

CT Diagnosis of COVID-19: A View through the PICOTS Lens

Manuscript Type: Editorial

Geoffrey D. Rubin, MD, MBA
University of Arizona Medical Center - University Campus - Department of Radiology
1501 N. Campbell Ave. Tucson Arizona 85724-5128
United States
grubin@arizona.edu

See also the article by Revel et al.

No other event has upended worldwide healthcare in the past 100 years like the COVID-19 pandemic. Over the almost 18 months since the SARS-CoV2 virus and its associated clinical manifestations emerged, 173 million COVID-19 cases have been confirmed, and 3.7 million have died worldwide (1). Scientific discovery and healthcare innovation have responded with unprecedented focus and accomplishment. The availability and widespread initial distribution of an effective vaccine less than one year after the emergence of the pathogen have been a triumph of modern science. Refinements to the diagnosis and management of COVID-19 have substantially improved outcomes for hospitalized patients since disease surges overtook healthcare delivery systems in New York, Wuhan, and Northern Italy in the late winter of 2020 (2, 3).

The impact of COVID-19 on the medical literature has been similarly disruptive, with 140,000 clinical research publications listed in PubMed. The effort to amass this body of knowledge over just 18 months has been heroic and represents a tremendous effort by thousands of scientists and physicians worldwide operating through the multi-dimensional uncertainty and turmoil of the pandemic. In this issue of *Radiology*, the work of the STOIC investigators to amass and systematically interpret over 10,000 CT scans acquired from 20 university hospitals in France during the first wave of the pandemic is emblematic of these herculean efforts. It represents the largest cohort evaluated for COVID-19 and studied with thoracic CT to date (4). Separate from the responsibilities of their clinical work, 20 thoracic radiologists each read up to 1,122 (mean of 537) of 10,735 CT scans, undoubtedly representing thousands of hours of effort to provide the medical community with a snapshot of COVID-19 manifestations on thoracic CT during two months in the Spring of 2020 in mostly Paris, France.

Looking back upon this huge cohort, the STOIC investigators pose and answer two primary questions:

- (1) During the initial surge of the COVID-19 pandemic among patients hospitalized in Paris with dyspnea or blood oxygen desaturation, what was the diagnostic performance of CT for SARS-Cov-2 infection relative to prevailing RT-PCR methods for virus detections?
- (2) Was the extent of lung abnormalities detected with CT associated with severe versus non-severe outcomes one month after presentation, and how did the magnitude of that association compare to age, sex, use of oxygen supplementation, hypertension, and coronary artery disease?

The PICOT framework has been widely endorsed to prospectively structure research questions and retrospectively analyze published research. A PICOTS question defines the Population, Intervention, Comparator, Outcome, Timeframe, and Setting (5, 6).

The first question asked by the STOIC investigators centers on the diagnosis of COVID-19 by thoracic CT. The study Population comprises patients presenting to a hospital with dyspnea or desaturation as measured by pulse oximetry. The Intervention was the classification of a CT scan as either positive or negative for COVID-19. The Comparator was RT-PCR results for an Outcome of COVID-10 diagnosis. The Timeframe was upon initial hospitalization for COVID-19, and the Setting was inpatients within major urban hospitals in France. Framing the diagnostic

question informs the applicability of the results to clinical decision-making within future clinical scenarios.

Relevant to the patient population and setting, early in the pandemic, the community-based pretest probability of COVID-19 was called out together with the severity of disease manifestations, patient-specific risk factors, and resource constraints as critical factors for determining the utility of imaging of the lungs in COVID-19 (7).

The STOIC setting included a prevalence or pretest probability for COVID-19 of 60%, which is very high relative to outpatient and typical non-surge inpatient settings. By comparison, in April 2021, India reported an almost 20% COVID-19 test positivity, among the highest community-based COVID-19 prevalence since the pandemic began (1). At the same time, test positivity in the United States was 8.25% (8).

The high pretest probability of the STOIC setting has a profound influence on the generalizability of its results. The likelihood that a patient has COVID-19 given either a positive or negative CT expressed as positive (PPV) and negatives (NPV) predictive values, respectively, relates to disease prevalence as follows:

$$PPV = \frac{(Se \times Pr)}{(Se \times Pr) + [(1 - Sp) \times (1 - Pr)]} \quad \text{and} \quad NPV = \frac{(Sp \times (1 - Pr))}{(Sp \times (1 - Pr)) + [(1 - Se) \times (Pr)]}$$

where Se, Sp, and Pr represent sensitivity, specificity, and prevalence, respectively.

STOIC reported a PPV of 85.6%. In other words, 14.4% of individuals had positive CTs despite testing negative with RT-PCR. However, recalculating PPV using STOIC sensitivity and specificity values applied to the COVID-19 prevalence in late April 2021, populations tested with RT-PCR in India and the U.S. results in PPVs of 49.7% and 26.2%, respectively, for CT to diagnose COVID-19. Thus, based upon the sensitivity and specificity of CT determined through STOIC, when applied to a population with pretest probability akin to those undergoing testing in the United States in late April 2021, almost 3 of every 4 would receive a false-positive diagnosis of COVID-19. A clear definition of pretest probability is critical to understanding the applicability of CT for the determination of a COVID-19 diagnosis. While the size of the STOIC cohort is large by diagnostic imaging standards, the generalizability of its findings must be considered relative to disease prevalence.

Of note, early reports comparing CT diagnosis to RT-PCR reported high rates of false-positive CT results. These high false-positive rates were attributed to imperfections in RT-PCR techniques, inconsistencies in the reporting of CT findings, and timing of CT acquisition and RT-PCR testing relative to disease onset (9, 10). Reporting consistency supported by the use of structured CT reporting and what can be presumed to be reliable RT-PCR technique likely contributed to the high specificity reported by STOIC. However, assuming this high specificity could be reproduced in an outpatient setting, it is not high enough to avoid an overwhelming number of false-positive results in low prevalence environments.

Turning from the “P” in PICOT to the “T,” the STOIC investigators provided a sub-analysis of 3,141 patients who underwent CT five days after the onset of symptoms, and COVID-19

prevalence was 92.1%. Measured at this second timepoint, PPV rose to 96.7%, but NPV fell to 32.1%, meaning that two of every three negative CTs were in RT-PCR positive patients. From these data, we might reasonably conclude that a positive CT is highly predictive of COVID-19 when prevalence is high, and a negative CT is highly predictive of the absence of COVID-19 when prevalence is low. Conversely, a negative CT result is unreliable when disease prevalence is high, and a positive CT is unreliable for COVID-19 diagnosis when prevalence is low.

The preceding analysis has important implications. Consider if STOIC's 10,735 CT scans and associated interpretations were used to train a machine-learning algorithm to facilitate the diagnosis of COVID-19. Perhaps a well-designed algorithm trained using the STOIC cohort may achieve superior diagnostic performance than the 20 chest radiologists from the STOIC project when validated against a portion of the STOIC cohort withheld from training. That would be impressive. However, the algorithm would be useful only for cohorts with disease prevalence approximating the STOIC cohort. For cohorts with a similar prevalence of COVID-19 to that encountered in communities tested worldwide today, validation of the algorithm would need to occur using a matched cohort. The risk of skipping this step would result in an overwhelming amount of overdiagnosis. Performing on par with the 20 thoracic radiologists of STOIC, an AI algorithm would falsely diagnose COVID-19 in four of every five CT scans when applied to a cohort where 10% were RT-PCR positive.

In a second analysis using multivariable regression, STOIC found that CT was significantly more predictive of severe outcomes than age, sex, the use of oxygen supplementation, hypertension, past coronary artery disease, or severe coronary calcium. Specifically, CT documented pneumonia involving $\geq 50\%$ of the lungs was more predictive of endotracheal intubation or death. While this result supports published recommendations to perform CT for worsening respiratory status (7), the actionability of this information is limited for two reasons. First, the independent variables for the regression model did not include biometric or laboratory data points commonly available in hospitalized patients, leaving open the possibility that prediction may be equally or more effective when considering other frequently available clinical data elements. Second, while disease extent predicts severity of outcome, its effectiveness for specific clinical decision-making is unknown.

Informed by PICOTS, a focus on clinical actionability, and knowledge of interval scientific discoveries, therapeutic innovations, and societal trends, we may formulate the following questions when confronted with an opportunity to study the utility of CT in COVID-19:

- (1) Can COVID-19 manifestations measured by CT uniquely guide when the application of intervention X improves the likelihood of outcome Y when considered in association with other available patient characteristics Z? In this question, X may be a therapeutic maneuver or agent such as immunotherapy, Y may be the duration and nature of disease manifestations, need for hospitalization, chronic sequelae, disability, mortality, financial toxicity, or healthcare resource utilization, and Z may be co-morbidities or biometric, demographic, or social factors.
- (2) If so, then what is the best way to measure disease extent with CT to effectively guide the application of the intervention?

Reliable answers to these questions lie in the prospective construction of controlled trials where data elements are collected comprehensively, avoidable biases are controlled, and unavoidable biases are articulated.

While the retrospective construction and heterogeneous data availability of STOIC limits the immediate derivation of actionable clinical insights, the publication is a landmark, nevertheless. It provides a valuable historical record regarding the use of thoracic CT at a highly volatile moment during the early days of the pandemic. The creation and curation of a 10,735 patient cohort presenting with suspected COVID-19 during a two-month window was a vast undertaking to pursue amidst the disruption of the pandemic. Once publicly available, these data will be in demand by a variety of investigators seeking to better understand the manifestations of COVID-19 and inform future healthcare practice. Those investigators will be wise to consider PICOTS when framing their research questions to ensure their efforts support specific actions that enhance the effectiveness of healthcare interventions.



Geoffrey D. Rubin, MD, MBA, FACR, FAHA, FSABI, FNASCI is Professor and Chair of Medical Imaging at the University of Arizona. He was previously the George B. Geller Distinguished Professor for Cardiovascular Research at Duke University and professor of radiology at Stanford University. He is past president of the Fleischner Society and currently chairs the Fleischner Publication Development and Oversight Committee.

References

1. World Health Organization. (2021). COVID-19 Explorer. Retrieved May 28, 2021 from <https://worldhealthorg.shinyapps.io/covid/>
2. Narain, S., Stefanov, D. G., Chau, A. S., et al. Comparative Survival Analysis of Immunomodulatory Therapy for Coronavirus Disease 2019 Cytokine Storm. *Chest*. 2021;159(3):933-948
3. Lam, C., Siefkas, A., Zelin, N. S., et al. Machine Learning as a Precision-Medicine Approach to Prescribing COVID-19 Pharmacotherapy with Remdesivir or Corticosteroids. *Clin Ther*. 2021S0149-2918(21)00128
4. Revel, M. P., Chassagnon, G., Lederlin, L., et al. Study of Thoracic CT in COVID-19: The STOIC Project. *Radiology* (in press)
5. Huang, X., Lin, J., & Demner-Fushman, D. Evaluation of PICO as a knowledge representation for clinical questions. *AMIA. Annual Symposium proceedings. AMIA Symposium*. 2006;2006:359-363
6. Matchar, D. B. (2012). Introduction to the Methods Guide for Medical Test Reviews. In S. M. Chang & D. B. Matchar (Eds.), *Methods Guide for Medical Test Reviews* (Review ed.). AHRQ Rockville (MD): Agency for Healthcare Research and Quality (US). Retrieved from <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22834022&retmode=ref&cmd=prlinks>
7. Rubin, G. D., Ryerson, C. J., Haramati, L. B., et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. *Radiology*. 2020;296(1):172-180
8. Centers for Disease Control and Prevention. (2021). COVID Data Tracker. Retrieved May 28, 2021 from https://covid.cdc.gov/covid-data-tracker/#testing_positivity7day
9. Kovács, A., Palásti, P., Veréb, D., et al. The sensitivity and specificity of chest CT in the diagnosis of COVID-19. *Eur Radiol*. 2021;31(5):2819-2824
10. De Smet, K., De Smet, D., Ryckaert, T., et al. Diagnostic Performance of Chest CT for SARS-CoV-2 Infection in Individuals with or without COVID-19 Symptoms. *Radiology*. 2021;298(1):E30-E37