[Original Research]

CHEST

Accuracy of High-Speed Video Analysis to Diagnose Primary Ciliary Dyskinesia

Q26 *Bruna Rubbo, MSc; Amelia Shoemark, PhD; Claire L. Jackson, PhD; Robert Hirst, PhD; James Thompson; Joseph Hayes, BSc; Emily Frost, MSc; Fiona Copeland; Claire Hogg, MBChB; Christopher O'Callaghan, MD, PhD; Isabel Reading, PhD; Jane S. Lucas, MD, PhD; on behalf of the National PCD Service, UK*

Q1 Q2 Q3 Q4

BACKGROUND: Diagnosis of primary ciliary dyskinesia (PCD) relies on a combination of tests. High-speed video microscopy analysis (HSVA) is widely used to contribute to the diagnosis. It can be analyzed on the day of diagnostic consultation, but the qualitative analyses are subjective. Diagnostic accuracy and reliability of assessing ciliary function have not been robustly evaluated. We aimed to establish the accuracy of HSVA to diagnose PCD compared with a combination of tests, and to assess the interobserver reliability of HSVA analysis.

METHODS: We randomly selected and anonymized archived videos from 120 patients seen at three UK PCD centers. Three experienced scientists independently reviewed six videos per patient, using a standardized proforma, blinded to diagnostic and clinical data. We compared study outcomes with two references: (1) a combination of diagnostic tests in accordance with the European Respiratory Society PCD diagnostic guidelines and (2) original clinical outcome determined by all available diagnostic tests.

RESULTS: HSVA had excellent sensitivity and specificity to diagnose PCD: (1) 100% and 96%, respectively, compared with ERS guidelines, and (2) 96% and 91% compared with diagnostic outcomes. There was high interobserver agreement for "PCD-positive" outcomes ($\kappa = 0.7$).

CONCLUSIONS: Specialist scientists accurately diagnosed PCD using HSVA, with high interobserver agreement. HSVA can be reliably used to counsel patients and commence treatment on the day of testing while awaiting confirmatory investigations.

CHEST 2019; ■(■):■-■

KEY WORDS: accuracy; diagnostic tests; microscopy; primary ciliary dyskinesia; sensitivity and specificity

Q8

ABBREVIATIONS: ALI = air-liquid interface; CBP = ciliary beat pattern; ERS = European Respiratory Society; HSVA = high-speed video microscopy analysis; MDT = multidisciplinary team; nNO = nasal nitric oxide; PCD = primary ciliary dyskinesia; TEM = transmission electron microscopy

AFFILIATIONS: From the Primary Ciliary Dyskinesia Centre (Ms Rubbo, Dr Jackson, Mr Thompson, and Dr Lucas), NIHR Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK; the Academic Unit of Clinical and Experimental Medicine (Ms Rubbo, Dr Jackson, Mr Thompson, and Dr Lucas), University of Southampton Faculty of Medicine, Southampton, UK; the Primary Ciliary Dyskinesia Centre (Dr Shoemark), Paediatric Respiratory Medicine, Royal Brompton Hospital, London UK; the Division of Molecular and Clinical Medicine (Dr Shoemark), Scottish Centre for Respiratory Research, University of Dundee, Dundee, UK; the Centre for PCD Diagnosis and Research (Dr Hirst, Mr Hayes, and Dr O'Callaghan), Department of Infection, Immunity and Inflammation, University of Leicester, and Robert Kilpatrick Clinical Sciences Building, Leicester Royal Infirmary, Leicester, UK; the Primary Ciliary Dyskinesia Centre (Ms Frost and Ms Hogg), Department of Paediatrics and Department of Paediatric Respiratory Medicine, Imperial College and Royal Brompton Hospital, London UK; the PCD Family Support Group (Ms Copeland), Milton Keynes, UK; the Infection, Immunity, Inflammation and Physiological Medicine Programme (Dr O'Callaghan), Institute of Child Health, University College London, London, UK; Research Design Service South Central (Dr Reading), National Institute for Health Research,

FLA 5.5.0 DTD ■ CHEST2171_proof ■ 22 March 2019 ■ 8:21 pm ■ EO: CHEST-18-2329

Primary ciliary dyskinesia (PCD) is a rare (?1:10,000-20,000), heterogeneous disease, usually inherited as an autosomal recessive condition. Impaired function of motile cilia leads to neonatal respiratory distress in term infants, persistent wet cough, bronchiectasis, chronic rhinosinusitis, fertility issues, and conductive hearing impairment. Approximately 50% of patients have situs inversus, and congenital heart disease has been reported in 5% of children.[1]

In the absence of a single "gold standard" test, guidelines recommended that diagnosis requires access to a number of tests.[2,3] In our centers, a multidisciplinary team (MDT) of clinical and laboratory staff determines whether patients have PCD using clinical history, nasal nitric oxide (nNO), high-speed video microscopy analysis (HSVA), transmission electron microscopy (TEM), and more recently air-liquid interface (ALI) cell culture, immunofluorescence, and genetic analysis.[4] TEM can confirm a diagnosis, but is normal in 15% to 20% of patients with PCD and therefore cannot be used to exclude a diagnosis.[2,5,6] Similarly, poor sensitivity (?0.65) means that genotyping cannot be used in isolation, but pathogenic biallelic mutations in known genes confirm a diagnosis.[7,8]

HSVA is a technique where the respiratory cilia are visualized ex vivo with a light microscope, and recorded with a high-speed video camera. Videos are assessed for multiple parameters including ciliary beat frequency, ciliary beat pattern (CBP), and effective mucociliary clearance. HSVA is the only widely used test that assesses ciliary function, and results are available on the day of testing. In comparison, TEM and genetic analyses may take weeks or even months to get a definitive result. HSVA is used frequently at European and Australian PCD centers, but less so in North America.[2,3] Previous retrospective studies have suggested high sensitivity and specificity of HVSA as a diagnostic test[9,10]; however, both studies risked bias due to study design.[11] In addition, there has never been a study to assess the intra- and interobserver agreement of HSVA. If confirmed to be accurate, with good reliability, clinicians could make informed decisions on whether to initiate treatment on the day of a patient's clinic appointment whilst awaiting TEM and genetics results, reducing time to diagnosis and potentially limiting disease progression.

We hypothesized that (1) scientists using HSVA would accurately diagnose PCD and (2) there would be good interobserver reliability of the test.

Materials and Methods

Local and national research and ethics approvals were adhered to (Southampton and South West Hampshire Research Ethics 07/Q1702/109).

Patient Population and Diagnostic Decisions in the Clinical Setting

Patients were referred to one of three UK PCD diagnostic centers between January 2015 and April 2017. Testing included a combination of clinical history, nNO, HSVA, and TEM. With selected cases, we additionally included reanalysis following ALI culture and immunofluorescence staining; genetic testing was conducted on selected patients for research. For HSVA, diagnostic scientists report the sample to be compatible with PCD, unlikely to be PCD, or inconclusive; they base this decision on analysis of at least six videos from the same sample, including five side views and one top view. Investigations are detailed in the online article. Teams from the three centers share diagnostic protocols and frequently discuss difficult cases.

Diagnostic results were reviewed at MDT meetings, including a clinician, HSV microscopist, and TEM technician. All clinical and diagnostic data were considered when agreeing on the MDT diagnostic outcome as "PCD positive," "PCD highly likely," "PCD highly unlikely," or "inconclusive," based on clinical experience. An inconclusive diagnosis was reported when abnormalities not attributed to secondary defects were seen after repeated testing of adequate samples, but not sufficiently or consistently throughout the repeat testing to be deemed "PCD highly likely," or when further testing was still needed to rule in or rule out a PCD diagnosis.

Selection of Reference Standards

There is no "gold standard" reference for PCD diagnostics, and we therefore compared the scientists' study outcomes with two imperfect references[12]: (*a*) outcomes defined using European Respiratory Society (ERS) guidelines for the diagnosis of PCD[2] (Fig 1A); and (*b*) the clinical MDT outcome for the patient, extracted from contemporary MDT meeting reports (Fig 1B). For reference *a*, diagnostic test results were retrospectively used to define the patient outcome as "PCD positive" or "PCD highly unlikely." Both "PCD highly likely" and "inconclusive" outcomes were considered as indeterminate for accuracy calculations, as they do not provide a

and University Hospital Southampton NHS Foundation Trust, Southampton, UK.

FUNDING/SUPPORT: This report is independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference No. PB-PG-1215-20014). The views expressed are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health. The PCD Diagnostic Centres in Southampton, Leicester, and London are funded by NHS England.

DISCLAIMER: We shall make study data available to the scientific community with as few restrictions as feasible, while retaining exclusive use until the publication of major outputs. Anonymized study data will be available from the corresponding author. Ethical approvals for the use of the video archive currently restrict access to the UK PCD Centres; J. S. L., C. O'C., and C. H. are custodians of these clinical data.

CORRESPONDENCE TO: Jane Lucas, MD, PhD, Southampton University Hospital, Mailpoint 803 F level, Tremona Rd, Southampton SO16 6YD, UK; e-mail: jlucas1@soton.ac.uk

DOI: https://doi.org/10.1016/j.chest.2019.01.036

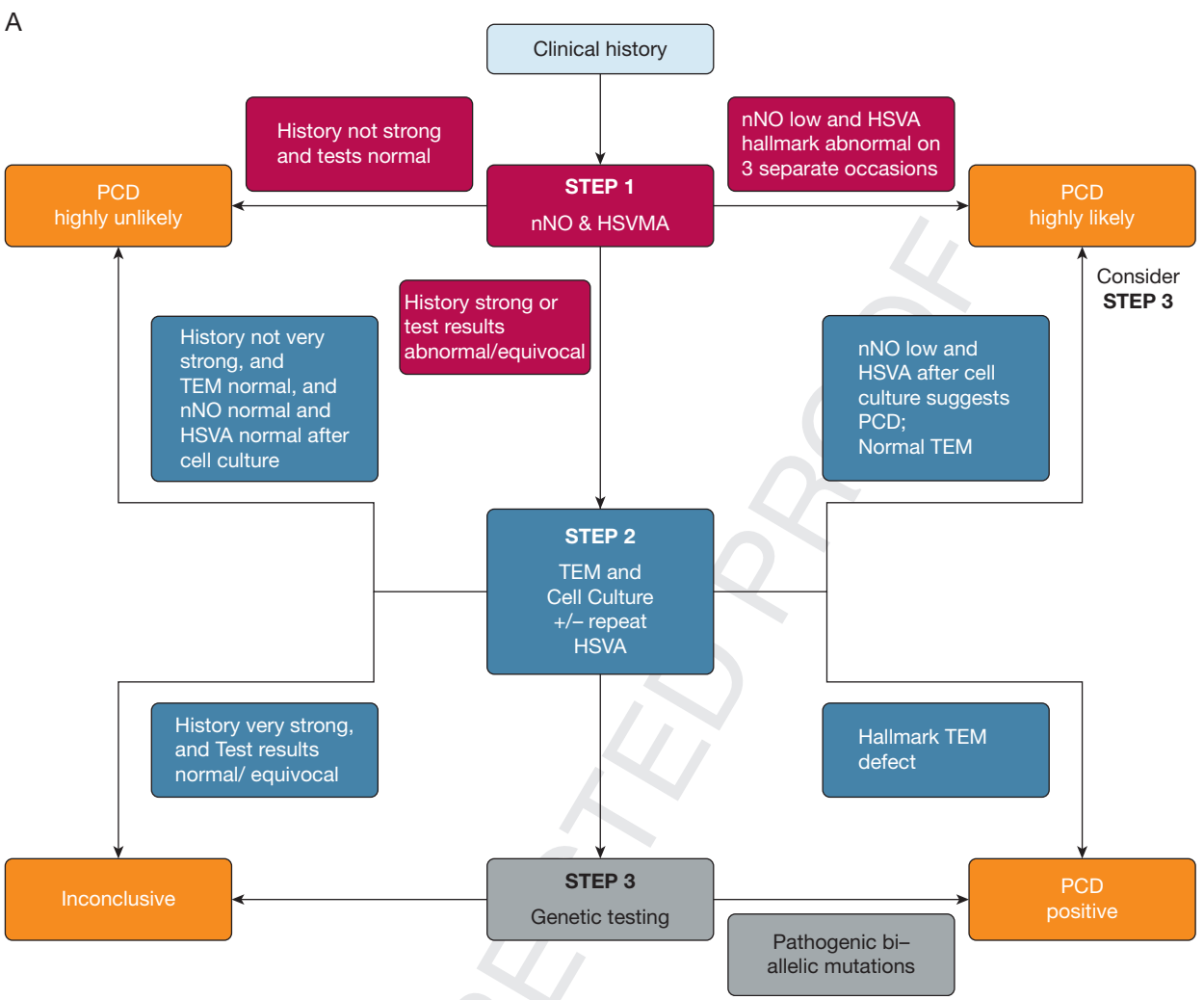

Figure 1 – *The two reference standards used to assess the accuracy of high-speed video microscopy (HSVA) analysis. A, ERS guidelines recommended a standardized terminology to describe the diagnostic outcome.*[2] *If the patient has a hallmark transmission electron microscopy (TEM) defect or biallelic or X-linked causative mutation in a primary ciliary dyskinesia (PCD) gene, they are diagnosed as "PCD positive." If nasal nitric oxide (nNO) is repeatedly low and hallmark HSVA alterations are found on three separate occasions or after air-liquid interface (ALI) cell culture, but TEM and genetics testing are normal, the patient is labeled as "PCD highly likely"; PCD is considered "highly unlikely" if all tests produce normal results. Despite repeated testing, the diagnostic category for a proportion of patients remains inconclusive; these were patients who did not satisfy the criteria to be included in any of the other categories. B, In the clinical setting, all available clinical and diagnostic data were reviewed at a multidisciplinary meeting to define the diagnostic outcome, based on the opinion of the expert team. ERS = European Respiratory Society.* Q27 Q17

definitive outcome.[2] Patients with diagnostic test results that did not fulfill criteria for "PCD positive," "PCD highly likely," or "PCD highly unlikely" were deemed "inconclusive." The strength of using this reference is that it follows an evidence-based international guideline, and that the "PCD positive" outcome is based only on "hallmark" TEM and/or pathogenic biallelic mutations in PCD genes, and therefore does not include HSVA in the reference standard. TEM and genetics are believed to have excellent specificity (?1.0), but the limitation is that both tests have poor sensitivity (0.8 and 0.7, respectively) and will therefore "miss" a significant proportion of patients with true PCD. Moreover, genotyping was undertaken only in a small subset of patients, as it is not readily available in the English National Health Service (NHS).

For reference *b*, diagnostic outcomes were extracted from the contemporary clinical MDT meeting reports. The strength of using this reference is that it was based on all data available to an expert MDT at the time of the meeting and it represents a clinical decision on how to manage patients; however, the limitation is that HSVA is included in the reference.

Analysis of Archived Videos

One hundred and twenty patients were randomly selected for inclusion in the study. Inclusion and exclusion criteria for sample selection are detailed in the online article.

Clinical data were extracted from local clinical databases: clinical symptoms, nNO results, TEM, genetic analysis (where available), and final diagnostic outcome by MDT decision. Images were anonymized and uploaded to a central platform. The HSVA scientists were not aware of the study period and were not involved in data extraction or uploading.

Three scientists, each with over 8 years of experience in HSVA, one from each UK PCD diagnostic center, independently viewed 720 videos from 120 anonymized patient samples (six videos per sample,

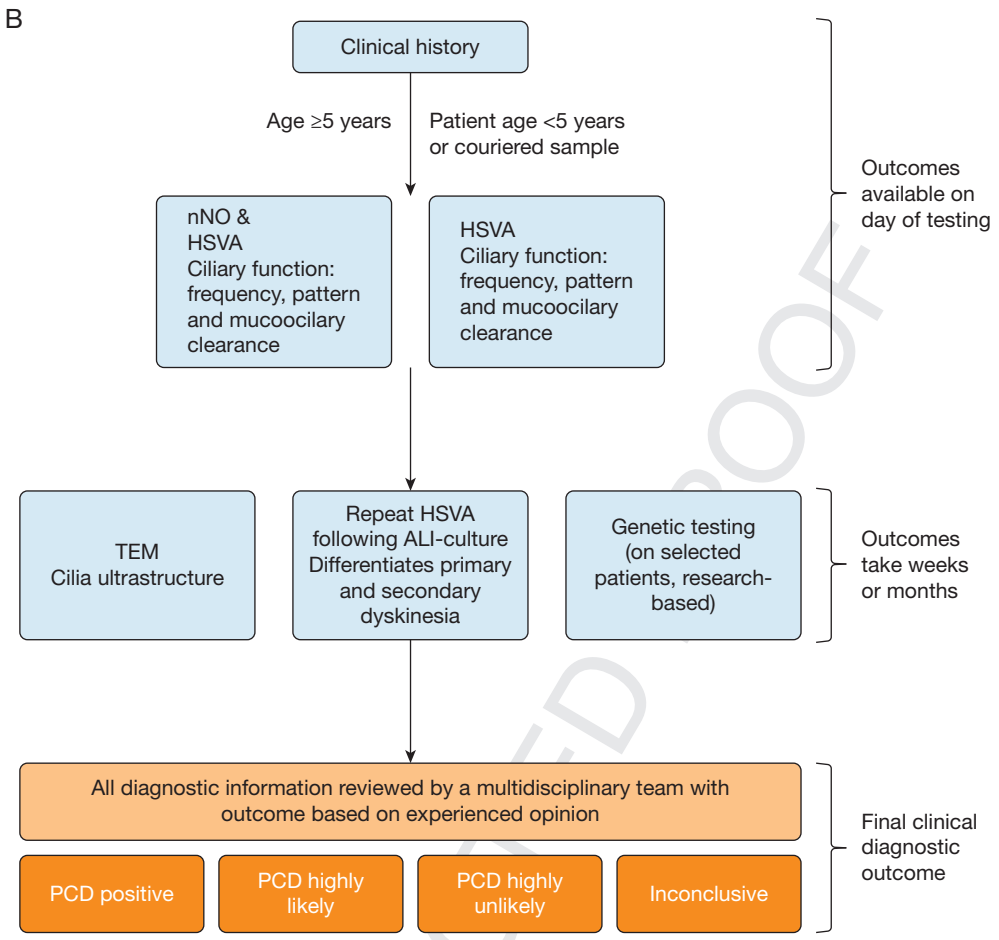

Figure 1 – *Continued*

according to the UK standard diagnostic protocol). Scientists scored the collection of six videos derived from each sample, blinded to other clinical or diagnostic data, to provide an a priori study outcome for each patient sample: "PCD positive," "PCD highly likely," "PCD highly unlikely," or "inconclusive," based on qualitative assessment of CBP and observed normality and abnormality in the samples analyzed. To calculate the intraobserver agreement after 1 year, each of the three scientists independently, and blinded to their initial assessment, reassessed 20 patient samples that were randomly selected.[13] We applied the same proportions of positive, negative, and inconclusive cases used in the selection of the original study sample (ie, 50%, 30%, and 20%, respectively).

Statistical Analyses

We stratified the total number of patients referred to each center during the study period by their clinical diagnostic outcome, based on the MDT final report: PCD positive (included PCD highly likely cases), PCD highly unlikely, and inconclusive. We used disproportional sampling in order to enhance the proportion of PCD-positive cases and obtain sufficient data on subgroups of interest.[14,15] Therefore 50% of our total cohort were randomly sampled from the PCD positive or PCD highly likely strata, 30% from the PCD highly unlikely stratum, and 20% from the inconclusive stratum. The sample size needed to detect a sensitivity of 90% with ± 0.9% confidence intervals was 90 patient samples.

To allow for missing data and indeterminate outcomes we randomly selected 120 patients from each outcome stratum: 59 "PCD positive," 36 "PCD highly unlikely," and 25 "inconclusive." Randomization for each stratum was performed in STATA (StataCorp).

To calculate the accuracy of HSVA, we compared the outcomes by each of the scientists with the patient reference outcome, using reference *a* (the ERS guidelines) and reference *b* (the original MDT report). For reference *a*, we defined true positive as ["PCD positive" by scientist] divided by ["PCD positive" by reference]. Similarly, true negative was defined as ["PCD highly unlikely" by scientist] divided by ["PCD highly unlikely" by reference]. For reference *b*, we grouped "PCD positive" and "PCD highly likely" outcomes, since these are clinically managed similarly, and the "PCD highly likely" group is likely to include patients with true PCD with normal TEM where the genotype has not yet been resolved (Fig 2). True positive was therefore defined as ["PCD positive" or "PCD highly likely" by scientist] divided by ["PCD positive" or "PCD highly likely" by MDT decision]. True negatives were defined as described for reference *a*. For both references, false positive or false negative was determined when HSVA scientists did not agree with reference.

We calculated the interobserver repeatability using the Fleiss κ coefficient for each diagnostic outcome.[16] We calculated the intraobserver repeatability for each of the scientists using the Cohen κ coefficient, with bootstrapped confidence intervals (n = 5).[17]

Data were analyzed in STATA version 14.0. Continuous variables are presented as median and interquartile range (IQR), and categorical variables are reported as proportions. Sensitivity and specificity are presented with 95% CIs, where appropriate. We report on both

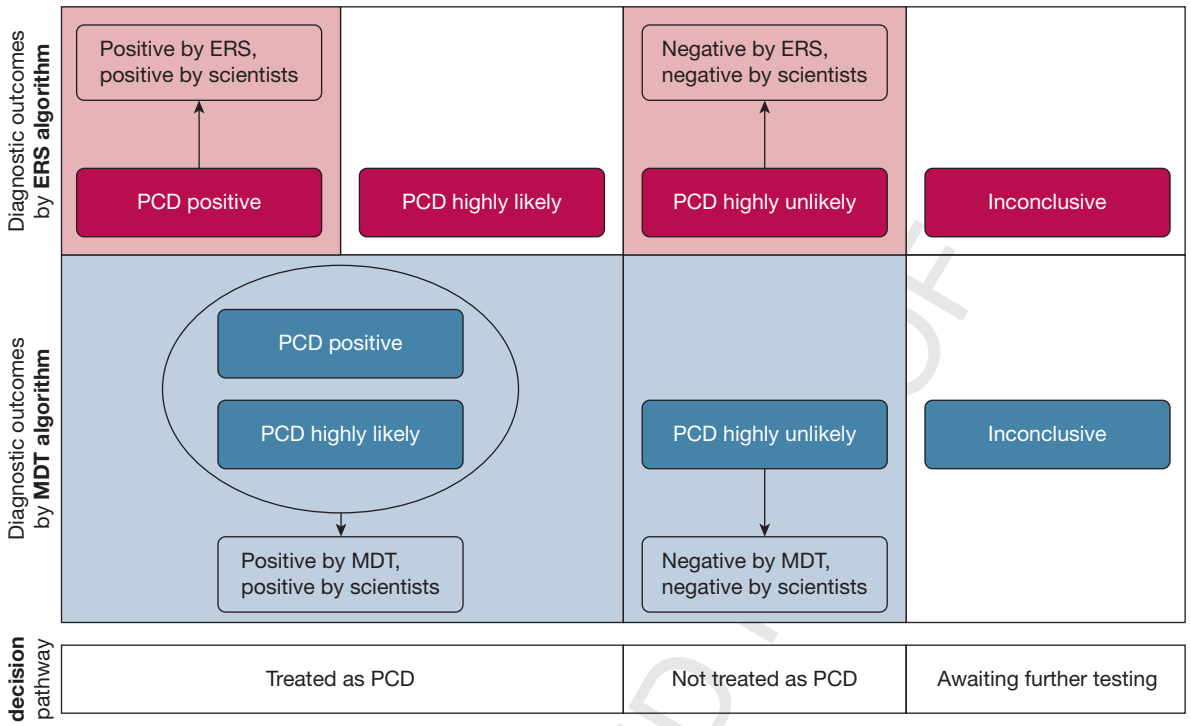

Figure 2 – *Diagnostic algorithms used to determine true positive and true negative outcomes for test accuracy calculations, according to the ERS guidelines (top half) and the clinical diagnostic decision by the multidisciplinary team (MDT) (bottom half). PCD = primary ciliary dyskinesia. See Figure 1 legend for expansion of other abbreviation.*

aggregate and individual (ie, each scientist) sensitivity and specificity of HSVA study outcomes compared with both reference standards. We obtained three outcomes for each sample: one from each scientist. To adjust for clustering of data and to provide robust confidence intervals, we used a generalized estimating equation (GEE) model when reporting on all aggregate diagnostic outcomes.[18]

To deal with "inconclusive" study outcomes, test accuracy was also calculated using the "worst-case scenario" approach, where "inconclusive" were recoded as either "false positives" or "false negatives" and adjusted for clustering using GEE modeling. Results are reported according to the STARD (Standards for the Reporting of Diagnostic Accuracy Studies) 2015 guidelines.[19]

Results

The three diagnostic centers received a total of 1,286 referrals from January 2015 to April 2017; 115 were PCD positive after review by the MDT, 852 were negative, and 305 were inconclusive. Thirteen nasal brushing samples were deemed insufficient for analysis. Characteristics of the patients whose videos were randomly selected for the study are outlined in Table 1. Clinical characteristics extracted were based on PICADAR, a PCD-specific diagnostic predictive tool.[20] Genetic results were available for 16 patients, of whom eight showed biallelic pathogenic mutations in a PCD-causative gene (three in *DNAH5*, two in *DNAH11*, two in *CCDC40*, one in *RSPH9*) and one in an X-linked PCD gene (*OFD1*).

Accuracy of HSVA Compared With the ERS-Defined Outcomes (Reference a)

Using the ERS PCD diagnostic guidelines, 36 patient samples were "PCD positive," 16 were "PCD highly likely," 26 were "PCD highly unlikely," and 42 were "inconclusive" (e-Table 2).

There was excellent sensitivity (100%) and specificity (96.2%; 95% CI, 91.7%-100%) when comparing the study decisions of HSVA scientists with the diagnostic outcome based on outcomes defined by the ERS PCD guidelines (Table 2). Specificity results were adjusted for clustering; however, it was not possible to adjust sensitivity as there were no "false negatives" observed. A "worst-case scenario" combined with GEE modeling showed that sensitivity remained high (93.3%; 95% CI, 92.0%-100%) but that specificity decreased from 96.2% to 67.9% (95% CI, 58.7%-77.2%).

Individual scientists had similarly good accuracy (e-Table 2). A proportion of samples was reported as "highly likely" or "inconclusive" when using either study HSVA results alone or the ERS guidelines, and these outcomes could not be included in the accuracy calculations.

TABLE 1] Clinical Characteristics of Study Participants Stratified by Diagnostic Outcome According to European Respiratory Society Guidelines

Characteristic	All Patients (n = 120)	PCD Positive (n = 36)	PCD Highly Likely (n = 16)	PCD Highly Unlikely (n = 26)	Inconclusive (n = 42)
Center for diagnostic tests					
UHS	40 (33.3%)	11 (30.6%)	3 (18.8%)	14 (53.9%)	12 (28.6%)
RBH	40 (33.3%)	12 (33.3%)	3 (18.8%)	3 (11.5%)	22 (52.4%)
LRI	40 (33.3%)	13 (36.1%)	10 (60.5%)	9 (34.6%)	8 (19.1%)
Age, y (median, IQR)	9.6 (2.8-16.7)	9.1 (3.0-20.9)	11.8 (8.9-12.6)	10 (2.0-29.5)	7.3 (2.9-14.8)
Preterm gestation	9 (8.9%)	0	3 (23.1%)	3 (14.3%)	3 (8.1%)
Chest symptoms in neonatal period	97 (82.9%)	26 (78.8%)	15 (93.8%)	18 (69.2%)	38 (90.5%)
Admission to neonatal unit	45 (41.3%)	17 (53.1%)	11 (78.6%)	7 (26.9%)	10 (27.0%)
Presence of situs abnormalities	22 (18.6%)	16 (45.7%)	3 (18.8%)	0	3 (7.3%)
Cardiac abnormality	5 (4.3%)	1 (2.9%)	0	3 (11.5%)	1 (2.5%)
Persistent perennial rhinitis	85 (72%)	28 (80.0%)	14 (93.3%)	13 (50.0%)	30 (71.4%)
Chronic ear or hearing symptoms	70 (60.3%)	20 (57.1%)	13 (86.7%)	13 (50.0%)	24 (60.0%)
nNO, nL/min, median (IQR); No. for whom data available	21.8 (7.2-105.0); n = 72	9.8 (4.8-15.9); n = 22	7.2 (3.0-63.6); n = 11	189.2 (69.2-218.0); n = 11	72.3 (19.9-117.8); n = 28
TEM results					
Normal	63 (52.5%)	2 (5.6%)[a]	7 (43.8%)	19 (73.1%)	35 (83.3%)
ODA alone	14 (11.7%)	13 (36.1%)	1 (6.25%)[b]	0	0
ODA + IDA	14 (11.7%)	14 (38.9%)	0	0	0
IDA alone	4 (3.3%)	0	4 (25.0%)	0	0
MTD + IDA	5 (4.2%)	5 (13.9%)	0	0	0
CC	5 (4.2%)	1 (2.8%)[c]	4 (25.0%)	0	0
Lack of cilia	2 (1.7%)	0	0	0	2 (4.8%)
Inconclusive	3 (2.5%)	1 (2.8%)[d]	0	0	2 (4.8%)
Not done	10 (8.3%)	0	0	7 (26.9%)	3 (7.1%)

CC = central complex defect; ERS = European Respiratory Society; IDA = inner dynein arm defect; IQR = interquartile range; LRI = Leicester Royal Infirmary; MTD = microtubular disarrangement; nNO = nasal nitric oxide; ODA = outer dynein arm defect; PCD = primary ciliary dyskinesia; RBH = Royal Brompton Hospital in London; TEM = transmission electron microscopy; UHS = University Hospital Southampton.

[a]Biallelic mutations in the *DNAH11* gene.

[b]TEM abnormality described as "thin ODA present," not a hallmark PCD defect according to the ERS guidelines.

[c]Biallelic mutations in the *RSPH9* gene.

[d]X-linked mutation in the *OFD1* gene.

Accuracy of HSVA Compared With MDT Decision (reference b)

Using the MDT diagnostic outcome as the reference standard, 59 patients were "PCD positive," 36 "PCD highly unlikely," and 25 had inconclusive test results (e-Table 3). There was excellent sensitivity (96.7%; 95% CI, 92.9%-100%) and specificity (91.1%; 95% CI, 85.3%-96.9%) of study HSVA analysis compared with the original MDT diagnostic outcome (Table 3). Sensitivity dropped to 85.3% (95% CI, 78.0%-92.6%) and specificity to 67.6% (95% CI, 58.4%-76.8%) when calculating accuracy using the "worse-case" approach. Individual scientist sensitivity ranged from 95.9% to 100% and specificity from 66.7% to 100% (e-Table 3).

FLA 5.5.0 DTD ■ CHEST2171_proof ■ 22 March 2019 ■ 8:21 pm ■ EO: CHEST-18-2329

Twenty-five cases remained "inconclusive" after review by MDT (Table 3). These were difficult clinical diagnostic cases that required further brushing and/or additional diagnostic testing. The scientists reported a similar number of samples as inconclusive (mean, 28 samples; range, 21-33) despite the fact that they had to rely on HSVA images alone while the MDT had the full range of clinical and diagnostic information at their disposal (e-Table 3).

Two cases were classified as "PCD highly likely" by both ERS guidelines and the MDT, but either "PCD highly unlikely" or "inconclusive" by the HSVA scientists (e-Table 3). The original clinical records indicated that one patient had an isolated inner dynein arm defect on TEM (ie, not a hallmark abnormality) and five repeat brushings. Ciliary beat frequency varied between low and normal on different occasions and CBP was described as "almost normal" in most brushing samples, some with observed mucociliary clearance. Two of the HSVA scientists classified this sample as "PCD highly unlikely" and one deemed it "inconclusive." The second patient had normal nNO, TEM, and genetics for known causative genes but was diagnosed as "PCD highly likely" based on "semirotating" CBP coupled with the observation of similar clinical symptoms and HSVA findings in the patient's sibling diagnosed with PCD. Two scientists classified this sample as "highly unlikely," while one said it was "inconclusive." Both patients are currently treated as having PCD (ie, receiving care by the PCD teams) but require further diagnostic testing before a definite diagnostic outcome can be determined.

Intra- and Interobserver Reliability

Using Fleiss κ agreement to compare scoring between the three scientists for each diagnostic outcome, we found substantial agreement ($\kappa = 0.70$) for "PCD positive" and moderate agreement ($\kappa = 0.44$) for "PCD highly unlikely." Agreement was low for "PCD highly likely" ($\kappa = 0.11$) and "inconclusive" ($\kappa = 0.20$).[21] The combined agreement for the overall diagnostic outcomes was moderate ($\kappa = 0.42$; 95% CI, 0.41-0.44).

The Cohen κ agreement for intraobserver reliability was $\kappa = 0.70$ (95% CI, 0.56-0.77) for scientist 1, $\kappa = 0.66$ (95% CI, 0.42-0.75) for scientist 2, and $\kappa = 0.78$ (95% CI, 0.61-0.85) for scientist 3. Importantly, none of the scientists changed the outcome from their original assessment from positive to negative or from negative to positive (e-Table 4).

Discussion

We have shown that HSVA has excellent accuracy and interobserver reliability for diagnosing PCD, when conducted by experienced scientists.

Accuracy of HSVA to Diagnose PCD

HSVA had excellent sensitivity and specificity to diagnose PCD. With lack of a "gold standard" reference, we used two imperfect references and found that sensitivity and specificity were 100% and 96%, respectively, when using diagnosis based on the ERS guidelines as a reference, and 96% and 91% when using the clinical diagnostic outcome as standard.

Independently analyzing 720 videos from 120 patients, HSVA scientists correctly identified all "PCD positive" cases using the ERS PCD guidelines as reference. Considering that these patients have either a hallmark TEM or pathogenic mutations, our findings suggest that HSVA approaches 100% accuracy to detect clear-cut PCD cases. If we were to consider those with an ERS-defined "highly likely diagnosis" (ie, lack of hallmark TEM or genetic confirmation but at least three HSVA abnormal results or two abnormal results plus abnormal ALI cell culture) as true PCD cases, we increase the detection rate by 15% in our study population. This increase matches the 15% to 20% PCD cases without a hallmark TEM defect reported in the literature, suggesting that HSVA can pick up cases that might have been otherwise "missed" by TEM, particularly if used in combination with nNO.[2]

Scientists reported two study samples as "highly unlikely" or "inconclusive," whereas both MDT and ERS guidelines had deemed the diagnostic outcome of the patients as "PCD highly likely." On further reviewing the diagnostic history of these patients, the clinical decisions were based on extensive repeat testing coupled with strong clinical and family histories, highlighting the complexity of some cases. Experts agree that some subtle beat pattern abnormalities are difficult to spot by HSVA, even with extensive training and years of experience.[11] In addition, secondary abnormities are common even in samples from healthy individuals, highlighting the need for experienced personnel analyzing the whole cilia strip to focus on the overall findings.[11,22-24] It is therefore not surprising that in our study population, a high proportion of patients had indeterminate outcomes according to both ERS guidelines (35%) and MDT decisions (21%). This was also reflected in the number of "inconclusive" outcomes by the scientists (23%). Sensitivity remained high even after reclassification of

FLA 5.5.0 DTD ■ CHEST2171_proof ■ 22 March 2019 ■ 8:21 pm ■ EO: CHEST-18-2329

TABLE 2] Aggregated Diagnostic Study Outcomes by Three Scientists Compared With Diagnostic Outcome Defined by ERS PCD Diagnostic Guidelines[a]

Study Outcomes by HSVA Scientists	Diagnostic Outcomes Based on ERS Guidelines				
	PCD Positive[b]	PCD Highly Unlikely	PCD Highly Likely	Inconclusive	Total
PCD Positive	**94**	**2**	25	13	134
PCD Highly Unlikely	**0**	**53**	4	42	99
PCD Highly Likely	10	4	11	17	42
Inconclusive	4	19	8	54	85
Total:	108	78	48	126	360
No. of Samples:	n = 36	n = 26	n = 16	n = 42	n = 120

HSVA = high-speed video microscopy analysis. See Table 1 legend for expansion of other abbreviations.
[a]See Lucas et al[2]; n = 360 scientists' outcomes from 120 patient samples. "PCD positive" and "PCD highly unlikely" outcomes contributed to the accuracy analyses. Individual scientists' results are shown in e-Table 2.
[b]"PCD positive" cases were those with a hallmark transmission electron microscopy defect and/or genotype.

"inconclusive" by HSVA to false negative. The drop in specificity is likely because the scientists were less confident to rule out PCD based on HSVA alone. This is expected, as scientists would normally have additional information at their disposal, and clinical decisions on whether to treat patients are based on HSVA coupled with clinical and nNO data. While the "worst-case scenario" calculations are reassuring, reclassifying the inconclusive outcome was probably overconservative because "inconclusive" is a legitimate clinical outcome; it is difficult to consider "inconclusive" as false positive or false negative, particularly as the management pathway includes further investigations for inconclusive outcomes.[2]

Reliability of HSVA to Diagnose PCD

We found high interobserver agreement for "PCD positive" and moderate agreement for "PCD highly unlikely" outcomes, as well as between pairs of scientists (see the online article). "PCD highly likely" and "inconclusive" had low agreement; this was due to the interchangeability of these outcomes, as some scientists felt more confident in assigning a "highly likely" outcome while others adopted a more cautious option (ie, "inconclusive"). In practice, samples labeled as "highly likely" or "inconclusive" would both require a repeat brushing from the patient and further testing.

We also found substantial intraobserver agreement of samples reassessed by each of the scientists 1 year after the original study outcome description. The fact that the scientists were able to discriminate between positive and negative outcomes, and agree on these between each other and with their own initial assessment, is key as these two extreme outcomes lead to different clinical management plans. These demonstrate reliability amongst experienced scientists when using HSVA to diagnose PCD.

Implications to Diagnostics and Clinical Practice

Following current guidelines, nasal brushings are taken from every patient referred to a PCD diagnostic center with a strong suspicion of PCD (ie, suggestive clinical history). Samples can be evaluated by scientists

TABLE 3] Aggregated Diagnostic Study Outcomes by Three Scientists Compared With Original Diagnostic Decision Made by MDT

Study Outcomes by HSVA Scientists	Diagnostic Outcomes Based on Original Expert MDT Decision			
	PCD Positive[a]	PCD Highly Unlikely	Inconclusive	Total
PCD Positive[a]	**151**	**7**	18	176
PCD Highly Unlikely	**4**	**73**	22	99
Inconclusive	22	28	35	85
Total:	177	108	75	360
No. of Samples:	n = 59	n = 36	n = 25	n = 120

n = 360 scientists' outcomes from 120 patient samples. "Inconclusive" outcomes were excluded from the accuracy analyses. Individual scientists' results are shown in e-Table 3. MDT = multidisciplinary team. See Table 1 and 2 legends for expansion of other abbreviations.
[a]Includes both "PCD positive" and "PCD highly likely" outcomes.

FLA 5.5.0 DTD ■ CHEST2171_proof ■ 22 March 2019 ■ 8:21 pm ■ EO: CHEST-18-2329

experienced in HSVA on the day of testing. The nasal sample is also sent for TEM analysis, but processing and analyses take weeks. Our study demonstrates that specialist scientists can reliably use HVSA to diagnose some patients with PCD on the day of testing. This provides the necessary evidence to counsel patients and initiate lifelong treatment in a "one-stop clinic" with the proviso that the final diagnostic outcome might change once all test results are available. Additional tests such as TEM, immunofluorescence, and genetic analysis will still be needed to confirm the diagnosis[2] and for deeper phenotyping.[7,25,26] The diagnosis remains inconclusive for a high proportion of patients following isolated HSVA, and these would need to wait for further diagnostic results; it is notable that our study also demonstrates that many patients have an indeterminate outcome even following comprehensive testing, as expected and discussed in the ERS PCD diagnostic guidelines.[2]

Strengths and Limitations

This is the first blinded study to assess the accuracy and reliability of HSVA to diagnose PCD. Previous literature has called for standardized methodology and reporting of diagnostic testing in PCD, in particular for HSVA.[2,3,9,11,23] In our study, diagnostic outcomes were prospectively assigned by three experienced scientists. Diagnostic outcomes were agreed a priori by the three scientists and applied in a standardized manner when independently scoring the video images.

However, our study has limitations. There is no "gold standard" reference to diagnose PCD; so, despite the use of combination testing as reference, we might have missed "difficult to diagnose" PCD cases, likely classified in this study as "inconclusive" by both MDT and the ERS guidelines. A second limitation was the use of HSVA in both comparator and the MDT reference; therefore, in our comparison of HSVA with a positive diagnosis according to ERS guidelines, we excluded HSVA from the reference for sensitivity analyses as only hallmark TEM and/or pathogenic mutations define a positive diagnosis. We had limited genetic information available for samples included in our study, which might have confirmed some of the "highly likely" or "inconclusive" cases as PCD. Equally, some of the "highly likely PCD" patients might not have PCD. Although we have good standardization of methods and reporting in the UK, our protocols differ from those used in many centers (eg, some centers measure HSVA at room temperature while we analyze samples at 37°C).

The use of disproportionate sampling allowed for the selection of a higher proportion of positive cases without having to review an unmanageable number of samples; however, because of this approach, negative cases were proportionally underrepresented. Fleiss κ performs poorly when the marginal classification probabilities are either very small or very large, underestimating the strength of agreement.[27] In addition, κ results rely on arbitrary convention for what are considered substantial, moderate, and low agreements. Therefore, we included e-Tables 2-4 to provide data on individual scientist's performances.

The study scientists are highly experienced in conducting HSVA. Accuracy and interobserver reliability would probably be lower if conducted by less experienced scientists. Scientists potentially recognized cases, but this is unlikely due to high diagnostic throughput and videos originating from analyses conducted some time ago. While we have demonstrated that it is technically reliable to provide same-day provisional feedback based on HSVA, the feasibility of achieving this will depend on local resources.

In conclusion, we found that when following standardized protocols HSVA has excellent sensitivity and specificity to diagnose PCD. We found good agreement between scientists on "PCD positive" and "PCD highly unlikely" outcomes, confirming that HSVA is a reliable diagnostic test. There is now a need for international standardization of analysis and reporting of HSVA.

FLA 5.5.0 DTD ■ CHEST2171_proof ■ 22 March 2019 ■ 8:21 pm ■ EO: CHEST-18-2329

Acknowledgments

Author contributions: J. S. L. had the concept for the study and takes responsibility for the integrity of the study and data; I. R., B. R., J. S. L.: designed the study and secured funding; B. R.: led standardized clinical and diagnostic data collection; randomization, data management and statistical analyses; and drafted the manuscript; A. S.: led standardization of study HSVA; C. L. J., R. H., A. S.: study analyses of video images; F. C.: PPI; J. T. led standardization of video selection and quality assessment; B. R., J. T., J. H., E. F., C. O'C., C. H., J. S. L.: extraction of clinical and diagnostic data from clinical records; J. T., J. H., E. F., B. R.: preparation of video images for scientist's review; C. H., C. O'C., J. S. L.: led multidisciplinary team decisions at diagnostic centers.

Financial/nonfinancial disclosures: J. S. L. reports grants, personal fees, and nonfinancial support from Aerocrine/Circassia; grants and personal fees from Vertex; and grants from Parion, outside the submitted work. None declared (B. R., A. S., C. L. J., R. H., J. T., J. H., E. F., F. C., C. H., C. O'C., I. R.).

Role of sponsors: The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

Additional information: The e-Appendix, e-Figure, and e-Tables can be found in the Supplemental Materials section of the online article.

Other contributions: The authors thank all laboratory and clinical members of the national PCD Diagnostic Service at UHS, RBH, and LRI who have contributed to the standardized clinical protocols and the clinical diagnostic testing of the patients; Hannah Mitchison, University College London, for conducting the genetic tests as part of a research project (London Bloomsbury Research Ethics Committee; 08/H0713/82); and patients who engage with research to further the understanding of PCD.

References

1. Lucas JS, Alanin MC, Collins S, et al. Clinical care of children with primary ciliary dyskinesia. *Expert Rev Respir Med*. 2017;11(10):779-790.
2. Lucas JS, Barbato A, Collins SA, et al. European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. *Eur Respir J*. 2017;49(1):1601090.
3. Shapiro AJ, Davis SD, Polineni D, et al. Diagnosis of primary ciliary dyskinesia: an official American Thoracic Society clinical practice guideline. *Am J Respir Crit Care Med*. 2018;197(12):e24-e39.
4. Lucas JS, Burgess A, Mitchison HM, Moya E, Williamson M, Hogg C. National PCD Service, UK. Diagnosis and management of primary ciliary dyskinesia. *Arch Dis Child*. 2014;99(9):850-856.
5. Kuehni CE, Lucas JS. Diagnosis of primary ciliary dyskinesia: summary of the ERS Task Force report. *Breathe (Sheff)*. 2017;13(3):166-178.
6. Schwabe GC, Hoffmann K, Loges NT, et al. Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by *DNAH11* mutations. *Hum Mutat*. 2008;29(2):289-298.
7. Davis SD, Ferkol TW, Rosenfeld M, et al. Clinical features of childhood primary ciliary dyskinesia by genotype and ultrastructural phenotype. *Am J Respir Crit Care Med*. 2015;191(3):316-324.
8. Marshall CR, Scherer SW, Zariwala MA, et al. Whole-exome sequencing and targeted copy number analysis in primary ciliary dyskinesia. *G3 (Bethesda)*. 2015;5(8):1775-1781.
9. Jackson CL, Behan L, Collins SA, et al. Accuracy of diagnostic testing in primary ciliary dyskinesia. *Eur Respir J*. 2016;47(3):837-848.
10. Papon JF, Bassinet L, Cariou-Patron G, et al. Quantitative analysis of ciliary beating in primary ciliary dyskinesia: a pilot study. *Orphanet J Rare Dis*. 2012;7:78.
11. Lucas JS, Evans HJ, Haarman EG, et al. Exploring the art of ciliary beating: the benefits of high-speed video analysis. *Chest*. 2017;152(6):1348-1349.
12. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard: a review of methods. *Health Technol Assess*. 2007;11(50):iii, ix-51.
13. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101-110.
14. Daniel J. *Sampling Essentials: Practical Guidelines for Making Sampling Choices [Internet]*. Thousand Oaks, CA: SAGE Publications; 2012. http://methods.sagepub.com/book/sampling-essentials. Accessed March 5, 2019.
15. Fox N, Hunn A, Mathers N. *Sampling and Sample Size Calculation [Internet]*. Sheffield, UK: NIHR Research Design Service for East Midlands, Yorkshire, and the Humber; 2007. https://www.coursehero.com/file/22207344/Sampling-and-sample-size-calculation/. Accessed April 5, 2018.
16. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76(5):378-382.
17. Reichenheim ME. Confidence intervals for the κ statistic. *Stata J*. 2004;4(4):421-428.
18. Genders TS, Spronk S, Stijnen T, Steyerberg EW, Lesaffre E, Hunink MG. Methods for calculating sensitivity and specificity of clustered data: a tutorial. *Radiology*. 2012;265(3):910-916.
19. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.
20. Behan L, Dimitrov BD, Kuehni CE, et al. PICADAR: a diagnostic predictive tool for primary ciliary dyskinesia. *Eur Respir J*. 2016;47(4):1103-1112.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
22. Hirst RA, Jackson CL, Coles JL, et al. Culture of primary ciliary dyskinesia epithelial cells at air-liquid interface can alter ciliary phenotype but remains a robust and informative diagnostic aid. *PLoS One*. 2014;9(2):e89675.
23. Kempeneers C, Seaton C, Chilvers MA. Variation of ciliary beat pattern in three different beating planes in healthy subjects. *Chest*. 2017;151(5):993-1001.
24. Peabody JE, Shei R-J, Bermingham BM, et al. Seeing cilia: imaging modalities for ciliary motion and clinical connections. *Am J Physiol Lung Cell Mol Physiol*. 2018;314(6):L909-L921.
25. Shah A, Shoemark A, MacNeill SJ, et al. A longitudinal study characterising a large adult primary ciliary dyskinesia population. *Eur Respir J*. 2016;48(2):441-450.
26. Noone PG, Leigh MW, Sannuti A, et al. Primary ciliary dyskinesia: diagnostic and phenotypic features. *Am J Respir Crit Care Med*. 2004;169(4):459-467.
27. Feinstein AR, Cicchetti DV. High agreement but low κ. I. The problems of two paradoxes. *J Clin Epidemiol*. 1990;43(6):543-549.