



Development and validation of a radiological diagnosis model for hypersensitivity pneumonitis

Margaret L. Salisbury¹, Barry H. Gross², Aamer Chughtai², Mohamed Sayyouh², Ella A. Kazerooni², Brian J. Bartholmai³, Meng Xia⁴, Susan Murray⁴, Jeffrey L. Myers⁵, Amir Lagstein⁵, Kristine E. Konopka⁵, Elizabeth A. Belloli¹, Jamie S. Sheth¹, Eric S. White¹, Colin Holtze¹, Fernando J. Martinez⁶ and Kevin R. Flaherty¹

Affiliations: ¹Division of Pulmonary and Critical Care, University of Michigan, Ann Arbor, MI, USA. ²Dept of Radiology, University of Michigan, Ann Arbor, MI, USA. ³Dept of Radiology, Mayo Clinic, Rochester, MN, USA. ⁴Dept of Biostatistics, University of Michigan, Ann Arbor, MI, USA. ⁵Dept of Pathology, University of Michigan, Ann Arbor, MI, USA. ⁶Division of Pulmonary and Critical Medicine, Cornell Medical College, New York, NY, USA.

Correspondence: Margaret L. Salisbury, Division of Pulmonary and Critical Care, University of Michigan, 3916 Taubman Center, 1500 East Medical Center Drive, Ann Arbor, MI 48109, USA. E-mail: msalisbu@med.umich.edu

 @ERSpublications

When HRCT shows more mosaic attenuation than reticulation and diffuse axial ILD, false hypersensitivity pneumonitis diagnosis risk is <10% <http://ow.ly/tthG30k3Vj2>

Cite this article as: Salisbury ML, Gross BH, Chughtai A, *et al.* Development and validation of a radiological diagnosis model for hypersensitivity pneumonitis. *Eur Respir J* 2018; 52: 1800443 [<https://doi.org/10.1183/13993003.00443-2018>].

ABSTRACT High-resolution computed tomography (HRCT) may be useful for diagnosing hypersensitivity pneumonitis. Here, we develop and validate a radiological diagnosis model and model-based points score.

Patients with interstitial lung disease seen at the University of Michigan Health System (derivation cohort) or enrolling in the Lung Tissue Research Consortium (validation cohort) were included. A thin-section, inspiratory HRCT scan was required. Thoracic radiologists documented radiological features.

The derivation cohort comprised 356 subjects (33.9% hypersensitivity pneumonitis) and the validation cohort comprised 424 subjects (15.5% hypersensitivity pneumonitis). An age-, sex- and smoking status-adjusted logistic regression model identified extent of mosaic attenuation or air trapping greater than that of reticulation (“MA-AT>Reticulation”; OR 6.20, 95% CI 3.53–10.90; $p < 0.0001$) and diffuse axial disease distribution (OR 2.33, 95% CI 1.31–4.16; $p = 0.004$) as hypersensitivity pneumonitis predictors (area under the receiver operating characteristic curve 0.814). A model-based score > 2 (1 point for axial distribution, 2 points for “MA-AT>Reticulation”) has specificity 90% and positive predictive value (PPV) 74% in the derivation cohort and specificity 96% and PPV 44% in the validation cohort. Similar model performance is seen with population restriction to those reporting no exposure (score > 2 : specificity 91%).

When radiological mosaic attenuation or air trapping are more extensive than reticulation and disease has diffuse axial distribution, hypersensitivity pneumonitis specificity is high and false diagnosis risk low ($< 10\%$), but PPV is diminished in a low-prevalence setting.

This article has supplementary material available from erj.ersjournals.com

Received: March 02 2018 | Accepted after revision: May 10 2018

Copyright ©ERS 2018

Introduction

Hypersensitivity pneumonitis is an interstitial lung disease (ILD) caused by exposure to a variety of antigens [1]. High-resolution computed tomography (HRCT) is commonly used in the diagnostic evaluation for ILDs. A specific HRCT pattern can be diagnostic of histopathological usual interstitial pneumonia (UIP), pathognomonic of idiopathic pulmonary fibrosis (IPF) in the correct clinical context [2]. Individuals without a specific ILD diagnosis after clinical and HRCT evaluation are often subjected to additional invasive diagnostic testing (e.g. bronchoscopy or surgical lung biopsy) to obtain a diagnosis.

Several studies have evaluated HRCT patterns associated with hypersensitivity pneumonitis. Two studies found that a radiologist's confident diagnosis of hypersensitivity pneumonitis is correct 88–92% of the time [3, 4]. JOHANSSON *et al.* [5] created a clinical prediction model combining patient age, environmental exposure history and radiological features of diffusely distributed ground glass or mosaic attenuation, and found that a high model-based score was associated with a high specificity for hypersensitivity pneumonitis. Unfortunately, these studies included limited alternative ILD diagnoses [3, 4] or had the potential for confirmation bias by modelling exposure history along with radiological features [5]. We sought to derive and externally validate a “rule-in” diagnostic model for hypersensitivity pneumonitis based solely on radiological findings.

Methods

Patient selection and clinical diagnosis assignment

The derivation cohort consisted of a retrospectively assembled cohort of consecutive adult patients undergoing diagnostic case review at the University of Michigan Health System (UMHS)'s multidisciplinary ILD conference between February 1, 2009 and August 31, 2014. Additional hypersensitivity pneumonitis cases were identified by “hypersensitivity pneumonitis” International Classification of Diseases (Ninth Revision) code (495.7, 495.8 and 495.9) electronic medical record search between January 1, 2004 and December 31, 2013. Clinical ILD diagnoses (the outcome variable) were verified after a detailed chart review. Supporting evidence sufficient for hypersensitivity pneumonitis diagnosis verification included classic findings on surgical lung biopsy, or at least two of: 1) bronchoalveolar lymphocytosis >20% [6, 7], 2) consistent findings on transbronchial or surgical biopsy (any of loose non-necrotising granulomas, giant cells, mononuclear inflammatory interstitial or peribronchiolar infiltrate), or 3) a plausible exposure history [1]. Patients with suspected drug-induced pulmonary hypersensitivity reaction were excluded. Non-hypersensitivity pneumonitis ILD diagnoses were verified in the setting of consistent histopathological evaluation and according to current guidelines [2, 8]. Documentation of ILD conference discussion [9] was available in 69 (57%) out of 122 identified hypersensitivity pneumonitis patients and 226 (94%) out of 240 identified subjects with non-hypersensitivity pneumonitis ILDs. Radiology (HRCT) reports were available to the multidisciplinary team and treating clinicians, but were not considered as part of the clinical diagnosis verification process for this study. All subjects had a baseline HRCT and unclassifiable ILDs were excluded. Figure 1 shows a flowchart with numbers of screened patients from each dataset, reasons for exclusion and final included clinical diagnoses for the UMHS derivation cohort. Demographics, clinical characteristics and pulmonary function measurements were collected from the electronic medical record.

The validation cohort consisted of patients with ILD undergoing clinically indicated lung tissue sampling procedures as part of the multicentre, National Institutes of Health-supported Lung Tissue Research Consortium (LTRC) study (www.ltrcpublic.com). Availability of a histopathological specimen obtained *via* diagnostic surgical lung biopsy and complete visual HRCT features scores documentation were required for inclusion. Baseline demographic and pulmonary function data were available for all patients. Methods of patient collection and assignment of clinical ILD diagnoses for the LTRC study are described in detail elsewhere [10]. Figure 2 shows a flowchart with numbers of included and excluded patients, and final included clinical diagnoses, for the LTRC validation cohort.

The Institutional Review Board at UMHS approved this study (HUM 00093978). The LTRC study was approved by participating centres and approved our use of de-identified data for this analysis. Some of these results have been presented as an abstract [11].

Derivation cohort radiological protocols

Images were obtained on a variety of CT scanners at UMHS and at referring centres, and viewed on a picture archiving and communication system. All HRCTs had a non-contrasted image series with thin (≤ 2 mm collimation) sections obtained at end-inspiration. Prone and expiratory images and a volumetric or image-sharpening protocol were available in most subjects, but not required for inclusion. HRCT scan quality was documented as excellent, diagnostic or non-diagnostic. Slice thickness, interslice interval (when appropriate), availability of a volumetric protocol, and expiratory and prone series were

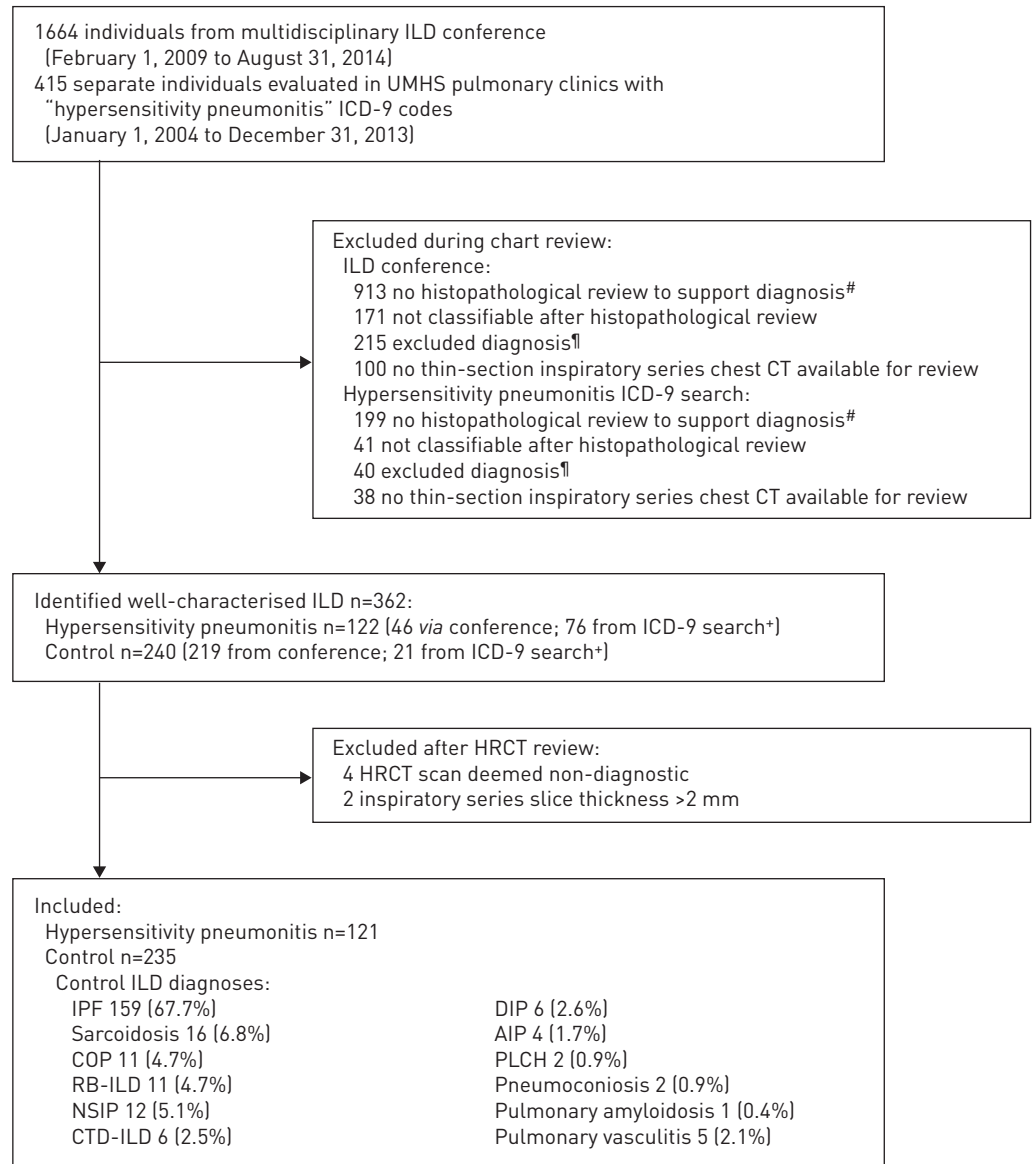


FIGURE 1 Patient flowchart: derivation cohort (University of Michigan Health System (UMHS)). ILD: interstitial lung disease; ICD-9: International Classification of Diseases (Ninth Revision); CT: computed tomography; HRCT: high-resolution CT; IPF: idiopathic pulmonary fibrosis; COP: cryptogenic organising pneumonia; RB-ILD: respiratory bronchiolitis-ILD; NSIP: non-specific interstitial pneumonitis; CTD-ILD: connective tissue disease-associated ILD; DIP: desquamative interstitial pneumonia; AIP: acute interstitial pneumonitis; PLCH: pulmonary Langerhans cell histiocytosis. #: includes those with ILD diagnosis confirmed by HRCT only (i.e. “definite usual interstitial pneumonia” HRCT pattern) and those unwilling or unable to undergo diagnostic biopsy; ¶: excluded diagnoses include primary cystic lung disease, pulmonary alveolar proteinosis, airway or pulmonary vascular disease without ILD, drug-induced lung disease, infection and other non-ILD diagnoses (congestive heart failure, etc.); *: of included patients identified from the ICD-9 search alone, 23 hypersensitivity pneumonitis and seven control patients had chart documentation of multidisciplinary case review that occurred prior to creation of a searchable database of conference minutes.

documented. Subjects with a non-diagnostic HRCT or inspiratory series slice thickness >2 mm were excluded (figure 1).

Three radiologists (B.H.G., A.C. and M.S.) blinded to clinical data independently interpreted each HRCT and documented findings on a standardised scoring sheet (supplementary figure E1). Using standard definitions [12], radiological features of honeycombing, reticular pattern, ground glass, mosaic attenuation and air trapping (when expiratory images were available) were scored semiquantitatively in each lobe (right upper lobe, right middle lobe, right lower lobe, left upper lobe/lingula and left lower lobe), with score 0 indicating no involvement, score 1 indicating <5% of lobe involved (present but minimal), score 2

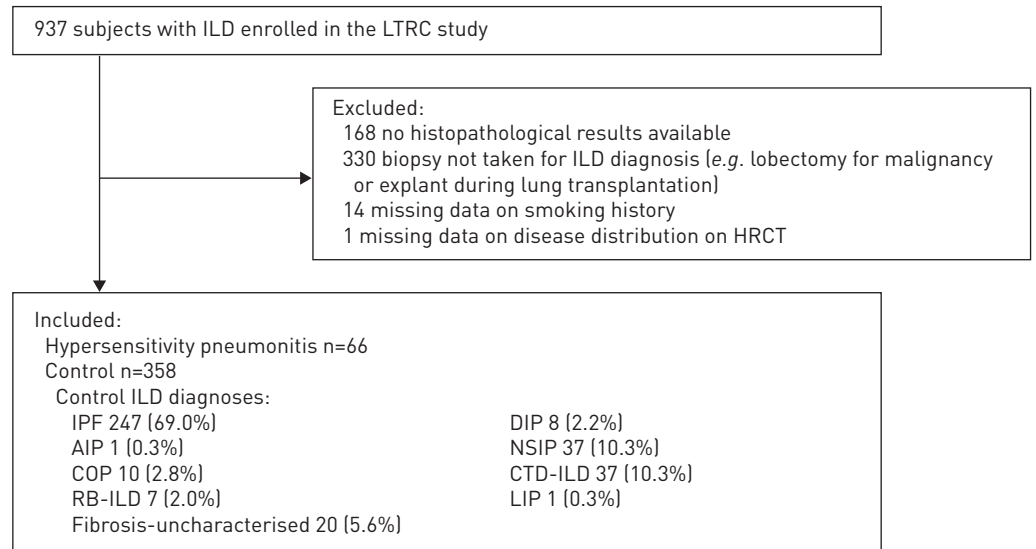


FIGURE 2 Patient flowchart: validation (Lung Tissue Research Consortium (LTRC)) cohort. ILD: interstitial lung disease; HRCT: high-resolution computed tomography; IPF: idiopathic pulmonary fibrosis; AIP: acute interstitial pneumonitis; COP: cryptogenic organising pneumonia; RB-ILD: respiratory bronchiolitis-ILD; DIP: desquamative interstitial pneumonitis; NSIP: non-specific interstitial pneumonitis; CTD-ILD: connective tissue disease-associated ILD; LIP: lymphoid interstitial pneumonia.

indicating 5–25% of lobe involved, score 3 indicating 25–49% of lobe involved, score 4 indicating 50–75% of lobe involved and score 5 indicating >75% of lobe involved. For each feature, the lobe scores were summed and divided by 5 to obtain an average on a scale of 0–5, representative of the proportion of total lung having the feature. Traction bronchiectasis and centrilobular nodules were recorded as present (score 1) or absent (score 0) in each of five lobes; lobar scores were summed to represent the number of lobes having the feature (scale 0–5). A dichotomous variable was created for each feature, considered “present” if the average or sum was >0.5. We selected this cut-point to avoid labelling a radiological feature as present when its extent was minimal. For example, a ground glass average of 0.4 means two lobes of the lung had ground glass present in ≤5% of the lobe ((1+1)/5=0.4) or one lobe had ground glass in 5–25% of the lobe (2/5=0.4). The predominant distribution of interstitial disease was assigned in the axial and craniocaudal dimensions. Axial distribution was noted as central (predominant parenchymal abnormality preferentially involves the central one-third of the lung), peripheral (abnormality preferentially involves the peripheral one-third of the lung), subpleural (abnormality preferentially involves the immediate subpleural region), peribronchovascular (abnormality preferentially involves the region adjacent to the peribronchovascular bundles) or diffuse (abnormality is widely distributed with no section involved more than any other). Multiple selections were allowed for axial distribution, but no reader made more than two selections. To be consistent with the methodology of the LTRC radiologist (see the following “Validation cohort radiological protocols” section), a selection of central, peripheral or diffuse disease was considered the primary distribution when multiple selections were made. When subpleural was the only selection, it was grouped with peripheral. The craniocaudal distribution of disease was noted as upper lung (predominant parenchymal abnormality is most prominent above the main carina), lower lung (abnormality is most prominent below the main carina) or diffuse (upper and lower lungs are relatively equally involved); only one selection was allowed. Three-reader consensus dichotomous and semiquantitative HRCT scores were created for each subject. Consensus semiquantitative scores were a three-reader average of the average/summed feature score. Dichotomised consensus variables were “present” when two of the three readers’ scores met “present” criteria. Several additional dichotomised variables were created: the “MA-AT” variable was “present” when a subject had consensus presence of mosaic attenuation or air trapping, the “GG>Reticulation” variable was “present” when the semiquantitative ground glass score was higher than the semiquantitative reticulation score and the “MA-AT>Reticulation” variable was “present” when the semiquantitative score for mosaic attenuation or air trapping was higher than the semiquantitative reticulation score. Consensus for the craniocaudal and axial disease distribution was the probability of a radiologist assigning the distribution category for each patient. A consensus dichotomous variable for a given category (*i.e.* upper, lower or diffuse) of the disease distribution was assigned as “present” when its probability was two-thirds or greater.

Validation cohort radiological protocols

The LTRC enrolment HRCT was assessed by one expert thoracic radiologist in the LTRC Radiology Core Laboratory (Mayo Clinic, Rochester, MN, USA). The semiquantitative scoring procedure and a list of radiological features analysed are described in detail elsewhere [10]. A dichotomous variable and a semiquantitative average (representative of the proportion of total lung having the feature) were created for the HRCT features of honeycombing, reticulation, ground glass, mosaic attenuation, air trapping, traction bronchiectasis and centrilobular nodules. The scale for semiquantitative average scores is 0–4 and represents quartiles of lung involvement with the feature (e.g. a score of 1 indicates 1–25% of lung and a score of 4 indicates 76–100% of lung). A feature was dichotomously “present” when the semiquantitative average was >0.5. Axial (none, diffuse/even, central and peripheral/subpleural) and craniocaudal (upper, lower and even/diffuse) disease distributions were also available.

Statistical methods

Analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC, USA) and R version 3.2.1 (www.r-project.org). Baseline patient and HRCT quality characteristics are shown as mean and standard deviation or number (percentage), as appropriate. Highly skewed variables are shown as median (range). Pairwise and three-way agreement of three radiologists on the dichotomised HRCT features are *via* κ and Light’s κ , respectively [13]. The bootstrap method was used to calculate the 95% confidence interval for Light’s κ . Pairwise agreement on semiquantitative features is *via* weighted κ , with the averaged or summed feature score (scale 0–5) rounded to the nearest integer. κ measures indicate that agreement is poor when $\kappa < 0.40$, intermediate when $0.40 < \kappa < 0.60$, good when $0.60 < \kappa < 0.75$ and excellent when $\kappa \geq 0.75$ [14]. To identify the nature of systematic differences in interpretation of continuous radiological features across the three readers, we also evaluated the mean and standard deviation of each feature for each pair of radiologists using paired t-tests. For dichotomous features, the number (percentage) of patients having the feature “present” by radiologists is shown, with significance of pairwise between-reader differences *via* McNemar’s test [15].

Logistic regression models identified clinical and radiological variables associated with hypersensitivity pneumonitis. The area under the receiver operating characteristic curve (AUC) gives model discrimination [16]. An “HP-HRCT Diagnosis Score” was created from the final multivariable model. To validate the model, we obtained the predicted probability of hypersensitivity pneumonitis from both the final regression formula developed in the UMHS derivation cohort and from the HP-HRCT Diagnosis Score for each subject in the LTRC dataset. HP-HRCT Diagnosis Score test characteristics are given as sensitivity and specificity at various points thresholds. To measure model calibration, LTRC subjects were categorised into discretised groups of predicted model-based probabilities of hypersensitivity pneumonitis diagnosis (using the UMHS final full model and adjusted HP-HRCT Diagnosis Score models) and then compared to observed frequencies of hypersensitivity pneumonitis diagnosis within each probability group. The resulting observed *versus* expected probabilities are displayed graphically with a superimposed line of perfect calibration shown for reference [17]. We also performed several subpopulation sensitivity analyses within the UMHS cohort: 1) with hypersensitivity pneumonitis subjects restricted to include only those with “classic” lung biopsy findings, 2) with hypersensitivity pneumonitis subjects restricted to “classic” lung biopsy findings and non-hypersensitivity pneumonitis subjects restricted to idiopathic interstitial pneumonias only (acute interstitial pneumonitis, cryptogenic organising pneumonia, desquamative interstitial pneumonia, IPF, non-specific interstitial pneumonitis (NSIP) and respiratory bronchiolitis-ILD), and 3) with all subjects restricted to those with no documented exposure history.

Results

Patient characteristics

In the UMHS cohort, the hypersensitivity pneumonitis group was slightly younger, had a higher proportion of female patients, a higher proportion of never-smokers and a higher proportion were treated with corticosteroids prior to the consultation at UMHS (table 1). In the LTRC cohort, the hypersensitivity pneumonitis group was slightly younger, had a higher proportion of never-smokers and had a higher diffusing capacity of the lung for carbon monoxide than the not hypersensitivity pneumonitis group (supplementary table E1).

HRCT characteristics

Among all included subjects, 60.4% had excellent scan quality, 72.8% had a volumetric protocol, 82.3% had expiratory series images and 87.4% had prone series images (supplementary table E2). The maximum inspiratory series slice thickness included was 1.4 mm.

Inter-rater agreement across three radiologists was good for dichotomised honeycombing, intermediate for mosaic attenuation, air trapping, traction bronchiectasis, and axial and craniocaudal distribution, and poor

TABLE 1 Patient characteristics at baseline: University of Michigan Health System (UMHS) (derivation) cohort

	All subjects	Hypersensitivity pneumonitis	Not hypersensitivity pneumonitis
Subjects	356	121	235
Age years	59.6±11.1	58.1±10.9	60.3±11.1
Male	175 (49.2)	39 (32.2)	136 (57.9)
White	308 (86.5)	108 (89.3)	200 (85.1)
Smoking status			
Current smoker	30 (8.4)	4 (3.3)	26 (11.1)
Ex-smoker	167 (46.9)	54 (44.6)	113 (48.1)
Never-smoker	157 (44.7)	63 (52.1)	96 (40.9)
Pack-years smoked (n=351)	14.6±19.9	11.0±17.6	16.4±20.8
Baseline physiology			
FVC % pred (n=345)	66.0±17.6	65.7±17.4	66.1±17.7
FEV ₁ % pred (n=345)	75.2±19.9	73.6±20.5	76.0±19.5
FEV ₁ /FVC ratio (n=345)	82.5±8.0	81.9±8.4	82.7±7.7
DL _{CO} % pred (n=248)	52.0±18.7	51.4±16.6	52.3±19.7
Time between HRCT and PFT days	8 (0–50)	7 (0–40)	11 (0–60)
Reported symptom duration months	25.3±33.2	27.5±43.9	24.2±26.0
Corticosteroid prior to UMHS evaluation	113 (32.0)	46 (38.0)	68 (28.9)
Diagnostic testing available			
Surgical lung biopsy performed	304 (85.4)	86 (71.1)	218 (92.8)
“Classic hypersensitivity pneumonitis” on surgical lung biopsy (among hypersensitivity pneumonitis)		64 (52.9)	
Transbronchial lung biopsy performed	111 (31.2)	63 (52.1)	48 (20.4)
BAL cell count and differential (of total with test)	91 (25.5)	54 (44.6)	37 (15.7)
Lymphocytes %	22.6±20.4	32.2±19.5	8.5±11.5
“Hypersensitivity pneumonitis panel” results (of total with test)	91 (25.6)	66 (54.5)	25 (10.6)
Negative (of those with test)	59 (64.8)	41 (62.1)	18 (72.0)
Bird (of those with test)	14 (15.4)	12 (18.2)	2 (8.0)
Microbe (of those with test)	11 (12.1)	7 (10.6)	4 (16.0)
Both (of those with test)	7 (7.7)	6 (9.1)	1 (4.0)
Reported exposures			
None reported	217 (61.0)	38 (31.4)	179 (76.2)
Asbestos	17 (4.8)	0 (0)	17 (7.2)
Beryllium	2 (0.6)	0 (0)	2 (0.9)
Bird	35 (9.8)	34 (28.1)	1 (0.4)
Chemical	5 (1.4)	1 (0.8)	4 (1.7)
Drug	1 (0.3)	1 (0.8)	0 (0)
Hot tub	10 (2.8)	9 (7.4)	1 (0.4)
Industrial dust	14 (3.9)	0 (0)	14 (6.0)
Microbe	35 (9.8)	25 (20.7)	10 (4.3)
Multiple	11 (3.1)	10 (8.3)	1 (0.4)
Wood dust	9 (2.5)	3 (2.5)	6 (2.6)

Data are presented as n, mean±SD, n (%) or median (interquartile range). FVC: forced vital capacity; FEV₁: forced expiratory volume in 1 s; DL_{CO}: diffusing capacity of the lung for carbon monoxide; HRCT: high-resolution computed tomography; PFT: pulmonary function testing; BAL: bronchoalveolar lavage.

for reticulation, ground glass and centrilobular nodules (table 2). There was systematic variation by reader, with reader 3 identifying more ground glass and less reticulation than reader 1 or 2, and reader 1 identifying less centrilobular nodules and traction bronchiectasis than reader 2 or 3 (supplementary table E3). Supplementary table E4 shows weighted κ estimates for each pair of readers. Table 3 and supplementary table E5 summarise the number and percentage of patients having dichotomised or categorical consensus features, and means and standard deviations for the three-reader average of semiquantitative scores, for the UMHS and LTRC cohorts, respectively.

Derivation of the radiological diagnostic model

In univariable analysis female sex, never-smoking history, fewer pack-years smoked, absence of dichotomised honeycombing, absent or less extensive reticulation, absent or less extensive traction

TABLE 2 Dichotomous[#] three-reader agreement: University of Michigan Health System cohort

	Two-way reader κ (95% CI)			Light's κ (95% CI)
	Reader 1 versus 2	Reader 1 versus 3	Reader 2 versus 3	
Honeycombing	0.61 [0.47–0.75]	0.59 [0.46–0.72]	0.63 [0.51–0.75]	0.61 [0.50–0.72]
Reticulation	0.53 [0.44–0.62]	0.26 [0.19–0.33]	0.14 [0.09–0.18]	0.31 [0.26–0.36]
Ground glass	0.49 [0.40–0.59]	0.18 [0.09–0.27]	0.20 [0.12–0.29]	0.29 [0.23–0.36]
Mosaic attenuation	0.59 [0.49–0.69]	0.66 [0.57–0.75]	0.53 [0.43–0.64]	0.59 [0.52–0.67]
Air trapping (n=293 with expiratory slice)	0.50 [0.41–0.59]	0.64 [0.56–0.73]	0.48 [0.38–0.58]	0.54 [0.47–0.61]
Centrilobular nodules	0.18 [0.09–0.27]	0.47 [0.30–0.64]	0.34 [0.23–0.45]	0.33 [0.24–0.42]
Traction bronchiectasis	0.50 [0.42–0.59]	0.59 [0.51–0.67]	0.66 [0.57–0.75]	0.59 [0.52–0.65]
Craniocaudal distribution	0.55 [0.48–0.63]	0.50 [0.43–0.57]	0.47 [0.40–0.54]	0.51 [0.46–0.56]
Axial distribution	0.45 [0.37–0.52]	0.47 [0.40–0.54]	0.34 [0.27–0.40]	0.42 [0.36–0.47]

[#]: a feature was considered “absent” when the average score (for honeycombing, reticulation, ground glass, mosaic attenuation or air trapping) or sum score (for centrilobular nodules or traction bronchiectasis) was <0.5, otherwise the feature was considered “present”.

bronchiectasis and diffuse axial or craniocaudal distributions of abnormality were associated with increased odds of hypersensitivity pneumonitis. Presence of more extensive ground glass, mosaic attenuation, air trapping, combined mosaic attenuation or air trapping (“MA-AT”), more extensive ground glass than reticulation (“GG>Reticulation”), more extensive mosaic attenuation or air trapping than reticulation (“MA-AT>Reticulation”) and centrilobular nodules were also associated with increased hypersensitivity pneumonitis odds (table 4). For simplicity, we limited candidate multivariable models to three categorical radiological predictors. Several candidate models, adjusted for age, sex and smoking status, were evaluated (supplementary table E6). All had good discrimination (AUC >0.790). As model 7 (including radiological MA-AT>Reticulation and diffuse axial distribution) had a high AUC, maximised sensitivity for hypersensitivity pneumonitis (sensitivity 65.3%) at the 90% specificity threshold and had good face value, it was selected. Next, an HP-HRCT Diagnosis Score was created based on model 7. In the adjusted model for the HP-HRCT Diagnosis Score (table 5 gives calculation parameters; score range 0–3), each 1-point score increase was associated with 2.45-fold increase in the odds of hypersensitivity pneumonitis (95% CI 1.99–3.02; p<0.0001; AUC 0.814). Figure 3 shows an example HRCT for each of two UMHS hypersensitivity pneumonitis patients with an HP-HRCT Diagnosis Score of 3.

Model validation and HP-HRCT Diagnosis Score test characteristics

As shown in table 6, when mosaic attenuation or air trapping are more extensive than reticulation and disease has a diffuse axial distribution (*i.e.* when 3 out of 3 points are assigned), there is a high specificity for hypersensitivity pneumonitis (UMHS specificity 90.2% and LTRC specificity 95.8%). Sensitivity for hypersensitivity pneumonitis when the HP-HRCT Diagnosis Score was 3 was 55.4% in the UMHS cohort and 18.2% in the LTRC cohort. Supplementary table E7 shows the final adjusted model in the LTRC cohort; the radiological variables maintained predictive ability for hypersensitivity pneumonitis in a similar strength and direction as the derivation model. Supplementary figure E2 shows complete and HP-HRCT Diagnosis Score-based receiver operating characteristic curves for the UMHS and LTRC cohorts; AUC was >0.70 in all models. The model was well calibrated, with most probability group points falling near the line of perfect calibration (supplementary figure E3). The results of sensitivity analyses (supplementary table E8) suggest that the HP-HRCT Diagnosis Score also identifies hypersensitivity pneumonitis when it is defined by “classic” surgical biopsy features among all comparators and among comparators with idiopathic interstitial pneumonias, and identifies hypersensitivity pneumonitis in the subgroup of subjects without a documented exposure history. For reference, supplementary table E9 gives summary demographic and HRCT characteristics for the subset of hypersensitivity pneumonitis patients (n=64) with a “classic” surgical lung biopsy.

Supplementary table E10 shows the regression formula allowing calculation of an adjusted model-based hypersensitivity pneumonitis probability and the positive predictive value (PPV) of various hypersensitivity pneumonitis probability thresholds in the UMHS and LTRC cohorts. Hypersensitivity pneumonitis prevalence is higher in the UMHS than the LTRC cohort and thus the same model-based hypersensitivity pneumonitis probability corresponds to a higher PPV in the UMHS cohort. When model-based probability is ≥80%, the PPV for hypersensitivity pneumonitis is 100% in the UMHS and LTRC cohorts; at a probability of ≥70%, UMHS PPV is 77% and LTRC PPV 46%. The PPV can be interpreted as the probability that a patient with a given test result (here, model-based hypersensitivity

TABLE 3 Consensus high-resolution computed tomography features: University of Michigan Health System cohort

	All subjects	Hypersensitivity pneumonitis	Not hypersensitivity pneumonitis
Subjects	356	121	235
Dichotomous scores[¶]			
Honeycombing	37 (10.4)	10 (8.3)	27 (11.5)
Reticulation	230 (64.6)	52 (43.0)	178 (75.7)
Ground glass	256 (74.4)	103 (85.1)	162 (68.9)
Mosaic attenuation	89 (25.0)	62 (51.2)	27 (11.5)
Air trapping (n=293 with expiratory slice)	127 (43.3)	75 (71.4)	52 (27.7)
"MA-AT"	145 (40.7)	85 (70.3)	60 (25.5)
"GG>Reticulation"	229 (64.3)	99 (81.8)	130 (55.3)
"MA-AT>Reticulation"	130 (36.5)	85 (70.3)	45 (19.2)
Centrilobular nodules	31 (8.7)	16 (13.2)	15 (6.4)
Traction bronchiectasis	252 (70.8)	60 (49.6)	192 (81.7)
Semiquantitative scores[#]			
Honeycombing	0.20±0.47	0.16±0.44	0.22±0.49
Reticulation	0.86±0.53	0.63±0.54	0.98±0.48
Ground glass	1.59±1.06	2.01±1.22	1.37±0.90
Mosaic attenuation	0.50±0.83	1.04±1.06	0.22±0.48
Air trapping (n=293 with expiratory slice)	0.82±0.87	1.41±0.93	0.49±0.63
Centrilobular nodules	0.46±0.98	0.71±1.15	0.34±0.87
Traction bronchiectasis	2.84±1.83	1.92±1.81	3.31±1.65
Cranio-caudal distribution probability			
Upper lung	0.14±0.28	0.18±0.30	0.12±0.27
Lower lung	0.55±0.43	0.31±0.38	0.67±0.40
Diffuse	0.31±0.36	0.50±0.37	0.22±0.30
Axial distribution probability			
Central lung	0.08±0.19	0.07±0.15	0.08±0.20
Peripheral/subpleural lung	0.52±0.42	0.30±0.38	0.63±0.40
Peribronchovascular	0.01±0.06	0.003±0.03	0.02±0.07
Diffuse	0.39±0.38	0.63±0.39	0.27±0.32

Data are presented as n (%) or mean±sd. "MA-AT": mosaic attenuation or air trapping; "MA-AT>Reticulation": extent of mosaic attenuation or air trapping greater than that of reticulation; "GG>Reticulation": extent of ground glass greater than that of reticulation. [#]: for honeycombing, reticulation, ground glass, mosaic attenuation and air trapping, the average score was determined by summing five lobe scores and dividing by 5 for each subject and each radiologist/reader; for centrilobular nodules and traction bronchiectasis, the sum was determined by adding the score for five lobes. Average and sum scores are scaled 0–5, corresponding to proportion of total lung having the feature (for honeycombing, reticulation, ground glass, mosaic attenuation and air trapping) or number of lobes having the feature (for centrilobular nodules and traction bronchiectasis). The scores of the three radiologists were then averaged. [¶]: a feature was designated as "absent" for each radiologist/reader when the average score (for honeycombing, reticulation, ground glass, mosaic attenuation and air trapping) or sum score (for centrilobular nodules and traction bronchiectasis) for that features was <0.5, otherwise the feature was considered "present". The consensus scores are shown in this table. A feature was considered "present" when two out of three radiologists called it present. Mosaic attenuation or air trapping was present when a subject had consensus presence of mosaic attenuation or air trapping. The "GG>Reticulation" variable was present when the semiquantitative ground glass score was higher than the semiquantitative reticulation score. The "MA-AT>Reticulation" variable was present when the semiquantitative score for mosaic attenuation or air trapping was higher than the semiquantitative reticulation score.

pneumonitis probability) has hypersensitivity pneumonitis. For comparison, an HP-HRCT Diagnosis Score of 3 corresponds to PPV 74% and negative predictive value (NPV) 80% in the UMHS cohort, and PPV 44% and NPV 86% in the LTRC cohort.

Discussion

We describe the development and validation of a radiological diagnosis model for hypersensitivity pneumonitis. When a combination of diffuse axial distribution of interstitial abnormality and the extent of mosaic attenuation or air trapping is greater than that of reticulation, the risk of making a false-positive hypersensitivity pneumonitis diagnosis is <10%. Misdiagnosed UMHS patients (n=23) had desquamative interstitial pneumonia (n=3), IPF (n=10), pneumoconiosis (n=1), respiratory bronchiolitis-ILD (n=5), pulmonary vasculitis (n=2), NSIP (n=1) and sarcoidosis (n=1). Misdiagnosed LTRC patients (n=15) had IPF (n=4), NSIP (n=4), desquamative interstitial pneumonia (n=2), respiratory bronchiolitis-ILD (n=2), uncharacterised fibrosis (n=1) and autoimmune disease (n=2).

Several previous analyses have evaluated the use of HRCT as a diagnostic test for hypersensitivity pneumonitis in various populations. LYNCH *et al.* [3] retrospectively evaluated the CTs of 19

TABLE 4 Univariable logistic regression predicting hypersensitivity pneumonitis diagnosis: University of Michigan Health System (UMHS) cohort

Predictor	OR (95% CI)	p-value
Age per +1 year	0.98 (0.96–1.002)	0.08
Male sex	0.35 (0.22–0.55)	<0.001
Race (White versus others)	1.45 (0.74–2.86)	0.28
Smoker status (ever versus never)	0.64 (0.41–0.99)	0.04
Pack-years smoked per +1 year	0.99 (0.97–0.997)	0.02
Symptom duration per +1 month	1.00 (0.996–1.01)	0.38
Corticosteroid prior to UMHS evaluation	1.51 (0.95–2.39)	0.08
Baseline physiology per +10%		
FVC % pred	0.99 (0.87–1.20)	0.83
FEV ₁ % pred	0.94 (0.84–1.06)	0.30
FEV ₁ /FVC ratio	0.86 (0.65–1.13)	0.28
DLco % pred	0.97 (0.85–1.12)	0.71
Honeycombing present	0.47 (0.25–0.89)	0.02
Honeycombing semiquantitative [#]	0.73 (0.44–1.23)	0.24
Reticulation present	0.24 (0.15–0.39)	<0.001
Reticulation semiquantitative [#]	0.26 (0.16–0.41)	<0.001
Ground glass present	2.58 (1.46–4.57)	0.001
Ground glass semiquantitative [#]	1.78 (1.43–2.22)	<0.001
Mosaic attenuation present	8.10 (4.73–13.84)	<0.001
Mosaic attenuation semiquantitative [#]	3.88 (2.71–5.55)	<0.001
Air trapping present	6.54 (3.85–11.11)	<0.001
Air trapping semiquantitative [#]	4.06 (2.85–5.79)	<0.001
“MA-AT” present	6.89 (4.23–11.22)	<0.001
“GG>Reticulation” present	3.64 (2.42–6.17)	<0.001
“MA-AT>Reticulation” present	9.97 (6.00–16.56)	<0.001
Centrilobular nodules present	2.24 (1.06–4.69)	0.03
Centrilobular nodules semiquantitative [#]	1.45 (1.16–1.81)	0.001
Traction bronchiectasis present	0.22 (0.14–0.36)	<0.001
Traction bronchiectasis semiquantitative [#]	0.65 (0.57–0.74)	<0.001
Craniocaudal distribution		
Upper lung	0.31 (0.12–0.78)	0.01
Lower lung	0.08 (0.04–0.16)	<0.001
Diffuse	Reference	
Axial distribution		
Central	0.14 (0.03–0.61)	0.01
Peripheral/subpleural	0.08 (0.04–0.15)	<0.001
Peribronchovascular	0.0007 (0–0.38)	0.02
Diffuse	Reference	

FVC: forced vital capacity; FEV₁: forced expiratory volume in 1 s; DLco: diffusing capacity of the lung for carbon monoxide; “MA-AT”: mosaic attenuation or air trapping; “MA-AT>Reticulation”: extent of mosaic attenuation or air trapping greater than that of reticulation; “GG>Reticulation”: extent of ground glass greater than that of reticulation. [#]: odds ratios for semiquantitative scores are per 1-unit increase in the score. Scores are on a scale of 0–5. The “MA-AT” variable was present when a subject had consensus presence of mosaic attenuation or air trapping. The “GG>Reticulation” variable was present when the semiquantitative ground glass score was higher than the semiquantitative reticulation score. The “MA-AT>Reticulation” variable was present when the semiquantitative score for mosaic attenuation or air trapping was higher than the semiquantitative reticulation score.

TABLE 5 HP-HRCT Diagnosis Score calculation

Feature	Points
Mosaic attenuation or air trapping more extensive than reticulation	2
Diffuse axial disease distribution	1

hypersensitivity pneumonitis and 33 IPF patients, and determined that an expert thoracic radiologist’s CT diagnosis of definite hypersensitivity pneumonitis was correct 92% of the time. A “definite hypersensitivity pneumonitis” pattern was not defined *a priori* in this study. HRCT features associated with

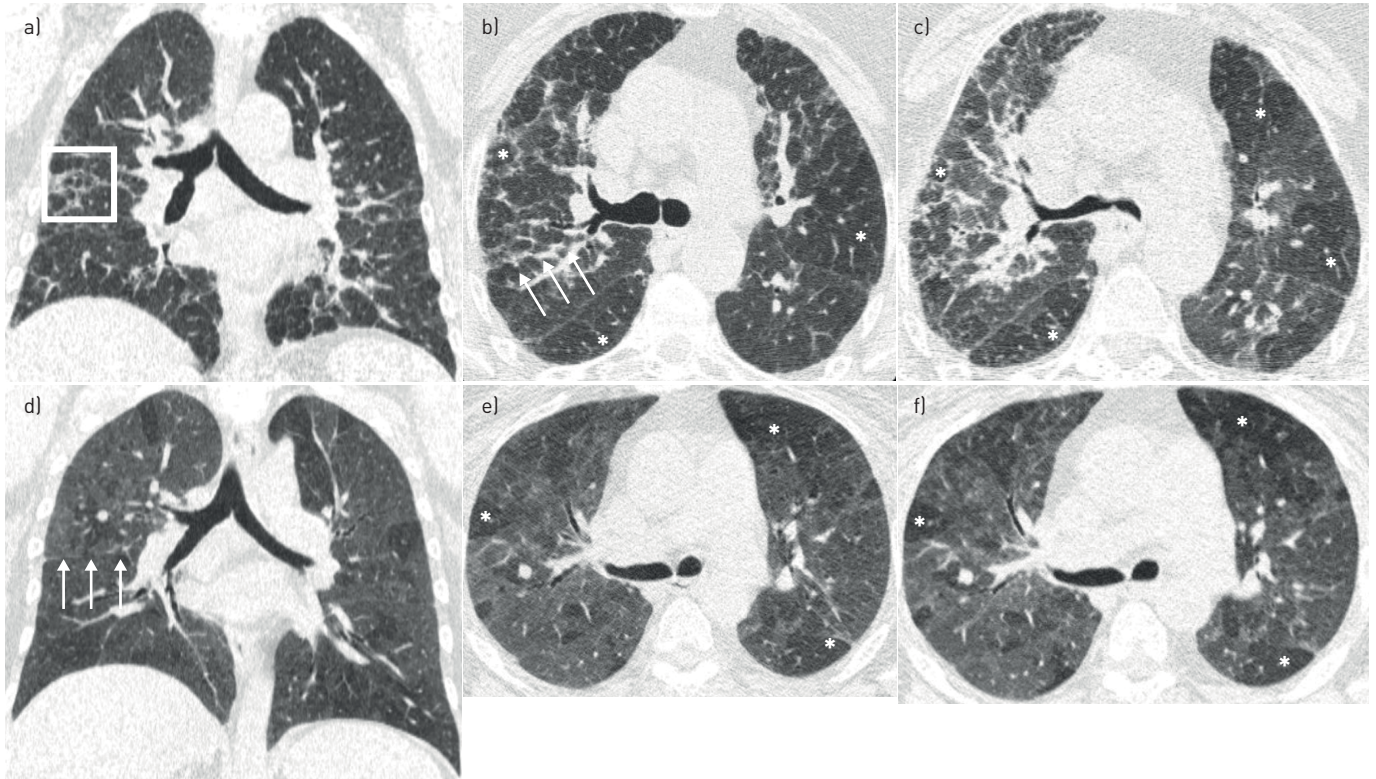


FIGURE 3 Representative images from two patients with hypersensitivity pneumonitis and an HP-HRCT Diagnosis Score of 3. a) Inspiratory coronal reconstruction, b) inspiratory axial supine and c) expiratory axial supine images from a subject with fibrotic hypersensitivity pneumonitis. Disease is diffusely distributed in the axial direction (white arrows in b). An example of reticulation is shown in the box in image (a), and asterisks mark regions of relatively decreased attenuation (mosaic attenuation) on the inspiratory image (b) and air trapping on the expiratory image (c). d) Inspiratory coronal reconstruction, e) inspiratory axial supine and f) expiratory axial supine images from a subject with non-fibrotic hypersensitivity pneumonitis. Disease is diffusely distributed in the axial direction (white arrows in d). Asterisks mark regions of relatively decreased attenuation (mosaic attenuation) on the inspiratory image (e) and air trapping on the expiratory image (f).

TABLE 6 HP-HRCT Diagnosis Score test characteristics: University of Michigan Health System (UMHS) and Lung Tissue Research Consortium (LTRC) cohorts

Threshold	Hypersensitivity pneumonitis	Not hypersensitivity pneumonitis	Sensitivity	Specificity
UMHS cohort				
>0	95	77	78.5 (71.2–85.8)	67.2 (61.2–73.2)
>1	85	45	70.2 (62.1–78.4)	80.9 (75.8–85.9)
>2	67	23	55.4 (46.5–64.2)	90.2 (86.4–94.0)
LTRC cohort				
>0	35	70	53.0 (41.0–65.1)	80.4 (76.3–84.6)
>1	22	29	33.3 (22.0–44.7)	91.9 (89.1–94.7)
>2	12	15	18.2 (8.9–27.5)	95.8 (93.7–97.9)

Data are presented as n or % [95% CI].

hypersensitivity pneumonitis included less honeycombing and traction bronchiectasis, and presence of micronodules and relative sparing of the lower half of the lower lung zone. Mosaic attenuation/perfusion was not evaluated. SILVA *et al.* [4] performed a similar analysis in a population of 16 hypersensitivity pneumonitis, 23 IPF and 25 NSIP patients, and found that a confident CT diagnosis of hypersensitivity pneumonitis was correct 88% of the time. Hypersensitivity pneumonitis patients had more lobular areas of decreased attenuation, air trapping, centrilobular nodules, relative sparing of the bases, random cranio-caudal disease distribution and peribronchovascular or random axial disease distribution. JOHANNSSON *et al.* [5] evaluated HRCTs of a broad range of ILD diagnoses (hypersensitivity pneumonitis prevalence 53%).

Clinical prediction models incorporating age, down feather exposure, bird exposure and either a radiologist's moderate to high confidence in an hypersensitivity pneumonitis diagnosis or HRCT features of diffuse craniocaudal distribution of ground glass abnormality and mosaic perfusion had good discriminative performance (c-statistic 0.758–0.778). Our model 4 in supplementary table E6 is similar to the model of JOHANSSON *et al.* [5], but with adjustment variables of sex and smoking history rather than bird/feather exposure. We found that ground glass as a dichotomous variable is not independently associated with hypersensitivity pneumonitis, but mosaic attenuation (alone and when combined with air trapping or relative extent of reticulation) and diffuse axial and craniocaudal distributions are. Our HRCT evaluation methodology does not allow for exact replication of the model of JOHANSSON *et al.* [5].

Our study expands upon the previously mentioned analyses in several ways. First, we include a broad range of non-hypersensitivity pneumonitis ILD diagnoses, including fibrotic and non-fibrotic interstitial lung diseases. The HP-HRCT Diagnosis Score can therefore be applied across ILD phenotypes (*i.e.* fibrotic and non-fibrotic). Roughly half of our hypersensitivity pneumonitis patients had fibrotic features of reticulation or traction bronchiectasis on HRCT. Second, the sensitivity analyses indicates that the HP-HRCT Diagnosis Score is also useful for identification of hypersensitivity pneumonitis diagnosed when surgical lung biopsy confirmed the presence of a classical/typical histopathological hypersensitivity pneumonitis pattern (as documented by the original interpreting UMHS pathologist). A challenging aspect of diagnosing a patient with ILD with hypersensitivity pneumonitis is the idea that hypersensitivity pneumonitis may present a variety of histopathological patterns [18–21]. Our HP-HRCT Diagnosis Score therefore identifies a subset of ILD likely to have typical histopathological findings of hypersensitivity pneumonitis where a biopsy to be performed. Third, a sensitivity analysis excluding subjects with a documented exposure history finds the model to be robust in the group without identified environmental exposure. Fourth, we have validated our final model and a simple model-based HP-HRCT Diagnosis Score in a separate, multicentre cohort of ILD subjects. The UMHS cohort HRCTs were evaluated by a different methodology than those of the LTRC cohort. This supports the idea that our model-based score is useful as a simple dichotomous checklist that can be applied in practice even when a complex HRCT scoring methodology is not in use.

Application of this type of score in practice should involve consideration of the sensitivity and specificity of the test as presented here, as well as post-test probability of disease which is dependent on disease prevalence. With this in mind, several limitations of our HP-HRCT Diagnosis Score should be noted. First, the sensitivity of a high HP-HRCT Diagnosis Score (>2 points) is relatively low (55.4% in the UMHS cohort and 18.2% in the LTRC cohort), indicating that a substantial fraction of patients with hypersensitivity pneumonitis are not identified by a high HP-HRCT Diagnosis Score. Second, in applying the HP-HRCT Diagnosis Score in practice, the evaluating clinician should consider the probability of having hypersensitivity pneumonitis given a positive test (*i.e.* the PPV of an HP-HRCT Diagnosis Score of 3) when deciding whether additional testing is needed to make a confident hypersensitivity pneumonitis diagnosis [1]. We present specificity of various HP-HRCT Diagnosis Score thresholds, which is not dependent on disease prevalence, and conclude that the risk of a false hypersensitivity pneumonitis diagnosis is <10% when a patient receives all 3 points. However, PPV at the same threshold depends on disease prevalence, which should be determined by clinicians using the test and based on knowledge of regional prevalence of hypersensitivity pneumonitis or other characteristics of the patient such as exposure history. In the UMHS cohort (hypersensitivity pneumonitis prevalence 34%), an HP-HRCT Diagnosis Score of 3 is associated with PPV 74%. In the LTRC cohort (hypersensitivity pneumonitis prevalence 16%), an HP-HRCT Diagnosis Score of 3 has PPV 44%. It is unknown how additional clinical variables or diagnostic test results such as identification of an hypersensitivity pneumonitis exposure or fluid lymphocytes during bronchoalveolar lavage will modify the probability/PPV of hypersensitivity pneumonitis in concert with the HP-HRCT Diagnosis Score. These questions should be the subject of future, ideally prospective, studies. Decisions regarding the need for additional diagnostic testing such as bronchoscopy or surgical lung biopsy after finding a “positive” result using the HP-HRCT Diagnosis Score should be individualised to the patient and practice setting.

Interestingly, the most recent iteration of multisociety diagnosis guidelines for IPF includes “GG>Reticulation” as a feature inconsistent with radiological UIP [2]. In our study, this variable was associated with hypersensitivity pneumonitis (and therefore against IPF, given the makeup of the non-hypersensitivity pneumonitis control population) in unadjusted analysis, but was not an independent predictor of hypersensitivity pneumonitis after adjusting for age, sex, smoking history, presence of mosaic attenuation or air trapping and axial disease distribution (supplementary table E6, model 5). Finding more extensive mosaic attenuation or air trapping than reticulation is consistently and strongly associated with hypersensitivity pneumonitis in our study.

Our study has several weaknesses. First, three thoracic radiologists scored each HRCT, generating robust data on agreement on HRCT features. This data alerted us to poor agreement on HRCT features of reticulation, ground glass and centrilobular nodules. We did include reticulation in our model *via* the combined “MA-AT>Reticulation” variable, raising potential concerns about reproducibility of results if broadly applied. Despite poor agreement on presence of reticulation among UMHS radiologists, this reticulation-based variable remained strongly associated with an hypersensitivity pneumonitis diagnosis in the LTRC cohort and the model appears to be valid despite measurement of reticulation using a different scoring system applied to CTs by different radiologists. Second, histopathological disease confirmation was required for inclusion in the control group of the UMHS cohort, and for hypersensitivity pneumonitis and control groups of the LTRC cohort. This methodology may select for subjects with an atypical HRCT, thereby potentially altering conclusions about what features distinguish hypersensitivity pneumonitis from other ILDs. The LTRC cohort included final diagnoses of unclassifiable fibrosis, mitigating this concern given good model performance in the validation cohort. Third, we cannot rule out the possibility that HRCT findings had some influence on the final clinical diagnoses. While we were careful to blind clinical diagnosis verification to the HRCT reports/findings, these reports were available to the clinicians and multidisciplinary teams assigning the original clinical diagnoses to the UMHS and LTRC patients. Fourth, UMHS was a participating centre in the LTRC study. Review of internal records indicates that 110 subjects from the UMHS cohort were enrolled in the LTRC study. The LTRC data was de-identified and did not include notation of the referring centre, so we are unable to remove the overlap. If all 110 UMHS cohort subjects were included among the 424 validation subjects, up to 26% of the validation cohort would overlap with the derivation cohort.

In conclusion, we have developed and validated a radiological diagnosis model for hypersensitivity pneumonitis such that when the extent of mosaic attenuation or air trapping is greater than reticulation and diffuse axial distribution of interstitial abnormality are present in combination, the specificity for hypersensitivity pneumonitis is <90% (*i.e.* <10% false-positive rate) for clinically diagnosed hypersensitivity pneumonitis.

Acknowledgements: This study utilised data provided by the Lung Tissue Research Consortium supported by the National Heart, Lung, and Blood Institute.

Author contributions: M.L. Salisbury, K.R. Flaherty, F.J. Martinez and E.A. Kazerooni conceived and designed the study; M.L. Salisbury and M. Xia analysed the data with supervision and assistance from S. Murray. K.R. Flaherty, B.J. Bartholmai, E.A. Kazerooni, F.J. Martinez, B.H. Gross, A. Chughtai, M. Sayyoub, J.L. Myers, A. Lagstein, K.E. Konopka, E.A. Belloli, J.S. Sheth, E.S. White and C. Holtze contributed data; M.L. Salisbury prepared the manuscript; all authors critically revised the manuscript for intellectual content, approved the final draft and agree to accountability for all aspects of the work.

Conflict of interest: M.L. Salisbury reports salary funding from a departmental National Institutes of Health (NIH) training grant. B.J. Bartholmai reports other support from the NIH/National Heart, Lung, and Blood Institute for research related to the Lung Tissue Research Consortium (LTRC), previous to the conduct of the study. S. Murray reports that NIH sponsored grants pay for statistical work carried out for the Pulmonary Division. F.J. Martinez has received grants for chronic obstructive pulmonary disease (COPD) and idiopathic pulmonary fibrosis (IPF) studies from NIH, has participated in steering committees for IPF studies for Bayer, Centocor, Gilead and Promedior, has received personal fees (IPF advisory board) from Ikaria, Genentech, Nycomed/Takeda, Pfizer and Vertex, personal fees (IPF CME programmes) from the American Thoracic Society, MedScape and National Association for Continuing Education, personal fees (IPF grand rounds) from Inova Health System, Spectrum Health System and University of Texas Southwestern, personal fees (IPF study DSMB) from Stromedix/Biogen, personal fees (IPF teleconference consultation) from Axon Communications, Johnson & Johnson and Genzyme, personal fees (IPF advisory board) from Boehringer Ingelheim, and personal fees (IPF diagnostic advisor) from Veracyte, during the conduct of the study; and has received personal fees (steering committee COPD study) from Forest, Janssen, GSK and Nycomed/Takeda, personal fees (COPD PRO development) from Amgen, personal fees (COPD advisory board) from Actelion, AstraZeneca, CSA Medical, Ikaria/Bellerophon, Forest, Genentech, GSK, Janssen, Merck, Pearl, Nycomed/Takeda, Pfizer, Roche and Sudler & Hennessey, personal fees (COPD CME programmes) from the American College of Chest Physicians, CME Incite, Center for Healthcare Education, MedScape, Miller Medical, National Association for Continuing Education, Paradigm, Peer Voice, Projects in Knowledge, UpToDate, Wayne State University and Annenberg, personal fees (COPD grand rounds) from Inova Health System, St John’s Hospital, St Mary’s Hospital and University of Illinois Chicago, personal fees (COPD study DSMB) from GSK, personal fees (COPD FDA mock presentation) from Boehringer Ingelheim, GSK and Ikaria, personal fees (European Respiratory Society (ERS) bronchiectasis presentation) from Bayer, personal fees (ERS COPD presentation) from Nycomed/Takeda, personal fees (COPD consulting teleconference) from Grey Healthcare and Merion, personal fees (COPD book royalties) from Informa, and personal fees (speaking on COPD) from GSK and Forest, outside the submitted work. K.R. Flaherty reports grants from NIH, during the conduct of the study; and reports personal fees for consultancy on IPF from Boehringer Ingelheim, Fibrogen, Genentech, Ikaria, ImmuneWorks, MedImmune, Novartis, Takeda, Vertex, Veracyte, Roche and Biogen, personal fees for consultancy and DSMB on IPF from Gilead, personal fees for employment from Pulmonary Fibrosis Foundation, personal fees for consultancy and grants for clinical trials on IPF from Intermune, and grants for clinical trials on IPF from Bristol-Myers Squibb, outside the submitted work.

Support statement: T32 HL00749-21 (Multidisciplinary Training Program in Lung Disease), National Institutes of Health K24 HL111316 (K.R. Flaherty), and National Institutes of Health/National Heart, Lung, and Blood Institute HHSN26820118C (Lung Tissue Research Consortium). Funding information for this article has been deposited with the Crossref Funder Registry.

References

- 1 Salisbury ML, Myers JL, Belloli EA, *et al.* Diagnosis and treatment of fibrotic hypersensitivity pneumonia. where we stand and where we need to go. *Am J Respir Crit Care Med* 2017; 196: 690–699.
- 2 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: 788–824.
- 3 Lynch DA, Newell JD, Logan PM, *et al.* Can CT distinguish hypersensitivity pneumonitis from idiopathic pulmonary fibrosis? *AJR Am J Roentgenol* 1995; 165: 807–811.
- 4 Silva CI, Muller NL, Lynch DA, *et al.* Chronic hypersensitivity pneumonitis: differentiation from idiopathic pulmonary fibrosis and nonspecific interstitial pneumonia by using thin-section CT. *Radiology* 2008; 246: 288–297.
- 5 Johannson KA, Elicker BM, Vittinghoff E, *et al.* A diagnostic model for chronic hypersensitivity pneumonitis. *Thorax* 2016; 71: 951–954.
- 6 Welker L, Jorres RA, Costabel U, *et al.* Predictive value of BAL cell differentials in the diagnosis of interstitial lung diseases. *Eur Respir J* 2004; 24: 1000–1006.
- 7 Ohshimo S, Bonella F, Cui A, *et al.* Significance of bronchoalveolar lavage for the diagnosis of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2009; 179: 1043–1047.
- 8 American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am J Respir Crit Care Med* 2000; 161: 646–664.
- 9 Flaherty KR, King TE Jr, Raghu G, *et al.* Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med*. 2004; 170: 904–910.
- 10 Salisbury ML, Xia M, Murray S, *et al.* Predictors of idiopathic pulmonary fibrosis in absence of radiologic honeycombing: a cross sectional analysis in ILD patients undergoing lung tissue sampling. *Respir Med* 2016; 118: 88–95.
- 11 Salisbury M, Gross BH, Chughtai A, *et al.* Utility of high-resolution computed tomography for diagnosis of hypersensitivity pneumonia. *Am J Respir Crit Care Med* 2017; 196: A1584.
- 12 Hansell DM, Bankier AA, MacMahon H, *et al.* Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; 246: 697–722.
- 13 Conger AJ. Integration and generalization of kappas for multiple raters. *Psychol Bull* 1980; 88: 322–328.
- 14 Marasini D, Quatto P, Ripamonti E. Assessing the inter-rater agreement for ordinal data through weighted indexes. *Stat Methods Med Res* 2016; 25: 2611–2633.
- 15 McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 1947; 12: 153–157.
- 16 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996; 15: 361–387.
- 17 Freeman EA, Moisen G. PresenceAbsence: an R package for presence absence analysis. *J Stat Softw* 2008; 23: 1–31.
- 18 Vourlekis JS, Schwarz MI, Cool CD, *et al.* Nonspecific interstitial pneumonitis as the sole histologic expression of hypersensitivity pneumonitis. *Am J Med* 2002; 112: 490–493.
- 19 Ohtani Y, Saiki S, Kitaichi M, *et al.* Chronic bird fancier's lung: histopathological and clinical correlation. An application of the 2002 ATS/ERS consensus classification of the idiopathic interstitial pneumonias. *Thorax* 2005; 60: 665–671.
- 20 Churg A, Sin DD, Everett D, *et al.* Pathologic patterns and survival in chronic hypersensitivity pneumonitis. *Am J Surg Pathol* 2009; 33: 1765–1770.
- 21 Chiba S, Tsuchiya K, Akashi T, *et al.* Chronic hypersensitivity pneumonitis with a usual interstitial pneumonia-like pattern: correlation between histopathologic and clinical findings. *Chest* 2016; 149: 1473–1481.