Diagnostic testing in primary ciliary dyskinesia

To the Editor:

The primary ciliary dyskinesia (PCD) group in Southampton, UK, led by J.S. Lucas is to be applauded for its recent efforts to improve the accuracy of PCD diagnosis. In two recent articles published by the group in the *European Respiratory Journal* [1, 2], a large number of PCD referral cases were analysed and various conclusions were drawn regarding diagnostic accuracy, as well as derivation of a clinical prediction rule that was suggested to aid in further referral and investigations for PCD.

The major problem with these data, however, is that the authors have not essentially evaluated diagnostic accuracy. In diagnostic accuracy studies, the ability of one or more tests to correctly identify patients with the target condition is typically expressed as the sensitivity and specificity of the test(s). These are estimated by comparing the test results to the results of the reference standard, the best available method for correctly identifying patients with the target condition.

Diagnostic accuracy studies can be at risk of bias and several sources of bias have been described [3, 4]. Studies at risk of bias because of design deficiencies typically overestimate the indicators of test performance but also limit the applicability of the study findings.

Unfortunately, these recent PCD studies are not at low risk of bias. The authors did not notify the readers of two design deficiencies. The first is that the diagnostic criteria in the study by JACKSON *et al.* [2] (the reference criteria) include the index tests measured. This was also done in the prediction rule study [1]. The resulting bias is known as incorporation bias, and is known to lead to inflated estimates of accuracy and to generous conclusions about test performance.

Furthermore, the calculated estimates of specificity and sensitivity were based on the number of participants who were diagnosed as PCD (or not) by the reference criteria. However, as no gold standard exists in PCD, exclusion of the disease (true negative) is almost impossible to achieve in that situation. This is exemplified in genetic tests, where a negative finding of a PCD-causing mutation does not necessarily confirm that the patient has no PCD, as only 60–70% of the mutations are known.

We feel that the best (and probably only) diagnostic performance indicator that can be meaningfully discussed in these studies is the positive predictive value of the various tests. To associate these tests with the conventional accuracy statistics, sensitivity and specificity, we still need to wait, as indeed suggested (for other reasons) by HAARMAN and SCHMIDTS [5] in their editorial accompanying one of the papers. Until we have some reference tests that are better than the current bronze or silver standards, diagnostic accuracy cannot be meaningfully evaluated or claimed.

@ERSpublications
Until we have better reference tests for PCD, diagnostic accuracy cannot be meaningfully evaluated or claimed http://ow.ly/4n6VrI

Israel Amirav[1] and Patrick M. Boussuyt[2]
[1]Pediatric Dept, University of Alberta, Edmonton, AB, Canada. [2]Dept of Clinical Epidemiology and Biostatistics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

Correspondence: Israel Amirav, Pediatric Dept, University of Alberta, Edmonton, AB, T6G 2C6, Canada.
E-mail: amirav@ualberta.ca

References

1   Behan L, Dimitrov BD, Kuehni CE, *et al.* PICADAR: a diagnostic predictive tool for primary ciliary dyskinesia. *Eur Respir J* 2016; 47: 1103–1112.
2   Jackson CL, Behan L, Collins SA, *et al.* Accuracy of diagnostic testing in primary ciliary dyskinesia. *Eur Respir J* 2016; 47: 837–848.
3   Lijmer JG, Mol BW, Heisterkamp S, *et al.* Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; 282: 1061–1066.
4   Whiting P, Rutjes AW, Reitsma JB, *et al.* Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004; 140: 189–202.

5   Haarman EG, Schmidts M. Accuracy of diagnostic testing in primary ciliary dyskinesia: are we there yet? *Eur Respir J* 2016; 47: 699–701.

From the authors:

We thank very much I. Amirav and P.M. 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 primary ciliary dyskinesia (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 evaluated the accuracy of different tests used to diagnose PCD: nasal nitric oxide, 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 by describing the lack of a single "gold reference standard" (second paragraph) as the major limitation of our study [1]. 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 correspondence 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 the 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 make the method transparent and easy to use.

An alternative approach is to use an imperfect reference standard and then adjust the calculated sensitivity and specificity based on existing data about the imperfections [4]. For example, we could have used TEM alone as a single reference standard and calculated the accuracy of the other tests compared with 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 (see online supplementary table S2 in [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], the 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 the 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 seven-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 standardised 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 bootstrapping testing of the receiver operator characteristic (ROC)