Review Article

Reporting of coronavirus disease 2019 prognostic models: the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement

Liuqing Yang[1,2], Qiang Wang[1,2], Tingting Cui[1,2], Jinxin Huang[1,2], Naiyang Shi[1,2], Hui Jin[1,2]

[1]Department of Epidemiology and Health Statistics, School of Public Health, Southeast University, Nanjing, China; [2]Key Laboratory of Environmental Medicine Engineering, Ministry of Education, School of Public Health, Southeast University, Nanjing, China

Contributions: (I) Conception and design: L Yang, H Jin; (II) Administrative support: H Jin; (III) Provision of study materials or patients: H Jin, Q Wang, T Cui; (IV) Collection and assembly of data: T Cui, Q Wang, J Huang; (V) Data analysis and interpretation: L Yang, N Shi, J Huang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Hui Jin. Department of Epidemiology and Health Statistics, School of Public Health, Southeast University, 87# Dingjiaqiao, Nanjing 210009, China. Email: jinhui_hld@163.com.

Abstract: Evaluation of the validity and applicability of published prognostic prediction models for coronavirus disease 2019 (COVID-19) is essential, because determining the patients' prognosis at an early stage may reduce mortality. This study was aimed to utilize the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) to report the completeness of COVID-19-related prognostic models and appraise its effectiveness in clinical practice. A systematic search of the Web of Science and PubMed was performed for studies published until August 11, 2020. All models were assessed on model development, external validation of existing models, incremental values, and development and validation of the same model. TRIPOD was used to assess the completeness of included models, and the completeness of each item was also reported. In total, 52 publications were included, including 67 models. Age, disease history, lymphoma count, history of hypertension and cardiovascular disease, C-reactive protein, lactate dehydrogenase, white blood cell count, and platelet count were the commonly used predictors. The predicted outcome was death, development of severe or critical state, survival time, and length-of-hospital stay. The reported discrimination performance of all models ranged from 0.361 to 0.994, while few models reported calibration. Overall, the reporting completeness based on TRIPOD was between 31% and 83% [median, 67% (interquartile range: 62%, 73%)]. Blinding of the outcome to be predicted or predictors were poorly reported. Additionally, there was little description on the handling of missing data. This assessment indicated a poorly-reported COVID-19 prognostic model in existing literature. The risk of over-fitting may exist with these models. The reporting of calibration and external validation should be given more attention in future research.

Keywords: Coronavirus disease 2019 (COVID-19); prognostic model; transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)

Submitted Oct 15, 2020. Accepted for publication Jan 17, 2021.
doi: 10.21037/atm-20-6933
View this article at: http://dx.doi.org/10.21037/atm-20-6933

Introduction

The novel coronavirus disease 2019 (COVID-19) poses an urgent threat to global health. As of August 28, 2020; 24,299,923 confirmed cases of COVID-19, including 827,730 deaths, were reported to the World Health Organization (WHO) (1). The huge number of infected cases brought tremendous pressure on the medical facilities. In addition to the high risk of infection to the medical

staff, effectively allocating resources, such as the number of intensive care unit (ICU) beds or other medical equipment, is also a challenge. According to existing reports, many infected patients show mild flu-like symptoms and can recover quickly (2). However, some rapidly develop acute respiratory distress syndrome, multiple organ failure, and death (3-6). Therefore, a current concern is to determine the patients' prognosis at an early stage, to reduce mortality. To provide the patients with the most reasonable level of treatment and care, many studies have combined multiple predictors to establish models, to predict the patients' prognosis in clinical practice, but the quality of these reports has not been evaluated (7-9). Complete reporting is benefit to study replication and assess the applicability to other individuals. Therefore, high-quality reporting about prediction model is essential. In 2015, multiple journals simultaneously published a study on how to improve the quality of reports on prediction model studies, namely transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement (10). TRIPOD is a list of 22 items involving title and abstract (items 1 and 2), background and objectives (item 3), methods (items 4 through 12), results (items 13 through 17), discussion (items 18 through 20), and other information (items 21 and 22). The TRIPOD statement covers the development and external validation of prediction models as well as studies with only external validation (updates with or without predictors).

A previous systematic review showed unsatisfactory level of quality of prediction models in various clinical fields (11). Wynants *et al.* also conducted a systematic review of the prediction models in COVID-19 (12). However, the results were qualitative, and no unified indicator to measure and compare the reporting integrity between different studies was reported. Our study provides a new evaluation method for model reporting, and summarizes the omissions commonly existing in current reporting, so that future research can focus on avoiding these problems to improve the quality of model reporting.

Our research aimed to use the TRIPOD tool to systematically review and critically evaluate the published models for predicting the prognosis or course of COVID-19 in patients. The results could provide the key for further improvement of the quality of COVID-19-related prognostic model reporting. We present the following article in accordance with the PRISMA reporting checklist (available at http://dx.doi.org/10.21037/atm-20-6933).

Methods

Search strategy

A search was conducted in PubMed and Web of Science databases until August 11, 2020, with no language restrictions. The terms related to COVID-19 (COVID-19, SARS-COV-2, novel corona, 2019-ncov) and prognostic model (prognostic, prediction model, regression) were searched in the databases. We also searched for reviews in this field and references of the original articles, to identify whether there were any missed studies. Only peer-reviewed studies on the prognostic model of COVID-19 were included in our research, and the preprint form was not considered.

Inclusion and exclusion criteria

We included articles on multivariate models or risk scores for predicting any prognostic outcomes of COVID-19. The exclusion criteria were as follows: (I) non-human research; (II) studies on the prediction model of disease transmission; (III) diagnostic model of COVID-19; (IV) studies on predictive factors but with no established prognostic models; (V) studies on prediction models using non-regression techniques; since TRIPOD does not support the evaluation of such methods (e.g., machine learning, neural networks) (13). Studies based on the above criteria were screened by two investigators (LQY and QW), and differences were resolved after discussion.

Data extraction

Two investigators (LQY and TTC) independently reviewed the titles and abstracts of all extracted articles. Any discrepancies were agreed upon through discussion and, if necessary, resolved by a consultant (HJ). Investigators used TRIPOD standard data extraction forms to determine the completeness of articles (www.tripod-statement.org). Additionally, the publications were grouped into four types of prediction models: development, external validation of existing models, incremental values, and development and validation of the same model. Publications could be classified into more than one type of prediction model.

In other words, for the development model, if different models were developed using the same data in one study, we extracted information from the primary model. For external validation of different existing models, information was extracted separately. Studies that reported both development and external validation of different models were classified

into both development and external validation models. The basic information of each study (study region, study design, sample size, and predicted outcomes) were extracted. In addition, information about predictors were addressed in the articles. Predictors refer to variables that are included in the model at the time of model construction and that build statistical relationships with predicted outcomes. Previous researchers encourage that age, sex, C-reactive protein, lactic dehydrogenase, lymphocyte count, and potentially features derived from CT-scoring should be included in the COVID-19 prognostic model (12). Similarly, we extracted the prediction performance, including discrimination and calibration and their standard error (SE) or 95% confidence interval (CI), if provided. Discrimination was usually measured by the area under the receiver operator characteristic curve (AUROC) or c-index, while calibration was usually quantified by calibration intercept and calibration slope. The closer the AUROC or c-index and calibration slope is to 1, the better the performance of the model. The performance data were extracted in the following order: external validation, internal validation, and original performance (if the two above were not included).

Analysis

To evaluate the completeness of included models, the number of TRIPOD items that were completely reported was divided by the total number of TRIPOD items in the article. Furthermore, to assess the overall reporting completeness of each item in the TRIPOD statement, we divided the number of models with complete reports for a specific TRIPOD item by the total number of models applicable to this item. To evaluate for completeness, if an item was not considered applicable to a study, the five items declared by TRIPOD included "if completed" or "if applicable" statements (items 5c, 10e, 11, 14b, and 17). Then, such items were excluded from both the numerator and denominator.

In validation, the random effect model was used to pool the presented prediction performance with their 95% CI in the meta-analysis. The I^2 statistic was used to assess the heterogeneity among the studies. When I^2 statistic was >50% (moderate heterogeneity), the random effect model was used for the analysis.

Results

After screening, a total of 52 publications were included in our study (*Figure 1*). From the 52 publications, we scored 67 models using the TRIPOD tool as follows: 37 (55%) development, 14 (21%) external validation of existing models, 3 (5%) incremental values, and 13 (19%) development and validation of the same model.

Primary information

Thirty-six studies used COVID-19 patients' data from China, four from Italy, and two from the United States. Britain, France, Norway, Turkey, Spain, and Mexico had one each. Four studies did not specify the country or region of the data. Regarding the study design, most (88%) were retrospective studies, while two were prospective studies. One study used retrospective data in model development, but prospective methods in a validation cohort to recruit patients. One study identified the race of the participants as Caucasian (8). In a total of 23 studies, the follow-up date was mentioned. All the studies reported the sample sizes (median sample size, 220.5 [interquartile range (IQR): 109.25, 459.25]. Detailed information is shown in *Table 1* and Appendix 1.

Prognostic predictors

In the final model, six studies used computed tomography (CT) or chest X-ray results to establish the scoring rules. The median number of prognostic predictors was five (IQR: 3, 6.25). The most frequently used predictors in the model (>10 times) were as follows: age, disease history, lymphocyte count, history of hypertension and cardiovascular disease, C reactive protein, lactate dehydrogenase, white blood cell count, and platelet count, reported 26 (50%), 17 (33%), 14 (27%), 12 (23%), 12 (23%), 11 (21%), 10 (19%), and 10 (19%) times, respectively. The commonly used predictors (>5 times) were as follows: lymphocyte ratio, procalcitonin, aspartate aminotransferase, and dyspnea reported 8 (15%), 5 (10%), 5 (10%), and 5 (10%) times, respectively (Appendix 2).

Prediction outcomes and performances

The prediction outcomes in 23, 17, 8, 2, and 2 studies were death, severe or critical state disease development, ICU admission/mechanical ventilation/death, survival time, and length-of-hospital stay, respectively (*Table 1*). For death, the reported discrimination performance ranged from 0.584 to 0.994. Another study reported the weighted kappa

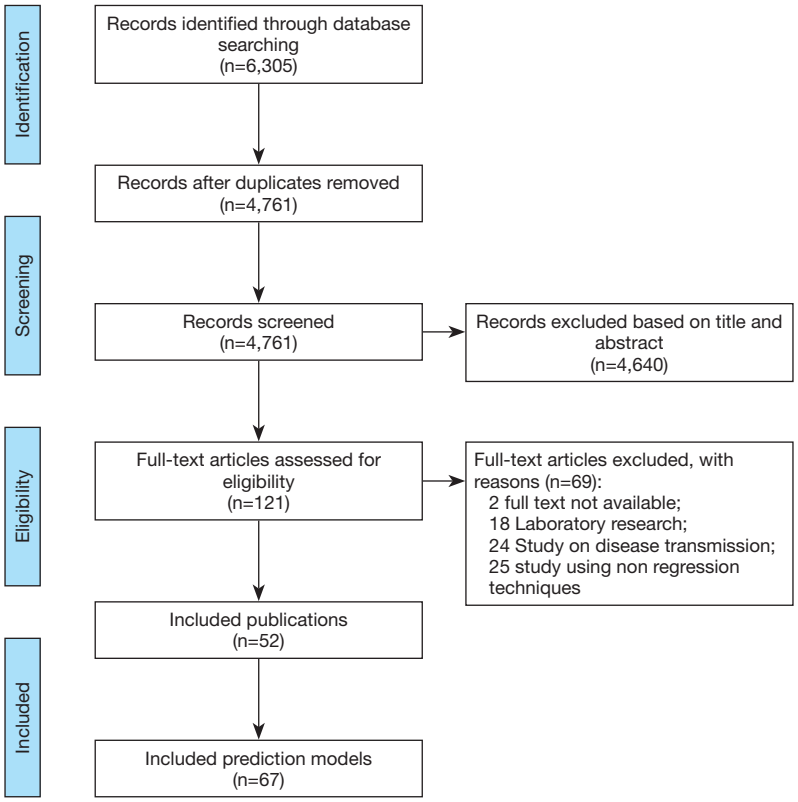


Figure 1 The flowchart of literature research. The flow chart is made according to PRISMA (the Preferred Reporting Items for Systematic Reviews and Meta-Analysis).

(k_w) and 95% CI (14). The calibration of the prediction models on mortality by Luo *et al.* showed good consistency between the prediction in the training cohort and actual observations (15). In two other studies, the model also fitted well (16,17). When the outcome was severe or critical progression of the disease, the discrimination ranged from 0.636 to 0.971. For ICU admission/mechanical ventilation/death, the discrimination varied between 0.712 and 0.900. Discrimination reported for the length-of-hospital stay outcome ranged from 0.361 to 0.848. For survival time, the discrimination was between 0.672 and 0.892.

Reporting completeness per model in TRIPOD

Figure 2 and the file (https://cdn.amegroups.cn/static/application/df0da0ff07a31a06aa1b1e1cf3b15d66/atm-20-6933-1.pdf) present the completeness of the model in TRIPOD. Overall, the reporting completeness was between 31% and 83%, with a median of 67% (IQR: 62%, 73%). The best completeness reporting was incremental value, with a median of 83%. This was followed by validation (70%, IQR: 64%, 74%). The development (66%, IQR: 62%, 70%) and the development and validation of the same model (62%, IQR: 56%, 71%) had similar reporting completeness.

Reporting completeness per TRIPOD items

We found that TRIPOD items in the discussion section were well completed (items 18–20); up to 100%. Supplementary information for item 21 and research funding for item 22 were well reported at 100%. The remaining 14 items were reported at ≥75% completeness, for all types of models (e.g., development, validation, development and validation of the same model, and incremental value). Four items reported <25%.

Information in the other parts of the TRIPOD items were described carefully below. Since there were three models in the incremental value that qualified and the sample size was small (hence not representative), we did not

Table 1 Primary information of prognostic models

No.	First author	Study region	Study design	Outcome	Sample size	Performance (discrimination)	Validation			
							Type of validation	Sample size	Performance	Calibration
1	Yuan	Wuhan, China	Retrospective	Death	27	0.901 (0.873, 0.928)	None	None	None	No
2	Osborne	Veterans, United States	Retrospective	Death	4,614	0.73	Internal validation (randomly split)	1,977	Not reported[‡]	No
3	Francone	Not reported	Retrospective	Death	130	0.672 (0.647, 0.877)	None	None	None[§]	No
4	Cozzi	Not reported	Retrospective	ICU admission[†]	234	ICC0.92 (0.88, 0.95)	None	None	None	No
5	Borghesi	Italy	Retrospective	Death	302	0.853	None	None	None	No
6	Wang	Wuhan, China	Retrospective	Death	296	0.88 (0.80, 0.95)	External validation	44	0.83 (0.75, 0.96)	No
7	Hong	Zhejiang, China	Retrospective	Prolonged length of stay in hospital	75	0.848 (0.753, 0.944)	None	None	None	No
8	Yu	Wuhan, China	Retrospective	Death	1,464	0.765 (0.725, 0.805)	None	None	None	No
9	Galloway	London, England	Not reported	Critical care admission and death	578	0.757 (0.713, 0.805)	Internal validation (randomly split)	579	0.712 (0.664, 0.759)	Yes
10	Liu	Wuhan, China	Retrospective	The development of severe/ critical disease	84	0.804 (0.702, 0.883)	External validation	71	0.881 (0.782, 0.946)	Yes
11	Borghesi	Italy	Not reported	Death	100	K_w 0.82 (0.79, 0.86)	None	None	None	No
12	Liu	Shanghai, China	Retrospective	Severe-event-free survival	134	0.78 (0.69, 0.88)	None	None	None	No
13	Yao	Wuhan, China	Retrospective	Death	248	0.85 (0.77, 0.92)	None	None	None	No
14	Zhou	Sichuan, China	Retrospective	Development of severe COVID-19	366	0.863 (0.801, 0.925)	Internal validation (bootstrap)	Not reported	0.839	Yes
15	Liang	China	Retrospective	Development of critical illness	1,590	0.88 (0.85, 0.91)	Internal validation (bootstrap)/external validation	Not reported/710	0.88 (0.85, 0.91)/0.88 (0.84, 0.93)	No
16	Dong	Wuhan, China	Retrospective	Survival time	377	0.901	Internal validation (randomly split)	251	0.892	Yes
17	Zheng	Hubei/Anhui, China	Retrospective	ICU admission, mechanical ventilation, or death	166	0.82 (0.76, 0.88)	External validation	72	0.89 (0.82, 0.96)	Yes
18	Zhang	Wuhan, China	Retrospective	Survival probability	516	0.886 (0.873, 0.899)	External validation	186	0.879 (0.856, 0.900)	Yes

Table 1 (*continued*)

Table 1 (*continued*)

No.	First author	Study region	Study design	Outcome	Sample size	Performance (discrimination)	Validation			
							Type of validation	Sample size	Performance	Calibration
19	Xiao	Hubei/Jiangxi, China	Retrospective	Severe state	231	0.861 (0.800, 0.922)	Internal validation (randomly split)/ external validation	101/110	0.871 (0.769, 0.972)/0.826 (0.746, 0.907)	Yes
20	Wang	Wuhan, China	Retrospective	Death	108	0.964 (0.909, 0.990)	None	None	None	No
21	Zheng	Zhejiang, China	Retrospective	Severe state	141	0.821 (0.746, 0.896)	None	None	None	Yes
22	Wu	Wuhan, China	Retrospective	Moderately ill and severely/ critically ill	210	0.955	Internal validation (randomly split)	60	0.945	No
23	Luo	Not reported	Retrospective	Death	1,018	0.907 (0.886, 0.928)	None	None	None	Yes
24	Huang	Hubei, China	Retrospective	Disease progression in mild cases	344	0.849	None	None	None	No
25	Liu	Wuhan, China	Retrospective	Death	336	0.994 (0.979, 0.999)	None	None	None	No
26	Hu	Wuhan, China	Retrospective	Death of severe or critical patients	105	0.864	None	None	None	No
27	Zhang	Wuhan, China	Retrospective	The death rate of critically patients in ICU	136	Not reported	None	None	None	No
28	Lorente-Ros	Not reported	Retrospective	Death	770	0.775	None	None	None	No
29	Myrstad	Oslo area, Norway	Prospective	Severe disease and in-hospital mortality	66	0.786 (0.659, 0.913)	None	None	None	No
30	Liu	Beijing, China	Prospective	Development of critical illness.	61	0.807 (0.676, 0.938)	External validation	54	0.882 (0.778, 0.986)	Yes
31	Nguyen	Paris, French	Retrospective	Unfavorable outcome	279	0.75	None	None	None	Yes
32	Zhang	Beijing, China	Retrospective	Severity of the disease	80	0.906	External validation	22	0.958	No
33	Satici	Istanbul, Turkey	Retrospective	30-day mortality	681	0.92 (0.89, 0.94)	None	None	None	No
34	Pascual Gómez	Madrid, Spain	Retrospective	Death rate	163	0.874 (0.816, 0.933)	None	None	None	No
35	Luo,	Wuhan, China	Retrospective	Death	1115	0.955 (0.941, 0.970)	None	None	None	No
36	Bello-Chavolla	Mexican	Retrospective	30-day death rate	41,306	0.822	Internal validation (randomly split)	10,327	0.83	No

Table 1 (*continued*)

Table 1 (*continued*)

No.	First author	Study region	Study design	Outcome	Sample size	Performance (discrimination)	Validation			
							Type of validation	Sample size	Performance	Calibration
37	Ji	Anhui/Beijing, China	Retrospective	Severe progression	208	0.86 (0.81, 0.91)	Internal validation (bootstrap)	Not reported	Not reported	Yes
38	Zhao	New York City, United states	Retrospective	ICU admission and death	454	0.87 (0.83, 0.92)	Internal validation	187	0.74 (0.63, 0.85)	No
39	Luo	Wuhan, China	Not reported	Death or survival	739	0.956 (0.928, 0.984)	None	None	None	No
40	Bi	Zhejiang, China	Retrospective	Occurrence of severe illness	113	0.712 (0.610, 0.814)	External validation	28	Not reported	Yes
41	Zheng	Zhejiang, China	Retrospective	Rehabilitation duration	90	R^2 0.361	None	None	None	No
42	Liu	Wuhan, China	Retrospective	Critical progression	88	0.971 (0.910, 0.995)	None	None	None	No
43	Gidari	Italy	Retrospective	ICU admission	71	0.90 (0.82, 0.97)	None	None	None	No
44	Vultaggio	Florence, Italy	Retrospective	Clinical deterioration	208	0.86	None	None	None	Yes
45	Yang	Chongqing, China	Retrospective	Critical progression	133	0.8842	None	None	None	No
46	Wang	Wuhan, China	Retrospective	Death of critical patients	104	0.893 (0.807, 0.98)	Internal validation (bootstrap)	Not reported	Not reported	No
47	Chen	China	Retrospective	Death	1,590	0.91 (0.85, 0.97)	Internal validation (bootstrap)	Not reported	Not reported	Yes
48	Shang	Wuhan, China	Retrospective	The death of severe cases	113	0.919 (0.870, 0.97)	External validation	339	0.938 (0.902, 0.973)	Yes
49	Li	Shanghai, China	Retrospective/ prospective	The development of severe disease	322	0.92 (0.88, 0.95)	External validation	317	0.92 (0.89, 0.95)	Yes
50	Zeng	Hunan, China	Retrospective	ICU admission	461	0.835 (0.742, 0.929)	None	None	None	Yes
51	Gong	Guangzhou, China	Retrospective	Severe progression	189	0.912 (0.846, 0.978)	Internal validation (3-fold cross-validation)/ external validation	165/18	Not reported/0.853 (0.790, 0.916)	Yes
52	Shang	Wuhan, China	Retrospective	Severe progression	443	0.774	None	None	None	No

[†], ICU is the abbreviation of intensive care unit; [‡], not reported means the information cannot be extracted; [§], none means this part is not appliable for this study.

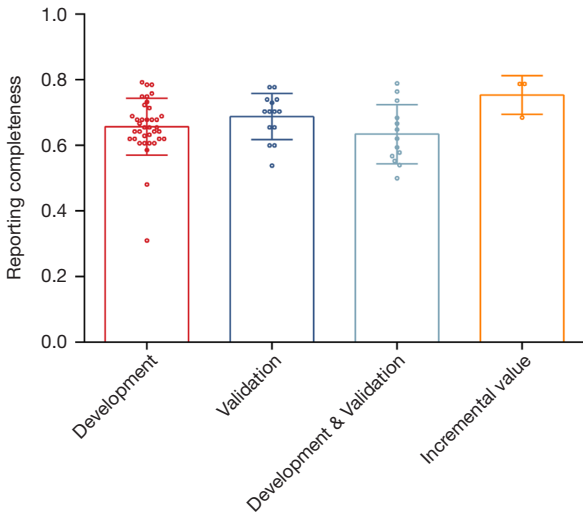


Figure 2 The reporting completeness of models in TRIPOD. Data are median [interquartile range (IQR)] and each point represents the completeness of one model; TRIPOD is the abbreviation of the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis.

include this type of model in the following elaboration. All details are shown in *Figure 3* and Appendix 3.

Items 1–3 (title/abstract/introduction)

In all types of models, the reporting completeness on the title and abstract section items was low, ranging from 5% to 36%. However, the completion of the introduction section (item 3) was high, both specifying the objectives, presenting the background, and including references to existing models.

In development, 5 (11%) of the 37 models explicitly identified the study as development and/or validation multivariable prediction models; then, they reported the target population and predicted the outcomes in the title. These completeness were 36% and 31% for the validation, and development and validation of the same model, respectively. Four models in the validation satisfied all the 12 elements in item 2. That is, the research objectives, study design, setting, participants, sample sizes, predictors, prediction outcomes, and statistical analyses were all provided in the abstract as well as brief results and conclusions. The completeness of item 2 was 5% and 23% in the development, and development and validation of the same model, respectively.

Items 4–12 (methods)

Items 4–5, 6a, 8, 10c, and 11 were highly reported among all the models; with all the values >80%. This meant that the sources of data, key study dates, and eligibility criteria for the participants were well reported. However, the reported completeness of how the missing data were handled (item 9) and the model-building procedures (item 10b) were low, at <15%.

In the development (57%) and development and validation of the same model (46%), the completeness of any blinding of the outcome to be predicted was not high. Assessment of the model performance (item 10d) had general completeness reporting of 24% in development, 43% in validation, and 54% in development and validation of the same model. These results were mainly due to the inadvertent reporting of the calibration element. In validation, very few (7%) noted the need to compare validation with data from development (item 12). However, item 12 was well reported in the development and validation of the same model; up to 77%.

Items 4–17 (results)

All types of models were highly completed in the reporting of the number of participants and outcome events in the analysis and the unadjusted association between candidate predictors and outcomes (items 14a and 14b); reaching more than 90%. However, only few models could consider all the four elements in item 13b, and the reporting completeness was <5%. This was due to the fact that researchers tended to ignore the number of participants with missing data in predictors and prediction outcomes when reporting information.

In the development, and development and validation of the same model, few studies reported adequate information in the final model (item15a), with the completeness of 32% and 8%, respectively. Although most models presented regression coefficients for each predictor, the intercept, or the cumulative baseline hazard (or baseline survival) for at least one time point was poorly reported.

In development, 46% of all models were fully reported for item 15b, and many researchers did not explain how to use the newly established prediction model. Whether in development, validation, or development and validation of the same model, the reporting of the prediction model

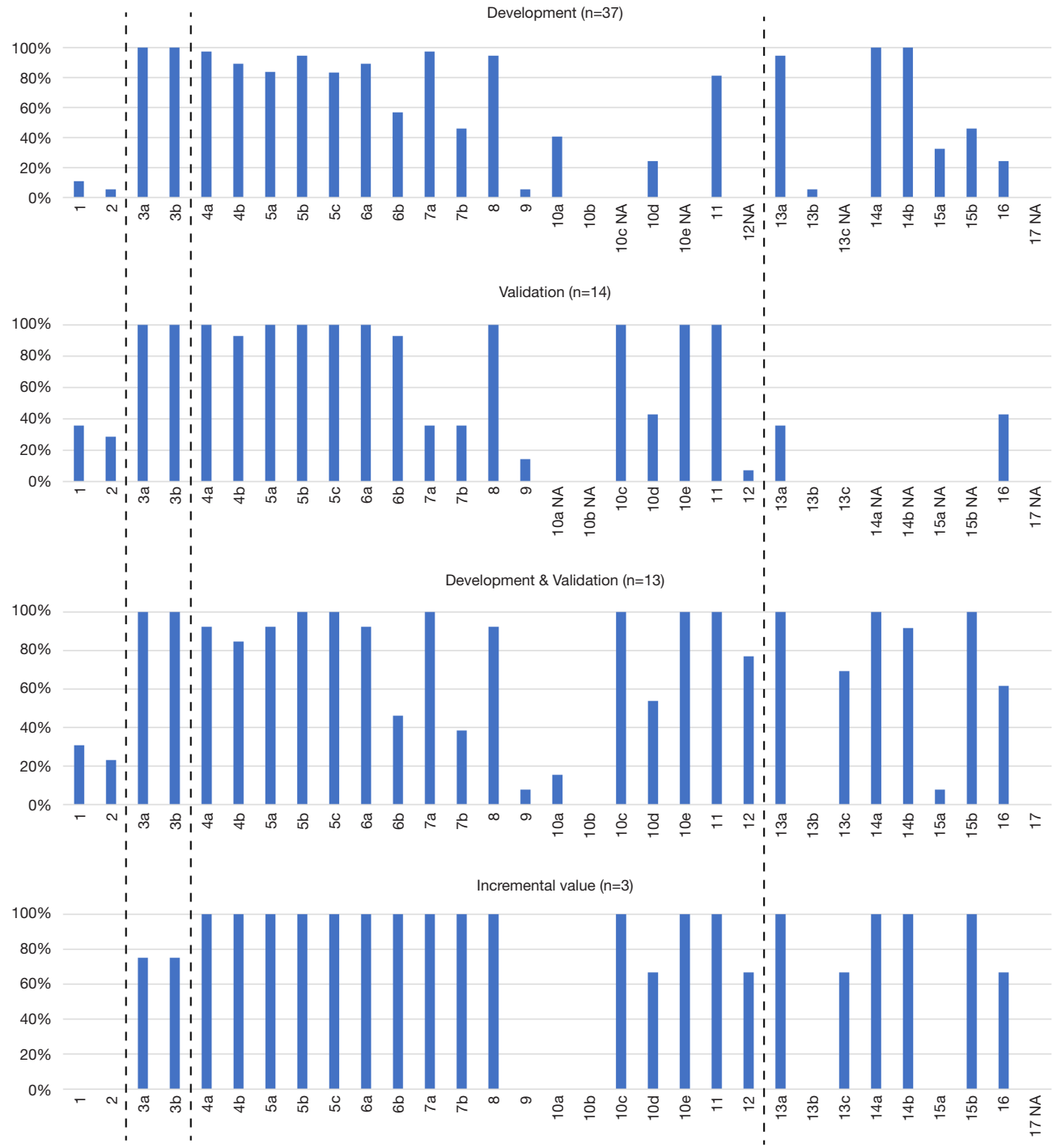


Figure 3 Reporting of the items in TRIPOD. The combination of numbers and letters in the abscissa represents the items in TRIPOD; TRIPOD is the abbreviation of the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis. NA is the abbreviation of not applicable and it means that the item does not apply to this type of models.

performance measures (item 16) was not ideal at 24%, 43%, and 62%, respectively. These were due to the inability of many models to adhere to one of these elements that reported model calibration, which also corresponded to the low reporting of item 10d in the methods section.

Meta-analysis

In the meta-analysis, we screened five studies for the included validation from which the discrimination of CURB-65 could be extracted. The CURB-65 score is a prediction model used to divide patients with community-acquired pneumonia into different treatment patient groups (18). The pooled performance of CURB-65 in COVID-19 infectious patients was 0.768 (95% CI, 0.694, 0.841). The forest plot is shown in Appendix 4.

Discussion

In this systematic review of prognostic models related to COVID-19, we included a total of 67 models from 52 studies. The main prediction outcomes were as follows: death, development of severe/critical state, ICU admission/mechanical ventilation/death, survival time, and length-of-hospital stay. There was a mix between outcomes. The predicted outcome of some studies were the indicators of the outcomes predicted in some other studies. Zeng *et al.* focused on identifying patients with a high risk of progression and who would require transfer to the ICU (19). On the other hand, many other studies listed ICU admission as one of the indicators of their prediction outcomes (i.e. severe or critical progression and mortality) (20-22). Additionally, the same outcome was defined differently in different studies; the definition of severe and critical cases was not uniform. Liu *et al.* assessed the status of patients according to the American Thoracic Society guidelines (23). Liang *et al.* also defined the severity based on the American Thoracic Society guidelines for community-acquired pneumonia, given the extensive acceptance of this guideline (24). However, Xiao *et al.* used the Diagnosis and Treatment Protocol for Novel Coronavirus Pneumonia (Trial Version 7) as the guideline for the spectrum of severity (25). The blind evaluation of the prediction outcome and prediction factors were ignored in the models. For the all-cause mortality, it was well-defined and not affected by subjective factors, while in other instances such as in severe state progression, an explicit mention about the judgement of outcome was expected.

Potential for popularizing clinical practice

Optimistic discrimination performance was reported for all the models. However, the existing models had the risk of over-fitting, because the number of available samples and events which were used for developing the new prediction model were limited by the sample sizes. In addition to the above reasons, most studies directly excluded the missing data from the original data, which reduced the sample sizes greatly. Multiple imputation may be used to address this challenge. The overfitting can also be alleviated by calibration, which has rarely been evaluated in models. In future prediction model research, attention should be paid to the disposal of missing values, and multiple interpolation should be carried out for missing values when appropriate. In addition, emphasis should be placed on calibration results in reporting model performance. Similarly, there were few (only 13) external validations of the newly established models, so these were insufficient to promote the existing models directly in clinical practice. In addition, there were few internal validations of the newly established models. Random splitting was the most frequently used method instead of bootstrap or k-fold cross-validation, which enhanced the limitation of the small sample size in the model prediction. Based on our findings, we encourage researchers to count age, disease history, lymphocyte count, history of hypertension and cardiovascular disease, C reactive protein, lactate dehydrogenase, white blood cell count and platelet count into the prediction model, rather than simply selecting the predictors in a data-driven manner, which may put the model at risk of overfitting.

Research participants should be adequately described in the development data, which is beneficial to popularize newly established models in the real world. Borghesi *et al.* identified Caucasians as participants in a study (8). Osborne clarified that their model was aimed at veterans in the United States (26). Pascual determined that the setting of their study was the hospital emergency room (27). However, the applicability of the model among most of the studies was not of great importance. Although we realized that due to the particularity of COVID-19, the time and space for the completion of these studies were limited.

Moreover, the reporting completeness of the final model presentation was poor. Although the regression coefficient (or a derivative such as hazard ratio, odds ratio, and risk ratio) for each predictor in the model was reported in a large number of models. The intercept or the cumulative

baseline hazard for at least one time point was ignored, which will make future research to re-validate the developed model and recalibrate it difficult. All of the above hindered the improvement of the prediction model and its promotion in clinical practice.

In our study, moderate or even excellent degree of discrimination ability was found when the existing CURB-65 model was used to predict the prognosis of COVID-19 patients. In future research, we may consider adding the prediction variables or recalibrating the model to achieve better prediction results. What's more, with the development of vaccine trials worldwide, whether vaccination will have an impact on the prediction model, that is, whether vaccination can also become a new predictor is also the direction that researchers need to focus on.

Limitations

The number of studies was relatively small. However, these evaluation results may be improved with the promotion of COVID-19 prognosis model research. In particular, the number of incremental value studies was few, so it may not be appropriate to use the quantitative method converted by the TRIPOD statement for the evaluation. Secondly, due to the limitation of the applicability of TRIPOD, we were unable to evaluate models that were established by artificial intelligence. Thirdly, some hospitals provided data for different studies at the same time, which made it unclear to us how much overlap we included from the studies. Moreover, most of the articles we included were from China, especially Wuhan; and there was no description of demographic variables such as race, economic status, and educational level that might affect patient outcomes. All of these factors may have potential impacts on our results.

Conclusions

In the present study, the prognostic prediction models for COVID-19 were evaluated according to the TRIPOD statement; we found the reporting completeness to be poor. The potential for the clinical promotion of the model is low due to over-fitting and the lack of calibration and external validation. Overall, we need to focus our research in the future on the validation and improvement of existing models. The premise for this was a high-quality research, following the TRIPOD reporting guidelines.

Acknowledgments

Funding: The work was supported by grants from the National Natural Science Foundation of China (81573258); and the Jiangsu Provincial Major Science & Technology Demonstration Project (BE2017749); and the Southeast University COVID-19 Fund (3225002001C1)

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist (available at http://dx.doi.org/10.21037/atm-20-6933).

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/atm-20-6933). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. No any human experiments or animals' experiments were involved in studies.

References

1. COVID19.WHO.int. WHO Coronavirus Disease (COVID-19) Dashboard; c2020. Available online: https://covid19.who.int/
2. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020;395:507-13.
3. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA

2020;323:1061-9.
4. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395:497-506.
5. Grasselli G, Zangrillo A, Zanella A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. JAMA 2020;323:1574-81.
6. CDC COVID-19 Response Team. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12-March 16, 2020. MMWR Morb Mortal Wkly Rep 2020;69:343-6.
7. Bello-Chavolla OY, Bahena-López JP, Antonio-Villa NE, et al. Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico. J Clin Endocrinol Metab 2020;105:dgaa346.
8. Borghesi A, Zigliani A, Golemi S, et al. Chest X-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: A study of 302 patients from Italy. Int J Infect Dis 2020;96:291-3.
9. Dong YM, Sun J, Li YX, et al. Development and Validation of a Nomogram for Assessing Survival in Patients with COVID-19 Pneumonia. Clin Infect Dis 2021;72:652-60.
10. Collins GS, Reitsma JB, Altman DG, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. BMC Med 2015;13. doi: https://doi.org/10.1186/s12916-014-0241-z
11. Heus P, Damen JAAG, Pajouheshnia R, et al. Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. BMC Med 2018;16:120.
12. Wynants L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m1328.
13. Moons KGM, 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. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. Radiol Med 2020;125:509-13.
15. Luo M, Liu J, Jiang W, et al. IL-6 and CD8(+) T cell counts combined are an early predictor of in-hospital mortality of patients with COVID-19. JCI Insight 2020;5:e139024.
16. Chen R, Liang W, Jiang M, et al. Risk Factors of Fatal Outcome in Hospitalized Subjects with Coronavirus Disease 2019 From a Nationwide Analysis in China. Chest 2020;158:97-105.
17. Shang Y, Liu T, Wei Y, et al. Scoring systems for predicting mortality for severe patients with COVID-19. EClinicalMedicine 2020;24:100426.
18. 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.
19. Zeng Z, Ma Y, Zeng H, et al. Simple nomogram based on initial laboratory data for predicting the probability of ICU transfer of COVID-19 patients: Multicenter retrospective study. J Med Virol 2021;93:434-40.
20. Myrstad M, Ihle-Hansen H, Tveita AA, et al. National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19-a prospective cohort study. Scand J Trauma Resusc Emerg Med 2020;28:66.
21. Liu J, Liu Y, Xiang P, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. J Transl Med 2020;18:206.
22. Wang B, Zhong F, Zhang H, et al. Risk factors analysis and nomogram construction of non-survivors in critical patients with COVID-19. Jpn J Infect Dis 2020;73:452-8.
23. Liu YP, Li GM, He J, et al. Combined use of the neutrophil-to-lymphocyte ratio and CRP to predict 7-day disease severity in 84 hospitalized patients with COVID-19 pneumonia: a retrospective cohort study. Ann Transl Med 2020;8:635.
24. Liang W, Liang H, Ou L, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020;180:1081-9.
25. Xiao LS, Zhang WF, Gong MC, et al. Development and validation of the HNC-LL score for predicting the severity of coronavirus disease 2019. EBioMedicine 2020;57:102880.
26. Osborne TF, Veigulis ZP, Arreola DM, et al. Automated EHR score to predict COVID-19 outcomes at US Department of Veterans Affairs. PLoS One

2020;15:e0236554.
27. Pascual Gómez NF, Monge Lobo I, Granero Cremades I, et al. Potential biomarkers predictors of mortality in COVID-19 patients in the Emergency Department. Rev Esp Quimioter 2020;33:267-73.

Cite this article as: Yang L, Wang Q, Cui T, Huang J, Shi N, Jin H. Reporting of coronavirus disease 2019 prognostic models: the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement. Ann Transl Med 2021;9(5):421. doi: 10.21037/atm-20-6933

Supplementary

Appendix 1 Included literature information

No.	First author	Article name	The journal
1	Yuan, M.L.	Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China	PLoS One
2	Osborne, T. F.	Automated EHR score to predict COVID-19 outcomes at US Department of Veterans Affairs	PLoS One
3	Francone, M.	Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis	European Radiology
4	Cozzi, D.	Chest X-ray in new Coronavirus Disease 2019 (COVID-19) infection: findings and correlation with clinical outcome	Radiologia Medica
5	Borghesi, A.	Chest X-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: A study of 302 patients from Italy	International Journal of Infectious Diseases
6	Wang, K.	Clinical and laboratory predictors of in-hospital mortality in patients with COVID-19: a cohort study in Wuhan, China	Clinical Infectious Diseases
7	Hong, Y.	Clinical characteristics of Coronavirus Disease 2019 and development of a prediction model for prolonged hospital length of stay	Annals of Translational Medicine
8	Yu, C.	Clinical Characteristics, Associated Factors, and Predicting COVID-19 Mortality Risk: A Retrospective Study in Wuhan, China	American Journal of Preventive Medicine
9	Galloway, J. B.	A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study	Journal of Infection
10	Liu, Y. P.	Combined use of the neutrophil-to-lymphocyte ratio and CRP to predict 7-day disease severity in 84 hospitalized patients with COVID-19 pneumonia: a retrospective cohort study	Annals of Translational Medicine
11	Borghesi, A.	COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression	Radiologia Medica
12	Liu, F. J.	CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients	Theranostics
13	Yao, Y.	D-dimer as a biomarker for disease severity and mortality in COVID-19 patients: a case control study	Journal of Intensive Care
14	Zhou, Y. W.	Development and validation a nomogram for predicting the risk of severe COVID-19:Amulti-center study in Sichuan, China	PLoS One
15	Liang, W. H.	Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19	JAMA Internal Medicine
16	Dong, Y. M.	Development and Validation of a Nomogram for Assessing Survival in Patients with COVID-19 Pneumonia	Clinical Infectious Diseases
17	Zheng, Y.	Development and Validation of a Prognostic Nomogram Based on Clinical and CT Features for Adverse Outcome Prediction in Patients with COVID-19	Korean Journal of Radiology
18	Zhang, S.	Development and validation of a risk factor-based system to predict short-term survival in adult hospitalized patients with COVID-19: a multicenter, retrospective, cohort study	Critical Care
19	Xiao, L. S.	Development and validation of the HNC-LL score for predicting the severity of coronavirus disease 2019	Ebiomedicine
20	Wang, F.	Establishing a model for predicting the outcome of COVID-19 based on combination of laboratory tests	Travel Medicine and Infectious Disease
21	Zheng, Y. F.	The hemocyte counts as a potential biomarker for predicting disease progression in COVID-19: a retrospective study	Clinical Chemistry and Laboratory Medicine
22	Wu, S.	Identification and validation of a novel clinical signature to predict the prognosis in confirmed COVID-19 patients	Clinical Infectious Diseases
23	Luo, M.	IL-6 and CD8+ T cell counts combined are an early predictor of in-hospital mortality of patients with COVID-19	JCI Insight
24	Huang, J. F.	Individualized prediction nomograms for disease progression in mild COVID-19	Journal of Medical Virology
25	Liu, Q.	Laboratory findings and a combined multifactorial approach to predict death in critically ill patients with COVID-19: a retrospective study	Epidemiology and Infection
26	Hu, K.	Logistic regression analysis of death risk factors of patients with severe and critical coronavirus disease 2019 and their predictive value	Zhonghua wei zhong bing ji jiu yi xue
27	Zhang, P.	The modified NUTRIC score can be used for nutritional risk assessment as well as prognosis prediction in critically ill COVID-19 patients	Clinical Nutrition
28	Lorente-Ros, A.	Myocardial injury determination improves risk stratification and predicts mortality in COVID-19 patients	Cardiology Journal
29	Myrstad, M.	National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19 - a prospective cohort study	Scandinavian Journal of Trauma Resuscitation & Emergency Medicine
30	Liu, J. Y.	Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage	Journal of Translational Medicine
31	Nguyen, Y.	A nomogram to predict the risk of unfavorable outcome in COVID-19: a retrospective cohort of 279 hospitalized patients in Paris area	Annals of Medicine
32	Zhang, C.	A Novel Scoring System for Prediction of Disease Severity in COVID-19	Frontiers in Cellular and Infection Microbiology
33	Satici, C.	Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19	International Journal of Infectious Diseases
34	Pascual Gómez, N. F.	Potential biomarkers predictors of mortality in COVID-19 patients in the Emergency Department	Revista espanola de quimioterapia
35	Luo, Y.	Prealbumin as a Predictor of Prognosis in Patients with Coronavirus Disease 2019	Frontiers in Medicine
36	Bello-Chavolla, O. Y.	Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico	The Journal of Clinical Endocrinology and Metabolism
37	Ji, D.	Prediction for Progression Risk in Patients with COVID-19 Pneumonia: the CALL Score	Clinical Infectious Diseases
38	Zhao, Z.	Prediction model and risk scores of ICU admission and mortality in COVID-19	PLoS One
39	Luo, Y.	Prediction Model Based on the Combination of Cytokines and Lymphocyte Subsets for Prognosis of SARS-CoV-2 Infection	Journal of Clinical Immunology
40	Bi, X. J.	Prediction of severe illness due to COVID-19 based on an analysis of initial Fibrinogen to Albumin Ratio and Platelet count	Platelets
41	Zheng, Q. N.	Prediction of the Rehabilitation Duration and Risk Management for Mild-Moderate COVID-19	Disaster Medicine and Public Health Preparedness
42	Liu, X.	Prediction of the severity of Corona Virus Disease 2019 and its adverse clinical outcomes	Japanese Journal of Infectious Diseases
43	Gidari, A.	Predictive value of National Early Warning Score 2 (NEWS2) for intensive care unit admission in patients with SARS-CoV-2 infection	Infectious Diseases
44	Vultaggio, A.	Prompt Predicting of Early Clinical Deterioration of Moderate-to-Severe COVID-19 Patients: Usefulness of a Combined Score Using IL-6 in a Preliminary Study	The Journal of Allergy and Clinical Immunology
45	Yang, P.	A retrospective study on the epidemiological characteristics and establishment of early warning system of severe COVID-19 patients	Journal of Medical Virology
46	Wang, B.	Risk factors analysis and nomogram construction of non-survivors in critical patients with COVID-19	Japanese Journal of Infectious Diseases
47	Chen, R. C.	Risk Factors of Fatal Outcome in Hospitalized Subjects with Coronavirus Disease 2019From a Nationwide Analysis in China	CHEST
48	Shang, Y.	Scoring systems for predicting mortality for severe patients with COVID-19	EClinicalMedicine
49	Li, Q.	A simple algorithm helps early identification of SARS-CoV-2 infection patients with severe progression tendency	INFECTION
50	Zeng, Z.	Simple nomogram based on initial laboratory data for predicting the probability of ICU transfer of COVID-19 patients: Multicenter retrospective study	Journal of Medical Virology
51	Gong, J.	A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19): A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China	Clinical Infectious Diseases
52	Shang, W. F.	The value of clinical parameters in predicting the severity of COVID-19	Journal of Medical Virology

http://dx.doi.org/10.21037/atm-20-6933

Appendix 2 Predictors information

No.	First author	Predictors
1	Yuan, M.L.	computed tomography (CT) score
2	Osborne, T. F.	CAN score
3	Francone, M.	computed tomography (CT) score
4	Cozzi, D.	chest X-ray performance
5	Borghesi, A.	Brixia score, patient age, and conditions that induced immunosuppression
6	Wang, K.	clinical: age, history of hypertension and coronary heart disease (CHD) laboratory: baseline age, peripheral capillary oxygen saturation (SpO_2), neutrophil count, lymphocyte count, hsCRP, D-dimer, AST and GFR
7	Hong, Y.	procalcitonin, heart rate, epidemiological history, lymphocyte count and cough
8	Yu, C.	age, male sex, history of diabetes, lymphopenia, and increased procalcitonin
9	Galloway, J. B.	respiratory rate, pulse oximetry saturations and oxygen requirement, higher neutrophil counts, higher CRP, lower albumin and renal impairment
10	Liu, Y. P.	underlying disease, age, NLR, RDW, PLT and CRP
11	Borghesi, A.	chest X-ray score
12	Liu, F. J.	artificial intelligence algorithms, representing the percentages of ground-glass opacity volume (PGV), semi-consolidation volume (PSV), and consolidation volume (PCV) in both lungs
13	Yao, Y.	D-dimer
14	Zhou, Y. W.	body temperature on admission, cough, dyspnea, hypertension, cardiovascular disease, chronic liver disease, and chronic kidney disease
15	Liang, W. H.	chest radiography abnormality, age, hemoptysis, dyspnea, unconsciousness, number of comorbidities, cancer history, neutrophil-to-lymphocyte ratio, lactate dehydrogenase, and direct bilirubin
16	Dong, Y. M.	hypertension, higher neutrophil-to-lymphocyte ratio and increased NT-proBNP
17	Zheng, Y.	underlying comorbidity, lymphocyte count and crazy-paving sign
18	Zhang, S.	older age, high lactate dehydrogenase level, evaluated neutrophil-to-lymphocyte ratio, and high direct bilirubin level
19	Xiao, L. S.	HNC-LL (Hypertension, Neutrophil count, C-reactive protein, Lymphocyte count, Lactate dehydrogenase) score
20	Wang, F.	neutrophils, lymphocytes, platelets, IL–2R
21	Zheng, Y. F.	neutrophil count, lymphocyte count and platelet count
22	Wu, S.	namely neutrophil count, lymphocyte count, procalcitonin, older age, and C-reactive protein
23	Luo, M.	IL-6 (>20 pg/mL) and CD8+ T cell counts (<165 cells/μL)
24	Huang, J. F.	model 1: male gender, age, hypertension, DM, CHD, CVD, COPD, cancer, sputum, T≥38.5℃, onset time model2: male gender, age, DM, CHD, sputum, T≥38.5℃, NLR, WBC
25	Liu, Q.	Lymr, BUN and DD at admission
26	Hu, K.	gender, age, concomitant coronary heart disease and hypertension, complicated with myocardial damage and thrombocytopenia
27	Zhang, P.	mNUTRIC score (consisted of five variables: age, APACHE II score at admission, SOFA score at admission, number of comorbidities and pre-ICU hospital length of stay (LOS))
28	Lorente-Ros, A.	cTnI, Charlson comorbidity index (CCI)
29	Myrstad, M.	the external validation of NEWS2/qSOFA/SIRS/CRB-65
30	Liu, J. Y.	NLR
31	Nguyen, Y.	age, overweight, polypnoea, fever, high C-reactive protein, troponin, and lymphopenia
32	Zhang, C.	age, WBC, NEU, GFR, and Myoglobin
33	Satici, C.	the external validation of CURB-65/PSI/ PSI, CRP
34	Pascual Gómez, N. F.	age, total WBC, glucose, creatinine
35	Luo, Y.	PAB, LYM, NEU, PCT, hsCRP, PT, LDH, CR, hs-cTnI
36	Bello-Chavolla, O. Y.	age > 65 years, diabetes mellitus, obesity, CKD, COPD, immunosuppression, and hypertension, SARS-CoV-2 severity
37	Ji, D.	comorbidity, Age, Lymphocyte, LDH
38	Zhao, Z.	ICU: LDH, procalcitonin, smoking history, ever smoker, SpO2, lymphocyte count death: heart failure, procalcitonin, LDH, COPD, SpO2, heart rate, age
39	Luo, Y.	IL-8, CD4 T cells, NK cells
40	Bi, X. J.	FAR, PLT
41	Zheng, Q. N.	white blood cell [WBC], partial pressure of carbon dioxide [PaCO2], serum potassium [K], total bilirubin [TBIL], and aspartate aminotransaminase [AST]
42	Liu, X.	IL-6 level, absolute lymphocyte count, age
43	Gidari, A.	the external validations of (NEWS2) National Early Warning Score 2
44	Vultaggio, A.	IL-6, C-reactive protein, SaO2/FiO2
45	Yang, P.	age, shortness of breath, lymphocyte count, PCT level, LDH level, APTT level, and CRP level
46	Wang, B.	Age, chest tightness, AST and BUN
47	Chen, R. C.	age, coronary heart disease, cerebrovascular disease, dyspnea, procalcitonin level, and aspartate aminotransferase level
48	Shang, Y.	old age, CHD, LYM%, PCT and DD
49	Li, Q.	age, lactate dehydrogenase (LDH), and CD4 count
50	Zeng, Z.	lymphocyte count, platelet count, AST level, LDH level and CRP level
51	Gong, J.	older age; higher lactate dehydrogenase (LDH), C-reactive protein (CRP), coefficient of variation of red blood cell distribution width (RDW), direct bilirubin (DBIL), and blood urea nitrogen (BUN); and lower albumin (ALB)
52	Shang, W. F.	NLR, CRP, and platelets

http://dx.doi.org/10.21037/atm-20-6933

Included studies references

1. Yuan M, Yin W, Tao Z, et al. Association of radiologic findings with mortality of patients infected with 2019 novel coronavirus in Wuhan, China. Plos One 2020;15.
2. Osborne TF, Veigulis ZP, Arreola DM, et al. Automated EHR score to predict COVID-19 outcomes at US Department of Veterans Affairs. Plos One 2020;15.
3. Francone M, Iafrate F, Masci GM, et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. European Radiology 2020;30:6808-17.
4. Cozzi D, Albanesi M, Cavigli E, et al. Chest X-ray in new Coronavirus Disease 2019 (COVID-19) infection: findings and correlation with clinical outcome. Radiologia Medica 2020;125:730-7.
5. Borghesi A, Zigliani A, Golemi S, et al. Chest X-ray severity index as a predictor of in-hospital mortality in coronavirus disease 2019: A study of 302 patients from Italy. International Journal of Infectious Diseases 2020;96:291-3.
6. Wang K, Zuo P, Liu Y, et al. Clinical and Laboratory Predictors of In-hospital Mortality in Patients With Coronavirus Disease-2019: A Cohort Study in Wuhan, China. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America 2020;71:2079-88.
7. Hong Y, Wu X, Qu J, et al. Clinical characteristics of Coronavirus Disease 2019 and development of a prediction model for prolonged hospital length of stay. Annals of Translational Medicine 2020;8.
8. Yu C, Lei Q, Li W, et al. Clinical Characteristics, Associated Factors, and Predicting COVID-19 Mortality Risk: A Retrospective Study in Wuhan, China. American Journal of Preventive Medicine 2020;59:168-75.
9. Galloway JB, Norton S, Barker RD, et al. A clinical risk score to identify patients with COVID-19 at high risk of critical care admission or death: An observational cohort study. Journal of Infection 2020;81:282-8.
10. Liu Y-P, Li G-M, He J, et al. Combined use of the neutrophil-to-lymphocyte ratio and CRP to predict 7-day disease severity in 84 hospitalized patients with COVID-19 pneumonia: a retrospective cohort study. Annals of Translational Medicine 2020;8.
11. Borghesi A, Maroldi R. COVID-19 outbreak in Italy: experimental chest X-ray scoring system for quantifying and monitoring disease progression. Radiologia Medica 2020;125:509-13.
12. Liu F, Zhang Q, Huang C, et al. CT quantification of pneumonia lesions in early days predicts progression to severe illness in a cohort of COVID-19 patients. Theranostics 2020;10:5613-22.
13. Yao Y, Cao J, Wang Q, et al. D-dimer as a biomarker for disease severity and mortality in COVID-19 patients: a case control study. Journal of Intensive Care 2020;8.
14. Zhou YW, He YQ, Yang H, et al. Development and validation a nomogram for predicting the risk of severe COVID-19: A multi-center study in Sichuan, China. Plos One 2020;15.
15. Liang W, Liang H, Ou L, et al. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. Jama Internal Medicine 2020;180:1081-9.
16. Dong Y-M, Sun J, Li Y-X, et al. Development and Validation of a Nomogram for Assessing Survival in Patients with COVID-19 Pneumonia. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America 2020.
17. Zheng YY, Xiao AL, Yu XR, et al. Development and Validation of a Prognostic Nomogram Based on Clinical and CT Features for Adverse Outcome Prediction in Patients with COVID-19. Korean Journal of Radiology 2020;21:1007-17.
18. Zhang S, Guo M, Duan L, et al. Development and validation of a risk factor-based system to predict short-term survival in adult hospitalized patients with COVID-19: a multicenter, retrospective, cohort study. Critical Care 2020;24.
19. Xiao L-s, Zhang W-F, Gong M-c, et al. Development and validation of the HNC-LL score for predicting the severity of coronavirus disease 2019. Ebiomedicine 2020;57.
20. Wang F, Hou H, Wang T, et al. Establishing a model for predicting the outcome of COVID-19 based on combination of laboratory tests. Travel Medicine and Infectious Disease 2020;36.
21. Zheng Y, Zhang Y, Chi H, et al. The hemocyte counts as a potential biomarker for predicting disease progression in COVID-19: a retrospective study. Clinical Chemistry and Laboratory Medicine 2020;58:1106-15.
22. Wu S, Du Z, Shen S, et al. Identification and validation of a novel clinical signature to predict the prognosis in confirmed COVID-19 patients. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America 2020.
23. Luo M, Liu J, Jiang W, et al. IL-6 and CD8(+) T cell counts combined are an early predictor of in-hospital mortality of patients with COVID-19. Jci Insight 2020;5.
24. Huang J, Cheng A, Lin S, et al. Individualized prediction nomograms for disease progression in mild COVID-19. Journal of Medical Virology 2020;92:2074-80.
25. Liu Q, Song NC, Zheng ZK, et al. Laboratory findings and a combined multifactorial approach to predict death in critically ill patients with COVID-19: a retrospective study. Epidemiology and Infection 2020;148.
26. Hu K, Li B. Logistic regression analysis of death risk factors of patients with severe and critical coronavirus disease 2019 and their predictive value. Zhonghua wei zhong bing ji jiu yi xue 2020;32:544-7.
27. Zhang P, He Z, Yu G, et al. The modified NUTRIC score can be used for nutritional risk assessment as well as prognosis prediction in critically ill COVID-19 patients. Clinical nutrition (Edinburgh, Scotland) 2020.
28. Lorente-Ros A, Monteagudo Ruiz JM, Rincon LM, et al. Myocardial injury determination improves risk stratification and predicts mortality in COVID-19 patients. Cardiology journal 2020;27:489-96.

 http://dx.doi.org/10.21037/atm-20-6933

29. Myrstad M, Ihle-Hansen H, Tveita AA, et al. National Early Warning Score 2 (NEWS2) on admission predicts severe disease and in-hospital mortality from Covid-19-a prospective cohort study. Scandinavian Journal of Trauma Resuscitation & Emergency Medicine 2020;28.
30. Liu J, Liu Y, Xiang P, et al. Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage. Journal of Translational Medicine 2020;18.
31. Nguyen Y, Corre F, Honsel V, et al. A nomogram to predict the risk of unfavourable outcome in COVID-19: a retrospective cohort of 279 hospitalized patients in Paris area. Annals of Medicine 2020;52:367-75.
32. Zhang C, Qin L, Li K, et al. A Novel Scoring System for Prediction of Disease Severity in COVID-19. Frontiers in Cellular and Infection Microbiology 2020;10.
33. Satici C, Demirkol MA, Altunok ES, et al. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. International Journal of Infectious Diseases 2020;98:84-9.
34. Pascual Gomez NF, Monge Lobo I, Granero Cremades I, et al. Potential biomarkers predictors of mortality in COVID-19 patients in the Emergency Department. Revista espanola de quimioterapia: publicacion oficial de la Sociedad Espanola de Quimioterapia 2020;33:267-73.
35. Luo Y, Xue Y, Mao L, et al. Prealbumin as a Predictor of Prognosis in Patients With Coronavirus Disease 2019. Frontiers in Medicine 2020;7.
36. Yaxmehen Bello-Chavolla O, Paola Bahena-Lopez J, Eduardo Antonio-Villa N, et al. Predicting Mortality Due to SARS-CoV-2: A Mechanistic Score Relating Obesity and Diabetes to COVID-19 Outcomes in Mexico. Journal of Clinical Endocrinology & Metabolism 2020;105:2752-61.
37. Ji D, Zhang D, Xu J, et al. Prediction for Progression Risk in Patients With COVID-19 Pneumonia: The CALL Score. Clinical Infectious Diseases 2020;71:1393-9.
38. Zhao Z, Chen A, Hou W, et al. Prediction model and risk scores of ICU admission and mortality in COVID-19. Plos One 2020;15.
39. Luo Y, Mao LY, Yuan X, et al. Prediction Model Based on the Combination of Cytokines and Lymphocyte Subsets for Prognosis of SARS-CoV-2 Infection. Journal of Clinical Immunology 2020;40:960-9.
40. Bi X, Su Z, Yan H, et al. Prediction of severe illness due to COVID-19 based on an analysis of initial Fibrinogen to Albumin Ratio and Platelet count. Platelets 2020;31:674-9.
41. Zheng Q-N, Xu M-Y, Zheng Y-L, et al. Prediction of the Rehabilitation Duration and Risk Management for Mild-Moderate COVID-19. Disaster medicine and public health preparedness 2020:1-6.
42. Liu X, Shi S, Xiao J, et al. Prediction of the Severity of the Coronavirus Disease and Its Adverse Clinical Outcomes. Jpn J Infect Dis 2020;73:404-10.
43. Gidari A, De Socio GV, Sabbatini S, et al. Predictive value of National Early Warning Score 2 (NEWS2) for intensive care unit admission in patients with SARS-CoV-2 infection. Infectious Diseases 2020;52:698-704.
44. Vultaggio A, Vivarelli E, Virgili G, et al. Prompt Predicting of Early Clinical Deterioration of Moderate-to-Severe COVID-19 Patients: Usefulness of a Combined Score Using IL-6 in a Preliminary Study. Journal of Allergy and Clinical Immunology-in Practice 2020;8:2575-+.
45. Yang P, Wang P, Song Y, et al. A retrospective study on the epidemiological characteristics and establishment of an early warning system of severe COVID-19 patients. Journal of Medical Virology 2020;92:2173-80.
46. Wang B, Zhong F, Zhang H, et al. Risk Factor Analysis and Nomogram Construction for Non-Survivors among Critical Patients with COVID-19. Jpn J Infect Dis 2020;73:452-8.
47. Chen R, Liang W, Jiang M, et al. Risk Factors of Fatal Outcome in Hospitalized Subjects With Coronavirus Disease 2019 From a Nationwide Analysis in China. Chest 2020;158:97-105.
48. Shang Y, Liu T, Wei Y, et al. Scoring systems for predicting mortality for severe patients with COVID-19. EClinicalMedicine 2020;24:100426-.
49. Li Q, Zhang J, Ling Y, et al. A simple algorithm helps early identification of SARS-CoV-2 infection patients with severe progression tendency. Infection 2020;48:577-84.
50. Zeng Z, Ma Y, Zeng H, et al. Simple nomogram based on initial laboratory data for predicting the probability of ICU transfer of COVID-19 patients: Multicenter retrospective study. Journal of Medical Virology 2020.
51. Gong J, Ou J, Qiu X, et al. A Tool for Early Prediction of Severe Coronavirus Disease 2019 (COVID-19): A Multicenter Study Using the Risk Nomogram in Wuhan and Guangdong, China. Clin Infect Dis 2020;71:833-40.
52. Shang W, Dong J, Ren Y, et al. The value of clinical parameters in predicting the severity of COVID-19. J Med Virol 2020;92:2188-92.

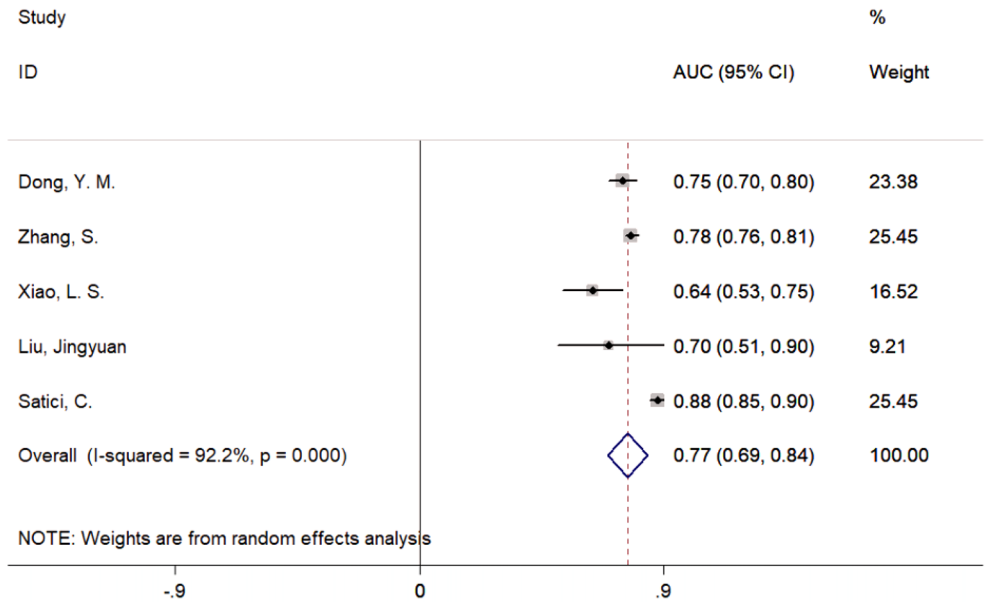
 http://dx.doi.org/10.21037/atm-20-6933

Appendix 3 Completeness of reporting TRIPOD items

Complete reporting ≥ 75%			Complete reporting ≤ 25%		
TRIPOD Items	content	Ratio	TRIPOD Items	content	Ratio
Item 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.	100% (D) 100% (V) 100% (D&V) 75% (IV)	Item 9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	5% (D) 14% (V) 8% (D&V) 0% (IV)
Item 3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	100% (D) 100% (V) 100% (D&V) 75% (IV)	Item 10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	0% (D) 0% (D&V) 0% (IV)
Item 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.	97% (D) 100% (V) 92% (D&V) 100% (IV)	Item 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.	5% (D) 0% (V) 0% (D&V) 0% (IV)
item 4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	89% (D) 93% (V) 85% (D&V) 100% (IV)	Item 17	If done, report the results from any model updating (i.e., model specification, model performance, recalibration). If updating was not done, score this TRIPOD item as 'Not applicable'.	0% (V) 0% (D&V)
item 5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centers.	84% (D) 100% (V) 92% (D&V) 100% (IV)			
item 5b	Describe eligibility criteria for participants.	95% (D) 100% (V) 100% (D&V) 100% (IV)			
item 5c	Give details of treatments received, if relevant. (i.e. notably for prognostic studies with long term follow-up)	83% (D) 100% (V) 100% (D&V) 100% (IV)			
Item 6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	89% (D) 100% (V) 92% (D&V) 100% (IV)			
Item 10c	For validation, describe how the predictions were calculated.	100% (V) 100% (D&V) 100% (IV)			
Item 10e	Describe any model updating (e.g., recalibration) arising from the validation, if done.	100% (V) 100% (D&V) 100% (IV)			
Item 11	Provide details on how risk groups were created, if done. If risk groups were not created, score this item as Yes.	81% (D) 100% (V) 100% (D&V) 100% (IV)			
Item 14a	Specify the number of participants and outcome events in each analysis.	100% (D) 100% (D&V) 100% (IV)			
Item 14b	If done, report the unadjusted association between each candidate predictor and outcome.	100% (D) 92% (D&V) 100% (IV)			

(D), (V), (D&V), (IV) stands for the following types of prediction models, namely development, external validation of existing models, development & validation of the same model and incremental values.

http://dx.doi.org/10.21037/atm-20-6933



AUC is the abbreviation of the area under the receiver operator characteristic curve.

Appendix 4 The forest plot of CURB-65 discrimination

http://dx.doi.org/10.21037/atm-20-6933