ORIGINAL RESEARCH LETTER

Evaluation of inter-observer variation for computed tomography identification of childhood interstitial lung disease

To the Editor:

Interstitial lung diseases (ILDs) that present in childhood (chILD) are seen far less frequently than ILDs presenting in adults which themselves constitute rare disorders [1]. Histopathological [2, 3] and imaging [4] characterisation of chILD disease subtypes therefore lags behind adult ILDs. The field has also been constrained by comparisons with disease morphology in adults, despite the developmental differences in terms of growth and healing in the paediatric lung, which may alter disease patterns and distributions. The American Thoracic Society [5] and European [1] chILD management guidelines both specify a pivotal role for computed tomography (CT) imaging in the work-up of chILD patients to: 1) determine whether a chILD is present or not; and 2) where possible, to make a specific diagnosis of the underlying cause. For the second aim to be achieved, diagnostic reviews need to be reproducible between experts. Our study uniquely examined agreement between observers of varying experience in the CT evaluation of chILD to inform whether the current status of CT imaging and knowledge can be diagnostic of specific chILDs. We hypothesised that observer agreement for chILD groups and diagnoses would be limited. The study was not designed to relate CT agreement to final diagnosis. As a secondary analysis, we examined how CT interpretation differed between observers in children under and over 2 years of age.

84 patients (<2 years, n=35) from 13 countries, referred to the Royal Brompton Hospital paediatric ILD unit from 1999 to 2014 were re-evaluated by 10 observers: three chILD sub-specialist radiologists with >25 years' experience (C.M. Owens, P. Garcia-Peña, A.S. Brody); four chILD pulmonologists with >20 years' experience (S. Cunningham, T.J. Vece, P. Aurora, A. Moreno-Galdó); and three radiologists with 5–10 years' chILD experience (A. Calder, P. Toma, T.A. Watson). Consecutive patients were chosen and included in the study if clinically indicated DICOM format CTs were retrievable and of acceptable quality (high-resolution, ⩽2 mm slice thickness) as determined by an independent radiologist (J. Jacob). Observers were given details of patient sex and age. Each scorer provided up to three choices when assigning a chILD diagnostic group or individual chILD diagnosis from a preselected list. For each choice, a measure of group/diagnostic likelihood was also assigned with confidence scores ranging from 0% to 100%: 100%=pathognomonic, 70–90%=high confidence, 40–60%=moderate confidence, 10–30%=low confidence, 0%=no confidence. The choices were based on previous histopathological work [2]. **Q1**

The Fleiss Kappa evaluated observer agreement for first-choice chILD groups/individual diagnoses. For chILD groups/individual diagnoses where agreement was at least fair, observer confidence was examined using linear weighted kappas [6]. Results were reanalysed according to patient age (<2 years *versus* >2 years, the conventional dichotomy in chILD classifications [7]) given the variable prevalence of chILD disease groups/individual diagnoses across these age ranges. Agreement was also examined in observer subgroups (senior *versus* junior radiologists *versus* senior pulmonologists). Statistical analyses were performed with R: A language and environment for statistical computing [8]. Approval for the study was obtained from the Institutional Research Committee of the Royal Brompton Hospital (Project 1157).

TABLE 1 Assignations of first-choice individual childhood interstitial lung disease (chILD) group or diagnosis in children under 2 years of age (n=35) and over 2 years of age (n=49) by 10 observers

chILD category	chILD group/diagnosis	Numbers of first-choice assignations	
		<2 years	>2 years
chILD groups	Normal	3	18
	Airways disease[#]	77	170
	Interstitial pneumonia	126	200
	Developmental/undefined aetiology disorders[¶]	99	48
	Unclassifiable	4	13
	Bronchopulmonary dysplasia	29	10
	Other	12	31
chILD Diagnoses	Exudative bronchiolitis	4	2
	Constrictive obliterative bronchiolitis	10	48
	Infection	48	86
	Chronic pneumonitis of infancy	47	12
	Pulmonary alveolar proteinosis	53	24
	Desquamative interstitial pneumonia	8	13
	Nonspecific interstitial pneumonia (cellular)	16	34
	Diffuse alveolar damage	3	8
	Fibrotic nonspecific interstitial pneumonia	1	47
	Organising pneumonia	0	30
	Diffuse alveolar haemorrhage	1	11
	Alveolar microlithiasis	0	3
	Sarcoidosis	0	6
	Langerhans cell histiocytosis	1	12
	Lung growth abnormality/chronic disease of prematurity	83	12
	Pulmonary interstitial glycogenosis	11	0
	Neuroendocrine cell hyperplasia of infancy	30	15
	Lymphoid interstitial pneumonia	1	23
	Diffuse lymphoid hyperplasia	1	2
	Lymphomatoid granulomatosis	1	4
	Follicular bronchiolitis	0	6
	Bronchus-associated lymphoid tissue	0	1
	Primary pulmonary arterial hypertension	1	4
	PVOD/PCH	3	5
	Thromboembolic disease	1	3
	Lymphangiomatosis	1	1
	Lymphangiectasia (primary and secondary)	7	7
	Secondary vasculopathies (e.g. due to cardiac disease)	6	2
	Primary and secondary pulmonary vasculitis	1	14
	Neoplasms	0	9
	Alternate diagnosis	11	24

PVOD: pulmonary veno-occlusive disease; PCH: pulmonary capillary haemangiomatosis. [#]: airways disease included infection and obliterative and exudative bronchiolitis; [¶]: developmental/undefined aetiology disorders included growth disorders such as alveolar hypoplasia, neuroendocrine cell hyperplasia of infancy and pulmonary interstitial glycogenosis.

840 first-choice chILD group/diagnostic assignations were made by the 10 observers (table 1). chILD groups/diagnoses whose frequency was <10% of the total number of first-choice diagnoses in patients under and over 2 years of age, were grouped as "others", in accord with previous studies [9]. Four chILD groups were analysed: airways disease (n=247), interstitial pneumonia (n=326), developmental/undefined aetiology disorders (n=147) and others (n=120). Five chILD diagnoses could be analysed in patients <2 years: infection (n=48), chronic pneumonitis of infancy (CPI; n=47), pulmonary alveolar proteinosis (PAP, n=53), developmental/undefined aetiology disorders (n=83) and others (n=119). Four chILD diagnoses could be analysed in patients >2 years: obliterative bronchiolitis (n=48), infection (n=86), fibrotic nonspecific interstitial pneumonia (fNSIP; n=47) and others (n=309).

Patient groups	Observer groups	All cases	Airways disease[#]	Interstitial pneumonia	Developmental/undefined[¶]
All patients	All observers	0.30 (0.28–0.32)	0.37 (0.34–0.41)	0.36 (0.33–0.40)	0.18 (0.15–0.21)
	Senior radiologists	0.35 (0.27–0.42)	0.46 (0.34–0.58)	0.44 (0.31–0.56)	0.16 (0.04–0.29)
	Senior pulmonologists	0.28 (0.23–0.33)	0.34 (0.25–0.42)	0.36 (0.27–0.45)	0.21 (0.12–0.29)
	Junior radiologists	0.31 (0.23–0.38)	0.45 (0.33–0.57)	0.39 (0.27–0.51)	0.07 (−0.06–0.19)
Patients <2 years of age	All observers	0.22 (0.19–0.25)	0.28 (0.23–0.33)	0.30 (0.25–0.35)	0.14 (0.09–0.19)
	Senior radiologists	0.36 (0.24–0.48)	0.61 (0.42–0.80)	0.46 (0.26–0.65)	0.15 (−0.05–0.34)
	Senior pulmonologists	0.18 (0.10–0.26)	0.24 (0.11–0.38)	0.26 (0.12–0.39)	0.16 (0.02–0.29)
	Junior radiologists	0.13 (0.01–0.24)	0.35 (0.16–0.54)	0.19 (−0.00–0.38)	−0.05 (−0.24–0.14)
Patients >2 years of age	All observers	0.33 (0.31–0.36)	0.40 (0.36–0.45)	0.40 (0.36–0.44)	0.11 (0.07–0.15)
	Senior radiologists	0.31 (0.21–0.41)	0.36 (0.20–0.53)	0.42 (0.26–0.58)	0.12 (−0.04–0.29)
	Senior pulmonologists	0.32 (0.25–0.40)	0.36 (0.24–0.47)	0.43 (0.32–0.55)	0.09 (−0.02–0.21)
	Junior radiologists	0.42 (0.32–0.53)	0.52 (0.36–0.68)	0.48 (0.32–0.64)	0.09 (−0.07–0.25)

TABLE 2 Fleiss kappa values for observer agreement for first choice group assignation in the 84 study cases. Sub-analyses are shown for patients under 2 years of age (n=35) and over 2 years of age (n=49)

[#]: airways disease included infection and obliterative and exudative bronchiolitis; [¶]: developmental/undefined aetiology disorders included growth disorders such as alveolar hypoplasia, neuroendocrine cell hyperplasia of infancy and pulmonary interstitial glycogenosis.

When subanalysed according to patient age, observer agreement for first-choice chILD groups was moderate for airways disease and interstitial pneumonia in patients >2 years, but only fair in patients <2 years (table 2). There was generally less observer variation between senior radiologists than pulmonologists. In patients <2 years senior radiologists demonstrated better agreement than other observers. High confidence chILD group assignations (>70%) were more common in children >2 years and increased with observer seniority (data not shown). Encouragingly, agreement for diagnostic groups was best for infection/airways disease where conservative management without biopsy may be preferable.

Observer agreement in assigning first-choice individual chILD diagnoses was moderate for obliterative bronchiolitis (in children >2 years), and fair for other diagnoses except CPI where agreement was poor. Further analyses were only performed in individual diagnoses where agreement was at least fair. Observer agreement for diagnostic likelihoods were moderate for obliterative bronchiolitis (children >2 years of age), but were at best only fair for other diagnoses, regardless of patient age (data not shown). Observer agreement for diagnostic likelihood of infection in children <2 years was poor across all observer subgroups.

We report poor agreement between even experienced thoracic radiologists for the identification of most chILD groups and individual diagnoses across ages. The lower observer agreement for individual chILD diagnoses compared with chILD diagnostic groups may reflect differing methods by which radiologists classify CT appearances. Some observers might classify according to histological pattern and distinguish features of NSIP from PAP in a child >2 years, while others might classify according to suspected aetiology and recognise both patterns as suggesting a surfactant protein disorder. Our findings stress the need for a new common uniform descriptor language for CT appearances, because the present system is not reproducible and therefore confusing.

The current classification system in chILD is based on histopathological diagnoses [2, 7], and as occurred with adult fibrosing lung diseases, it is hoped that histopathological descriptors will inform those CT features that are of importance in delineating childhood disease. Two early studies examining diagnostic accuracy using radiological–histopathological correlations in a small subset of chILD patients demonstrated that CT features could identify histopathological diagnoses with acceptable accuracy [4, 10]. However, prior to the evaluation of CT diagnostic accuracy it is essential to understand whether expert readers can coalesce on a single diagnosis with sufficient frequency that a single CT diagnosis can be compared to a single histopathological diagnosis.

Study limitations included the time frame over which cases were acquired, where CT imaging protocols and quality varied widely. However, imaging heterogeneity is an unavoidable constraint when examining real-world paediatric data and we actively sought to draw real-world conclusions on chILD observer agreement. Limiting clinical information to patient age and sex might have increased observer disagreement. However rather than examining diagnostic accuracy, our *a priori* study aim was to understand the limitations of CT interrogation of chILD and identify those conditions where pattern recognition alone is insufficient to make a confident diagnosis. Though the study case mix from a tertiary centre is likely to have been weighted towards more challenging presentations of disease, chILD cases are only seen with sufficient frequency to allow analysis in such centres.

Our study is the largest examination of CT imaging in chILD cases, scored by the largest number of observers, to date, who importantly were from different specialties. We demonstrate that making chILD diagnoses on CT is difficult, even amongst sub-specialists. Considering the current state of knowledge in the field, we found that agreement for chILD diagnostic groups was generally better than for individual diagnoses and was highest amongst senior radiologists. Further detailed work is needed to understand if, with studies and training, improvements in CT diagnostic agreement are possible. Additionally, the ability of the multidisciplinary team to improve diagnostic confidence by augmenting radiology with clinical, functional and haematological results, as per guideline recommendations [1, 5] requires further study.

Joseph Jacob [1,2], **Catherine M. Owens**[3], **Alan S. Brody**[4], **Thomas Semple** [5], **Tom A. Watson**[3], **Alistair Calder**[3], **Pilar Garcia-Peña**[6], **Paolo Toma** [7], **Anand Devaraj**[5], **Henry Walton**[8], **Antonio Moreno-Galdó**[9], **Paul Aurora**[10], **Alexandra Rice**[11], **Timothy J. Vece**[12], **Steve Cunningham**[13], **Andre Altmann** [2], **Athol U. Wells**[14], **Andrew G. Nicholson**[11] and **Andrew Bush**[15]

[1]Dept of Respiratory Medicine, University College London, London, UK. [2]Centre for Medical Image Computing, University College London, London, UK. [3]Dept of Radiology, Great Ormond Street Hospital, London, UK. [4]Dept of Radiology, University of Cincinnati College of Medicine, and Cincinnati Children's Hospital, Cincinnati, OH, USA. [5]Dept of Radiology, Royal Brompton and Harefield NHS Foundation Trust, London, UK. [6]Dept of Pediatric Radiology, University Hospital Vall d'Hebron, Barcelona, Spain. [7]Dept of Radiology, Ospedale Pediatrico Bambino Gesù, Rome, Italy. [8]Dept of Radiology, Royal Free London NHS Foundation Trust, London, UK. [9]Dept of Paediatric Respiratory Medicine, University Hospital Vall d'Hebron, Universitat Autònoma de Barcelona, Barcelona, Spain; CIBERER, Spain. [10]Dept of Paediatric Respiratory Medicine, Great Ormond Street Hospital, London, UK. [11]Dept of Hisopathology, Royal Brompton and Harefield NHS Foundation Trust and National Heart and Lung Institute, Imperial College, London, UK. [12]Dept of Pediatrics, Division of Pediatric Pulmonology, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA. [13]Centre for Inflammation Research, University of Edinburgh, Edinburgh, UK. [14]Interstitial Lung Disease Unit, Royal Brompton and Harefield NHS Foundation Trust, London, UK. [15]Depts of Paediatrics and Paediatric Respiratory Medicine, Imperial College and Royal Brompton Hospital, London, UK.

Correspondence: Joseph Jacob, Dept of Respiratory Medicine, First Floor, Rayne Institute, University Street, London, WC1E 6JF. E-mail: j.jacob@ucl.ac.uk

References

1   Bush A, Cunningham S, de Blic J, *et al.* European protocols for the diagnosis and initial treatment of interstitial lung disease in children. *Thorax* 2015; 70: 1078–1084.

2   Rice A, Tran-Dang M-A, Bush A, *et al.* Diffuse lung disease in infancy and childhood: expanding the chILD classification. *Histopathology* 2013; 63: 743–755.

3   Fan LL, Dishop MK, Galambos C, *et al.* Diffuse lung disease in biopsied children 2 to 18 years of age. Application of the chILD classification scheme. *Ann Am Thorac Soc* 2015; 12: 1498–1505.

4   Copley SJ, Coren M, Nicholson AG, *et al.* Diagnostic accuracy of thin-section CT and chest radiography of pediatric interstitial lung disease. *AJR* 2000; 174: 549–554.

5   Kurland G, Deterding RR, Hagood JS, *et al.* An Official American Thoracic Society Clinical Practice Guideline: classification, evaluation, and management of childhood interstitial lung disease in infancy. *Am J Respir Crit Care Med* 2013; 188: 376–394.

6   Walsh SL, Calandriello L, Sverzellati N, *et al.* Hansell DM, null n. Interobserver agreement for the ATS/ERS/JRS/ALAT criteria for a UIP pattern on CT. *Thorax* 2016; 71: 45–51.

7   Deutsch GH, Young LR, Deterding RR, *et al.* Diffuse lung disease in young children: application of a novel classification scheme. *Am J Respir Crit Care Med* 2007; 176: 1120–1128.

8   R Core Team. R: A Language and Environment for Statistical Computing. Vienna, R Foundation for Statistical Computing, 2016.

9   Aziz ZA, Wells AU, Hansell DM, *et al.* HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax* 2004; 59: 506–511.

10 Lynch DA, Hay T, Newell JD, Jr., *et al.* Pediatric diffuse lung disease: diagnosis and classification using high-resolution CT. *AJR* 1999; 173: 713–718.