



# No need for pulmonologists to interpret pulmonary function tests

Christophe Delclaux<sup>1,2,3</sup>

**Affiliations:** <sup>1</sup>AP-HP, Hôpital Robert Debré, Service de Physiologie Pédiatrique – Centre Pédiatrique des Pathologies du Sommeil, Paris, France. <sup>2</sup>Université Paris Diderot, Sorbonne Paris Cité, Faculté de Médecine, Paris, France. <sup>3</sup>INSERM – UMR1141, Paris, France.

**Correspondence:** Christophe Delclaux, Service de Physiologie Pédiatrique, Hôpital Robert Debré, 48, boulevard Sérurier, 75019 Paris, France. E-mail: christophe.delclaux@aphp.fr

 @ERSpublications

The interpretation of pulmonary function tests can be easily automated, but whether artificial intelligence is warranted for diagnosis remains open to debate, even if this aid seems already available <http://bit.ly/2x93kQB>

**Cite this article as:** Delclaux C. No need for pulmonologists to interpret pulmonary function tests. *Eur Respir J* 2019; 54: 1900829 [<https://doi.org/10.1183/13993003.00829-2019>].

The interpretation of pulmonary function tests (PFTs) is an everyday task for every pulmonologist. This interpretation relies on simple rules that were standardised in 2005 by the American Thoracic Society (ATS)/European Respiratory Society (ERS) task force [1]. This task force pointed out that the definition of abnormal patterns (obstructive, restrictive, and mixed defects) would logically be based on a statistical approach; namely, an obstructive defect is a forced expiratory volume in 1 s (FEV<sub>1</sub>)/vital capacity (VC) ratio <5th percentile (or a z-score of FEV<sub>1</sub> <−1.645) and a restrictive defect is a total lung capacity (TLC) <5th percentile (or a z-score of TLC <−1.645). It must be emphasised that hyperinflation was not defined by the task force, probably due to the difficulty in reaching a common definition for both static and dynamic hyperinflations. Computers can easily make these functional diagnoses based on the availability of z-scores, but one may wonder if it is necessary given the simple rules of interpretation. Thus, almost 100% of pulmonologists would give a correct PFT interpretation, at least in university centres that are responsible for medical student teaching. However, it must be pointed out that a FEV<sub>1</sub>/VC ratio <5th percentile is not necessarily abnormal; it is only atypical because 5% of the normal subjects exhibit this atypical feature. It is up to the physician interpreting the functional exploration to provide this kind of nuance depending on the clinical context. These interpretation difficulties have previously been discussed by some authors [2].

To arrive at a disease diagnosis, the results of PFTs are combined with patient information, symptoms and, often, the results of other tests. A Belgian study demonstrated that expert panels could reach 77% accuracy when predicting the diagnosis based on PFTs and clinical history alone [3]. Interestingly, in the latter study, the authors also demonstrated the added value of spirometry of airway resistance, static lung volumes, and diffusing capacity of the lung for carbon monoxide measurements to obtain the final diagnosis of patients with respiratory symptoms [3]. It stresses that experts were able to distinguish subtle abnormalities or combinations of abnormalities that are probably poorly described by the simplified algorithm of assessing lung function in clinical practice given by the ATS/ERS task force [1]. This issue is an avenue for artificial intelligence. The success of deep learning has been shown mainly in problems with inputs of image data, as shown in medical image analysis, speech recognition and board game playing. In pulmonary medicine, artificial intelligence has been used to diagnose respiratory sounds, which may be

---

Received: April 26 2019 | Accepted after revision: May 29 2019

Copyright ©ERS 2019

useful given the large interobserver variability of auscultation diagnosis [4, 5]. More recently, in the same Belgian pulmonary function study, investigators showed that while the ATS/ERS algorithm resulted in a correct diagnostic label in 38% of their 968 subjects, an unbiased machine learning framework integrating lung function with clinical variables improved the general accuracy to 68% [6]. Nevertheless, deep learning lacks explanatory power; deep neural networks cannot explain how a diagnosis is reached and the features enabling discrimination are not easily identifiable [7]. Replacing the doctor with an intelligent medical robot is a recurring theme in science fiction; whether it adds value in the interpretation of PFTs remains open to debate.

The study by TOPALOVIC *et al.* [8], reported in a recent issue of the *European Respiratory Journal*, demonstrates that their artificial intelligence-based software perfectly matched the PFT pattern interpretations (100%) and assigned a correct diagnosis in 82% of all cases, while pattern recognition of PFTs by pulmonologists matched the guidelines in 74.4% of the cases and correct diagnoses were made in 44.6% of the cases.

The first issue deals with PFT pattern interpretation. There is no need for artificial intelligence to ensure a 100% pattern recognition that is based on a simple algorithm. More surprisingly, pulmonologists were not able to reach such a success rate, even in university centres. It stresses that a lot of the pulmonologists do not apply the 2005 ATS/ERS recommendations, which is in accordance with their inability to apply other guidelines [9–11]. Overall, it may be an argument to support the use of the artificial intelligence system that knows that a FEV<sub>1</sub>/FVC z-score <−1.64 is obstructive and a TLC z-score <−1.64 is restrictive. When looking more precisely at the results of this study, the authors show that out of the misclassified normal patterns falsely labelled as an obstructive pattern, 76% were related to four cases having an FEV<sub>1</sub>/FVC ratio just above the normal limits but still below the 0.70 fixed cut-off for positive labelling. It emphasises that the arbitrary definition of chronic obstructive pulmonary disease (FEV<sub>1</sub>/FVC <0.70 after bronchodilation) proposed by the Global Initiative for Chronic Obstructive Lung Disease consensus is a source of confusion for many pulmonologists, which may seem expected. The study also demonstrates that the identification of a restrictive pattern was a more difficult task, which emphasises that the 2005 definition is not used by most pulmonologists. So, a software analysis may be warranted because it may seem unrealistic to assume that the rules of PFT interpretation will be correctly applied by 100% of the pulmonologists. Another complementary approach to improve compliance is by regularly disseminating information about these “new” rules through relevant journals [12, 13].

Finally, it should be interesting to assess whether the interpretation given by the software is correct in classical traps of PFTs such as the small airway obstructive pattern. This pattern, as emphasised by the ATS/ERS task force, is defined by a concomitant decrease (z-scores <−1.645) in FEV<sub>1</sub> and VC with a normal FEV<sub>1</sub>/VC ratio (z-score ≥1.645) while static lung volumes, measured by plethysmography, are normal or even increased (z-scores >+1.645) [1]. Again, computers will be able to make this diagnosis based on z-scores, which is not uncommon in a PFT unit [14], and that is another argument to support additional software analysis for more difficult diagnoses. Moreover, software analysis might further standardise the classification of defect severities according to the ATS/ERS recommendations [1], which was not investigated in the study by TOPALOVIC *et al.* [8]. One may hypothesise that pulmonologists do not comply with the recommended classification of these defect severities.

The second issue deals with the clinical diagnosis; the artificial intelligence-based software assigned a correct diagnosis in 82% of all cases, which is the rate previously obtained by experts in Belgium [3], as compared to a ~45% success rate for pulmonologists in this study. Nevertheless, as stated by the authors “the further usefulness of the artificial intelligence software will be demonstrated if it decreases the time to final diagnosis, reduces the number of tests needed for a final diagnosis and, if by standardising PFT interpretation, a number of misdiagnoses can be avoided.” Like everyone, I wondered what this artificial intelligence software would bring me in my daily practice, knowing that most of the time, I practice PFTs in the framework of pathologies already identified and so they are rarely diagnostic. In addition, the diagnostic problem in pulmonary medicine rarely arises in the face of PFT results alone since other tests are often required for the final diagnosis.

Over the past decade, machine learning techniques have made substantial advances in many domains. There have been suggestions that machine learning will drive changes in healthcare within a few years, specifically in medical disciplines that require more accurate prognostic models and those based on pattern recognition, such as radiology and pathology. Nevertheless, a major issue related to the incorporation of artificial intelligence aids in medicine could be the overreliance on the capabilities of automation. Although the phenomenon of overreliance on technology could be tempting to users in the short term for the convenience and efficiency of automated aids, in the long term these tools can lead to the related phenomenon of de-skilling. When some or all components of a task are partly automated then there is a

consequent reduction in the level of skills required to complete the task, and this may lead to serious disruptions of performance or inefficiencies whenever the technology fails or breaks down.

As a consequence, is pattern recognition by software useful? I would respond yes, because of the inability of all pulmonologists to learn the rules of interpretation. However, is clinical diagnosis aided by artificial intelligence warranted? I would respond that making a diagnosis after a single set of PFTs is not the main issue. Certainly, artificial intelligence will perform better than pulmonologists in suggesting a diagnosis, but its ability to reach the final diagnosis more rapidly or more efficiently remains to be demonstrated.

At this time, artificial intelligence commands a certain level of fascination in the medical world; there is no medical field that seems to escape it. In the field of pulmonary medicine, the future challenge for artificial intelligence is to answer complex clinical questions that are real healthcare issues.

Conflict of interest: None declared.

## References

- 1 Pellegrino R, Viegi G, Brusasco V, *et al*. Interpretative strategies for lung function tests. *Eur Respir J* 2005; 26: 948–968.
- 2 Berry CE, Wise RA. Interpretation of pulmonary function test: issues and controversies. *Clin Rev Allergy Immunol* 2009; 37: 173–180.
- 3 Decramer M, Janssens W, Derom E, *et al*. Contribution of four common pulmonary function tests to diagnosis of patients with respiratory symptoms: a prospective cohort study. *Lancet Respir Med* 2013; 1: 705–713.
- 4 Pasterkamp H, Kraman SS, Wodicka GR. Respiratory sounds: advances beyond the stethoscope. *Am J Respir Crit Care Med* 1997; 156: 974–987.
- 5 Bokov P, Mahut B, Flaud P, *et al*. Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population. *Comput Biol Med* 2016; 70: 40–50.
- 6 Topalovic M, Laval S, Aerts J-M, *et al*. Automated interpretation of pulmonary function tests in adults with respiratory complaints. *Respiration* 2017; 93: 170–178.
- 7 Tarassenko L, Watkinson P. Artificial intelligence in health care: enabling informed care. *Lancet* 2018; 391: 1260.
- 8 Topalovic M, Das N, Burgel P-R, *et al*. Artificial intelligence outperforms pulmonologists in the interpretation of pulmonary function tests. *Eur Respir J* 2019; 53: 1801660.
- 9 Visentin E, Nieri D, Vagaggini B, *et al*. An observation of prescription behaviors and adherence to guidelines in patients with COPD: real world data from October 2012 to September 2014. *Curr Med Res Opin* 2016; 32: 1493–1502.
- 10 Burgel P-R, Deslée G, Jebrak G, *et al*. Real-life use of inhaled corticosteroids in COPD patients versus the GOLD proposals: a paradigm shift in GOLD 2011? *Eur Respir J* 2014; 43: 1201–1203.
- 11 Grimshaw JM, Thomas RE, MacLennan G, *et al*. Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technol Assess* 2004; 8: iii–iv, 1–72.
- 12 Similowski T, Straus C. [Pulmonary function testing: dissemination in French of the European recommendations]. *Rev Mal Respir* 2007; 24: 2S5–2S7.
- 13 Bokov P, Delclaux C. [Interpretation and use of routine pulmonary function tests: spirometry, static lung volumes, lung diffusion, arterial blood gas, methacholine challenge test and 6-minute walk test]. *Rev Médecine Interne* 2016; 37: 100–110.
- 14 Chevalier-Bidaud B, Gillet-Juvin K, Callens E, *et al*. Non specific pattern of lung function in a respiratory physiology unit: causes and prevalence: results of an observational cross-sectional and longitudinal study. *BMC Pulm Med* 2014; 14: 148.