Diagn Interv Radiol 2022; DOI: 10.5152/dir.2022.21488

CHEST IMAGING

ORIGINAL ARTICLE



# Structured report improves radiology residents' performance in reporting chest high-resolution computed tomography: a study in patients with connective tissue disease

Lorenzo Cereser
Filippo Marchesini
Emma Di Poi
Luca Quartuccio
Alen Zabotti
Chiara Zuiani
Rossano Girometti

**PURPOSE**
The study aimed to evaluate the performance of radiology residents (RRs) when using a dedicated structured report (SR) template for chest high-resolution computed tomography (HRCT) in patients with suspected connective tissue disease–interstitial lung disease (CTD–ILD), compared to the traditional narrative report (NR).

**METHODS**
We retrospectively evaluated 50 HRCT exams in patients with suspected CTD–ILD. A chest-devoted radiologist reported all HRCT exams as the reference standard, pointing out pulmonary fibrosis findings (i.e., honeycombing, traction bronchiectasis, reticulation, and volume loss), the presence and pattern of ILD, and possible other diagnoses. We divided 4 RRs into 2 groups according to their expertise level. In each group, RRs reported all HRCT examinations alternatively with NR or SR, noting each report's reporting time. The Cohen's Kappa, Wilcoxon, and McNemar tests were used for statistical analysis.

**RESULTS**
Regarding the pulmonary fibrosis findings, we found higher agreement between RRs and the reference standard reader when using SR than NR, regardless of their expertise level, except for volume loss.

RRs' accuracy for "other diagnosis" was higher when using SR than NR, moving from 0.48 to 0.66 in the novel group ($P = .035$) and from 0.44 to 0.80 in the expertise group ($P < .001$). No differences in accuracy were found between ILD presence and ILD pattern. The reporting time was significantly lower ($P = .001$) when using SR than NR.

**CONCLUSION**
SR is of value in increasing the reporting of critical chest HRCT findings in the complex CTD–ILD scenario and should be used early and systematically during residency.

Interstitial lung diseases (ILD) are heterogeneous lung pathologies classified as idiopathic or secondary.[1] Among the latter, many cases are framed as connective tissue disease (CTD)-associated ILD (CTD–ILD). The term CTD refers to a heterogeneous group of autoimmune disorders causing immune-mediated organ dysfunction. This group includes rheumatoid arthritis, systemic sclerosis, Sjogren's syndrome, antisynthetase syndrome, polymyositis, dermatomyositis, systemic lupus erythematosus, and mixed connective tissue disease.[2] Notably, nearly 40% of patients with CTD develop ILD, thus testifying this condition's epidemiological relevance.[3]

Chest high-resolution computed tomography (HRCT) is considered the reference standard for imaging CTD-ILD.[4,5] CTD-ILD may manifest with different typical HRCT patterns of lung involvement, including usual interstitial pneumonia (UIP), nonspecific interstitial pneumonia (NSIP), organizing pneumonia (OP), NSIP overlap OP (NSIP–OP), and lymphocytic interstitial pneumonitis (LIP).[6]

The structured report (SR) is a radiology report model aimed to improve the readability of conventional narrative reports (NR) by referring clinician and, ultimately, lead to higher clinical usefulness.[7,8] Various benefits of SR have been identified in training environments,

From the Department of Medicine (L.C., ✉ lcereser@sirm.org, F.M., C.Z., R.G.), Institute of Radiology, University of Udine, Udine, Italy; and Department of Medicine (E.D.P., L.Q., A.Z.), Rheumatology Clinic, University of Udine, Udine, Italy.

Received 1 August 2021; revision requested 16 September 2021; last revision received 4 October 2021; accepted 13 October 2021.

Publication date: 1 December 2022.

DOI: 10.5152/dir.2022.21488

You may cite this article as: Cereser L, Marchesini F, Di Poi E, et al. Structured report improves radiology residents' performance in reporting chest high-resolution computed tomography: a study in patients with connective tissue disease. *Diagn Interv Radiol.* 2022;28(6):569-575.

for example, it aids in training report organization and teaching critical anatomy, key findings, and complications, with improved reporting uniformity and efficiency.[9] These advantages led international institutions supporting high-quality post-graduation education (i.e., the European Union of Medical Specialists and the Accreditation Council for Graduate Medical Education) to promote understanding and using SR as a requirement for radiology residents (RRs).[10,11] Various studies explored the value of SR as a teaching tool for RRs regarding diverse diagnostic settings and techniques (e.g., chest radiograph, head and neck ultrasound, cervical spine computed tomography [CT], CT angiography, maxillofacial CT, and CT enterography).[12–17] Indeed, in an article demonstrating higher completeness and effectiveness of SR than NR for chest radiograph, Marcovici and Taylor[12] argued that SR might help RRs by promoting and reinforcing a rational cognitive approach at the time of imaging interpretation. Of note, while generally promoting SR implementation during residency programs, some authors warned about SR's potential downsides due to a questionable impact on accuracy compared to NR.[14,17,18]

To the best of our knowledge, no published studies have evaluated the impact of chest HRCT structured reporting on RRs' performance. The pleomorphism of the many possible lung presentations at HRCT, sometimes with overlapping imaging features, may be difficult to interpret for a non-chest-devoted radiologist and, even more, for RRs. In this light, an SR template for HRCT and CTD–ILD has recently proved to be a valuable tool for radiologists, with high referring clinicians' perceived report quality.[19] We believe this SR model may provide a valuable framework for RRs approaching the HRCT study of CTD–ILD. Thus, the study's purpose was to evaluate RRs' performance when using a dedicated SR model for chest HRCT in patients with suspected CTD–ILD compared to the traditional NR.

**Main points**

- Structured report helps radiology residents to detect fibrosis findings.
- Structured report aids radiology residents not to overlook alternative diagnoses.
- Radiology residents' reporting time is shorter when the report is structured.

## Methods

### Study population

The Institutional Review Board of the Department of Medicine - University of Udine approved the study (protocol no. 049/2020_IRB). The need for written informed consent was waived due to the retrospective design.

By performing a computerized search, we identified 180 consecutive patients aged ≥18 years with CTD who underwent chest HRCT in our institution from July 2012 to November 2020. Patients were referred from the Rheumatology Clinic of the same center. We included in the analysis the first examination in the case of multiple exams performed on the same patient. We excluded 7 patients whose clinical data at the time of HRCT were not available. Finally, we randomly selected 25 cases with confirmed CTD–ILD and 25 cases with no CTD–ILD for a total of 50 HRCT cases that formed the study population. Randomization was performed with freely available software (https://www.randomizer.org).

### Chest HRCT exams

All included chest HRCT exams were performed on a 64-row multidetector CT scanner (Discovery HD750, GE Healthcare). Exams were acquired volumetrically with patients in the supine position at the end of full inspiration. The main technical parameters were tube potential, 100-120 kV (according to patient size); tube current modulation range, 150-350 mA; detector configuration, 64 × 0.625 mm; reconstructed slice thickness, 1.25 mm; reconstructed interval, 1.25 mm. Additional end-expiratory volumetric scans, prone scans, and postcontrast (iobitridol 350 mgI/mL, Xenetix, Guerbet or iomeprol 400 mgI/mL, Iomeron, Bracco Imaging) acquisitions were disposable in 28/50, 6/50, and 8/50 patients, respectively. Images were reconstructed using a high-spatial-frequency algorithm and parenchymal windowing (level, −500 HU; width, 1700 HU) for the lungs and a soft tissue algorithm and windowing (level, 50 HU; width, 350 HU) for the mediastinum and chest wall.

### Standard of reference

The multidisciplinary team opinion is regarded as the gold standard for the final diagnosis of ILD.[20,21] However, since our study focused on radiological interpretation, we used a surrogate standard or reference, that is, the interpretation provided by a single radiologist with 12 years of experience in chest HRCT. The radiologist interpreted the images without knowing the clinical and follow-up data.

For each examination, the reference reader established the presence/absence of 4 fibrosis findings (i.e., honeycombing, traction bronchiectasis, reticulation, and volume loss).[22] On these bases, he also provided a final impression regarding the presence/absence of CTD-ILD, HRCT pattern associated with ILD (i.e., NSIP, OP, NSIP–OP, LIP), and possible other diagnoses.

### Residents' selection process and groups' allocation

A study coordinator, not involved in image reading, selected 4 RRs (RR_1-4), attending the 4-year training program of our institution, and divided them into 2 groups, according to the expertise level in CT imaging, that is, the number of reported CT examinations during the residency at the time of the study's beginning. Group 1 included 2 RRs who reported 200-300 CT examinations each (RR_1 and RR_2, novice residents), while group 2 included 2 RRs who reported 600-700 CT examinations each (RR_3 and RR_4, more experienced residents). The study coordinator randomly assigned the reporting strategy (SR or NR) to readers within each group: NR for RR_1 and RR_3 and SR for RR_2 and RR_4.

The NR was defined as a freely written report. The study coordinator gave no specific indication to the RRs about the content organization, leaving the readers free to provide a final impression or not.

Our SR template was a recently published one,[19] derived from an SR for fibrosing lung disease based on a Delphi survey involving chest-devoted radiologists.[23] All the included items follow the Fleischner Society glossary definitions[24] (Table 1). According to Nobel et al.,[7] the SR we used can be defined as a "level 1 SR," designed with an itemized format. The study coordinator provided RR reporting with SR (i.e., RR_2 and RR_4), an electronic document (Microsoft Word document) with our template.

### Image analysis

The RRs performed the readings independently, blinded to the study aim. Concerning clinical information, RRs were only aware of patient's gender, age, and a standardized indication for imaging, provided under the form of "suspected ILD in a patient with CTD." No other clinical data were disclosed.

**Table 1.** Structured report template for chest HRCT in patients with connective tissue disease and known or suspected interstitial lung disease

| | | |
|---|---|---|
| Clinical information | | |
| Ref. number | | |
| Age and gender | | |
| Clinical indication | | |
| HRCT findings | | |
| Lung fibrosis findings | Honeycombing | Yes/no |
| | Traction bronchiectasis | Yes/no |
| | Signs of volume loss | Yes/no |
| | Reticulation | Yes/No |
| | Distribution: *cranio-caudal* | Upper lung/middle lung lower lung |
| | Distribution: *axial* | Subpleural/peribronchovascular/ diffuse/subpleural sparing |
| | Estimated total extent of fibrosis (%) | |
| Lung parenchyma findings | Ground-glass opacities | Yes/no (cranio-caudal and axial distribution); *free text* |
| | Nodules | Yes/no (cranio-caudal and axial distribution); *free text* |
| | Cysts | Yes/no (cranio-caudal and axial distribution); *free text* |
| | Emphysema | Yes/no (cranio-caudal and axial distribution); *free text* |
| | Other pulmonary findings | *Free text (e.g., consolidation, findings suggesting OP or LIP)* |
| Other findings | Pleural effusion | Yes/no *(quantification)* |
| | Pericardial effusion | Yes/no *(quantification)* |
| | Lymph nodes enlargement | Yes/no *(station and size)* |
| | Pulmonary artery enlargement | Yes/no *(specify diameter in mm)* |
| Conclusions | | |
| Is ILD present? | Yes/no | |
| If yes, specify the HRCT pattern | NSIP; OP; NSIP–OP; UIP; acute complication in UIP; LIP; DAD; others | |
| Other diagnoses and recommendations | *Free text* | |

HRCT, high-resolution computed tomography; –OP, organizing pneumonia; LIP, lymphocytic interstitial pneumonia; ILD, interstitial lung disease; NSIP, nonspecific interstitial pneumonia; NSIPOP, nonspecific interstitial pneumonia overlap organizing pneumonia; UIP, usual interstitial pneumonia; LIP, lymphocytic interstitial pneumonia; DAD, diffuse alveolar damage.

Readers evaluated all HRCT examinations on a Picture Archiving and Communication System workstation (Suitestensa Ebit Srl, Esaote Group Company).

For each reading, the study coordinator measured the time needed to make the report (reporting time), defined as the time (expressed in seconds) between image presentation and the end of report writing.

### Statistical analysis

Descriptive statistics of the data are presented with n (%) and for non-normalized variables, data are shown as median (25-75 percentiles). Data normality was checked with the Shapiro–Wilk test. Each RR's performance using the SR for reporting the HRCT exams was assessed on 3 domains: description, impression, and reporting time.

#### *Description*

Cohen's Kappa (κ) was used to evaluate each RR and the reference reader's agreement in assessing the presence or absence of any fibrosis findings on a per-patient basis. The agreement was also evaluated between RRs on an intergroup basis, namely RR_1 versus RR_3 and RR_2 versus RR_4.

Interpretation of κ coefficient was as follows: <0.00, poor; 0.00-0.20, slight; 0.21-0.40, fair; 0.41-0.60, moderate; 0.61-0.80, substantial; 0.81-1.00, almost perfect.[25]

#### *Impression*

By comparing RRs' impression with that of the reference reader, we calculated RRs' accuracy in assessing ILD presence, HRCT pattern (when ILD is present), and "other diagnosis." Per-feature accuracy was defined as the ratio between the number of patients with the correct impression (true positives + true negatives) over the total number of cases. When NR or SR was ambiguous regarding the impression parameters or an impression was not provided, the report was not computed as a true positive nor true negative case. For each impression parameter, we used the McNemar test to assess the differences between accuracy of RRs in the same group, that is, RR_1 versus RR_2 and RR_3 versus RR_4.

#### *Reporting time*

The difference in reporting time on an intergroup basis (NR vs. SR) was assessed with the Wilcoxon test. For the analysis, we merged the reporting times of RR_1 + RR_3 and RR_2 + RR_4. By performing the Sieve-bootstrap Mann–Kendall test and the Student t-test for a linear trend,[26] we also tested if the reporting time of the residents using SR (i.e., RR_2 and RR_4) varied as the number of reporting activities increased.

The α level was set at 0.05. All statistical analyses were performed using commercially available software (MedCalc Software bvba, version 18.11.6, Ostend, Belgium; R x64 v.4.0.2, R Core Team 2013, R: A language and environment for statistical computing. R Foundation for Statistical Computing).

## Results

The study population (n = 50) included 41 (82%) women and 9 (18%) men, with a median age of 66.5 (52-71). CTDs were as follows: 29 (58%) systemic sclerosis, 6 (12%) rheumatoid arthritis, 5 (10%) Sjögren syndrome, 6 (12%) polymyositis–dermatomyositis, and 4 (8%) systemic lupus erythematosus.

Concerning the 25 cases with CTD–ILD, the HRCT patterns were 18 (72%) NSIP, 1 (4%) OP, 3 (12%) NSIP–OP, 2 (8%) UIP, and 1 (4%) LIP.

**Table 2.** Inter-reader agreement between each radiology resident and the standard of reference for each fibrosis finding

| | Group 1—*Novice RRs* | | Group 2—*More experienced RRs* | |
|---|---|---|---|---|
| | NR<br>RR_1<br>κ (95% CI)<br>*P* | SR<br>RR_2<br>κ (95% CI)<br>*P* | NR<br>RR_3<br>κ (95% CI)<br>*P* | SR<br>RR_4<br>κ (95% CI)<br>*P* |
| Honeycombing | Fair<br>0.40 (−0.02 to 0.81)<br>.003 | Moderate<br>0.46 (0.02-0.90)<br><.001 | Fair<br>0.34 (−0.05 to 0.74)<br>.007 | Moderate<br>0.46 (0.02-0.90)<br><.001 |
| Traction bronchiectasis | Moderate<br>0.59 (0.37-0.82)<br><.001 | Substantial<br>0.67 (0.46-0.87)<br><.001 | Moderate<br>0.55 (0.31-0.78)<br><.001 | Substantial<br>0.71 (0.51-0.90)<br><.001 |
| Reticulation | Moderate<br>0.60 (0.38-0.82)<br><.001 | Substantial<br>0.76 (0.58-0.94)<br><.001 | Moderate<br>0.60 (0.38-0.82)<br><.001 | Substantial<br>0.76 (0.58-0.94)<br><.001 |
| Volume loss | Substantial<br>0.63 (0.41-0.85)<br><.001 | Substantial<br>0.70 (0.49-0.91)<br><.001 | Slight<br>0.11 (-0.11 to 0.32)<br>.285 | Substantial<br>0.71 (0.50-0.91)<br><.001 |

RRs, radiology residents; NR, narrative report; SR, structured report; RR_1-4, radiology resident_1-4; κ, Cohen's Kappa.

"Other diagnosis" category was assigned in 33 (66%) cases, including 21 pulmonary findings other than ILD (i.e., 4 nodules, 1 mass, 7 signs of active infection, and 9 stigmata of previous infections); 1 pleural effusion; 4 mediastinal findings (i.e., 1 thymic lesion, 1 esophageal ectasia, and 2 lymphadenopathies); 4 vascular findings (i.e., 4 signs of pulmonary hypertension); and 3 other abnormal findings (i.e., 2 thyroid goiter and 1 thoracic vertebral fracture).

The per-finding inter-reader agreement between each RR and the reference reader is reported in Table 2. Except for volume loss, we found higher agreement with the standard of reference when using SR than NR, regardless of the findings and the group.

As detailed in Table 3, the inter-group agreement between RRs was higher with SR than NR. The only exception was the agreement for reticulation, which was moderate when using both NR and SR.

To determine whether the agreement was due to chance, we also compared the *P*-values to the significance level. In most cases, the *P*-value was far less than the significance level, confirming that the agreement is significantly different from what would be achieved by chance (Tables 2 and 3).

**Table 3.** The inter-reader agreement between radiology residents reporting with NR vs. SR for each fibrosis finding.

| | Narrative report<br>RR_1 vs. RR_3<br>κ (95% CI)<br>*P* | Structured report<br>RR_2 vs. RR_4<br>κ (95% CI)<br>*P* |
|---|---|---|
| Honeycombing | Moderate<br>0.56 (0.21-0.90)<br><.001 | Substantial<br>0.78 (0.48-1.00)<br><.001 |
| Traction bronchiectasis | Fair<br>0.38 (0.12-0.64)<br>.006 | Substantial<br>0.77 (0.59-0.96)<br><.001 |
| Reticulation | Moderate<br>0.43 (0.18-0.68)<br>.002 | Moderate<br>0.60 (0.38-0.82)<br><.001 |
| Volume loss | Fair<br>0.28 (-0.04 to 0.59)<br>.275 | Almost perfect<br>0.83 (0.67-0.99)<br><.001 |

RRs, radiology residents; RR_1-4, radiology resident_1-4; κ, Cohen's Kappa.

An example case is illustrated in Figure 1.

Table 4 represents the per-reader accuracy values relative to the 3 impression parameters, that is, ILD presence, HRCT pattern, and other diagnoses.

Concerning the parameters "ILD presence" and "HRCT pattern," there were no significant differences in accuracy when NR or SR was used in both groups (accuracy range 0.80-0.84 with $P=1.000$ for "ILD presence" in groups 1 and 2; accuracy range 0.56-0.72 with $P=.146$ and $P=.211$ for "HRCT pattern" in group 1 and group 2, respectively). However, readers were more accurate in assessing "other diagnosis" when using SR than NR, moving from 0.48 to 0.66 in group 1 ($P=.035$) and from 0.44 to 0.80 in group 2 ($P<.001$). Figure 2 illustrates an example case.

The reporting time was significantly lower ($P=.001$) when using SR (743 seconds (95% CI: 655-934 seconds)) than NR (1200 seconds (95% CI: 1080-1200 seconds)). Concerning the reporting times of both the RRs using SR (i.e., RR_2, median time 490 seconds [430-661] and RR_4, median time 1217 seconds [944-1760]), we demonstrated a decreasing, monotonic ($P=.010$ [Mann–Kendall's tau = -0.23611, *P*-value = .01; sample estimates: AR_order [1] 0; AR_coefficients numeric (0)] and $P<.001$ [Mann–Kendall's tau = −0.36653, *P*-value < 2.2e-16; sample estimates: AR_order [1] 0; AR_coefficients numeric(0)] for RR_2 and RR_4, respectively), and linear ($P=.02$ [Student t value = −2.4487, *P*-value = .02 sample estimates: AR order [1] 0; AR coefficients numeric(0)] and $P<.001$ [Student t value = −4.6468, *P*-value < 2.2e-16; sample estimates: AR order [1] 0; AR coefficients numeric(0)] for RR_2 and RR_4, respectively) trend when moving from the first to the last reading (Supplementary Figure 1).

## Discussion

We observed that, compared to NR, the use of SR increased the agreement between RRs and a standard of reference radiologist in assessing most fibrosis findings by HRCT. This did not translate into greater accuracy in providing the final impression of ILD presence and pattern, although it helped make an alternative diagnosis. These results are in line with previous works, in which the introduction of an SR template did not improve the RRs' accuracy in evaluating cervical spine CT in the emergency department[13] and brain

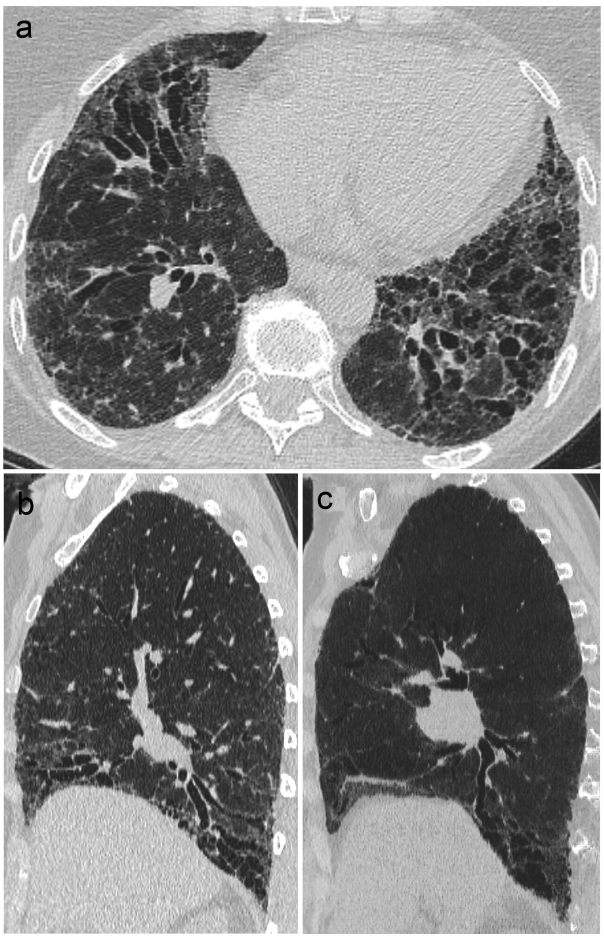



**Figure 1.** Case example 1. Fibrotic HRCT pattern in a 68-year-old woman with systemic sclerosis and worsening of dyspnea. Axial **(a)** and sagittal (**b** and **c**) chest HRCT images show reticulation, traction bronchiectasis, and mild ground-glass opacities, with bilateral, subpleural, and predominantly basal distribution. Fissure's distortion is shown in **(b)**. Only the RRs using SR reported volume loss signs in the description part of the report. HRCT, high-resolution computed tomography; RR, radiology resident; SR, structured report.

magnetic resonance imaging in suspected stroke.[18] We believe that a reasonable explanation for our findings is that inexperienced readers find it more difficult to connect different radiological signs within the unifying frame of a pattern or even a general diagnosis such as CTD–ILD, regardless of the consistency with which they can identify those radiological signs. Our data suggest that while SR cannot be expected to increase the diagnostic accuracy in terms of final impression, it is a tool to improve the consistency and completeness of what is reported as individual radiological signs (e.g., fibrosis or other diagnosis findings). Stated this improvement, the SR could be beneficial when reporting follow-up HRCT examination or evaluating the potential predictive role of a particular imaging finding.

**Table 4.** Accuracy values relative to the 3 impression parameters considered

| | Group 1—*novice RRs* | | | Group 2—*more experienced RRs* | | |
|---|---|---|---|---|---|---|
| | NR RR_1 | SR RR_2 | $P^a$ | NR RR_3 | SR RR_4 | $P^a$ |
| ILD presence | 0.82 | 0.80 | 1.000 | 0.82 | 0.84 | 1.000 |
| HRCT pattern | 0.56 | 0.70 | .146 | 0.60 | 0.72 | .211 |
| Other diagnosis | 0.48 | 0.66 | *.035* | 0.44 | 0.80 | *<.0001* |

[a]McNemar test.
Italics values are statistically significant ($P < 0.05$). There is no need to write them in italics.
RRs, radiology residents; NR, narrative report; SR, structured report; RR_1-4, radiology resident_1-4; ILD, interstitial lung disease; HRCT, high-resolution computed tomography.

Concerning the "description" domain, the SR increased the RRs–standard of reference agreement for honeycombing, traction bronchiectasis, and reticulation. Such an improvement was limited in the case of honeycombing (from fair to moderate in both groups), coherently with the rarity of this finding in our series (6% of HRCT examinations) and the limited reliability even in expert chest radiologists.[27] On the contrary, SR set the agreement of traction bronchiectasis and reticulation at a higher and substantial level than NR, regardless of RRs' expertise. Our results suggest the potential impact of SR in reporting fibrosis findings consistently as the first step toward patterns identification and, in turn, triggering an appropriate treatment option.[28-30] Although we did not measure a direct effect of SR on pattern recognition, this is reasonably a long-term effect of increasing experience, as further longitudinal studies might evaluate.

Of note, agreement improved to a similar extent when using SR compared to NR, regardless of the group of RRs' expertise. In particular, more experienced RRs (600-700 CT examinations) performed similarly to novice RRs (200-300 CT examinations). The only exception was volume loss, which increased more substantially in the group of greater expertise. We have no definite explanation for this finding, representing an outlier and then being interpreted with caution. Overall, our results indirectly suggest that experience in reporting HRCT in CTD patients forms over a prolonged time spent reading a significant number of CT examinations. Different from other diagnostics fields, such as prostatic imaging,[31] chest HRCT imaging has no definite quantitative criteria for qualifying a reader as an expert. However, our results suggest that SR might contribute to increase experience in this field, reasonably acting with additional factors such as participating in multidisciplinary team meetings, specific educational programs, and the number of reported HRCT rather than body CT examinations. Limited to the topic we investigated, we believe that SR should be used early and systematically during residency, in line with previous studies on different clinical scenarios, suggestive that it is of value in increasing the reporting of key findings of complex disease processes.[14,32] However, further longitudinal studies should investigate the cut-off of SRs to be read for being qualified as an expert.

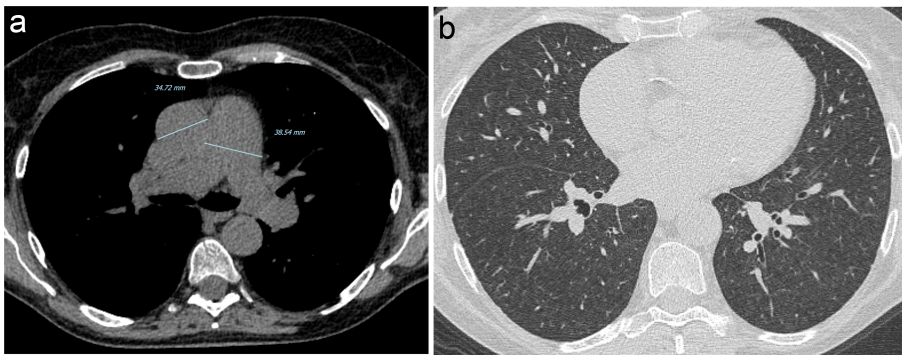



**Figure 2.** Case example 2. HRCT findings in a 67-year-old woman with known systemic sclerosis and pulmonary hypertension. CT image on the axial plane with soft tissue windowing in **(a)** shows pulmonary artery and ascending aorta dilation (39 and 35 mm, respectively, with a pulmonary artery to ascending aorta caliber ratio >1). Axial HRCT image with lung windowing in **(b)** shows multiple faint centrilobular ground-glass nodules, which are considered severe pulmonary hypertension manifestations. Signs of pulmonary hypertension were reported as "other diagnosis" in the impression part of the report only by the RRs using SR. CT, computed tomography.

SR has been advocated as a means to increase institutional report uniformity.[9] Our findings confirm this assumption, showing that for 3 of 4 fibrosis findings (i.e., honeycombing, traction bronchiectasis, and volume loss), the agreement between the 2 RRs using SR was higher than NR, ranging from moderate to almost perfect.

The "impression" part of a radiology report has been defined as the reader's synthesis of findings into a clear clinical assessment,[9] thus framing the subjective interpretation of objective findings in the radiological medical act's final step.[33] For evaluating the RRs' performance on "impression," we measured the per-reader accuracy values compared to our reference standard concerning 3 parameters, namely "ILD presence," "HRCT pattern," and "other diagnosis." Similar to previous works,[18,26] we found no significant differences in accuracy for "ILD presence" and "HRCT pattern" for RRs using NR or SR. A possible explanation is that accuracy may be most influenced not by the report's template style but by the individual level of training, experience, and attitude on a particular topic. The specific task of delineating an HRCT pattern of ILD proved challenging even for thoracic radiologists, with inter-reader agreement widely ranging from fair to substantial.[34] Another factor potentially influencing the report's diagnostic accuracy is the knowledge of patients' clinical information[35] which we do not investigate since RRs were not aware of patients' clinical data at the time of HRCT reporting. Of note, when testing the parameter "other diagnosis," we found a significantly greater accuracy for RRs reporting with SR. This result is in line with previous studies reporting that, despite not improving overall accuracy, SR was beneficial not to overlook collateral findings and alternative diagnoses,[13] thus improving the reporting completeness.

RRs' reporting time using SR was significantly shorter than NR (about 12 minutes vs. 20 minutes). Interestingly, we found a monotonic, linear decreasing trend in the reporting time, suggesting that readers needed some initial "learning period" to achieve faster reporting with SR. Knowing the increasingly high workload in radiology departments, we believe that a shorter reporting time may represent a clinical advantage favoring the implementation of SR for everyday RRs' clinical practice for chest HRCT reporting, similar to other subspeciality radiological activities.[15]

This study has some limitations, besides the monocentric and retrospective design. First, one may argue that, in such a complex clinical scenario, splitting RR's expertise into 2 ranks based on the total number of body CT examinations may not be sufficient for defining the actual level of skill in HRCT imaging of each RR, thus potentially limiting the generalizability of our results. Nevertheless, the number range of body CT examinations we used to define novice and more experienced RRs (200-300 and 600-700, respectively) reflects a corresponding comparable number of rotations in modality-driven sections, clinical radiology meetings, and obligatory formal teaching sessions followed during the residency training program in our institution, thus making the 2 groups representative of different level of expertise. Second, the low number of readers for each expertise group may have highlighted differences due to individual factors (e.g., specific interest in the topic, general knowledge, and personal attitude to use SR templates) than the reporting strategy itself. Nevertheless, we designed the study to allow pairwise comparisons but with the prime aim to consider the results as a whole rather than for single groups of expertise. Third, we could not test intra-reader agreement as a second reading might have been confounded by the experience made during the first reading, given that the readers were residents. We believe that intrareader agreement related to our SR model should be better tested in future studies with experienced readers. Fourth, one may argue that "interlobular septal thickening" and "architectural distortion" should have been considered individual items. Since we did not modify the SR template of reference 19, the RRs reported them as part of the "other pulmonary findings" component. Finally, additional prone imaging was available only for a minority of patients (6/50), thus potentially limiting the RRs, particularly the novice ones, in discriminating subtle interstitial abnormalities from dependent, subpleural atelectasis. Nevertheless, despite the low number of additional prone scans, the inter-reader agreement between each RR and the standard of reference for the finding "reticulation" was substantial, suggesting a reasonably limited impact on our results.

In conclusion, when using a dedicated SR model for chest HRCT in patients with suspected CTD–ILD, RRs demonstrated a higher agreement with the standard or reference radiologist in assessing most fibrosis findings. Although the use of SR did not increase the RRs' accuracy in providing the final impression of ILD presence and pattern, it helped make an alternative diagnosis. These results suggest that SR is valuable in increasing the reporting of critical HRCT findings in the complex CTD-ILD scenario and should be used early and systematically during the residency.

## Acknowledgements

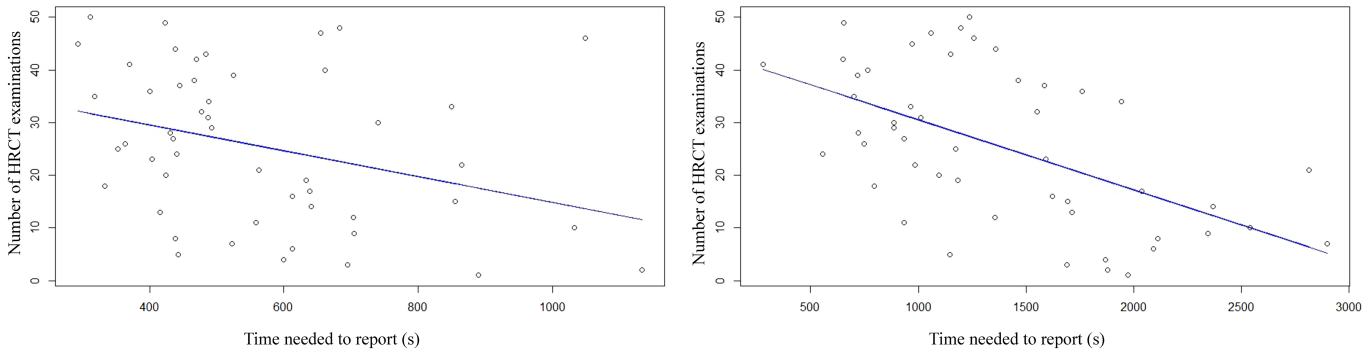
The authors gratefully acknowledge Dr. Annarita Tullio for having provided the requested additional statistical analysis.

## Conflict of interest disclosure

The authors declared no conflicts of interest.

## References


1. Travis WD, Costabel U, Hansell DM, et al. An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med*. 2013;188(6):733-748. [CrossRef]
2. Jeganathan N, Sathananthan M. Connective tissue disease-related interstitial lung disease: prevalence, patterns, predictors, prognosis, and treatment. *Lung*. 2020;198(5):735-759. [CrossRef]
3. Mira-Avendano I, Abril A, Burger CD, et al. Interstitial lung disease and other pulmonary manifestations in connective tissue diseases. *Mayo Clin Proc*. 2019;94(2):309-325. [CrossRef]
4. Nishimura K, Izumi T, Kitaichi M, Nagai S, Itoh H. The diagnostic accuracy of high-resolution computed tomography in diffuse infiltrative lung diseases. *Chest*. 1993;104(4):1149-1155. [CrossRef]
5. Walsh SL, Sverzellati N, Devaraj A, Keir GJ, Wells AU, Hansell DM. Connective tissue disease related fibrotic lung disease: high resolution computed tomographic and pulmonary function indices as prognostic determinants. *Thorax*. 2014;69(3):216-222. [CrossRef]
6. Ahuja J, Arora D, Kanne JP, Henry TS, Godwin JD. Imaging of pulmonary manifestations of connective tissue diseases. *Radiol Clin North Am*. 2016;54(6):1015-1031. [CrossRef]
7. Nobel JM, Kok EM, Robben SGF. Redefining the structure of structured reporting in radiology. *Insights Imaging*. 2020;11(1):10. [CrossRef]
8. European Society of Radiology (ESR). ESR paper on structured reporting in radiology. *Insights Imaging*. 2018;9(1):1-7. [CrossRef]
9. Burns J, Catanzano TM, Schaefer PW, et al. Structured reports and radiology residents: friends or foes? *Acad Radiol*. 2020;S1076-6332(20):30601-30602. [CrossRef]
10. European training curriculum for radiology. Framework for training in Radiology in Europe. 2018. Available at: https://www.uems.eu/__data/assets/pdf_file/0018/72270/UEMS-2018.37-European-Training-Requirements-in-Radiology.pdf. Accessed May 4, 2021.
11. The Accreditation Council for Graduate Medical Education (ACGME). Diagnostic radiology milestones. 2019. Available at: https://www.acgme.org/Portals/0/PDFs/Milestones/DiagnosticRadiologyMilestones.pdf?ver=2020-09-01-145824-837. Accessed May 4, 2021.
12. Marcovici PA, Taylor GA. Journal Club. Journal Club: Structured radiology reports are more complete and more effective than unstructured reports. *AJR Am J Roentgenol*. 2014;203(6):1265-1271. [CrossRef]
13. Lin E, Powell DK, Kagetsu NJ. Efficacy of a checklist-style structured radiology reporting template in reducing resident misses on cervical spine computed tomography examinations. *J Digit Imaging*. 2014;27(5):588-593. [CrossRef]
14. Wildman-Tobriner B, Allen BC, Bashir MR, et al. Structured reporting of CT enterography for inflammatory bowel disease: effect on key feature reporting, accuracy across training levels, and subjective assessment of disease by referring physicians. *Abdom Radiol (NY)*. 2017;42(9):2243-2250. [CrossRef]
15. Ernst BP, Strieth S, Katzer F, et al. The use of structured reporting of head and neck ultrasound ensures time-efficiency and report quality during residency. *Eur Arch Otorhinolaryngol*. 2020;277(1):269-276. [CrossRef]
16. Johnson TF, Brinjikji W, Doolittle DA, Nagelschneider AA, Welch BT, Kotsenas AL. Structured head and neck CT angiography reporting reduces resident revision rates. *Curr Probl Diagn Radiol*. 2019;48(2):114-116. [CrossRef]
17. Powell DK, Lin E, Silberzweig JE, Kagetsu NJ. Introducing radiology report checklists among residents: adherence rates when suggesting versus requiring their use and early experience in improving accuracy. *Acad Radiol*. 2014;21(3):415-423. [CrossRef]
18. Johnson AJ, Chen MY, Swan JS, Applegate KE, Littenberg B. Cohort study of structured reporting compared with conventional dictation. *Radiology*. 2009;253(1):74-80. [CrossRef]
19. Cereser L, Marchesini F, Di Poi E, et al. Structured report for chest high-resolution computed tomography in patients with connective tissue disease: impact on the report quality as perceived by referring clinicians. *Eur J Radiol*. 2020;131:109269. [CrossRef]
20. Raghu G, Collard HR, Egan JJ, et al. An Official ATS/ERS/JRS/ALAT Statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med*. 2011;183(6):788-824. [CrossRef]
21. Cottin V, Crestani B, Valeyre D, et al. Diagnosis and management of idiopathic pulmonary fibrosis: French practical guidelines. *Eur Respir Rev*. 2014;23(132):193-214. [CrossRef]
22. Jacob J, Hansell DM. HRCT of fibrosing lung disease. *Respirology*. 2015;20(6):859-872. [CrossRef]
23. Sverzellati N, Odone A, Silva M, et al. Structured reporting for fibrosing lung disease: a model shared by radiologist and pulmonologist. *Radiol Med*. 2018;123(4):245-253. [CrossRef]
24. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology*. 2008;246(3):697-722. [CrossRef]
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174. [CrossRef]
26. Noguchi K, Gel YR, Duguay CR. Bootstrap-based tests for trends in hydrological time series, with application to ice phenology data. *J Hydrol*. 2011;410(3-4):150-161. [CrossRef]
27. Watadani T, Sakai F, Johkoh T, et al. Interobserver variability in the CT assessment of honeycombing in the lungs. *Radiology*. 2013;266(3):936-944. [CrossRef]
28. Sambataro G, Ferro F, Orlandi M, et al. Clinical, morphological features and prognostic factors associated with interstitial lung disease in primary Sjögren's syndrome: a systematic review from the Italian Society of Rheumatology, Autoimmun. *Autoimmun Rev*. 2020;19(2):102447. [CrossRef]
29. Flaherty KR, Wells AU, Cottin V, et al. Nintedanib in progressive fibrosing interstitial lung diseases. *N Engl J Med*. 2019;381(18):1718-1727. [CrossRef]
30. Distler O, Highland KB, Gahlemann M, et al. Nintedanib for systemic sclerosis–associated interstitial lung disease. *N Engl J Med*. 2019;380(26):2518-2528. [CrossRef]
31. de Rooij M, Israël B, Tummers M, et al. ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training. *Eur Radiol*. 2020;30(10):5404-5416. [CrossRef]
32. Brook OR, Brook A, Vollmer CM, Kent TS, Sanchez N, Pedrosa I. Structured reporting of multiphasic CT for pancreatic cancer: potential effect on staging and surgical planning. *Radiology*. 2015;274(2):464-472. [CrossRef]
33. Tamburrini O, Dalla Palma F. The Radiological Medical Act approved by the SIRM Executive Committee on July 2, 2007. *Radiol Med*. 2008;113(3):319-328. [CrossRef]
34. Kwon BS, Choe J, Do KH, Hwang HS, Chae EJ, Song JW. Computed tomography patterns predict clinical course of idiopathic pulmonary fibrosis. *Respir Res*. 2020;21(1):295. [CrossRef]
35. Castillo C, Steffens T, Sim L, Caffery L. The effect of clinical information on radiology reporting: a systematic review. *J Med Radiat Sci*. 2021;68(1):60-74. [CrossRef]




**Supplementary Figure 1.** Relationship between the number of HRCT examinations progressively read by residents (RR_2 in the right plot; RR_4 in the left plot) in the y-axis and the time needed to report them on the x-axis. The plots show a significantly decreasing, monotonic, and linear trend, suggesting that the reporting time decreased as the readers familiarized themselves with the SR template.