EUROPEAN RESPIRATORY journal
OFFICIAL SCIENTIFIC JOURNAL OF THE ERS

Accuracy of Diagnostic Testing for Primary Ciliary Dyskinesia

Journal:	*European Respiratory Journal*
Manuscript ID	Draft
Manuscript Type:	Correspondence
Date Submitted by the Author:	n/a
Complete List of Authors:	Lucas, Jane; University of Southampton, Clinical and Experimental Sciences Dimitrov, Borislav ; University of Southampton and University Hospital Southampton NHS Foundation Trust, NIHR Southampton Respiratory Biomedical Research Unit,; University of Southampton, 5. Primary Care and Population Sciences, Faculty of Medicine Behan, Laura; University Hospital Southampton NHS Foundation Trust, Clinical & Experimental Sciences Academic Unit; University College Cork, School of Applied Psychology Kuehni, Claudia; University of Bern, Institute of Social and Preventive Medicine
Key Words:	primary ciliary dyskinesia, Diagnostic methods, epidemiology

Accuracy of Diagnostic Testing for Primary Ciliary Dyskinesia: Response to correspondence from Israel Amirav and Patrick Boussuyt

Authors: Jane S Lucas (1,2,3) Borislav D Dimitrov (3,4), Laura Behan (1,2,3), Claudia E Kuehni (5)

1. Primary Ciliary Dyskinesia Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK
2. Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK
3. NIHR Southampton Respiratory Biomedical Research Unit, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK
4. Primary Care and Population Sciences, Faculty of Medicine, University of Southampton, Southampton, UK
5. Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland

(guidelines: 800 words plus 1 table; 10 references **currently** 937 words, no table 7 references)

We thank very much Drs Amirav and Boussuyt for their interest in our manuscripts [1, 2]. We agree that our two studies have some limitations, caused by the lack of a "gold standard" test for diagnosing PCD, but we strongly disagree with their claim that we "did not notify the readers of these design deficiencies". Instead, we had taken great care to highlight these uncertainties and risks of bias.

Our first manuscript evaluates the accuracy of different tests used to diagnose PCD: nasal nitric oxide (nNO), high-speed video microscopy analysis (HSVMA) and transmission electron microscopy (TEM), used alone or in combination, against an expert diagnostic consensus based on information from all available tests. [1] This was clearly described in the methods section; and several sensitivity analyses tested the robustness of our results by varying different parameters. Most notably, we began the discussion section [1] by describing the lack of a single 'gold reference standard' (second paragraph) as the major limitation of our study. We highlighted that we used a surrogate reference standard, which was an expert multidisciplinary consensus, based on results from all available diagnostic tests. We also cautioned that since each test contributed to the final diagnostic decision, our sensitivity and specificity estimates of the single tests might be overestimated.

Their letter is helpful as it highlights the challenges of investigating diagnostic accuracy for diseases where there is no 'gold standard'. This is typical for many diseases, including rare ones like PCD[3], and common ones like asthma. It complicates research, but should not impede it; or else we will never progress. Guidance recommends that in situations without a 'gold standard' researchers can consider constructing a reference standard from multiple test results or use an imperfect reference[4]. In our clinical practice, a multidisciplinary panel of specialists considers results of multiple tests to develop a consensus diagnosis of PCD, based on pre-determined rules. In the paper, we used this composite diagnostic outcome as the study's reference standard. Pre-specified rules for deciding the composite diagnostic outcome makes the method transparent and easy to use.

An alternative approach[4] is to use an imperfect reference standard and then adjust the calculated sensitivity and specificity based on existing data about the imperfections. For example, we could have used TEM alone as a single reference standard, and calculated the accuracy of the other tests compared to TEM, taking into consideration that the latter has excellent specificity but limited sensitivity (70-90%). For completeness, we did this for TEM and for HSVMA (online supplementary table S2 [1]), but given that this was not our primary approach and that pre-existing data about the degree of imperfection were highly variable, we decided not to make any adjustments.

As discussed in our manuscript[1] and the accompanying editorial by Haarman and Schmidts[5], generalisability of our findings should be considered with caution. Accurate analysis of ciliary function by HSVMA and of ultrastructure by TEM depends on expertise. The UK PCD reference centres regularly audit each other's analyses, and discuss difficult cases[6]. We routinely analyse *de novo* cilia following culture of the original sample at air liquid interface allowing us to differentiate primary and secondary functional and structural defects [7]. These methods are technically demanding and not available in many centres. In centres without similar infrastructure subtle cases of PCD are more likely to be missed, and secondary defects incorrectly attributed to PCD; even with these facilities, we may misdiagnose some patients. Introduction of new tests is likely to advance the accuracy of diagnostic decisions; we did not include data on genotype and immunofluorescence because these methods were introduced relatively recently and were not available during the study period.

The second manuscript that Amirav and Boussuyt refer to describes the PICADAR tool, a 7-item prediction rule aimed at identifying patients needing referral for diagnostic testing. [2] The tool uses information on clinical symptoms at presentation to predict the likelihood (or, risk) of a final PCD diagnosis. As diagnostic outcome, we used the same composite reference standard as in the first study [1] based on results from all available diagnostic tests. A standardized clinical history was taken from all referred patients, prior to performing any of the diagnostic tests. As detailed in the manuscript, diagnosis was based on positive test results, but not on the clinical history. Symptoms were not part of the composite diagnosis, so there was no incorporation bias in this study. In addition to the published analyses, as strategy to assess potential model overfitting, we performed a bootstrapping testing of the receiver operator characteristic curves of the derivation population, which indicated an expected shrinkage of <3% (data available from the authors). This suggests that there is no significant overestimation for the ROC curves as produced within the predictive logistic regression models (results not shown). We further published the external validation using an independent patient cohort [2] but pointed out that PICADAR should be further validated, and if necessary modified, in different study populations, in general respiratory clinics and different countries.

In summary, we agree that current methods for diagnosing PCD and for assessing diagnostic accuracy are imperfect and require further development and scrutiny. However, we confirm that our manuscripts, which were very transparent about diagnostic pathways, diagnostic and statistical methods, and their limitations, are an acceptable way forward whilst we establish better methods. The two publications are not intended as the definitive answer, but as significant steps in the plight to develop an evidence-base to diagnostic testing. Waiting for a 'gold standard' will not allow us to move forward in the near future.

References

1. 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. *European Respiratory Journal* 2016: 47(3): 837-848.
2. 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. *European Respiratory Journal* 2016: 47(4): 1103-1112.
3. Lucas JS, Leigh MW. Diagnosis of primary ciliary dyskinesia: searching for a gold standard. *European Respiratory Journal* 2014: 44(6): 1418-1422.
4. Rutjes A, Reitsma J, Coomarasamy A, Khan K, Bossuyt P. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technology Assessment* 2007: 11(50): 72.
5. Haarman EG, Schmidts M. Accuracy of diagnostic testing in primary ciliary dyskinesia: are we there yet? *European Respiratory Journal* 2016: 47(3): 699-701.
6. Lucas JS, Burgess A, Mitchison HM, Moya E, Williamson M, Hogg C. Diagnosis and management of primary ciliary dyskinesia. *ArchDisChild* 2014: 99(9): 850-856.
7. 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.

BOOTSTRAPING TESTING of ROC Curves in DERIVATION SAMPLE

1) ALL VARIABLES (all, with siblings)

Logistic regression Number of obs = 650
LR chi2(8) = 275.09 Prob > chi2 = 0.0000
Log likelihood = -118.3186 Pseudo R2 = 0.5376

PCDPOS_NEG	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SCBU	1.019719	.0091101	2.19	0.029	1.002019	1.037732
NEO_CHES_SYMP	.9638965	.0236908	-1.50	0.135	.918564	1.011466
FH_PCD_SIB	1.007328	.0080422	0.91	0.360	.9916885	1.023215
SITUS_ABNORM	.9875172	.0422878	-0.29	0.769	.9080175	1.073977
RHINITIS	.9813244	.0418874	-0.44	0.659	.9025671	1.066954
EAR_HEAR_LOSS	360.591	383.6506	5.53	0.000	44.81047	2901.685
Fullterm	4.782844	1.915656	3.91	0.000	2.181511	10.48613
CARDIO_ANOMALY_DETECTED	103.7103	113.3973	4.25	0.000	12.16507	884.1569
_cons	.0250385	.008817	-10.47	0.000	.0125565	.0499286

Note: 0 failures and 6 successes completely determined.

. lroc, nograph

Logistic model for PCDPOS_NEG
number of observations = 650
area under ROC curve = 0.8738

.end of do-file

. * take sample of n=150 out of n=650(4,699) to better illustrate effect of overfitting
. set seed 999 **. sample 160, count** (490 observations deleted)

Logistic regression Number of obs = 152
LR chi2(6) = 6.36 Prob > chi2 = 0.3842
Log likelihood = -33.697042 Pseudo R2 = 0.0862

PCDPOS_NEG	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SCBU	.9798301	.0419005	-0.48	0.634	.9010541	1.065493
NEO_CHES_SYMP	.9754884	.0394524	-0.61	0.539	.9011485	1.055961
FH_PCD_SIB	1.016631	.0127282	1.32	0.188	.991988	1.041887
SITUS_ABNORM	.9918081	.0791381	-0.10	0.918	.8482206	1.159702
RHINITIS	.990847	.0822594	-0.11	0.912	.842055	1.165931
EAR_HEAR_LOSS	1	(omitted)				
Fullterm	3.884731	2.861091	1.84	0.065	.9171787	16.45387
CARDIO_ANOMALY_DETECTED	1	(omitted)				
_cons	.0388896	.0233193	-5.42	0.000	.0120069	.1259612

. lroc, nograph

Logistic model for PCDPOS_NEG
number of observations = 152
area under ROC curve = 0.5813

Logistic regression Number of obs = 650
LR chi2(8) = 275.09 Prob > chi2 = 0.0000
Log likelihood = -118.3186 Pseudo R2 = 0.5376

PCDPOS_NEG	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SCBU	1.019719	.0091101	2.19	0.029	1.002019	1.037732
NEO_CHES_SYMP	.9638965	.0236908	-1.50	0.135	.918564	1.011466
FH_PCD_SIB	1.007328	.0080422	0.91	0.360	.9916885	1.023215
SITUS_ABNORM	.9875172	.0422878	-0.29	0.769	.9080175	1.073977
RHINITIS	.9813244	.0418874	-0.44	0.659	.9025671	1.066954
EAR_HEAR_LOSS	360.591	383.6506	5.53	0.000	44.81047	2901.685
Fullterm	4.782844	1.915656	3.91	0.000	2.181511	10.48613
CARDIO_ANOMALY_DETECTED	103.7103	113.3973	4.25	0.000	12.16507	884.1569
_cons	.0250385	.008817	-10.47	0.000	.0125565	.0499286

Note: 0 failures and 6 successes completely determined.

Van Howelingen and le Cessie heuristic shrinkage estimator

. display " estimated shrinkage factor: " (e(chi2)-8)/e(chi2)*100 "%"

estimated shrinkage factor: 97.091819%

. display " % of fit due to noise: " (1-(e(chi2)-8)/e(chi2))*100 "%"

% of fit due to noise: 2.9081814%

. end of do-file

2) SELECTED VARIABLES (WITHOUT SIBLINGS)

Logistic regression Number of obs = 650
LR chi2(7) = 274.36 Prob > chi2 = 0.0000
Log likelihood = -118.6805 Pseudo R2 = 0.5362

PCDPOS_NEG	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SCBU	1.02036	.00914	2.25	0.024	1.002602	1.038432
NEO_CHES_SYMP	.9654272	.0233445	-1.46	0.146	.9207401	1.012283
SITUS_ABNORM	.9900271	.0422388	-0.23	0.814	.9106074	1.076373
RHINITIS	.9805204	.0426834	-0.45	0.651	.9003319	1.067851
EAR_HEAR_LOSS	347.4145	368.7305	5.51	0.000	43.39304	2781.479
Fullterm	4.641552	1.846216	3.86	0.000	2.128584	10.12129
CARDIO_ANOMALY_DETECTED	100.7376	109.9079	4.23	0.000	11.87146	854.8289
_cons	.0263708	.0090618	-10.58	0.000	.0134469	.0517157

Note: 0 failures and 6 successes completely determined.

. lroc, nograph

Logistic model for PCDPOS_NEG
number of observations = 650
area under ROC curve = 0.8612

.
end of do-file

. * take sample of n=150 out of n=650(4,699) to better illustrate effect of overfitting
. set seed 999
. sample 550, count (100 observations deleted)

Logistic regression Number of obs = 550
LR chi2(7) = 201.82 Prob > chi2 = 0.0000
Log likelihood = -112.55063 Pseudo R2 = 0.4727

PCDPOS_NEG	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SCBU	1.014445	.0114786	1.27	0.205	.9921947	1.037194
NEO_CHES_SYMP	.9696404	.0232385	-1.29	0.198	.925147	1.016274
SITUS_ABNORM	.9876309	.042862	-0.29	0.774	.9070965	1.075315
RHINITIS	.9832708	.0413217	-0.40	0.688	.9055275	1.067689
EAR_HEAR_LOSS	199.3468	213.4436	4.95	0.000	24.44641	1625.561
Fullterm	4.364512	1.746819	3.68	0.000	1.991853	9.563439
CARDIO_ANOMALY_DETECTED	78.08049	85.78838	3.97	0.000	9.063851	672.6239
_cons	.0322767	.0110554	-10.02	0.000	.0164944	.0631598

Note: 0 failures and 4 successes completely determined.

. lroc, nograph

Logistic model for PCDPOS_NEG
number of observations = 550
area under ROC curve = 0.8330
. end of do-file

Logistic regression **Number of obs = 650**
LR chi2(7) = 274.36 Prob > chi2 = 0.0000
Log likelihood = -118.6805 Pseudo R2 = 0.5362

PCDPOS_NEG	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
SCBU	1.02036	.00914	2.25	0.024	1.002602	1.038432
NEO_CHES_SYMP	.9654272	.0233445	-1.46	0.146	.9207401	1.012283
SITUS_ABNORM	.9900271	.0422388	-0.23	0.814	.9106074	1.076373
RHINITIS	.9805204	.0426834	-0.45	0.651	.9003319	1.067851
EAR_HEAR_LOSS	347.4145	368.7305	5.51	0.000	43.39304	2781.479
Fullterm	4.641552	1.846216	3.86	0.000	2.128584	10.12129
CARDIO_ANOMALY_DETECTED	100.7376	109.9079	4.23	0.000	11.87146	854.8289
_cons	.0263708	.0090618	-10.58	0.000	.0134469	.0517157

Note: 0 failures and 6 successes completely determined.

Van Howelingen and le Cessie heuristic shrinkage estimator

. display " estimated shrinkage factor: " (e(chi2)-8)/e(chi2)*100 "%"

estimated shrinkage factor: 97.084146%

. display " % of fit due to noise: " (1-(e(chi2)-8)/e(chi2))*100 "%"

% of fit due to noise: 2.9158536%

.
.
end of do-file