# Title page

**Title:** Accuracy of high-speed video analysis to diagnose primary ciliary dyskinesia

**Authors:** Bruna Rubbo, Amelia Shoemark, Claire L Jackson, Robert Hirst, James Thompson, Joseph Hayes, Emily Frost, Fiona Copeland, Claire Hogg, Christopher O’Callaghan, Isabel Reading, Jane S Lucas; on behalf of the National PCD Service, UK.

**Affiliations:**

**Bruna Rubbo MSc**  B.Rubbo@soton.ac.uk [1] Primary Ciliary Dyskinesia Centre, NIHR Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK [2]University of Southampton Faculty of Medicine, Academic Unit of Clinical and Experimental Medicine, Southampton, UK

**Amelia Shoemark PhD** a.shoemark@dundee.ac.uk **[1]** Primary Ciliary Dyskinesia Centre, Paediatric Respiratory Medicine, Royal Brompton Hospital, London UK [2] Scottish Centre for Respiratory Research, Division of Molecular and Clinical Medicine, University of Dundee, Dundee, UK

**Claire L Jackson PhD**  clj@soton.ac.uk [1] Primary Ciliary Dyskinesia Centre, NIHR Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK [2]University of Southampton Faculty of Medicine, Academic Unit of Clinical and Experimental Medicine, Southampton, UK

**Robert A Hirst** **PhD** rah9@leicester.ac.uk Centre for PCD Diagnosis and Research, Department of Infection, Immunity and Inflammation, University of Leicester, Robert Kilpatrick Clinical Sciences Building, Leicester Royal Infirmary, Leicester, UK

**James Thompson** jt3v12@soton.ac.uk [1] Primary Ciliary Dyskinesia Centre, NIHR Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK [2]University of Southampton Faculty of Medicine, Academic Unit of Clinical and Experimental Medicine, Southampton, UK

**Joseph Hayes BSc** joh9@le.ac.uk Centre for PCD Diagnosis and Research, Department of Infection, Immunity and Inflammation, University of Leicester, Robert Kilpatrick Clinical Sciences Building, Leicester Royal Infirmary, Leicester, UK

**Emily Frost MSc** emilyhowieson23@gmail.com Primary Ciliary Dyskinesia Centre, Departments of Paediatrics and Paediatric Respiratory Medicine, Imperial College and Royal Brompton Hospital, London UK

**Fiona Copeland** chair@pcdsupport.org.uk PCD Family Support Group, Milton Keynes, UK

**Claire Hogg MBChB** C.Hogg@rbht.nhs.uk Primary Ciliary Dyskinesia Centre, Departments of Paediatrics and Paediatric Respiratory Medicine, Imperial College and Royal Brompton Hospital, London UK

**Christopher O’Callaghan MD PhD** c.ocallaghan@ucl.ac.uk [1] Centre for PCD Diagnosis and Research, Department of Infection, Immunity and Inflammation, University of Leicester, Robert Kilpatrick Clinical Sciences Building, Leicester Royal Infirmary, Leicester, UK [2] Infection, Immunity, Inflammation and Physiological Medicine, Institute of Child Health, University College London WC1N 1EH, UK

**Isabel Reading PhD** I.C.Reading@soton.ac.uk National Institute of Health Research Design Service South Central, University Hospital Southampton NHS Foundation Trust, Southampton, UK

**Jane S Lucas MD PhD** jlucas1@soton.ac.uk [1] Primary Ciliary Dyskinesia Centre, NIHR Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK [2]University of Southampton Faculty of Medicine, Academic Unit of Clinical and Experimental Medicine, Southampton, UK

**Correspondence:** Jane Lucas, Southampton University Hospital, Mailpoint 803 F level, Tremona Road, Southampton, SO16 6YD, UK. E-mail: jlucas1@soton.ac.uk

**Funding:**

This report is independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit RfPB) Programme (Grant Reference Number 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.

**Data Sharing Statement**

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; JSL, CO’C and CH are custodians of this clinical data.

**Conflict of interest declaration**

JSL reports grants, personal fees and non-financial support from Aerocrine/ Circassia, grants and personal fees from Vertex, grants from Parion, outside the submitted work. All other authors have declared no competing interest.

**Short running head:** Accuracy of High-speed Video in PCD (current 35, max 50)

**Word Count : 3682**

This article has supplementary material available.

# Abstract

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 analysed 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 to combination of tests, and assess the inter-observer reliability of HSVA analysis.

Methods: We randomly selected and anonymised archived videos from 120 patients seen at three UK PCD centres. Three experienced scientists independently reviewed six videos per patient using a standardised proforma, blinded to diagnostic and clinical data. We compared study outcomes to two references: (a) combination of diagnostic tests following the ERS PCD diagnostic guidelines and (b) original clinical outcome determined by all available diagnostic tests.

Results: HSVA had excellent sensitivity and specificity to diagnose PCD; (a) 100% & 96% compared to ERS guidelines, and (b) 96% & 91% compared to diagnostic outcomes. There was high inter-observer agreement for ‘PCD positive’ outcomes (κ=0.7).

Conclusions: Specialist scientists accurately diagnosed PCD using HSVA, with high inter-observer agreement. HSVA can be reliably used to counsel patients and commence treatment on the day of testing whilst awaiting confirmatory investigations.

**Abstract Word Count**= 213

**Keywords:** primary ciliary dyskinesia; sensitivity and specificity; diagnostic tests; accuracy; microscopy

# Introduction

Primary ciliary dyskinesia (PCD) is a rare (≈1:10-20,000), 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 centres, 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 (IF) and genetic analysis [4]. TEM can confirm a diagnosis, but is normal in 15-20% 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 bi-allelic mutations in known genes confirms a diagnosis [7, 8].

HSVA is a technique where the respiratory cilia are visualised ex-vivo with a light microscope, and recorded using a high-speed video camera. Videos are assessed for multiple parameters including ciliary beat frequency (CBF), 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 centres, 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]. Additionally, there has never been a study to assess the intra- and inter-observer 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 the patients’ clinic appointment whilst awaiting TEM and genetics results, reducing time-to-diagnosis and potentially limiting disease progression.

We hypothesized that a) scientists using HSVA would accurately diagnose PCD and b) there would be good inter-observer reliability of the test.

# Materials and methods

Local and national Research and ethical 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 centres 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 6 videos from the same sample, including five side views and one top view. Investigations are detailed in the online supplementary file. Teams from the three centres share diagnostic protocols and frequently discuss difficult cases.

Diagnostic results were reviewed at MDT meetings, including a clinician, HSV-microscopist and TEM-microscopist. All clinical and diagnostic data were considered when agreeing 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 to two imperfect references [12]: **(*a*) outcomes defined using European Respiratory Society (ERS) guidelines for the diagnosis of PCD** [2] (Figure 1a); and **(**b**) the clinical MDT outcome for the patient**, extracted from contemporary MDT meeting reports (Figure 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 definitive outcome [2]. Patients with diagnostic test results that did not fulfil 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 bi-allelic 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 true PCD patients. Moreover, genotyping was only undertaken 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 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 supplementary material.

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 anonymised 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 eight years’ experience in HSVA, one from each UK PCD diagnostic centres, independently viewed 720 videos from 120 anonymised patient samples (six videos per sample, 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 a-priori study outcome for each patient sample: ‘PCD positive’, ‘PCD highly likely’, ‘PCD highly unlikely’ and ‘inconclusive’, based on qualitative assessment of CBP and observed normality and abnormality in the samples analysed. To calculate the intra-observer agreement after one year, each of the three scientists’ independently, and blinded to their initial assessment, re-assessed 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 (i.e. 50%, 30% and 20% respectively).

## Statistical analyses

We stratified the total number of patients referred to each centre 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 inconclusive. 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’. Randomisation for each stratum was performed in STATA.

To calculate the accuracy of HSVA, we compared the outcomes by each of the scientists to the patient reference outcome, using ***a)*** the ERS guidelines, and ***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 true PCD patients with normal TEM where the genotype has not yet been resolved (Figure 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 were determined when HSVA scientists did not agree with reference.

We calculated the inter-observer repeatability using Fleiss kappa coefficient for each diagnostic outcome [16]. We calculated the intra-observer repeatability for each of the scientists using Cohen kappa coefficient, with bootstrapped confidence intervals (*n*=5) [17].

Data were analysed in STATA version 14.0. Continuous variables were presented as median and inter-quarter range (IQR) and categorical variables were reported as proportions. Sensitivity and specificity were presented with 95% confidence intervals (CI95%), where appropriate. We report on both aggregate and individual (i.e. each scientist) sensitivity and specificity of HSVA study outcomes compared to both reference standards. We obtained three outcomes for each sample, one from each scientist. To adjust for clustering of data and provide robust confidence intervals, we used a generalised 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 modelling. Results were reported according to the STARD 2015 guidelines on reporting of diagnostic test accuracy studies [19].

# Results

The three diagnostic centres received a total of 1286 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 which eight showed bi-allelic pathogenic mutations in a PCD-causative gene (three in *DNAH5*, two in *DNAH11*, two in *CCDC40*, one in *RSPH9*) and one in x-linked PCD gene (*OFD1*).

## Accuracy of HSVA compared to 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’ (see Table S2 in the online data supplement).

There was excellent sensitivity (100%) and specificity (96.2%, CI95% 91.7 to 100%) when comparing the study decisions of HSVA scientists to 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. ‘Worst-case scenario’ combined with GEE modelling showed that sensitivity remained high (93.3%, CI95% 92.0 to 100%) but specificity decreased from 96.2% to 67.9% (CI95% 58.7 to 77.2%).

Individual scientists had similarly good accuracy (see Table S2 in the online data supplement). A proportion of samples were reported as ‘highly likely’ or ‘inconclusive’ when using either study HSVA results alone or using the ERS guidelines and these outcomes could not be included in the accuracy calculations.

## Accuracy of HSVA compared to 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 (see Table S3 in the online data supplement). There was excellent sensitivity (96.7%, CI95% 92.9 to 100%) and specificity (91.1%, CI95% 85.3 to 96.9%) of study HSVA analysis compared to the original MDT diagnostic outcome (Table 3). Sensitivity dropped to 85.3% (CI95% 78.0 to 92.6%) and specificity to 67.6% (CI95% 58.4 to 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% (see Table S3 in the online data supplement).

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 (see Table S3 in the online data supplement).

Two cases were classified as ‘PCD highly likely’ by both ERS guideline and the MDT, but either ‘PCD highly unlikely ’or ‘inconclusive’ by the HSVA scientists (see Table S3 in the online data supplement). The original clinical records indicated that one case had an isolated inner dynein arm defect on TEM (i.e. not a hallmark abnormality) and five repeat brushings. CBF 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 case had normal nNO, TEM and genetics for known causative genes but was diagnosed as ‘PCD highly likely’ based on “semi-rotating” CBP coupled with a sibling diagnosed with PCD exhibiting similar clinical symptoms and HSVA findings to the patient in question. Two scientists classified this sample as ‘highly unlikely’, while one said it was ‘inconclusive’. Both cases are currently treated as PCD (i.e. under management care by the PCD teams) but require further diagnostic testing before a definite diagnostic outcome can be determined.

## Intra- and inter-observer reliability

Using Fleiss kappa agreement to compare scoring between the three scientists for each diagnostic outcome, we found substantial agreement (κ=0.70) for ‘PCD positive’ and moderate agreement (κ=0.44) for ‘PCD highly unlikely’. Agreement was low for ‘PCD highly likely’ (κ=0.11) and ‘inconclusive’ (κ=0.20) [21]. The combined agreement for the overall diagnostic outcomes was moderate (κ=0.42, CI95%0.41 to 0.44).

The Cohen kappa agreement for intra-observer reliability was κ=0.70, CI95%0.56 to 0.77 for scientist 1, κ=0.66, CI95%0.42 to 0.75 for scientist 2 and κ=0.78, CI95%0.61 to 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 (see Table S4 in the online data supplement).

# Discussion

We have shown that HSVA has excellent accuracy and inter-observer 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 sensitivity and specificity were 100% & 96% when using diagnosis based on the ERS guideline as a reference, and 96% & 91% when using the clinical diagnostic outcome as standard.

Independently analysing 720 videos from 120 patients, HSVA scientists correctly identified all ‘PCD positive’ cases using the ERS PCD guideline 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’ (i.e. 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 15-20% 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’. Further reviewing the diagnostic history of these cases, 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 using HSVA, even with extensive training and years of experience [11]. Additionally, secondary abnormities are common even in samples from healthy individuals, highlighting the need for experienced personnel analysing 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 guideline (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 ‘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. Whilst the ‘worse-case scenario’ calculations are reassuring, reclassifying the inconclusive outcome was probably over conservative 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 a high inter-observer agreement for ‘PCD positive’ and moderate agreement for ‘PCD highly unlikely’ outcomes, as well as between pairs of scientists (see supplementary material). ‘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 (i.e. ‘inconclusive’). In practice, samples labelled as ‘highly likely’ or ‘inconclusive’ would both require a repeat brushing from the patient and further testing.

We also found substantial intra-observer agreement of samples re-assessed by each of the scientists one 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 demonstrates 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 centre with a strong suspicion of PCD (i.e. suggestive clinical history). Samples can be evaluated by scientists 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 PCD patients 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, IF 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 standardised 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 standardised 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 ERS guideline. 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 ‘highly likely’ or ‘inconclusive’ cases as PCD. Equally, some of the ‘highly likely PCD’ patients might not have PCD. Although we have good standardisation of methods and reporting in the UK, our protocols differ with those used in many centres (e.g. some centres measure HSVA at room temperature whilst we analyse samples at 37oC).

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 kappa performs poorly when the marginal classification probabilities are either very small or very large, underestimating the strength of agreement [27]. Additionally, kappa results rely on arbitrary convention for what is considered substantial, moderate and low agreements. Therefore, we included supplementary tables S2-4 to provide data on individual scientist’s performances.

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

In conclusion, we found that when following standardised protocols HSVA has an excellent sensitivity and specificity to diagnose PCD. We found a 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 standardisation of analysis and reporting of HSVA.

## Acknowledgements

The authors thank all laboratory and clinical members of the national PCD Diagnostic Service at UHS, RBH and LRI who have contributed to the standardised clinical protocols and the clinical diagnostic testing of the patients; Dr 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.

**Contributions:**

JSL had the concept for the study and takes responsibility for the integrity of the study and data; IR, BR, JSL: designed the study and secured funding; BR: led standardised clinical and diagnostic data collection; randomisation, data management and statistical analyses; drafted the manuscript; AS: led standardisation of study HSVA; CLJ, RH, AS: Study analyses of video images; FC: PPI; JT led standardisation of video selection and quality assessment; BR, JT, JH, EF, CO’C, CH, JSL: extraction of clinical and diagnostic data from clinical records; JT, JH, EF, BR: preparation of video images for scientist’s review; CH, CO’C, JSL: led multidisciplinary team decisions at diagnostic centres

# Figure legends

**Figure 1.** The two reference standards used to assess the accuracy of high-speed video microscopy (HSVA) analysis. **A)** ERS guidelines recommended a standardised terminology to describe the diagnostic outcome [2]. If the patient has a hallmark transmission electron microscopy (TEM) defect or bi-allelic 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 labelled as ‘PCD highly likely’; PCD is considered ‘highly unlikely’ if all tests are normal. Despite repeated testing, the diagnostic category for a proportion of patients remains inconclusive; these were patients that 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 was reviewed at a multidisciplinary meeting to define the diagnostic outcome, based on the opinion of the expert team.

**Figure 2.** Diagnostic algorithms used to determine true positive and true negative outcomes for test accuracy calculations, according to the European Respiratory Society (ERS) guidelines (top half) and the clinical diagnostic decision by the multidisciplinary (MDT) team (bottom half).

**Figure S1.** Patient-samples excluded from the study (n=15) and reasons for exclusion, stratified by multidisciplinary team meeting (MDT) diagnostic outcome and centre of origin of patient-sample. Each symbol represents a patient-sample excluded; circles represent patient-samples from the Royal Brompton Hospital (RBH, n=9), squares from Leicester Royal Infirmary (LRI, n=5) and triangles from University Hospital Southampton (UHS, n=1).

**Figure S2.** Comparison of diagnostic outcomes from study samples between the original multidisciplinary team (MDT) clinical decision (reference b) and outcomes reclassified using the ERS guidelines (reference a).

# Footnotes

**Table 1.** ERS: European Respiratory Society; PCD: primary ciliary dyskinesia; UHS: University Hospital Southampton; RBH: Royal Brompton Hospital in London; LRI: Leicester Royal Infirmary; ODA: outer dynein arm defect; IDA: inner dynein arm defect; MTD: microtubular disarrangement; CC: central complex defect; αbi-allelic mutations in *DNAH11* gene; βbi-allelic mutations in *RSPH9* gene; \*X-linked mutation in *OFD1* gene; ¬ TEM abnormality described as “thin ODA present”, not a hallmark PCD defect according to the ERS Guidelines.

**Table 2.** ERS: European Respiratory Society; HSVA: high-speed video microscopy analysis; PCD: primary ciliary dyskinesia;\* ‘PCD positive’ cases were those with a hallmark transmission electron microscopy defect and/or genotype.

**Table 3.** MDT: multidisciplinary team; HSVA: high-speed video microscopy analysis; \* Includes both ‘PCD positive’ and ‘PCD highly likely’ outcomes.

**Table S2.** ERS: European Respiratory Society; HSVA: high-speed video microscopy analysis; PCD: primary ciliary dyskinesia;\* ‘PCD positive’ cases were those with a hallmark TEM defect and/or genotype.

**Table S3.** MDT: multidisciplinary team; HSVA: high-speed video microscopy analysis; \* Includes both ‘PCD positive’ and ‘PCD highly likely’ outcomes.

**Table S4.** \* Includes both ‘PCD positive’ and ‘PCD highly likely’ outcomes.

# Tables

**Table 1.** Clinical characteristics of study participants stratified by the diagnostic outcome according to the European Respiratory Society guideline.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | All patients (n=120) | PCD positive (n= 36) | PCD highly likely (n= 16) | PCD highly unlikely (n= 26) | Inconclusive (n= 42) |
| Centre 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 in years (median, IQR) | 9.6 (2.8 to 16.7) | 9.1 (3.0 to 20.9) | 11.8 (8.9 to 12.6) | 10 (2.0 to 29.5) | 7.3 (2.9 to 14.8) |
| Pre-term 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); *n* data available | 21.8 (7.2 to 105.0); *n*=72 | 9.8 (4.8 to 15.9); *n*=22 | 7.2 (3.0 to 63.6); *n*=11 | 189.2 (69.2 to 218.0); *n*=11 | 72.3 (19.9 to 117.8); *n*=28 |
| TEM results |  |  |  |  |  |
| *Normal* | 63 (52.5%) | 2 (5.6%)α | 7 (43.8%) | 19 (73.1%) | 35 (83.3%) |
| *ODA alone* | 14 (11.7%) | 13 (36.1%) | 1 (6.25%)¬ | 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%)β | 4 (25.0%) | 0 | 0 |
| *Lack of cilia* | 2 (1.7%) | 0 | 0 | 0 | 2 (4.8%) |
| *Inconclusive* | 3 (2.5%) | 1 (2.8%)\* | 0 | 0 | 2 (4.8%) |
| *Not done* | 10 (8.3%) | 0 | 0 | 7 (26.9%) | 3 (7.1%) |

**Table 2.** Aggregated diagnostic study outcomes by the three scientists compared to the diagnostic outcome defined by the ERS PCD diagnostic guidelines [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 supplementary Table S2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Diagnostic outcomes based on the ERS guideline** | | | | |
| PCD positive\* | PCD highly unlikely | PCD highly likely | Inconclusive | Total |
| **Study outcomes by HSVA scientists** | 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  (n samples) | 108  (n=36) | 78  (n=26) | 48  (n=16) | 126  (n=42) | 360 (n=120) |

**Table 3.** Aggregated diagnostic study outcomes by the three scientists compared to the original diagnostic decision made by the MDT (n=360 scientists’ outcomes from 120 patient samples). ‘Inconclusive’ outcomes were excluded from the accuracy analyses. Individual scientists results are shown in supplementary Table S3.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Diagnostic outcomes based on the original expert MDT decision** | | | |
| PCD Positive\* | PCD Highly unlikely | Inconclusive | Total |
| **Study outcomes by HSVA scientists** | Positive\* | 151 | 7 | 18 | 176 |
| Highly unlikely | 4 | 73 | 22 | 99 |
| Inconclusive | 22 | 28 | 35 | 85 |
|  | Total (n samples) | 177 (n=59) | 108 (n=36) | 75 (n=25) | 360 (n=120) |

**Supplementary table S1.** Equipment used at the PCD centres for high-speed video microscopy, nasal nitric oxide analysis and electron microscopy.

|  |  |  |  |
| --- | --- | --- | --- |
|  | University Hospital Southampton | Leicester Royal Infirmary | Royal Brompton London |
| Specimen slide | 0.5 mm coverwell imaging chamber (Sigma-Aldrich, Poole, UK) mounted onto a glass slide | Chamber created by the separation of a coverslip and a glass slide by two adjacent coverslips | 0.5 mm coverwell imaging chamber (Sigma-Aldrich, Poole, UK) mounted onto a glass slide |
| Microscopy | Olympus IX71 inverted microscope and condenser. Specimen inverted onto an x100 UPlan wide aperture oil objective. | Leitz Diaplan upright microscope with x100 interference contrast plan apochromat objective lens. | Leica DM-LB upright microscope with x100 oil plan objective lens |
| Environmental control | 37oC heated environmental chamber (Solent Scientific, Southampton, UK). | 37oC heated stage; anti-vibration table (Wentworth Laboratories Ltd, Sandy, UK). | 37oC heated stage; anti-vibration table (Wentworth Laboratories Ltd, Sandy, UK). |
| High-speed video imaging and analysis | Photron FASTCAM MC2 high-speed video digital camera and Photron software. | IDT X4 high speed camera. AVI images analysed using MotionPro software, IDT. | Troubleshooter TS-5 Fastec imaging |
| Nasal nitric oxide analysis | Ecomedics CLD 88 Exhalyzer; exhalation against resistance; sampling 0.33 l/min | Sievers 280i Chemiluminescence; exhalation against resistance; sampling 0.3 l/min | Logan LR5000 Chemiluminscence Analyser (Rochester Kent); breath hold sampling 0.25 l/min |
| Electron microscope | 60,000x magnification (minimum) by Hitachi H7000 | 60,000x magnification by Joel 1200 | 60,000x magnification (minimum) by Hitachi H7000 |

**Supplementary table S2.** Diagnostic outcomes decisions made by each of the three HSVA scientists (i.e. scientist 1-3, n=120 for each) compared to the ERS PCD diagnostic guidelines classification.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | **Outcomes based on the ERS guideline** | | | | |
| PCD positive\* | PCD highly unlikely | PCD highly likely | Inconclusive | Total |
| **Outcomes by scientist 1** | PCD positive | 32 | 0 | 7 | 4 | 43 |
| PCD highly unlikely | 0 | 21 | 2 | 13 | 36 |
| PCD highly likely | 2 | 0 | 4 | 2 | 8 |
| Inconclusive | 2 | 5 | 3 | 23 | 33 |
| Total | 36 | 26 | 16 | 42 | 120 |
|  | | PCD positive\* | PCD highly unlikely | PCD highly likely | Inconclusive | Total |
| **Outcomes by scientist 2** | PCD positive | 35 | 2 | 12 | 8 | 57 |
| PCD highly unlikely | 0 | 8 | 0 | 8 | 16 |
| PCD highly likely | 1 | 4 | 2 | 9 | 16 |
| Inconclusive | 0 | 12 | 2 | 17 | 31 |
| Total | 36 | 26 | 16 | 42 | 120 |
|  | | PCD positive\* | PCD highly unlikely | PCD highly likely | Inconclusive | Total |
| **Outcomes by scientist 3** | PCD positive | 27 | 0 | 6 | 1 | 34 |
| PCD highly unlikely | 0 | 24 | 2 | 21 | 47 |
| PCD highly likely | 7 | 0 | 5 | 6 | 18 |
| Inconclusive | 2 | 2 | 3 | 14 | 21 |
| Total | 36 | 26 | 16 | 42 | 120 |

**Supplementary table S3.** Diagnostic outcomes decisions made by each of the three HSVA scientists (i.e. scientist 1-3, n=120 for each) compared to the original diagnostic decision made by the MDT.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Outcomes based on the expert MDT decision** | | | |
| PCD positive\* | PCD highly unlikely | Inconclusive | Total |
| **Outcomes by scientist 1** | Positive\* | 49 | 0 | 2 | 51 |
| Highly unlikely | 2 | 26 | 8 | 36 |
| Inconclusive | 8 | 10 | 15 | 33 |
| Total | 59 | 36 | 25 | 120 |
|  | | PCD positive\* | PCD highly unlikely | Inconclusive | Total |
| **Outcomes by scientist 2** | Positive\* | 55 | 7 | 11 | 73 |
| Highly unlikely | 0 | 14 | 2 | 16 |
| Inconclusive | 4 | 15 | 12 | 31 |
| Total | 59 | 36 | 25 | 120 |
|  | | PCD positive\* | PCD highly unlikely | Inconclusive | Total |
| **Outcomes by scientist 3** | Positive\* | 47 | 0 | 5 | 52 |
| Highly unlikely | 2 | 33 | 12 | 47 |
| Inconclusive | 10 | 3 | 8 | 21 |
| Total | 59 | 36 | 25 | 120 |

**Supplementary table S4.** Diagnostic outcomes decisions from a random subset of patient-samples (*n*=20) by each of the three scientists (i.e. scientist 1-3) using high-speed video microscopy analysis at two different time points (i.e. original study analysis and re-analysis after 1 year from study completion), for calculation of intra-observer agreement.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Outcomes after re-assessment of samples by scientist 1** | | | |
| PCD positive\* | PCD highly unlikely | Inconclusive | Total |
| **Original outcomes by scientist 1** | Positive\* | 7 | 0 | 0 | 7 |
| Highly unlikely | 0 | 4 | 3 | 7 |
| Inconclusive | 1 | 0 | 5 | 6 |
| Total | 8 | 4 | 8 | 20 |
|  | | **Outcomes after re-assessment of samples by scientist 2** | | | |
| PCD positive\* | PCD highly unlikely | Inconclusive | Total |
| **Original outcomes by scientist 2** | Positive\* | 10 | 0 | 0 | 10 |
| Highly unlikely | 0 | 2 | 1 | 3 |
| Inconclusive | 2 | 1 | 4 | 7 |
| Total | 12 | 3 | 5 | 20 |
|  | | **Outcomes after re-assessment of samples by scientist 3** | | | |
| PCD positive\* | PCD highly unlikely | Inconclusive | Total |
| **Original outcomes by scientist 3** | Positive\* | 6 | 0 | 0 | 6 |
| Highly unlikely | 0 | 5 | 2 | 7 |
| Inconclusive | 1 | 0 | 6 | 7 |
| Total | 7 | 5 | 8 | 20 |

# References

1. Lucas JS, Alanin MC, Collins S, Harris A, Johansen HK, Nielsen KG, Papon JF, Robinson P, Walker WT. Clinical care of children with primary ciliary dyskinesia. Expert Rev Respir Med 2017;11:779-90.

2. Lucas JS, Barbato A, Collins SA, Goutaki M, Behan L, Caudri D, Dell S, Eber E, Escudier E, Hirst RA, Hogg C, Jorissen M, Latzin P, Legendre M, Leigh MW, Midulla F, Nielsen KG, Omran H, Papon JF, Pohunek P, Redfern B, Rigau D, Rindlisbacher B, Santamaria F, Shoemark A, Snijders D, Tonia T, Titieni A, Walker WT, Werner C, Bush A, Kuehni CE.. European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. Eur Respir J. 2017 Jan 4;49(1): 1601090

3. Shapiro AJ, Davis SD, Polineni D, Manion M, Rosenfeld M, Dell SD, Chilvers MA, Ferkol TW, Zariwala MA, Sagel SD, Josephson M, Morgan L, Yilmaz O, Olivier KN, Milla C, Pittman JE, Daniels MLA, Jones MH, Janahi IA, Ware SM, Daniel SJ, Cooper ML, Nogee LM, Anton B, Eastvold T, Ehrne L, Guadagno E, Knowles MR, Leigh MW, Lavergne V. Diagnosis of Primary Ciliary Dyskinesia. An Official American Thoracic Society Clinical Practice Guideline. Am J Respir Crit Care Med2018: 197(12): e24-e39.

4. Lucas JS, Burgess A, Mitchison HM, Moya E, Williamson M, Hogg C. Diagnosis and management of primary ciliary dyskinesia. Arch Dis Child. 2014 Sep 1;99(9):850.

5. Kuehni CE, Lucas JS. Diagnosis of primary ciliary dyskinesia: summary of the ERS Task Force report. Breathe (Sheff). 2017 Sep;13(3):166–78.

6. Schwabe Georg C., Hoffmann Katrin, Loges Niki Tomas, Birker Daniel, Rossier Colette, de Santi Margherita M., et al. Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by DNAH11 mutations. Hum Mutat. 2008 Jan 14;29(2):289–98.

7. Davis SD, Ferkol TW, Rosenfeld M, Lee HS, Dell SD, Sagel SD, Milla C, Zariwala MA, Pittman JE, Shapiro AJ, Carson JL, Krischer JP, Hazucha MJ, Cooper ML, Knowles MR, Leigh MW. 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, Lau L, Paton TA, Stockley T, Jobling RK, Ray PN, Knowles MR, Hall DA, Dell SD, Kim RH. Whole-Exome Sequencing and Targeted Copy Number Analysis in Primary Ciliary Dyskinesia. G3 (Bethesda, Md) 2015: 5(8): 1775-1781.

9. Jackson CL, Behan L, Collins SA, Goggin PM, Adam EC, Coles JL, Evans HJ, Harris A, Lackie P, Packham S, Page A, Thompson J, Walker WT, Kuehni C, Lucas JS. Accuracy of diagnostic testing in primary ciliary dyskinesia. Eur Respir J. 2016: 47(3): 837-848.

10. Papon JF, Bassinet L, Cariou-Patron G, Zerah-Lancner F, Vojtek AM, Blanchon S, Crestani B, Amselem S, Coste A, Housset B, Escudier E, Louis B. Quantitative analysis of ciliary beating in primary ciliary dyskinesia: a pilot study. Orphanet J Rare Dis 2012;7: 78–78.

11. Lucas JS, Evans HJ, Haarman EG, Hirst RA, Hogg C, Jackson CL, Nielsen KG, Omran H, Papon JF, Robinson P, Shoemark A, Walker WT. 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 technology assessment (Winchester, England) 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]. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc.; 2012 [cited 2018 Apr 5]. Available from: http://methods.sagepub.com/book/sampling-essentials

15. Sampling and sample size calculation - The NIHR Research Design Service for Yorkshire the Humber Sampling and Sample Size Calculation Authors Nick Fox [Internet]. [cited 2018 Apr 5]. Available from: https://www.coursehero.com/file/22207344/Sampling-and-sample-size-calculation/

16. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bulletin. 1971;76(5):378–82.

17. Reichenheim ME. Confidence intervals for the kappa 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. Radiology2012: 265(3): 910-916.

19. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, Irwig L, Levine D, Reitsma JB, de Vet HC, Bossuyt PM. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016: 6(11): e012799.

20. Behan L, Dimitrov BD, Kuehni CE, Hogg C, Carroll M, Evans HJ, Goutaki M, Harris A, Packham S, Walker WT, Lucas JS. 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, Williams G, Rutman A, Goggin PM, Adam EC, Page A, Evans HJ, Lackie PM, O'Callaghan C, Lucas JS. 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, Phillips SE, Turner B, Rowe SM, 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, Bhaludin B, Rogers A, Bilton D, Hansell DM, Wilson R, Loebinger MR. 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, Minnix SL, Carson JL, Hazucha M, Zariwala MA, Knowles MR. Primary ciliary dyskinesia: diagnostic and phenotypic features. Am J Respir Crit Care Med2004: 169(4): 459-467.

27. Feinstein AR, Cicchetti DV. High agreement but low Kappa: I. the problems of two paradoxes. J Clin Epidemiol. 1990 Jan 1;43(6):543–9.