Radiology

Disease Severity Scoring for COVID-19: A Welcome Semiquantitative Role for Chest Radiography

Brent P. Little, MD

Dr Brent P. Little is a senior associate consultant radiologist in the Division of Cardiothoracic Imaging in the Department of Radiology at Mayo Clinic Florida. He has research interests in diffuse lung disease, infections (including COVID-19), and lung cancer screening, and has many interests in medical education.

Interest in quantitative medical imaging has accelerated, fueled by emergence of reliable and efficient computer software, and assisted by machine learning and artificial intelligence. In thoracic imaging, examples of considerable research interest include tumor volume and texture analysis, lung and airway volumetrics, lung CT densitometry, and nodule measurement. Previous research regarding quantification and standardized grading of acute thoracic diseases has an important but almost niche role over the past several decades; for example, there is sporadic literature regarding CT quantification of acute respiratory distress syndrome and cardiogenic pulmonary edema (1). Whereas quantitative CT and MRI applications in cardiovascular imaging have been fully integrated into routine radiologist workflow and reporting, clinical application of quantitative imaging in the thoracic realm is lagging. Historically, even simple disease severity scoring or grading systems have had only minor incorporation into radiology practice, such as the Scadding stages of sarcoidosis, certain cystic fibrosis scoring systems, and various attempts at pneumonia and edema grading. However, the COVID-19 pandemic has prompted compelling developments in the use of thoracic imaging to grade severity of acute pulmonary disease. In this issue of *Radiology*, the study by Au-Yong et al (2) provides evidence of the feasibility and prognostic power of radiographic disease severity scoring systems in COVID-19.

Although devastating, the COVID-19 pandemic has initiated a renaissance in imaging research in pneumonia, including quantitative and artificial intelligence applications to assist in diagnosis and prognostication of severity of COVID-19. In addition to descriptive and diagnostic studies regarding imaging in COVID-19 pneumonia, an expanding literature, including the study by Au-Yong et al, investigates the role of imaging in predicting patient outcomes, such as mortality and the need for intensive care and mechanical ventilation. Interest in pneumonia research has intensified because of the exigencies of the COVID-19 pandemic, with often limited hospital resources and widely disparate patient disease course and outcomes. Groups have found compelling correlations between chest radiography and CT severity scores and risk of death in patients who present to the emergency department and are then hospitalized with COVID-19 (3). Other studies have found correlations between chest radiograph severity scores and clinical end points such as intensive care unit admission, intubation, and death (4,5).

The study by Au-Yong et al assesses the reproducibility and prognostic value of three key chest radiography disease severity scoring systems. These include the radiographic assessment of lung edema (RALE) score, Brixia score, and percentage lung opacification. Three radiologists scored admission chest radiographs in 751 patients with COVID-19 with the three systems; 50 were scored by all readers to assess intra- and interreader variability. Scores were compared with outcomes of intensive care unit admission and death within 60 days. The scores were reproducible, showing strong associations with the outcome measures and higher prognostic value when combined with clinical prognostication systems. The study replicates important previous investigations regarding chest radiograph scoring and addresses several key questions in the use of severity scoring in COVID-19. The two most prominent systems (RALE and Brixia) are compared with a more straightforward assessment of lung parenchymal involvement (percent opacification score), with good reproducibility and prognostic power shown for all three.

Two of the lung scoring systems assessed by Au-Yong and colleagues were recently developed and validated by other groups. In 2018, Warren et al (6) described the RALE score, the sum of the products of extent and severity grades for each of the four lung quadrants, with score ranges from 0 to 48. In a proof of concept by using a cohort of whole-lung specimens from 72 deceased lung transplant donors with contemporaneous chest radiographs, the group found a positive correlation between RALE scores and lung weights adjusted for body height. In a separate cohort of 174 patients from one of the ARDSNet trials, the group also found that higher RALE scores were independently associated with lower PaO_2/FiO_2 and higher mortality. In addition, with every

From the Division of Cardiothoracic Imaging, Department of Radiology, Mayo Clinic Florida, 4500 San Pablo Rd, Jacksonville, FL 32224. Received August 30, 2021; revision requested September 2; revision received September 3; accepted September 7. **Address correspondence to** the author (e-mail: *little.brent@mayo.edu*).

Conflicts of interest are listed at the end of this article.

See also the article by Au-Yong et al in this issue.

Radiology 2022; 302:470–472 • https://doi.org/10.1148/radiol.2021212212 • Content code: CH

five-point decrease in RALE score, the adjusted hazard of death decreased by 16%. The Brixia score proposed in 2020 by Borghesi et al (7) is similar but uses six zones and a slightly different grading of opacities, resulting in scores ranging from 0 to 18. In 302 patients with COVID-19, the Brixia score (along with age and immunosuppressed status) was one of the few variables to show an independent correlation with in-hospital mortality. Impressively, the scoring system was actually put into routine clinical use at the authors' institution, with a score appended to the chest radiography report. Importantly, additional permutations of severity scoring were studied during the Middle Eastern respiratory syndrome (known as MERS) (8) and the sudden acute respiratory syndrome (known as SARS) (9) outbreaks, with similarly promising results.

The study by Au-Yong et al provides further evidence of the potential value of severity scoring on chest radiographs in patients with COVID-19 and answers several key questions. First, what is the predictive power of severity scoring by using imaging alone? The authors found that all severity scoring systems stratified patients by survival and escalation-free survival. Second, could the integration of chest radiograph scoring with clinical risk scores incorporating data such as vital signs, oxygenation status, and blood markers of inflammation improve the predictive power of models? The authors found the combination of clinical severity scoring systems and chest radiograph severity scoring improved model discrimination, with best results for the combination of National Early Warning Score 2 (known as NEWS2) and RALE scores, concordant with previous studies (4). Third, and perhaps most important from a practical consideration, is severity scoring feasible in terms of ease of training, reproducibility, and interpretability? The authors found that all systems could be used to score chest radiographs in under a minute each and found good intra- and interreader correlations in scores from all systems. Limitations were certainly present, such as the inclusion of only cases severe enough to merit hospitalization, but other studies have already shown relationships between higher chest radiograph severity scores and higher risk of poor outcomes in mild disease (4). The study also raises questions about the balance between ease of use and predictive power of scoring systems. One wonders if even simpler systems (eg, unifocal vs multifocal and unilateral vs bilateral) could have similar results. The generalizability of the prognostic power for COVID-19 variants of concern, such as the Delta variant, is also uncertain.

The study by Au-Yong et al is part of a body of research supporting the status of chest radiograph severity scoring as an important prognostic marker that provides an index of the pulmonary effects of COVID-19. Au-Yong et al also provide good evidence that severity scoring is reproducible and relatively rapid and capable of being integrated into a clinical workflow. Scoring was consistent across observers in different imaging subspecialties, with good interreader agreement between the scores of subspecialists in chest, breast, and gastrointestinal radiology. Although the quoted average times for scoring a single radiograph with all scoring systems were under 1 minute, the times could potentially be shorter with increasing reader experience.

In spite of the promising literature on severity scoring in COVID-19 and other acute pulmonary diseases, including the work by Au-Yong et al, chest radiograph and CT severity scoring are generally absent from most clinical radiology practices and a devastating pandemic has so far done little to change this. Why? A primary reason might be that despite years of research interest, potential clinical decision-making uses for such scoring systems are still unstudied and unproven. Should severity scores be used for triage, to make admission decisions, or to predict need to escalation of care to the intensive care unit? Should severity scores be used to help assess need for intubation? During peak admissions in hospital systems with limited resources, could chest radiograph severity assessment serve as an early warning system to predict need for resources hours or days in advance? Although clinical studies providing guidance about the effects of incorporating chest radiography (or CT) disease severity scoring into clinical decision making are lacking, this may be a "first mover" problem. Routine use of severity scoring could trigger and facilitate such research, but skeptics await proof of clinical uses before implementation. Clinical use of such systems might be expected to grow in proportion to availability; clinical insights might arise in the same way in which laboratory testing and patient telemetry can provide clinical guidance for patient management.

Implementation of routine reporting of radiographic severity scores for diseases such as COVID-19 pneumonia may meet skepticism in radiology. Adding tasks to growing workloads is seldom popular. However, Au-Yong et al show that severity scoring may be practical, potentially adding only seconds to a subset of chest radiography reports. In addition, some groups have explored the use of artificial intelligence to provide automated severity scoring or to assist radiologists in scoring, with good correlation to human scoring and equivalent prognostic power (10). Regardless of how scoring is performed, the practice might provide a welcome inclusion of quantitative or semiquantitative information in the chest radiograph report, potentially replacing subjective terms like *extensive*, *severe*, *dense*, *mild*, *patchy*, and *hazy*. Perhaps studies from the COVID-19 pandemic like that of Au-Yong et al might finally encourage adoption of validated standardized severity scoring systems, bringing fresh clinical relevance of a more quantitative role for the chest radiograph in the assessment and management of acute pulmonary disease.

Disclosures of Conflicts of Interest: B.P.L. Royalties from Elsevier.

References

1. Gattinoni L, Caironi P, Cressoni M, et al. Lung recruitment in patients with the acute respiratory distress syndrome. N Engl J Med 2006;354(17):1775–1786.
2. Au-Yong I, Higashi Y, Giannotti E, et al. Chest radiograph scoring alone or combined with other risk scores for predicting outcomes in COVID-19. Radiology 2022;302(2):460–469.
3. Francone M, Iafrate F, Masci GM, et al. Chest CT score in COVID-19 patients: correlation with disease severity and short-term prognosis. Eur Radiol 2020;30(12):6808–6817.
4. Toussie D, Voutsinas N, Finkelstein M, et al. Clinical and Chest Radiography Features Determine Patient Outcomes in Young and Middle-aged Adults with COVID-19. Radiology 2020;297(1):E197–E206.
5. Schalekamp S, Huisman M, van Dijk RA, et al. Model-based Prediction of Critical Illness in Hospitalized Patients with COVID-19. Radiology 2021;298(1):E46–E54.

6. Warren MA, Zhao Z, Koyama T, et al. Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ARDS. Thorax 2018;73(9):840–846.
7. 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;96:291–293.
8. Cha MJ, Chung MJ, Kim K, Lee KS, Kim TJ, Kim TS. Clinical implication of radiographic scores in acute Middle East respiratory syndrome coronavirus pneumonia: Report from a single tertiary-referral center of South Korea. Eur J Radiol 2018;107:196–202.
9. Chau TN, Lee PO, Choi KW, et al. Value of initial chest radiographs for predicting clinical outcomes in patients with severe acute respiratory syndrome. Am J Med 2004;117(4):249–254.
10. Li MD, Arun NT, Gidwani M, et al. Automated Assessment and Tracking of COVID-19 Pulmonary Disease Severity on Chest Radiographs using Convolutional Siamese Neural Networks. Radiol Artif Intell 2020;2(4):e200079.