulse rate (/min)	89.3	82.8	29.6 (13.8-50.2)	80.5 (70.3-88.4)	1.5 (0.7-3.2)	0.9 (0.7-1.1)	1.7 (0.7-4.6)	100/min from the model by van Vugt[11]
SpO2 (%)	93.7	96.4	51.9 (31.9-71.3)	91.5 (83.2-96.5)	6.1 (2.7-13.5)	0.5 (0.4-0.8)	11.5 (4.0-33.4)	95% as normal value
respiratory rate (/min)	21.3	19	3.7 (0.1-19.0)	100 (95.5-100)	-	1.0 (0.9-1.0)	-	30/min from CURB65[15]
body temperature (℃)	37.6	36.7	29.6 (13.8-50.2)	91.5 (83.2-96.5)	3.5 (1.4-8.7)	0.8 (0.6-1.0)	4.5 (1.5-13.6)	38°C from the model by Hopstaken[10]
decreased respiratory sound (%)	18.5	4.9	18.5 (6.3-38.1)	95.1 (87.8-98.6)	3.8 (1.1-13.0)	0.9 (0.7-1.0)	4.4 (1.2-16.4)	-
crackle (%)	48.2	14.8	48.1 (28.7-68.1)	85.2 (75.6-92.1)	3.3 (1.7-6.2)	0.6 (0.4-0.9)	5.3 (2.1-14.0)	-
wheeze (%)	3.7	3.7	3.7 (0.1-19.0)	96.3 (89.6-99.2)	1.0 (0.1-9.2)	1.0 (0.9-1.1)	1.0 (0-7.4)	-
rhonchi (%)	3.7	6.2	3.7 (0.1-19.0)	93.8 (86.2-98.0)	0.6 (0.1-4.9)	1.0 (0.9-1.1)	0.6 (0-4.0)	-

CAP: community-acquired pneumonia; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; DOR: diagnostic odds ratio; SBP: systolic blood pressure; DBP: diastolic blood pressure; SpO2: oxygen saturation; ROC: receiver operating characteristic; CURB65: Confusion, urea > 7mmol/L, respiratory rate >= 30 breaths/min, low blood pressure (systolic < 90mmHg or diastolic <= 60 mmHg)

For peer review only

Table 3. Comparison of diagnostic performance among existing models and appetite loss

	Calibration (HL test)		Discrimination	
Model	$\chi 2$	value	AUC	95% CI
Diehr[6]	134.9	<0.001	0.75	0.64-0.86
Singal[7]	39.2	<0.001	0.76	0.65-0.87
Heckerling[8]	19.8	0.003	0.75	0.64-0.86
Melbye[9]	6104.2	<0.001	0.64	0.51-0.77
Hopstaken[10]	48.2	<0.001	0.54	0.42-0.66
van Vugt[11]	18.7	0.017	0.75	0.63-0.88
Appetite loss	1.91	0.928	0.76	0.65-0.87

HL, Hosmer-Lemeshow; AUC, area under the curve; CI, confidence interval

For peer review only

Figure legends

Figure 1. Calibration plots of existing models

The X-axis represented predicted proportion and the y-axis represented observed proportion, with a reference for perfect calibration (red line). The size of the circle indicates the number of patients. The model by van Vugt visually demonstrated better fit than the other models.

Figure 2. Comparison of the calibration plots among the model of appetite loss, the model by van Vugt, and the extended model

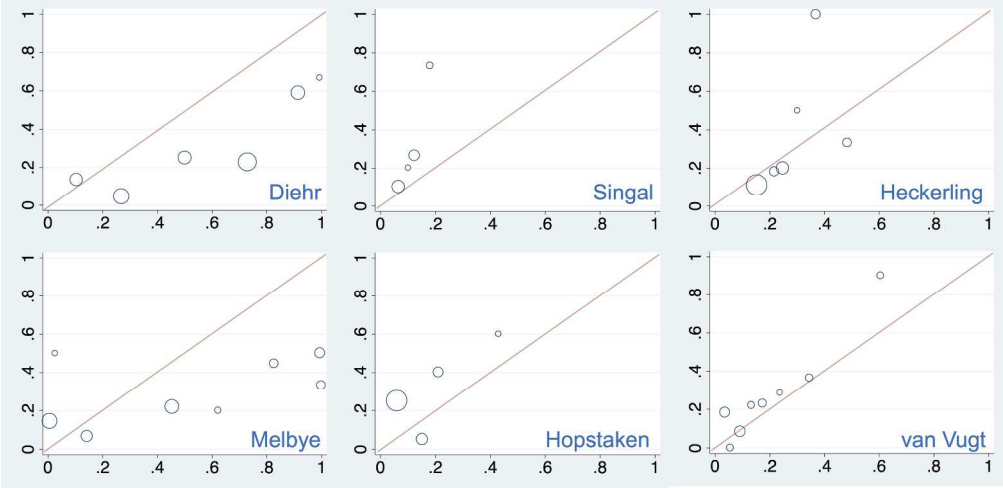
The extended model comprised the model by van Vugt with added information on appetite loss

Figure 3. Decision curve for the model by van Vugt and the extended model

The black line showed the net benefit for the strategy "CXR for all patients". The black dashed line for the strategy "No CXR", the grey dashed line for the model by van Vugt, and the yellow line for the extended model, which comprised the model by van Vugt with added information on appetite loss.

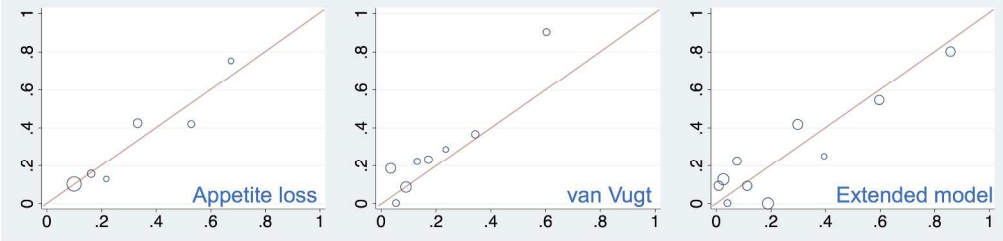
At a threshold of 10% to 30%, the extended model demonstrated higher net benefit than the original model.

For peer review only



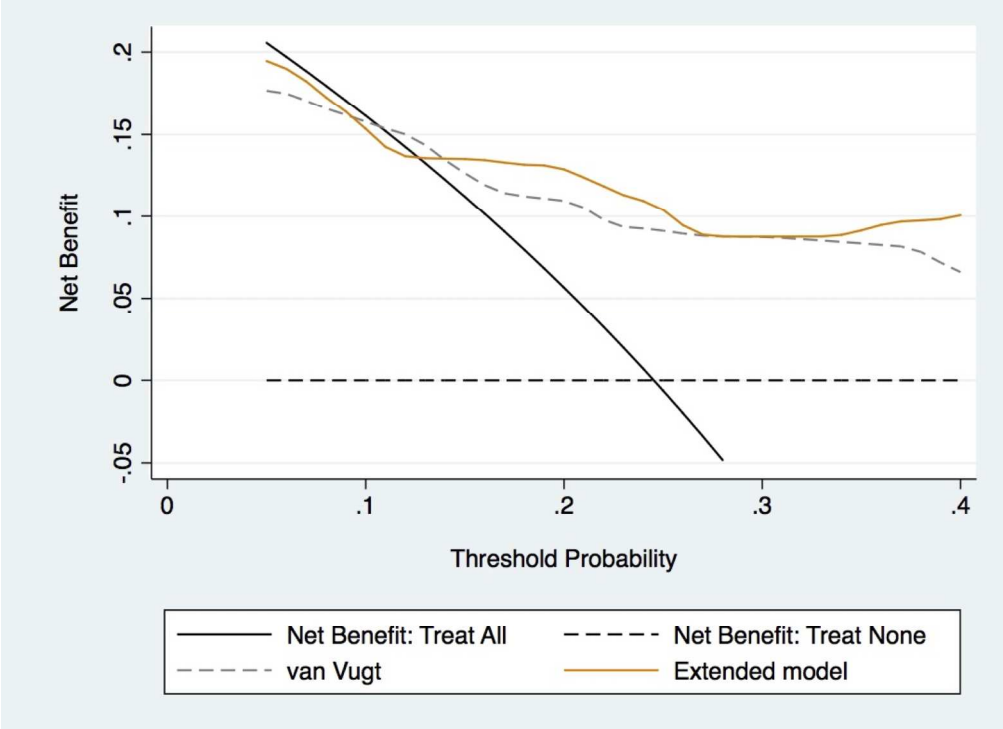
291x141mm (300 x 300 DPI)

For peer review only



291x70mm (300 x 300 DPI)

For peer review only



190x138mm (300 x 300 DPI)

For peer review only

TRIPOD Checklist: Prediction Model Validation

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	5
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	6
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	6
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	6
	5b	Describe eligibility criteria for participants.	6
	5c	Give details of treatments received, if relevant.	N/A
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7
	6b	Report any actions to blind assessment of the outcome to be predicted.	7
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	6
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	Explain how the study size was arrived at.	8
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	8
Statistical analysis methods	10c	For validation, describe how the predictions were calculated.	7
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	7,8
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	8
Risk groups	11	Provide details on how risk groups were created, if done.	N/A
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	6
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	9 Table 1
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Table 1
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	N/A
Model performance	16	Report performance measures (with CIs) for the prediction model.	Table 3, Figure 1
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance).	11
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	12,13
Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	11,12
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	11,12
Implications	20	Discuss the potential clinical use of the model and implications for future research.	11,12
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	14
Funding	22	Give the source of funding and the role of the funders for the present study.	14

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only

BMJ Open

Diagnostic utility of appetite loss in addition to existing prediction models for community-acquired pneumonia in the elderly: a prospective diagnostic study in acute care hospitals in Japan

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-019155.R1
Article Type:	Research
Date Submitted by the Author:	26-Sep-2017
Complete List of Authors:	Takada, Toshihiko; Fukushima Kenritsu Ika Daigaku, Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR) Yamamoto, Yosuke; Kyoto University, Department of Healthcare Epidemiology Terada, Kazuhiko; Kimitsu Chuo Byoin, Department of General Medicine Ohta, Mitsuyasu; Kanagawa Kenritsu Ashigarakami Byoin, Department of Internal Medicine Mikami, Wakako; Kanagawa-ken Keiyukai Keiyu Byoin, Department of Radiology Yokota, Hajime; Chiba Daigaku Igakubu Fuzoku Byoin, Department of Radiology Hayashi, Michio; Fukushima Kenritsu Ika Daigaku, Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR) Miyashita, Jun ; Fukushima Kenritsu Ika Daigaku, Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR); Kyoto University, Department of Healthcare Epidemiology Azuma, Teruhisa; Fukushima Kenritsu Ika Daigaku, Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR) Fukuma, Shingo; Kyoto University, Department of Healthcare Epidemiology Fukuhara, Shunichi; Kyoto University, Department of Healthcare Epidemiology
Primary Subject Heading:	Diagnostics
Secondary Subject Heading:	Infectious diseases, Geriatric medicine, Respiratory medicine
Keywords:	GENERAL MEDICINE (see Internal Medicine), INFECTIOUS DISEASES, PRIMARY CARE, Respiratory infections < THORACIC MEDICINE

SCHOLARONE™
Manuscripts

Diagnostic utility of appetite loss in addition to existing prediction models for community-acquired pneumonia in the elderly: a prospective diagnostic study in acute care hospitals in Japan

Toshihiko Takada,[1,2] Yosuke Yamamoto,[2] Kazuhiko Terada,[3] Mitsuyasu Ohta,[4] Wakako Mikami,[5] Hajime Yokota,[6] Michio Hayashi,[1] Jun Miyashita,[1,2] Teruhisa Azuma,[1] Shingo Fukuma,[1,2] Shunichi Fukuhara[1,2]

1 Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR), Fukushima Medical University, Fukushima, Japan

2 Department of Healthcare Epidemiology, School of Public Health in the Graduate School of Medicine, Kyoto University, Kyoto, Japan

3 Department of General Medicine, Kimitsu Chuo Hospital, Chiba, Japan

4 Department of Internal Medicine, Ashigarakami Hospital, Kanagawa, Japan

5 Department of Radiology, Keiyu Hospital, Kanagawa, Japan

6 Department of Radiology, Chiba University Hospital, Chiba, Japan

Corresponding author:

Toshihiko Takada, MD, MPH, PhD

Department of General Medicine, Shirakawa Satellite for Teaching And Research (STAR), Fukushima Medical University, 2-1 Toyochi Kamiyajiro, Shirakawa, Fukushima 961-0005, Japan

Tel: +81-(0)248-22-2211; Fax: +81-(0)248-22-2211

E-mail: ttakada@water.ocn.ne.jp

For peer review only

Abstract

Objective

Diagnosis of community-acquired pneumonia (CAP) in the elderly is often delayed because of atypical presentation and nonspecific symptoms, such as appetite loss, falls, and disturbance in consciousness. The aim of this study was to investigate the external validity of existing prediction models and the added value of the nonspecific symptoms for the diagnosis of CAP in elderly patients.

Design

Prospective cohort study.

Setting

General medicine departments of 3 teaching hospitals in Japan.

Participants

A total of 109 elderly patients who consulted for upper respiratory symptoms between October 1, 2014 and September 30, 2016.

Main outcome measures

The reference standard for CAP was chest radiograph evaluated by two certified radiologists. The existing models were externally validated for diagnostic performance by calibration plot and discrimination. To evaluate the additional value of the nonspecific symptoms to the existing prediction models, we developed an extended logistic regression model. Calibration, discrimination, category-free net reclassification improvement (NRI), and decision curve analysis (DCA) were investigated in the extended model.

For peer review only

Results

Among the existing models, the model by van Vugt demonstrated the best performance, with an area under the curve (AUC) of 0.75 [95% confidence interval (CI) 0.63–0.88]; calibration plot showed good fit despite a significant Hosmer–Lemeshow test (P = 0.017). Among the nonspecific symptoms, appetite loss had positive likelihood ratio of 3.2 (2.0–5.3), negative likelihood ratio of 0.4 (0.2–0.7), and odds ratio of 7.7 (3.0–19.7). Addition of appetite loss to the model by van Vugt led to improved calibration at P = 0.48, NRI of 0.53 (P = 0.019), and higher net benefit by DCA.

Conclusions

Information on appetite loss improved the performance of an existing model for the diagnosis of CAP in the elderly.

For peer review only

Strengths and limitations of this study

This was the first study to investigate the external validity of existing prediction models for the diagnosis of CAP in the elderly.

To evaluate the additional value of appetite loss to existing prediction models, we developed an extended logistic regression model, which was evaluated by net reclassification improvement and decision curve analysis.

To explore the external validity of our findings, a similar investigation on a larger sample size in a primary care setting should be further conducted.

For peer review only

Introduction

In elderly patients, community-acquired pneumonia (CAP) could be more severe than that in younger patients and could result in poor prognosis when treatment is delayed.[1] Therefore, several predictors of mortality in elderly patients with CAP have been reported to identify individuals at risk for poor outcomes and to initiate early intervention.[2] However, in the elderly, the diagnosis of CAP is often delayed because of their atypical presentation and underlying comorbidities.[1 3] In particular, the common symptoms of cough, sputum production, fever, chills, rigors, and chest pain could be absent and be replaced by nonspecific deterioration of their general condition, which can manifest as appetite loss, falls, consciousness disturbance, and so on.[1] Likewise, physical examination is less reliable in the elderly than in younger individuals.[4] For example, asymptomatic elderly patients may have chest examination findings of crackles, which could be age-related.[5] In addition to such challenges in the diagnosis of CAP, routine use of chest x-ray (CXR) for all elderly patients with respiratory symptoms is time-consuming and might not be cost-effective.

Although several prediction models based on signs and symptoms are available for the diagnosis of CAP,[6–12] none had been developed or validated specifically for elderly patients. Because of the atypical presentation of CAP in the elderly, the performance of these models could be poor. We hypothesized that addition of the nonspecific symptoms that are characteristic of CAP in elderly patients may improve the diagnostic performance of these existing models.

Accordingly, we investigated the external validity of these existing models and evaluated the value of adding the nonspecific symptoms for the diagnosis of CAP in elderly patients.

Materials and Methods

For peer review only

This was a prospective observational study, that followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) statement for prediction model studies.[13] Written informed consent was obtained from all the subjects.

Setting

This study was conducted between October 1, 2014 and September 30, 2016 at the general medicine departments of 3 teaching hospitals: Shirakawa Kosei General Hospital in Fukushima, Japan (471-bed capacity); Kimitsu Chuo Hospital in Chiba, Japan (661-bed capacity); and Ashigarakami Hospital in Kanagawa, Japan (296-bed capacity).

Patients

We included ≥65-year-old outpatients who presented with a chief complaint of cough or sputum production. Exclusion criteria were 1) longer than 1 month duration of cough or sputum production; 2) current intake of antibiotics; 3) did not reside in "community" (Patients who resided in nursing homes or those transferred from another hospital were excluded because epidemiology has been reported to differ among CAP, health-care-associated pneumonia, and hospital-acquired pneumonia.[14]

Data collection

Using a structured collection form, physicians collected data upon examination of eligible patients. When patients themselves were unable to answer the questions, physicians collected data from their caregivers. The items in the form included the predictors in existing models: 1) history [cough, sputum production, sore throat, coryza, dyspnea, fever, chills, night sweats, myalgia, pleurisy, diarrhea, duration of symptoms, and past history of bronchial asthma (BA), chronic obstructive pulmonary disease

For peer review only

(COPD), chronic heart failure (CHF), or ischemic heart disease (IHD)] and 2) signs (blood pressure, pulse rate, oxygen saturation, respiratory rate, body temperature, diminished breath sounds, wheezes, crackles, and rhonchi).[6–11] We added items on appetite loss, falls, and consciousness disturbance. Appetite loss was based on self-assessment of 24-h food intake in proportion to the usual intake.

Outcome measurement

The reference standard for CAP diagnosis was CXR in the posteroanterior and lateral views or in the anteroposterior view for patients who could not stand up. Two certified radiologists (WK and HY) who were blinded to the history and physical examination data independently assessed the CXRs. Any disagreement was resolved by discussion.

Existing prediction models for CAP

Existing models were identified based on a previous external validation study on the prediction models for CAP in the primary care setting.[12] Inclusion criteria for these existing models were 1) use of logistic regression for the diagnosis of CAP in adult patients; 2) predictors were items about history and physical examination; 3) the diagnosis of CAP was made by CXR or computed tomography (CT).

Statistical analysis

The existing models in their original form and without adjustment of intercept and coefficients were externally validated for diagnostic performance by calibration plot and discrimination. The calibration plots were tested with the Hosmer–Lemeshow (HL) test and a *P* value of less than 0.05 indicated a lack of good fit.[13] Discrimination was assessed by area under the curve (AUC).

The diagnostic performance of each item in the existing models, as well as that of the nonspecific symptoms (appetite loss, falls and consciousness disturbance),

For peer review only

were assessed in terms of sensitivity, specificity, positive likelihood ratio (LR+), and negative likelihood ratio (LR−), all of which were reported as point estimates of diagnostic odds ratio (DOR) with 95% confidence interval (CI). DOR ranged from zero to infinity and was calculated as the ratio of LR+ to LR−, with higher values indicative of better discriminative performance; a value of 1 indicated that the test did not discriminate between people with and without the disease.[15] The cut-off points for continuous variables were determined according to existing prediction models[16] or the values with the best sensitivity and specificity in the receiver operating characteristic (ROC) curve. The diagnostic performance of the logistic regression model that included appetite loss as the sole predictor was evaluated by calibration plot.

To evaluate the additional value of these nonspecific symptoms to the existing prediction models, we developed an extended logistic regression model without changing the coefficients of the existing models. Calibration, discrimination, category-free net reclassification improvement (NRI), and decision curve analysis (DCA) were applied on the extended model. The NRI was the cumulative net proportion of events reclassified correctly plus the net proportion of non-events reclassified correctly.[17] The DCA was a graphical approach to evaluate the prediction models based on the principle that the relative harm of false positives and false negatives can be expressed in terms of a probability threshold. Net benefit was obtained by subtracting the proportion of false positives from the proportion of true positives, weighed by the relative harm of false positive and false negative results.[18] Net benefit was assessed by the probability threshold, which was decided on 10 physicians who worked in the participating hospitals and who were blinded to the results of our research.

To assess the agreement in the CXR interpretations of the 2 radiologists, Cohen's Kappa coefficient (κ) was calculated.[19] Imputation method was not applied for missing values because of the possibility of perfect prediction of categorical data. We

For peer review only

did not estimate sample size *a priori* because there was no available consensus on the adequate sample size among the external validation studies on the prediction models or evaluation studies on the diagnostic performance of the added predictors.[13] Statistical analyses were performed with a commercial software program (STATA, version 14.2 SE; StataCorp LP, College Station, Texas).

Results

Patient characteristics

A total of 109 patients were included in the study population. The baseline characteristics of the patients are listed in Table 1; mean age was 76.7 years [standard deviation (SD) 7.8] and 58.7% were men. CAP was diagnosed by CXR in 22% of patients. There was moderate agreement (κ = 0.58) between the 2 radiologists who read the CXRs.

External validation of existing models

Six existing prediction models for CAP diagnosis were externally validated in our cohort. The formula for each model is presented in the supplementary table.

The diagnostic performance of each predictor item in the history and physical examination is presented in Table 2. For each model, a calibration plot was drawn by comparing the predicted probability with the observed probability in our dataset (Fig. 1). All existing models demonstrated significant *P* values in the HL test, indicating poor fit (Table 3). On the other hand, the calibration plot for the model by van Vugt[11] visually demonstrated good fit between the predicted probability and the observed prevalence of CAP. Discrimination was highest for the model by Singal[7] (AUC = 0.76; 95% CI 0.65–0.87), followed by the models by van Vugt[11] (AUC = 0.75; 95% CI 0.63–0.88); Diehr[6] (AUC = 0.75; 95% CI 0.64–0.86); and Heckerling[8] (AUC = 0.75; 95% CI 0.64–0.86) (Table 3).

For peer review only

Diagnostic performance of the nonspecific symptoms

The sensitivity, specificity, LR+, and LR− for appetite loss, falls, and consciousness disturbance are listed in Table 2. The cut-off point for appetite loss was determined at 50% by the ROC curve. The AUC of appetite loss for CAP diagnosis was 0.76 (95% CI 0.65–0.87). The calibration plot of the model for appetite loss showed better fit than the model by van Vugt,[11] which showed the best calibration among the existing models (Fig. 2). A non-significant HL test suggested good fit of the model for appetite loss (Table 3).

Added value of the nonspecific symptoms

In the evaluation of the diagnostic performance of each nonspecific symptom, falls and consciousness disturbance seemed less useful; therefore, we investigated the added value of appetite loss to the model by van Vugt,[11] which showed the best performance among the existing models.

The formula for the extended model included the following: −5.258, −0.135 for appetite, +0.446 for dyspnea, +0.698 for absence of runny nose, +0.596 for diminished vesicular breathing, +1.404 for crackles, +0.961 for tachycardia, and +0.980 for temperature >37.8 °C. The extended model showed improved performance on the calibration plot, especially in the subgroup with high probability of CAP (Fig. 2), and good fit by the HL test ($P = 0.48$; $\chi^2 = 8.6$). Compared with the original model, the extended model tended to have an improved AUC, but this change was not significant [0.77 (95% CI 0.65–0.89) vs. 0.75 (95% CI 0.63–0.88); $P = 0.48$]. The NRI of the extended model was 0.53 (95% CI 0.08–0.97; $P = 0.019$), which meant that 53% of patients correctly obtained increased probability of CAP.

Fig. 3 illustrates the decision curves for the model by van Vugt[11] and the extended model. Among the 10 physicians, the mean threshold for ordering CXR for the diagnosis of CAP in the elderly was 22.5% (SD, 7.1%; range 10–30%). At a threshold of 10% to 30%, a higher net benefit was more frequently observed in the

For peer review only

extended model than in the original model. At a threshold of 20%, the net benefit of the extended model increased by 0.02 compared with the original model.

Discussion

Our study assessed the external validity of existing models for the diagnosis of CAP in an elderly population at acute care hospitals in Japan. These existing models had been validated in a meta-analysis of individual data of patients who were at least 18 years old.[12] Although the existing models were supposed have poor external validity in the elderly population, the model by van Vugt[11] showed relatively good calibration and discrimination. In the evaluation of the diagnostic performance of the nonspecific symptoms of CAP in the elderly, falls and consciousness disturbance showed high specificity but low sensitivity, leading to clinically irrelevant LR+ and LR−. On the other hand, information about appetite showed good performance for the diagnosis of CAP in the elderly. Nurse-assessed food consumption has been reported to predict bacteremia with a sensitivity of 0.92 and negative predictive value of 0.98.[20] Therefore, appetite loss could be an important predictor of infection. In the assessment of the added value of appetite loss to the model by van Vugt,[11] calibration was improved, especially in the population with high probability; however, the improvement in discrimination was not significant. NRI was significant as more than 50% of the patients were correctly assessed to have increased probabilities. Results of the DCA showed higher net benefit of the extended model at a threshold appetite loss of 10% to 30%. For example, applying the extended model at a threshold appetite loss of 20% could correctly detect CAP in 2 additional patients per 100 patients.

The strength of our study was that this was the first study that evaluated the external validity of existing models for CAP in the elderly. Based on our results, the model by van Vugt[11] can be applied in the elderly population despite its relatively poor calibration. To the best of our knowledge, this was the first study to investigate and

For peer review only

identify the value of information on appetite loss, independently and in addition to the model by van Vugt,[11] for the diagnosis of CAP in the elderly. Because routine use of CXR might not be time-efficient or cost-effective, estimating the diagnostic probability using information on appetite loss could be an effective strategy in elderly patients with respiratory symptoms. The current study was conducted in hospital settings where CXR was easily available, but 3 of the 6 existing models were derived in the hospital setting.[6–8] Therefore, we believe that there was a need to correctly estimate the diagnostic probability of CAP in both hospital and primary care settings.

This study had several limitations. First, our sample size was relatively small. Because this study was conducted in acute care hospitals, many patients were ineligible because they were already prescribed antibiotics in the primary care setting. Therefore, a similar investigation on a larger sample size in a primary care setting should be further explored. Second, the frequency of CAP was relatively higher than that previously reported,[12] probably because the number of patients with CAP was higher in the acute care setting than in the primary care setting. Therefore, the external validity of our results should be applied carefully in the primary care setting. Healthcare in Japan is a free-access system that allows people to be examined and treated at the medical institutions (i.e., clinics, secondary hospitals, and university hospitals) of their choice.[21] Therefore, the differences in patient characteristics between the hospital and primary care settings might be less prominent in Japan than those in other countries where patients have to see their primary care physicians first. Third, although CT has been reported to be superior in terms of diagnostic accuracy,[22] CXR was used as the reference standard of CAP in this study because performing CT on all patients was not feasible and ethical due to difficult access, higher radiation exposure, and cost.

In conclusion, the model by van Vugt[11] demonstrated relatively high performance among the existing models for the diagnosis of CAP in the elderly.

For peer review only

Information about appetite loss independently demonstrated diagnostic utility for CAP in the elderly and improved the performance of the model by van Vugt.[11]

For peer review only

Contributors:

TT had full access to all of the data in the study; he takes responsibility for the integrity of the data and accuracy of the data analysis and wrote the first draft. YY designed the study, interpreted the data, and drafted the paper. KT, MO, WM, and HY collected and interpreted the data. MH, JM, and TA designed the study, and collected and interpreted the data. S Fukuma and S Fukuhara supervised the research and revised the work critically for important intellectual content.

Funding:

This work was supported by a Grant-in-Aid for Epidemiological Research of St. Luke's Life Science Institute.

Competing interests:

None declared.

Ethics approval:

Approval was granted by the ethics committees of Kyoto University Graduate School and Faculty of Medicine, Shirakawa Kosei General Hospital, Kimitsu Chuo Hospital, and Ashigarakami Hospital.

Data sharing statement:

No additional data are available.

For peer review only

References

1. Marrie TJ. Community-acquired pneumonia in the elderly. *Clin Infect Dis* 2000;**31**:1066–78.
2. Cataudella E, Giraffa CM, Di Marca S, et al. Neutrophil-To-Lymphocyte Ratio: An Emerging Marker Predicting Prognosis in Elderly Adults with Community-Acquired Pneumonia. *J Am Geriatr Soc* 2017;**65**:1796–801.
3. Berman P, Hogan DB, Fox RA. The atypical presentation of infection in old age. *Age Ageing* 1987;**16**:201–7.
4. Thiem U, Heppner HJ, Pientka L. Elderly patients with community-acquired pneumonia: optimal treatment strategies. *Drugs Aging* 2011;**28**:519–37.
5. Kataoka H, Matsuno O. Age-related pulmonary crackles (rales) in asymptomatic cardiovascular patients. *Ann Fam Med* 2008;**6**:239–45.
6. Diehr P, Wood RW, Bushyhead J, Krueger L, Wolcott B, Tompkins RK. Prediction of pneumonia in outpatients with acute cough--a statistical approach. *J Chronic Dis* 1984;**37**:215–25.
7. Singal BM, Hedges JR, Radack KL. Decision rules and clinical prediction of pneumonia: evaluation of low-yield criteria. *Ann Emerg Med* 1989;**18**:13–20.
8. Heckerling PS, Tape TG, Wigton RS, et al. Clinical prediction rule for pulmonary infiltrates. *Ann Intern Med* 1990;**113**:664–70.
9. Melbye H, Straume B, Aasebo U, Dale K. Diagnosis of pneumonia in adults in general practice. Relative importance of typical symptoms and abnormal chest signs evaluated against a radiographic reference standard. *Scand J Prim Health Care* 1992;**10**:226–33.
10. Hopstaken RM, Muris JW, Knottnerus JA, Kester AD, Rinkens PE, Dinant GJ. Contributions of symptoms, signs, erythrocyte sedimentation rate, and C-reactive protein to a diagnosis of pneumonia in acute lower respiratory tract infection. *Br J Gen Pract* 2003;**53**:358–64.

For peer review only

11. van Vugt SF, Broekhuizen BD, Lammens C, et al. Use of serum C reactive protein and procalcitonin concentrations in addition to symptoms and signs to predict pneumonia in patients presenting to primary care with acute cough: diagnostic study. *BMJ* 2013;**346**:f2450.
12. Schierenberg A, Minnaard MC, Hopstaken RM, et al. External Validation of Prediction Models for Pneumonia in Primary Care Patients with Lower Respiratory Tract Infection: An Individual Patient Data Meta-Analysis. *PLoS One* 2016;**11**:e0149895.
13. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;**162**:W1–73.
14. Kollef MH, Shorr A, Tabak YP, et al. Epidemiology and outcomes of health-care-associated pneumonia: results from a large US database of culture-positive pneumonia. *Chest* 2005;**128**:3854–62.
15. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003;**56**:1129–35.
16. 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.
17. Pencina MJ, D'Agostino RB, Sr., Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;**30**:11–21.
18. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565–74.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;**1**:307–10.

For peer review only

20. Komatsu T, Onda T, Murayama G, et al. Predicting bacteremia based on nurse-assessed food consumption at the time of blood culture. *J Hosp Med* 2012;**7**:702–5.

21. Toyabe S. Non-specialized inpatient care provided by university hospitals in Japan. *Risk Manag Healthc Policy* 2008;**1**:23–9.

22. Syrjala H, Broas M, Suramo I, Ojala A, Lahde S. High-resolution computed tomography for the diagnosis of community-acquired pneumonia. *Clin Infect Dis* 1998;**27**:358–63.

For peer review only

Tables

Table 1. Patients' characteristics

Age (Mean ± SD)	76.7±7.8
Male (%)	64 (58.7)
Past history of BA/COPD (%)	6 (5.5)
Past history of IHD/CHF (%)	14 (12.9)
Clinical diagnosis (%)	
CAP	32 (29.4)
Upper respiratory infection	61 (56.0)
Sinusitis	9 (8.2)
Others	7 (6.4)
CAP diagnosed by CXR (%)	24 (22.0)

SD, standard deviation; BA, bronchial asthma; COPD, chronic obstructive pulmonary disease; IHD, ischemic heart disease; CHF, chronic heart failure; CAP, community-acquired pneumonia; CXR, chest x ray

For peer review only

Table 2. Diagnostic performance of the predictors of CAP in the elderly

	CAP+ (N=27)	CAP- (N=82)	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)	Cut-off point
cough (%)	100	98.8	100 (87.2-100)	1.2 (0.3-6.6)	1.0 (1.0-1.0)	0	-	-
sputum (%)	88.9	74.1	88.9 (70.8-97.6)	25.9 (16.8-36.9)	1.2 (1.0-1.4)	0.4 (0.1-1.3)	2.8 (0.8-9.6)	-
sore throat (%)	22.2	40.2	22.2 (8.6-42.3)	59.8 (48.3-70.4)	0.6 (0.3-1.2)	1.3 (1.0-1.7)	0.4 (0.2-1.1)	-
rhinorrhea (%)	34.6	60.5	34.6 (17.2-55.7)	39.5 (28.8-51.0)	0.6 (0.3-1.0)	1.7 (1.1-2.4)	0.3 (0.1-0.9)	-
dyspnea (%)	40.7	34.2	40.1 (22.4-61.2)	65.9 (54.6-76.0)	1.2 (0.7-2.1)	0.9 (0.6-1.3)	1.3 (0.6-3.2)	-
fever (%)	74.1	35.4	74.1 (53.7-88.9)	64.6 (53.3-74.9)	2.1 (1.5-3.0)	0.4 (0.2-0.8)	5.2 (2.0-13.5)	-
chill (%)	29.6	27.2	29.6 (13.8-50.2)	71.6 (60.5-81.1)	1.0 (0.5-2.1)	0.98 (0.7-1.3)	1.1 (0.4-2.7)	-
night sweats (%)	33.3	24.4	33.3 (16.5-54.0)	75.6 (64.9-84.4)	1.4 (0.7-2.6)	0.9 (0.7-1.2)	1.5 (0.6-3.9)	-
myalgia (%)	7.4	18.3	7.4 (0.9-24.3)	81.7 (71.6-89.4)	0.4 (0.1-1.7)	1.1 (1.0-1.3)	0.4 (0-1.5)	-
pleural pain (%)	7.4	7.4	7.4 (0.9-24.3)	92.6 (84.6-97.2)	1.0 (0.2-4.7)	1.0 (0.9-1.1)	1.0 (0-4.7)	-
diarrhea (%)	7.4	7.4	7.4 (0.9-24.3)	92.6 (84.6-97.2)	1.0 (0.2-4.7)	1.0 (0.9-1.1)	1.0 (0-4.7)	-
duration of symptoms (days)	6.1	7.2	42.3 (23.4-63.1)	62.5 (51.0-73.1)	1.1 (0.7-1.9)	0.9 (0.6-1.3)	1.2 (0.5-3.0)	7 days from the model by Melbye[9]
coma (%)	15.4	6.1	15.4 (4.4-34.9)	93.9 (86.3-98.0)	2.5 (0.7-8.7)	0.9 (0.8-1.1)	2.8 (0.7-10.6)	-
fall (%)	11.1	1.2	11.1 (2.4-29.2)	98.8 (93.4-100.0)	9.1 (1.0-84.0)	0.9 (0.8-1.0)	10.1 (1.4-∞)	-
appetite loss (%)	48.8	80.1	66.7 (46.0-83.5)	79.3 (68.9-87.4)	3.2 (2.0-5.3)	0.4 (0.2-0.7)	7.7 (3.0-19.7)	50% determined by the ROC curve
SBP(mmHg)	129.9	130.1	3.7 (0.1-19.0)	100 (95.6-100)	-	1.0 (0.9-1.0)	-	90mmHg from CURB65[16]
DBP (mmHg)	72.5	73.9	22.2 (8.6-42.3)	84.1 (74.4-91.3)	1.4 (0.6-3.3)	0.9 (0.7-1.2)	1.4 (0.5-4.4)	60mmHg from CURB65[16]
pulse rate (/min)	89.3	82.8	29.6 (13.8-50.2)	80.5 (70.3-88.4)	1.5 (0.7-3.2)	0.9 (0.7-1.1)	1.7 (0.7-4.6)	100/min from the model by van Vugt[11]
SpO2 (%)	93.7	96.4	51.9 (31.9-71.3)	91.5 (83.2-96.5)	6.1 (2.7-13.5)	0.5 (0.4-0.8)	11.5 (4.0-33.4)	95% as normal value
respiratory rate (/min)	21.3	19.0	3.7 (0.1-19.0)	100 (95.5-100)	-	1.0 (0.9-1.0)	-	30/min from CURB65[16]
body temperature (℃)	37.6	36.7	29.6 (13.8-50.2)	91.5 (83.2-96.5)	3.5 (1.4-8.7)	0.8 (0.6-1.0)	4.5 (1.5-13.6)	38°C from the model by Hopstaken[10]
decreased respiratory sound (%)	18.5	4.9	18.5 (6.3-38.1)	95.1 (87.8-98.6)	3.8 (1.1-13.0)	0.9 (0.7-1.0)	4.4 (1.2-16.4)	-
crackle (%)	48.2	14.8	48.1 (28.7-68.1)	85.2 (75.6-92.1)	3.3 (1.7-6.2)	0.6 (0.4-0.9)	5.3 (2.1-14.0)	-
wheeze (%)	3.7	3.7	3.7 (0.1-19.0)	96.3 (89.6-99.2)	1.0 (0.1-9.2)	1.0 (0.9-1.1)	1.0 (0-7.4)	-
rhonchi (%)	3.7	6.2	3.7 (0.1-19.0)	93.8 (86.2-98.0)	0.6 (0.1-4.9)	1.0 (0.9-1.1)	0.6 (0-4.0)	-

CAP: community-acquired pneumonia; CI: confidence interval; LR+: positive likelihood ratio; LR-: negative likelihood ratio; DOR: diagnostic odds ratio; SBP: systolic blood pressure; DBP: diastolic blood pressure; SpO2: oxygen saturation; ROC: receiver operating characteristic; CURB65: Confusion, urea > 7mmol/L, respiratory rate >= 30 breaths/min, low blood pressure (systolic < 90mmHg or diastolic <= 60 mmHg)

For peer review only

Table 3. Comparison of diagnostic performance among existing models and appetite loss

	Calibration (HL test)		Discrimination	
Model	$\chi 2$	*P* value	AUC	95% CI
Diehr[6]	134.9	<0.001	0.75	0.64-0.86
Singal[7]	39.2	<0.001	0.76	0.65-0.87
Heckerling[8]	19.8	0.003	0.75	0.64-0.86
Melbye[9]	6104.2	<0.001	0.64	0.51-0.77
Hopstaken[10]	48.2	<0.001	0.54	0.42-0.66
van Vugt[11]	18.7	0.017	0.75	0.63-0.88
Appetite loss	1.91	0.928	0.76	0.65-0.87

HL, Hosmer-Lemeshow; AUC, area under the curve; CI, confidence interval

For peer review only

Figure legends

Figure 1. Calibration plots of existing models

The X-axis represents the predicted proportion and the y-axis represents the observed proportion, with a reference for perfect calibration (red line). The size of the circle indicates the number of patients. The model by van Vugt visually demonstrated better fit than the other models.

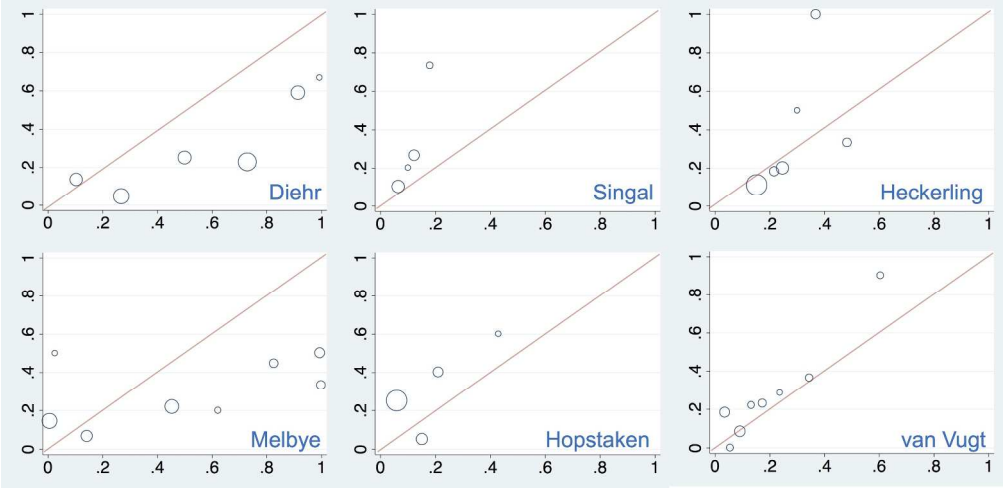
Figure 2. Comparison of the calibration plots among the model of appetite loss, the model by van Vugt, and the extended model

The extended model comprised the model by van Vugt with additional information on appetite loss

Figure 3. Decision curve for the model by van Vugt and the extended model

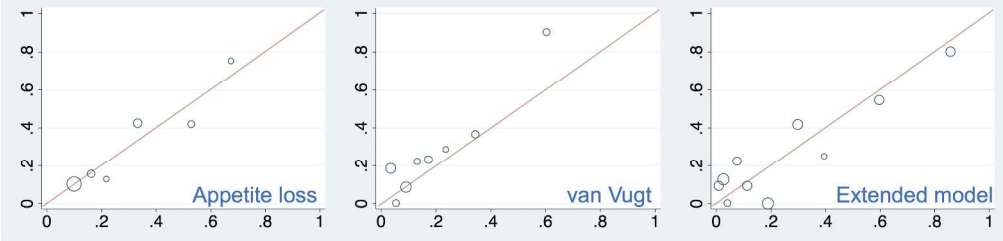
The net benefits for the strategy "CXR for all patients" (black line), for the strategy "No CXR" (black dashed line), for the model by van Vugt (gray dashed line), and for the extended model (yellow line), are shown. At a threshold of 10% to 30%, the extended model demonstrated higher net benefit than the original model.

For peer review only



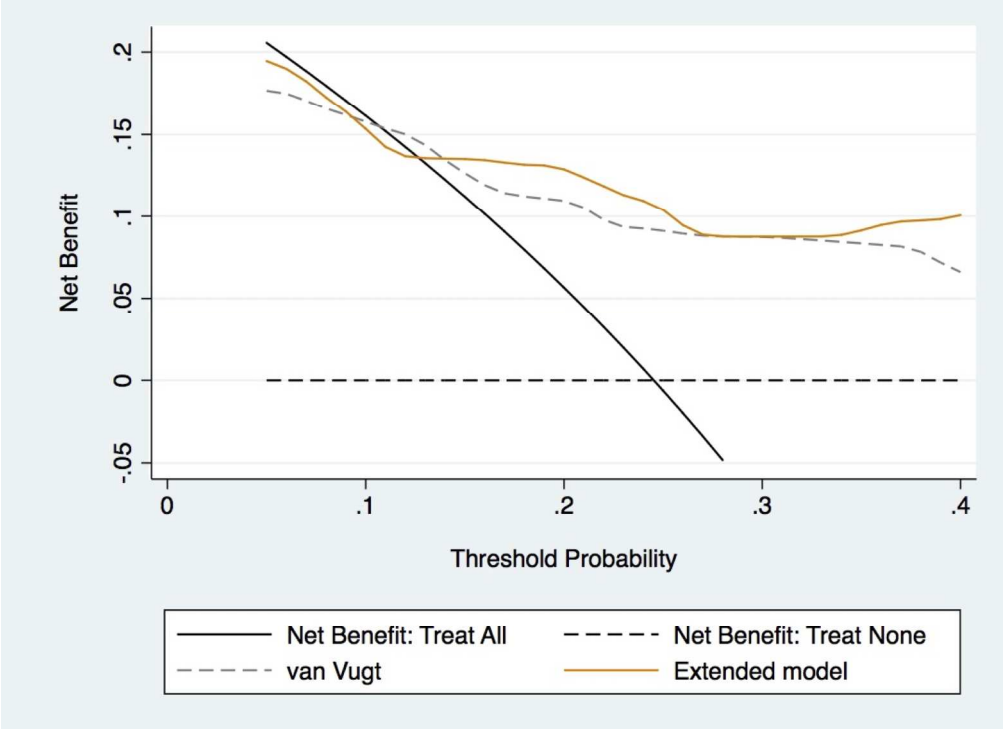
291x141mm (300 x 300 DPI)

For peer review only



291x70mm (300 x 300 DPI)

For peer review only



190x138mm (300 x 300 DPI)

For peer review only

Supplementary Table

The formula of existing prediction models for diagnosis of CAP

Diehr, 1984[6]

−2 for rhinorrhoea, −1 for sore throat, +1 for night sweats, +1 for myalgia, +1 for sputum, +2 for respiratory rate >25, +2 for temperature >37.7°C

Singal, 1989[7]

–3.539, +0.884 for cough, +0.681 for fever, +0.464 for crackles, +0.030 for 5.0 (pretest probability of pneumonia)

Heckerling, 1990[8]

–1.705, +0.494 for temperature >37.7°C, +0.428 for pulse >100 beats/min, +0.658 for rales, +0.638 for decreased breath sounds, +0.691 for absence of asthma

Melbye, 1992[9]

+ 4.7 for fever (reported by patient) with duration of illness of >1 week,
–4.5 for coryza, –2.1 for sore throat, +5.0 for dyspnea, +8.2 for lateral chest pain, +0.9 for crackles

Hopstaken, 2003[10]

–2.74, +1.02 for dry cough, +1.78 for diarrhea, +1.13 for temperature ≥38°C

van Vugt, 2013[11]

−3.984, +0.446 for dyspnea, +0.698 for absence of runny nose, +0.596 for diminished vesicular breathing, +1.404 for crackles, +0.961 for pulse>100 beats/min, +0.980 for temperature >37.8°C

For peer review only

TRIPOD Checklist: Prediction Model Validation

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	5
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	5
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	6
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	6
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	6
	5b	Describe eligibility criteria for participants.	6
	5c	Give details of treatments received, if relevant.	N/A
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	7
	6b	Report any actions to blind assessment of the outcome to be predicted.	7
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	6
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	Explain how the study size was arrived at.	8
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	8
Statistical analysis methods	10c	For validation, describe how the predictions were calculated.	7
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	7,8
	10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	8
Risk groups	11	Provide details on how risk groups were created, if done.	N/A
Development vs. validation	12	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	6
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	9 Table 1
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Table 1
	13c	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	N/A
Model performance	16	Report performance measures (with CIs) for the prediction model.	Table 3, Figure 1
Model-updating	17	If done, report the results from any model updating (i.e., model specification, model performance).	11
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	12,13
Interpretation	19a	For validation, discuss the results with reference to performance in the development data, and any other validation data.	11,12
	19b	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	11,12
Implications	20	Discuss the potential clinical use of the model and implications for future research.	11,12
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	14
Funding	22	Give the source of funding and the role of the funders for the present study.	14

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No.	Recommendation	Page No.
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	6
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-7
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	6
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	N/A
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	6-7
Bias	9	Describe any efforts to address potential sources of bias	7
Study size	10	Explain how the study size was arrived at	8-9

Continued on next page

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	7-8
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	7-8
		(*b*) Describe any methods used to examine subgroups and interactions	N/A
		(*c*) Explain how missing data were addressed	8
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	N/A
		(*e*) Describe any sensitivity analyses	N/A
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9
		(b) Give reasons for non-participation at each stage	N/A
		(c) Consider use of a flow diagram	N/A
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	9, Table 1
		(b) Indicate number of participants with missing data for each variable of interest	N/A
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	N/A
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	N/A
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	N/A
		Cross-sectional study—Report numbers of outcome events or summary measures	9, Table 1
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	9-11, Table 2, 3
		(*b*) Report category boundaries when continuous variables were categorized	Table 2
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N/A

Continued on next page

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	N/A
Discussion			
Key results	18	Summarise key results with reference to study objectives	11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	12
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	12-13
Generalisability	21	Discuss the generalisability (external validity) of the study results	12
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	14

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml