Accuracy of nasal nitric oxide measurement as a diagnostic test for primary ciliary dyskinesia: A systematic review and meta-analysis

Adam J. Shapiro, MD1, Maureen Josephson, MD2, Margaret Rosenfeld, MD3, Ozge Yilmaz, MD4, Stephanie D. Davis, MD5, Deepika Polineni, MD, MPH6, Elena Guadagno,7 Margaret W. Leigh, MD8 and Valery Lavergne, MD9

1Division of Pediatric Respiratory Medicine, Montreal Children’s Hospital, McGill University Health Centre Research Institute, 1001 Boulevard Décarie, BRC.5016, Montreal, Quebec, H4A 3J1, Canada. Adam.shapiro@muhc.mcgill.ca

2Division of Pediatric Pulmonology, Children’s Hospital of Philadelphia, University of Pennsylvania, Philadelphia, Pennsylvania, USA.

3Division of Pediatric Pulmonology, Seattle Children’s Hospital and Regional Medical Center, University of Washington, Seattle, Washington, USA.

4Department of Pediatric Allergy and Pulmonology, Celal Bayar University Medical Faculty, Manisa, Turkey

5Section of Pediatric Pulmonology, Allergy and Sleep Medicine; Riley Children’s Hospital, Indiana University School of Medicine, Indianapolis, Indiana, USA.

6Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, University of Kansas, Kansas City, Kansas, USA

7Medical Library, McConnell Resource Center, McGill University Health Centre, Montreal, Quebec, Canada

8Department of Pediatrics, Division of Pediatric Pulmonology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

9Department of Medical Biology, Hôpital du Sacré-Coeur de Montréal, Université de Montréal, Montreal, Quebec, Canada

Running head: Nasal nitric oxide in Primary Ciliary Dyskinesia

Descriptor number: 14.6 Rare Pediatric Lung Disease

MeSH Key words: Primary Ciliary Dyskinesia, Kartagener Syndrome, nitric oxide

Funding: This study was supported through a Pediatric Assembly grant from the American Thoracic Society for creation of clinical diagnostic guidelines for PCD.

Adam Shapiro is the primary author of this manuscript and should be contacted for all correspondence at the email address above. Dr. Shapiro accepts full responsibility for this work. EG was the master librarian and designed the document search. AS, VL, SDD and MWL drafted and edited this manuscript. AS, MJ, MR, OY, DP, and VL performed the systematic document review and edited this manuscript. VL served as the project methodologist and performed the meta-analysis and evidence grading. All authors agree with all aspects of this manuscript.

This is the author's manuscript of the article published in final edited form as:
Abstract Word Count: 397

Total Body Word Count: 4000

This article has a data supplement, which is accessible from this issue's table of contents online at www.atsjournals.org
ABSTRACT

Background: Primary Ciliary Dyskinesia (PCD) is a rare disorder causing chronic oto-sino-pulmonary disease, generally diagnosed through evaluation of respiratory cilia ultrastructure and/or genetic testing. Nasal nitric oxide (nNO) measurement is a PCD screening test, as PCD patients have low nNO levels, but its value as a diagnostic test remains unknown.

Objective: Perform a systematic review assessing the utility of nNO measurement (index test) as a diagnostic tool compared to the reference standard of electron microscopy (EM) evaluation of ciliary defects and/or detection of biallelic mutations in PCD genes.

Data sources: Ten databases from inception through July 29, 2016

Data extraction: Study inclusion was limited to publications with rigorous nNO index testing, reference standard diagnostic testing with EM and/or genetics, and calculable diagnostic accuracy information for cooperative patients (generally >5 years old), highly suspected of PCD.

Synthesis: Meta-analysis provided a summary estimate for sensitivity and specificity and a hierarchical summary receiver operator curve. The QUADAS-2 tool assessed study quality and GRADE assessed diagnostic test accuracy of studies to evaluate the certainty of evidence. In twelve study populations (1,344 patients: 514 PCD, 830 non-PCD), using a reference standard of EM alone or EM and/or genetic testing, summary sensitivity was 97.6% (92.7-99.2), and specificity was 96.0% (87.9-98.7), with a positive likelihood ratio of 24.3 (7.6-76.9), a negative likelihood ratio of 0.03 (0.01-0.08) and a diagnostic odds ratio of 956.8 (141.2-6481.5) for nNO measurements. Excluding studies using EM alone as the reference standard, the seven studies using an extended reference standard of EM and/or genetic testing show a summary sensitivity of nNO measurements as 96.3% (88.7-98.9), and specificity as 96.4% (85.1-99.2), with a positive
likelihood ratio of 26.5 (5.9-119.1), a negative likelihood ratio of 0.04 (0.01-0.12), and a
diagnostic odds ratio of 699.3 (67.4-7256.0). Certainty of the evidence was graded as moderate.

**Conclusions:** Nasal nitric oxide is a sensitive and specific test for PCD in cooperative patients
(generally >5 years old) with high clinical suspicion for this disease. With a moderate level of
evidence, this meta-analysis confirms that nNO testing using velum closure maneuvers has
similar diagnostic accuracy to EM and/or genetic testing for PCD, when cystic fibrosis is ruled
out. Thus, low nNO values, accompanied by an appropriate clinical phenotype, could be used as
a diagnostic PCD test, though EM and/or genetics will continue to provide confirmatory
information.

**Funding:** The American Thoracic Society for creation of clinical practice guidelines on
diagnostic testing for PCD.
INTRODUCTION

Primary Ciliary Dyskinesia (PCD) is a rare autosomal recessive disease resulting in impaired mucociliary clearance and chronic oto-sino-pulmonary infections. Nasal nitric oxide (nNO) levels are low in PCD, and since nNO results are immediately available, these measurements are often used as a screening tool for PCD, before proceeding to ciliary electron microscopy (EM), high speed videomicroscopy analysis (HSVA), or genetic analysis for confirmatory diagnostic testing. These latter tests are expensive ($550-$2,200 USD), can take months to complete, and sometimes yield non-diagnostic results. Inexperience in obtaining biopsy samples can lead to insufficient cilia for EM analysis, and inexperience in interpretation can lead to false positive or false negative EM results. Diagnostic HSVA testing can also be challenging, as there is no standardization of ciliary waveform analysis, multiple biopsies at separate visits or re-differentiation of ciliated cells in culture are required to insure permanence of diagnostic ciliary waveform abnormalities (i.e. not arising from secondary insults such as viral infection)(1), and interpretation of HSVA samples from healthy controls shows poor inter-observer agreement(2). Finally, genetic testing currently can only detect biallelic mutations in about two-thirds of patients with PCD(3).

Previous publications have examined the diagnostic testing accuracy of nNO in PCD, yet many incorporated methodological flaws in study design, which could affect diagnostic accuracy. These errors include using HSVA as a screening test for study entry (excluding all subjects with normal videomicroscopy from further PCD testing), incorporating nNO measurement into both index (the new test being evaluated) and reference (the chosen gold standard) standard testing(4, 5), or using imperfect reference standard testing, by enrolling some
subjects diagnosed with PCD through HSVA analysis alone, and not presenting data on permanence of ciliary waveform abnormalities on repeat HSVA testing or after cellular regrowth(6). Two previous meta-analyses examined the diagnostic testing accuracy of nNO in PCD, yet these analyses included studies with methodological flaws(7, 8). These methodological errors include: 1) not providing detailed information on tests used to diagnose patients with PCD(9-11), 2) inclusion of non-standard EM diagnoses in the reference standard (isolated inner dynein arm (IDA) defects without microtubule disorganization (MTD), and without repeat verification of isolated IDA defects on 2 separate biopsies)(12, 13), 3) inclusion of cystic fibrosis (CF) patients as disease controls, in whom nNO levels commonly fall below PCD cut-off values, impacting diagnostic accuracy(14-17), and 4) using non-standard technology or techniques for nNO measurement(6, 10, 16). Additionally, these meta-analyses did not routinely incorporate genetic results into their reference standard, even though commercial genetic testing is now a front-line clinical test for PCD.

The American Thoracic Society has supported creation of clinical diagnostic guidelines for PCD. As part of these guidelines, a robust systematic review and meta-analysis was performed, examining the diagnostic testing accuracy of nNO measurement for PCD, and results are presented here. This review uses strict inclusion and exclusion criteria to define acceptable index and reference standard testing for PCD. The objective of this analysis is to assess if nNO measurement can be used as a diagnostic test for PCD (as opposed to only a screening test), in cooperative patients (generally >5 years old), who have a high probability of having this disease based on a highly suggestive clinical phenotype(18), and in whom cystic fibrosis has been ruled out. Specifically, the usefulness of this tool is evaluated as a replacement for the diagnostic
METHODS

Data sources and searches

For the literature search, the consulted databases were: Africa-Wide Information (Ebsco), AMED (Ovid), BIOSIS (Ovid), Cochrane (Wiley), Embase (Ovid), Global Health (Ovid), MEDLINE (Ovid), PubMed (NLM), Scopus (Elsevier), and Web of Science (Thomson Reuters). We manually searched all references from included articles to identify other potential literature of interest. The search was performed from all database inceptions until July 29, 2016 (Supplemental material, Appendix 1).

Study selection

Eligible studies:

Selected studies evaluate the accuracy of nNO testing (index test) in cooperative patients (generally >5 years old), who were deemed at high probability for having PCD based on a compatible clinical phenotype, compared to the reference standards of classic EM ultrastructural ciliary defect (outer dynein arm defect, outer plus inner dynein arm defect, inner dynein arm defect with microtubule disorganization, radial spoke or central apparatus defect) and/or biallelic mutations in known PCD genes. Articles were not excluded on the basis of language or date of publication.

Exclusion criteria:
Articles were excluded if any of the following were present: 1) <10 PCD patients in the recruited population, 2) the index test was inadequate - nNO measurement used electrochemical technology (NIOX Mino), only used non-velum closure techniques (tidal breathing), and/or used nasal sampling flow rates outside of the American Thoracic Society/European Respiratory Society recommended range(19), 3) the reference standard relied only on a single HSVA for PCD confirmation (without a second positive PCD diagnostic test or without HSVA after cellular regrowth in culture) or ≥30% of subjects had non-standard EM defects (unrepeated, isolated IDA defects without MTD)(20), 4) diagnostic testing accuracy was either not provided, not accurate, or not calculable, and 5) index testing was incorporated in the reference standard.

Selection process:

After duplicate article exclusion, two independent reviewers (A.S., D.P.) screened titles and abstracts to exclude non-pertinent publications. Full texts of eligible articles were assessed for final eligibility by a team of three independent reviewers (M.J., M.R., O.Y.). Final selection was based on full text assessment with complementary information provided by authors, when needed. Three months were allowed for authors to answer email queries, after which, articles lacking crucial information were excluded. If the article was included, but was found to contain missing information, a worst-case scenario was assumed (e.g., for unconfirmed, isolated IDA defects, patients were assumed as not having PCD). Disagreements were resolved by discussion (A.S., V.L.).

Data abstraction

Two reviewers extracted data independently (A.S. & M.J., M.R. or O.Y.) and assessed data quality (A.S and V.L.). Disagreements were resolved through discussion with a third
reviewer (M.J.). Nasal NO values by exhalation against resistance (ER) and breath hold (BH) techniques were both collected; breath hold values were accepted if ER data were unavailable. If nNO measurement techniques were unclear, authors were contacted for clarification on techniques used and the number of subjects who performed ER or BH maneuvers. All nNO measurements are presented in nanoliters/minute (nL/min). Quality assessment data was collected, including blinding to reference or index tests, pre-specification of the PCD diagnostic nNO cut-off value, and index test results as compared to the reference standard (true positive (TP), false positive (FP), true negative (TN), false negative (FN), and inconclusive result).

Quality assessment

The Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool was used to evaluate the internal and external validity of each study(21). Risk of bias and applicability were assessed in four domains (patient selection, index test, reference standard, and flow/timing). Each item was graded as low, high, or unclear risk. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system for Diagnostic Test Accuracy(22-24) analyzed the certainty of evidence for each test result and for overall accuracy. Certainty of evidence considered the study design, risk of bias, precision, consistency, and directness.

Data synthesis and analysis

A bivariate model calculated summary estimates for sensitivity and specificity using a generalised linear mixed model approach. Summary likelihood ratios and diagnostic odd ratios were reported. A good discrimination was defined as a positive likelihood ratio >5.0 and
negative likelihood ratio <0.2(25). A hierarchical summary receiver operator curve was constructed describing the relationship between a continuous cut-off and accuracy. Analyses were performed using STATA (version IC 14, StataCorp, College Station, Texas, US) with the commands “metandi” and “metandiplot”(26). Different sources of heterogeneity, other than variation in thresholds between studies, were explored. A sensitivity analysis was performed to estimate the accuracy of nNO testing after excluding studies relying on EM alone as reference standard. Other sources of heterogeneity were explored using subgroup analyses. Analyses were performed in Review Manager 5.3 (Cochrane collaboration). Heterogeneity was assessed by visual inspection of the summary receiver operator curve.

**General methodology**

This manuscript follows the PRISMA-P reporting guidelines for systematic review and meta-analyses (Supplemental material, Appendix 2)(27) and Cochrane Handbook for diagnostic testing accuracy reviews (28).

**RESULTS**

**Study selection**

In total, 10,787 records were identified through a generalized search of all publications related to PCD, for use in comprehensive guidelines on PCD diagnosis. Results were not initially limited to articles investigating nNO testing. After removing duplicates, 6,204 records were screened by title and abstract, and 6,127 records not addressing nNO testing were excluded.
Seventy-six full-text articles were assessed for eligibility, from which 65 were excluded (Figure 1). Twelve study populations from 11 articles were included in the quantitative synthesis. (14-17, 29-35)

**Study characteristics**

All twelve included studies were published between 2003 and 2015, from the following countries: Italy (3), United States (3), France (2), United Kingdom (2), Belgium (1) and Canada (1). Sample sizes ranged from 28 to 373 patients (8-149 PCD patients, 15-153 non-PCD patients). Four studies were cohort designs (prospective investigation of consecutive symptomatic referrals for PCD) while eight studies were case-control designs (retrospective comparison of previously diagnosed PCD populations against healthy and/or disease controls).

**Population characteristics (Table 1)**

A total of 1,721 patients were included in these twelve studies. In two studies, 42 patients were excluded for technical difficulties (problems with the NO analyzer, nasal obstruction, high ambient NO, or incomplete data) (29, 33). We excluded 191 CF patients (14-17, 30, 34) to better reflect real practice, where CF should be ruled out before nNO testing for PCD and 88 uncooperative children who could not perform nNO with velum closure techniques (32, 33). We further excluded 56 patients who had inconclusive reference standard results (29). In total, 1,344 patients were analysed (514 PCD patients, 830 non-PCD patients). Half of the studies included mainly a pediatric population (under 18-25 years old) (14, 15, 29, 32, 33, 35) while half included patients of all ages (16, 17, 30, 31, 34). Prevalence of PCD patients in cohort studies ranged from
28% to 57% of patients included in the quantitative analysis(17, 29, 31, 33). Nine studies provided information on symptoms leading to clinical suspicion of PCD, which generally included at least one of the following: chronic rhino-sinusitis, chronic otitis media, chronic bronchitis, bronchiectasis, neonatal respiratory distress, and/or organ laterality defects (mainly situs inversus totalis). Six studies ruled out cystic fibrosis and five studies ruled out immunodeficiency prior to PCD testing.

**Index test characteristics (Table 2)**

Several different brands of chemiluminescence nitric oxide analyzers were used across the studies (NIOX Flex, Endono 8000, EcoPhysics CLD88, Sievers 280i, EVA4000, LR2000). Sampling flow rates ranged from 0.25 to 0.5 L/min, but only one study included regular verification (via standard operating procedures) of sampling flow rates with direct measurement using a Gilmont flowmeter(17). Most studies performed device calibration per device manufacturer recommendations. Six studies reported nNO measurement via ER and five studies used BH maneuvers (technique not fully reported in one study). Diagnostic nNO cut-off values ranged from 16.8 to 100 nL/min, with a median cut-off at 76.9 nL/min.

**Reference standards characteristics and strategies (Table 2)**

**Electronic microscopy (EM):**

All studies included ciliary EM as the sole or main reference standard. The majority followed standard EM methodology(36). Most isolated IDA defects were either confirmed upon
repeat EM study, associated with MTD on \textit{post-hoc} EM review, or confirmed \textit{post-hoc} by
disease-causing mutations in CCDC39 or CCDC40 genes. Nevertheless, Wodehouse and al.
reported twelve patients (28.6\%) as having isolated IDA defects without further specification by
the authors, which increased the level of bias for this included publication(34). One basal body
anomaly reported as PCD was excluded from analysis(35).

\textbf{Genetic testing:}

Three studies reported genetic testing as part of the original reference standard (usually as
a complementary tool when EM was non-diagnostic rather than a systematic test used on all
patients)(17, 29). After contacting authors, we found five additional cohorts(14-16, 30, 33) in
whom genetic testing was performed \textit{post hoc} in individuals with EM defects (n=24) or non-
diagnostic EM studies (n=32). Two cohorts tested only a single PCD gene (DNAH11), one
cohort tested two genes, one cohort tested at least six genes, and one cohort tested 12-32 PCD
genes (Table 2). In the meta-analysis, patients with biallelic mutations in a PCD-causing gene,
whether identified prospectively or \textit{post hoc}, were categorized as having PCD.

\textbf{Quality assessment (Figure 2)}

\textbf{Patient selection (risk of bias and applicability):}

Four studies were cohort type(17, 29, 31, 33), while eight were case-control type studies.
Among the case-control studies, five used disease controls(14-17, 34), while three used healthy
controls(17, 30, 32, 35). The populations examined in cohort studies were selected populations
considered at high risk for PCD (excluding CF patients) in whom PCD testing was being
pursued.
Index test (risk of bias and applicability):

In seven of twelve studies, the nNO cut-off was not pre-specified. Blinding of the index test was often not reported, but since nNO is an objective measurement, this was judged as having low impact on the risk of bias. In most studies, patients were tested when free of acute respiratory tract infection for >2 weeks and not around nasal instrumentation. Only nNO results from cooperative children, who could perform velum closure maneuvers (via breath hold or exhalation against resistance techniques), were evaluated.

Reference standard (risk of bias and applicability):

The majority of studies reported that reference standards were blinded to the nNO measurements. There was no major concern regarding the technical aspects of the reference standard testing except in one cohort study, in which 39.4% patients (56 out of 142) were left undiagnosed due to inconclusive reference standard results (29). The remaining 86 patients with conclusive reference standard testing were included in our meta-analysis.

Flow and timing (risk of bias):

Differential verification (EM or genetic testing was only performed in PCD patients and not in controls) and absence of simultaneous testing (index and reference tests were performed sequentially instead of simultaneously) were frequent, especially in case-control studies. Both of these factors may artificially increase sensitivity and specificity.

Data synthesis
When pooling the results of twelve studies, the bivariate analysis (average sensitivity and specificity for all thresholds) showed a summary sensitivity of 97.6% (92.7-99.2) and specificity of 96.0% (87.9-98.7) as well a positive likelihood ratio of 24.3 (7.6-76.9), a negative likelihood ratio of 0.03 (0.01-0.08), and a diagnostic odds ratio of 956.8 (141.2-6481.5) for nNO measurements. For this analysis, isolated IDA defects were reclassified as non-PCD when feasible. Assuming a pre-test probability of 35% (17, 18, 29, 31, 33), corresponding positive and negative predictive values were 92.9% (80.5-97.6) and 98.7% (95.7-99.6), respectively (Supplemental material, Appendix 3). A forest plot presenting studies in ascending order of thresholds is presented in Figure 3. Summary hierarchical receiver operator curve illustrating how sensitivity and specificity trade-off with each other as thresholds vary is presented in Figure 4.

**Heterogeneity**

**Subgroup analysis**

Sources of heterogeneity were explored using subgroup analyses. Studies presenting a lower risk of bias in different domains (such as using cohort-type design, disease controls over healthy controls, and pre-specified nNO cut-off values) showed slightly lower diagnostic test accuracy. Interestingly, studies that systematically excluded CF prior to PCD testing (15, 17, 30, 31, 33) showed a slightly higher diagnostic accuracy than studies that did not exclude CF (14, 29, 31, 32, 34, 35) (sensitivity of 97.7% vs 95.1%, and specificity of 98.5% vs 91.4%, respectively).

**Sensitivity analysis**: 


The most relevant source of heterogeneity was the strategy used for the reference standard of PCD disease (EM alone vs extended reference standard combining EM and/or genetic testing). Thus, we performed a sensitivity analysis including only the seven studies with the extended reference standard of EM defects and/or genetic diagnoses (14-17, 29, 30, 33), which included 1,086 patients (430 PCD patients, 656 non-PCD patients). Globally, these seven studies were at lower risk of bias than the whole group (Figure 5), with proportionally more cohort-type studies, less using asymptomatic patients as their control group, and more studies pre-specifying their nNO cut-off. Pooled analysis showed a summary sensitivity of 96.3% (88.7-98.9) and specificity of 96.4% (85.1-99.2) as well as a positive likelihood ratio of 26.5 (5.9-119.1), a negative likelihood ratio of 0.04 (0.01-0.12), and a diagnostic odds ratio of 699.3 (67.4-7255.9) when comparing nNO to the extended reference standard of EM defects and/or biallelic genetic mutations (Figure 6). Per GRADE methodology, the overall certainty of evidence was moderate, when evaluating studies comparing nNO to an extended reference standard of EM and/or genetics (see Table 3).
DISCUSSION

In this meta-analysis, the diagnostic testing accuracy of nNO is excellent when compared against EM, and only slightly lower in comparison to the extended reference standard of EM and/or genetic testing. Both EM and genetic analysis are imperfect reference standard PCD tests, with currently estimated sensitivities at 0.70 (3, 37), and each of these detecting PCD cases that can be missed by the other test. Additionally, these reference standard tests can frequently provide non-diagnostic results, with up to 40% of clinical biopsies showing inadequate cilia for EM analysis(38) and up to 43% of genetic testing detecting monoallelic mutations or variants of unknown significance(39). Conversely, nNO measurement is a highly feasible test in cooperative patients (generally >5 years old), with successful measurements accomplished in >90% of patients in this meta-analysis. Although nNO testing has been largely considered as a PCD screening test, this analysis shows that nNO has a similar diagnostic potential to the accepted confirmatory PCD tests of EM and genetic analysis. Thus, in populations with an appropriate clinical phenotype for PCD, where CF is ruled out, nNO measurement is a comparable PCD diagnostic test, with the added benefits of being highly feasible, painless, non-invasive, rapid, and relatively inexpensive ($25-85 USD) for patients. However, there are limitations to nNO testing for PCD, including high purchase cost of chemiluminescence machines, training of device operators, lack of clinical approval for nNO devices in the United States, and the inability to rigorously test uncooperative children (generally <5 years old).

Disease prevalence influences post-test probability, and this analysis assumes a PCD disease prevalence of 35%, as demonstrated when PCD is strongly suspected due to the presence of a highly suggestive clinical phenotype(18). This robust phenotype of 1) unexplained neonatal respiratory distress at term birth, 2) year-round wet cough starting before six months of age, 3)
year-round nasal congestion starting before six months of age, and 4) organ laterality defects, is highly predictive of PCD. While some of the studies included in this meta-analysis did not use these specific symptoms to select candidates for PCD diagnostic testing, most studies included variations of these clinical criteria. Thus, in a pre-selected population expressing these PCD-specific symptoms, nNO measurement is a highly accurate diagnostic test and can replace EM or genetic testing. If the prevalence is lower due to less stringent phenotype screening, the positive predictive value will be lower. For example, if the prevalence of PCD is 10% in a less-selected group, the positive predictive value of nNO testing for PCD is considerably lower at 73%. At this lower PCD prevalence, approximately one quarter of patients with a positive nNO test will not have PCD upon confirmatory testing. Therefore, it is critical that careful selection of patients for diagnostic evaluation by nNO testing be accomplished. Otherwise, in less enriched groups, nNO will be more useful as a triage test prior to PCD diagnostic testing, as opposed to a replacement diagnostic test. Clinicians must consider this point, and appropriately screen patients for PCD-specific clinical criteria before embarking on PCD diagnostic investigations, including nNO testing.

Two past meta-analyses have shown similar findings to this analysis, but neither publication used an extended reference standard incorporating genetic testing(7, 8). Rather, included studies used varying combinations of different reference standards, including clinical phenotype, HSVA, EM, and rarely genetics. Our analysis used rigorous criteria to define reference standard testing. By contacting authors, we eliminated studies with ≥30% isolated IDA defects and assigned greater bias to studies with 20-30% isolated IDA defects, as 25% of isolated IDA defects resolve on repeat EM testing(20). Through author communication, we also significantly increased reference standard data on genetic testing, which improves
generalizability of this analysis to current clinical practices in North America, where genetic testing is increasingly used in PCD diagnosis. Lastly, we discovered that some studies only performed EM testing if HSVA was first abnormal, and often did not repeat HSVA studies on separate occasions or after cell culture. Altogether, our rigorous definition of reference standard testing greatly increases the strength of this meta-analysis.

This analysis also used meticulous criteria to define the index test of nNO measurement. We restricted analysis to studies using chemiluminescence technology, as only this technology is recommended for nNO measurement in PCD(40). Next, we limited our data to nNO testing only through velum closure techniques. While tidal breathing nNO measurements are of clinical value in young children, PCD diagnostic cut-off values have not been defined for these techniques. Lastly, we excluded all CF patients, who can have nNO levels below PCD cut-off values, which could affect diagnostic accuracy.

Even with our robust inclusion and exclusion criteria, this analysis has some limitations. First, despite its increasing clinical recognition, PCD is still a relatively rare disease, and our patient numbers are limited. Second, the heterogeneity of PCD reference standards poses difficulties for study generalizability. Ciliary EM alone identifies more classic cases of PCD, while missing variant forms(41). The expense of genetic testing also creates differential verification, where reference genetic testing is mainly performed in suspected PCD patients and not in healthy controls, which affects diagnostic testing accuracy. Non-simultaneous PCD diagnostic testing (using nNO as an initial screening test, followed by EM and/or genetic testing) may also have affected diagnostic accuracy in the selected studies, although blinding of researchers should have minimized these effects. Due to the rapid discovery of novel PCD-causing gene mutations, most genetic panels are incomplete by the time of study publication,
which further decreases the diagnostic accuracy of PCD genetic testing. However, with future
discovery of novel PCD genes that result in normal ultrastructure with low nNO levels, the false-
positive rate of nNO testing may decrease and diagnostic accuracy may actually improve.
Lastly, studies in this analysis using EM alone as the reference standard were more often
designed as case-control-type studies, did not pre-specify diagnostic nNO cut-off values, or were
not blinded to nNO results during reference standard testing. Each of these factors is associated
with an overestimation of diagnostic testing accuracy. Thus, while it is possible that nNO testing
is actually less accurate when using an extended reference standard of EM and/or genetics, it
seems more likely that studies using EM alone as the reference standard are at higher risk of bias,
resulting in falsely increased diagnostic testing accuracy.

CONCLUSION

Nasal nitric oxide is a sensitive and specific test for diagnosing PCD in cooperative
patients (generally >5 years old), in whom cystic fibrosis has been ruled out, and who have a
robust clinical phenotype for PCD. The gold standard tests of EM and/or genetic analysis are
imperfect tests, as both lack sensitivity for PCD diagnosis. Although nNO was previously
considered a PCD screening test, with a moderate level of evidence, this meta-analysis confirms
that nNO testing has at least equivalent and likely better diagnostic testing accuracy than EM
and/or genetic testing for PCD. Thus, we propose that nNO be considered a diagnostic test rather
than a screening test in this population. Physicians must realize that normal nNO levels do not
rule out PCD, and patients with highly compatible PCD clinical phenotypes but normal nNO
levels should progress to further testing. In addition, even in individuals with a compatible
clinical phenotype and low nasal NO, confirmatory testing with EM or genetics will yield
additional diagnostic information. As more genetic causes of PCD are discovered, repeat meta-analysis will be required to evaluate the diagnostic testing accuracy of nNO measurement, and the upcoming ATS sponsored clinical practice guidelines on PCD diagnosis will further investigate the accuracy of other PCD diagnostic tests. Future study of tidal breathing nNO measurement is needed to evaluate the usefulness of this non-invasive, rapid, and inexpensive test for successful PCD diagnosis in uncooperative children <5 years old.
### Table 1: Study and patient characteristics

<table>
<thead>
<tr>
<th>Study, year (reference)</th>
<th>Location</th>
<th>Study design</th>
<th>Patients, total n*</th>
<th>Patient description</th>
<th>PCD patients, n (prevalence)</th>
<th>Age</th>
<th>Gender, n male (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beydon, 2015 (29)</td>
<td>France</td>
<td>Cohort</td>
<td>86 patients suspected of having PCD</td>
<td>Patients included children with chronic rhino-sinusitis, serous otitis media, bronchiectasis, chronic bronchitis, or situs inversus</td>
<td>49 PCD total; Only 44 PCD performed nNO test correctly 49/86 (57.0%)</td>
<td>PCD median = 11.4 yo (range 7-13.9) Non-PCD median = 7.9 yo (range 4.9-11.6)</td>
<td>81/142 (57.0%)</td>
</tr>
<tr>
<td>Boon, 2014 (14)</td>
<td>Belgium</td>
<td>Case-control</td>
<td>191 patients: -38 PCD -153 non-PCD (51 HC, 48 asthma, 54 humoral immunodeficiency)</td>
<td>PCD patients included children and adults with recurrent upper or lower respiratory tract infections +/- organ situs anomalies</td>
<td>38 (NA)</td>
<td>Range = 5 to 25 yo PCD = 14.3 yo (range 8.8-18.1) Non-PCD = HC 14.9 yo (range 10.8-20.4), asthma 12.1 yo (range 9.8-16.5), humoral immunodeficiency = 10.7 yo (range 8.2-15.6)</td>
<td>85/191 (44.5%)</td>
</tr>
<tr>
<td>Harris, 2014 (16)</td>
<td>United Kingdom</td>
<td>Case-control</td>
<td>44 patients: -13 PCD -31 non-PCD (16 with symptoms, 15 HC)</td>
<td>Unclear</td>
<td>13 (NA)</td>
<td>Range = 6 to 79 yo</td>
<td>Not given</td>
</tr>
<tr>
<td>Leigh (leading site), 2013 (17)</td>
<td>United States</td>
<td>Case-control</td>
<td>296 patients: -149 PCD -147 non-PCD (37 asthma, 32 COPD and 78 HC)</td>
<td>PCD patients included children and adults with respiratory features suggestive of PCD (unexplained neonatal respiratory distress, year-round nasal congestion, year-round wet cough, &gt;5 episodes of otitis media by 2 yo, or situs anomalies, usually after cystic fibrosis &amp; immunodeficiency excluded)</td>
<td>149 (NA)</td>
<td>PCD mean= 19.1 ± 14.8 yo Non-PCD mean = HC 20.9 ± 15.7 yo, asthma 14.8 ± 11.5 yo, COPD 61.1 ± 8.9 yo</td>
<td>139/296 (47.0%)</td>
</tr>
<tr>
<td>Leigh (other sites), 2013 (17)</td>
<td>United States</td>
<td>Cohort</td>
<td>155 patients suspected of having PCD</td>
<td>Patients included children and adults with respiratory features suggestive of PCD (unexplained neonatal respiratory distress, year-round nasal congestion, year-round wet cough, &gt;5 episodes of otitis media by 2 yo, or situs anomalies, usually</td>
<td>71/155 (45.8%)</td>
<td>PCD mean = 23.3 ± 18 yo Non-PCD mean = 31.8 ± 22.3 yo</td>
<td>64/155 (41.3%)</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Design</td>
<td>Number of Patients</td>
<td>Description</td>
<td>PCD/Total</td>
<td>PCD Mean/Range</td>
<td>Non-PCD Mean/Range</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------</td>
<td>------------</td>
<td>--------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------</td>
<td>-----------</td>
<td>----------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Mateos Coral, 2011 (15)</td>
<td>Canada</td>
<td>Case-control</td>
<td>53 patients: -20 PCD -33 non-PCD (14 with bronchiectasis, 19 HC)</td>
<td>PCD patients included children with sinopulmonary symptoms typical of PCD, with CF and immunodeficiency ruled out</td>
<td>20 (NA)</td>
<td>PCD mean = 11.4 ± 3.5 yo Bronchiectasis mean = 10.9 ±3.3 yo, HC mean = 11.0 ± 3.7 yo</td>
<td>26/53 (49.1%)</td>
</tr>
<tr>
<td>Noone, 2014 (30)</td>
<td>United States</td>
<td>Case-control</td>
<td>140 patients: -69 PCD -71 non-PCD (27 HC, 44 healthy heterozygotes)</td>
<td>PCD patients included children and adults with lower airway disease with productive cough, wheeze, or shortness of breath and chronic upper airway symptoms of rhinitis/sinusitis +/- situs inversus totalis.</td>
<td>69 (NA)</td>
<td>PCD children median = 8 yo (range 1-17) PCD adults median = 36 yo (range 19-73) Non-PCD means = HC 37 ± 2 yo, and healthy heterozygotes = 44 ± 2 yo</td>
<td>PCD: 36/78 (46.2%)</td>
</tr>
<tr>
<td>Papon, 2012 (31)</td>
<td>France</td>
<td>Cohort</td>
<td>34 patients suspected of having PCD</td>
<td>Patients included children and adults with chronic upper and/or lower respiratory tract infections, bronchitis, bronchiectasis, and sinusitis.</td>
<td>13/34 (38.2%)</td>
<td>Mean = 32.5 yo (range 10-72)</td>
<td>16/34 (47.1%)</td>
</tr>
<tr>
<td>Piacentini, 2008 (32)</td>
<td>Italy</td>
<td>Case-control</td>
<td>-35 patients: -8 PCD -27 non-PCD (HC)</td>
<td>PCD patients included children with situs inversus and/or bronchiectasis and/or sinusitis</td>
<td>10 PCD total; Only 8 performed nNO test correctly (NA)</td>
<td>PCD mean = 17 yo; Non-PCD = 27 school aged with mean of 7 yo</td>
<td>53/87 (60.9%)</td>
</tr>
<tr>
<td>Pifferi, 2011 (33)</td>
<td>Italy</td>
<td>Cohort</td>
<td>-173 patients suspected of having PCD</td>
<td>Patients included children with clinical history and symptoms of PCD, without cystic fibrosis, aspiration, gastro-esophageal reflux, or immunodeficiency.</td>
<td>48 PCD total; Only 40 PCD performed nNO test correctly</td>
<td>48/173 (27.7%)</td>
<td>Median = 6.2 yo (range 1 mo to 17.5)</td>
</tr>
<tr>
<td>Santamaria, 2008 (35)</td>
<td>Italy</td>
<td>Case-control</td>
<td>28 patients: -14 PCD -14 non-PCD (14 HC)</td>
<td>Unclear</td>
<td>14 (NA)</td>
<td>PCD mean = 15 yo (range = 7-27) HC mean = 16 yo (range = 7-27)</td>
<td>18/28 (64.3%)</td>
</tr>
<tr>
<td>Wodehouse, 2003 (34)</td>
<td>United Kingdom</td>
<td>Case-control</td>
<td>108 patients: -42 PCD -66 non-PCD (20 with bronchiectasis, 12 Young’s syndrome, 18 sinusitis, 16 HC)</td>
<td>Unclear</td>
<td>42 (NA)</td>
<td>PCD mean = 34.2 ± 10.9 yo Non-PCD range of means = 36.2 to 53.2 yo</td>
<td>48/108 (44.4%)</td>
</tr>
</tbody>
</table>
*Number of patients included in our final analysis after excluding patients experiencing technical difficulties with nNO testing (Beydon (n=39) and Pifferi (n=3)), CF subjects (Boon (n=50), Harris (n=6), Leigh (lead site) (n=77), Mateos Coral (n=32), Noone (n=11), and Wodehouse (n=15)), and patients with an inconclusive reference standard result (Beydon (n=56)). Additionally, uncooperative children who could only perform tidal breathing nNO measurements were excluded from analysis (Beydon (PCD n=5, non-PCD n=7), Piacentini (PCD n=2, Healthy controls n=50), and Pifferi (PCD n=8, non-PCD=28)).
<table>
<thead>
<tr>
<th>Study, year (reference)</th>
<th>Index test characteristics*</th>
<th>Reference standard characteristics*</th>
<th>PCD diagnosis not confirmed by EM and/or genetics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beydon, 2015** (29)</td>
<td>NIOX Flex, Endono 8000</td>
<td>0.30 Mainly ER, 5 PCD via TB were excluded</td>
<td>44 of 49 PCD analysed: EM (n=44) and/or genetics (n=22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DNAI1 (n=5) DNAI2 (n=1) RSPH1 (n=1) RSPH9 (n=1) RSPH4A (n=2) DTY1C1 (n=2) RPGR (n=1) -Unknown total number of genes tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IDA+MTD (n=9) CCDC39 (n=6) CCDC40 (n=3) -Unknown total number of genes tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IDA alone (n=3)</td>
</tr>
<tr>
<td>Boon, 2014** (14)</td>
<td>EcoPhysics CLD88</td>
<td>0.30 ER 90</td>
<td>38 PCD analysed: EM (n=23) or HSVA after ciliary culture regrowth (n=15), and/or post hoc confirmation by genetics (n=21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DNAHS (n=4) -Only DNAHS tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IDA+MTD (n=3) CCDC40 (n=3) -Only CCDC40 tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RSP (n=1) -Unknown total number of genes tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DNAH11 (n=10) -Exome sequence used for 10 cases HYDIN (n=2) CCDC65 (n=1) -Unknown total number of genes tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal EM with abnormal HSVA (n=15)</td>
</tr>
<tr>
<td>Harris, 2014** (16)</td>
<td>NIOX Flex</td>
<td>0.30 BH 38</td>
<td>13 PCD analysed: EM (n=11) or HSVA after ciliary culture regrowth in some cases with post hoc confirmation by genetics (n=2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DNAH11 (n=2) -Only DNAH11 tested</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Normal EM with abnormal HSVA (n=2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Leigh (leading site), 2013** (17)</td>
<td>Sievers 280i, EcoPhysics CLD88, NIOX Flex</td>
<td>0.50, 0.33, 0.30 ER 76.9</td>
<td>149 PCD analysed: EM (n=143) or genetics (n=6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DNA (n=87) ODA (n=5) ODA+IDA (n=28) IDA+MTD (n=23) CA (n=5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Study</td>
<td>Institution(s)</td>
<td>instrument(s)</td>
<td>PCD Analysed</td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td><strong>Leigh (other sites), 2013</strong> (17)</td>
<td>Sievers 280, EcoPhysics CLD88, NIOX Flex</td>
<td>0.50, 0.33, 0.30</td>
<td>ER 76.9</td>
</tr>
<tr>
<td><strong>Confirmed but not disclosed (n=6)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>-Unknown total number of genes tested</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mateos Coral, 2011</strong> (15)</td>
<td>EcoPhysics CLD88</td>
<td>0.33</td>
<td>ER 58.5</td>
</tr>
<tr>
<td><strong>Noone, 2014</strong> (34) (30)</td>
<td>Sievers 270B</td>
<td>0.50</td>
<td>BH 100</td>
</tr>
<tr>
<td><strong>Papon, 2012</strong> (31)</td>
<td>EVA4000</td>
<td>per ATS standards</td>
<td>per ATS standards 100</td>
</tr>
<tr>
<td><strong>Piacentini, 2008</strong> (32)</td>
<td>NIOX Flex</td>
<td>0.30</td>
<td>Mainly BH, 2 PCD via TB were excluded</td>
</tr>
<tr>
<td><strong>Pifferi, 2011</strong> (33)</td>
<td>EcoPhysics CLD88</td>
<td>0.33</td>
<td>Mainly ER, 8 PCD via TB were excluded</td>
</tr>
<tr>
<td>Study</td>
<td>Method</td>
<td>Tidal breathing</td>
<td>Breath hold</td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>-----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Santamaria, 2008 (35)</td>
<td>NIOX Flex</td>
<td>0.28</td>
<td>BH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wodehouse, 2003 (34)</td>
<td>LR2000</td>
<td>0.25</td>
<td>BH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ER: exhalation against resistance BH: breath hold, TB: tidal breathing
CA: Central apparatus defect; IDA+MTD: Inner dynein arm + microtubule disorganization defect; ODA: Outer dynein arm defect; ODA+IDA: Outer dynein arm + Inner dynein arm defect;
*All information in *italics* are from personal communication with the authors
**Studies considered as using a combination of EM and/or genetics as the reference standard
Table 3: Summary of findings table including the 7 studies comparing nNO to an extended reference standard of EM and/or genetics

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No of studies (No of patients)</th>
<th>Study design</th>
<th>Factors that may decrease quality of evidence</th>
<th>Effect per 100 patients tested</th>
<th>Test accuracy QoE</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Risk of bias</td>
<td>Indirectness</td>
<td>Inconsistency</td>
<td>Imprecision</td>
<td>Publication bias</td>
<td>pre-test probability of 35%</td>
</tr>
<tr>
<td>True positives (patients with PCD)</td>
<td>7 studies 423 patients</td>
<td>cohort &amp; case-control type studies</td>
<td>serious &quot;a&quot;</td>
<td>not serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>False negatives (patients incorrectly classified as not having PCD)</td>
<td>1 (0 to 4)</td>
<td>CRITICAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negatives (patients without PCD)</td>
<td>7 studies 636 patients</td>
<td>cohort &amp; case-control type studies</td>
<td>serious &quot;a&quot;</td>
<td>not serious</td>
<td>not serious</td>
<td>not serious</td>
</tr>
<tr>
<td>False positives (patients incorrectly classified as having PCD)</td>
<td>2 (1 to 10)</td>
<td>IMPORTANT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inconclusive</td>
<td>7 studies 27 patients</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

a. 4 studies were case-control studies from which one study included only healthy patients in the control group. 2 studies did not pre-specify the nNO cut-off before performing measurements while not being blinded to the reference standard.
FIGURES

Figure 1: Summary of evidence search and selection

Figure 2: Assessment of validity of individual studies with QUADAS-2 tool for the 12 included studies. QUADAS-2 tool is designed to assess the quality of primary diagnostic accuracy studies and consists of 4 key domains evaluating the methods used in regard to patient selection, index test, reference standard, and flow of patients through the study and timing of the index tests and reference standard. The results presented here show several studies with high risk of bias in regard to the index test domain, especially in case-control studies.

Figure 3: Forest plot (in ascending order of nNO cut-off value in nL/min)

Figure 4: Summary ROC for the 12 included studies

Figure 5: Assessment of validity of individual studies with QUADAS-2 tool for the 7 included studies comparing nNO to an extended reference standard of EM and/or genetics. QUADAS-2 tool is designed to assess the quality of primary diagnostic accuracy studies and consists of 4 key domains evaluating the methods used in regard to patient selection, index test, reference standard, and flow of patients through the study and timing of the index tests and reference standard. The results presented here show that the 7 selected studies were at lower risk of bias and concerns regarding applicability as compared to the initial 12 analyzed studies presented in Figure 2.

Figure 6: Summary ROC for the 7 studies comparing nNO to an extended reference standard of EM and/or genetics
Supplemental Material

Appendix 1: Initial search strategy for PCD articles

Appendix 2: PRISMA 2009 checklist

Appendix 3: Fagan normogram - Assuming a 35% pre-test probability (in blue, based upon prevalence data in several large PCD cohort studies (17, 18, 29, 31, 33)) and a 10% pre-test probability (in red) with the corresponding post-test probabilities for a pooled PLR of 24.3 and NLR of 0.03.


