Algorithm for Automatic Forced Spirometry Quality Assessment: Technological Developments

Umberto Melia[1]*, Felip Burgos[2,3], Montserrat Vallverdú[1], Filip Velickovski[4,5,6], Magí Lluch-Ariet[4,5], Josep Roca[2,3], Pere Caminal[1]

1. Dept. d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial (ESAII), Centre for Biomedical Engineering Research (CREB), Universitat Politècnica de Catalunya, CIBER of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Barcelona, Spain, **2.** Department of Pulmonary Medicine. Hospital Clínic de Barcelona (ICT). IDIBAPS, Universitat de Barcelona, Barcelona, Spain, **3.** Centro de Investigación en Red de Enfermedades Respiratorias (CibeRes), Palma de Mallorca, Spain, **4.** Barcelona Digital Technology Centre, Barcelona, Spain, **5.** Dept. d'Enginyeria Telemàtica (ENTEL), Universitat Politècnica de Catalunya, Barcelona, Spain, **6.** ViCOROB, Universitat de Girona, Girona, Spain

*umberto.melia@upc.edu

Abstract

We hypothesized that the implementation of automatic real-time assessment of quality of forced spirometry (FS) may significantly enhance the potential for extensive deployment of a FS program in the community. Recent studies have demonstrated that the application of quality criteria defined by the ATS/ERS (American Thoracic Society/European Respiratory Society) in commercially available equipment with automatic quality assessment can be markedly improved. To this end, an algorithm for assessing quality of FS automatically was reported. The current research describes the mathematical developments of the algorithm. An innovative analysis of the shape of the spirometric curve, adding 23 new metrics to the traditional 4 recommended by ATS/ERS, was done. The algorithm was created through a two-step iterative process including: (1) an initial version using the standard FS curves recommended by the ATS; and, (2) a refined version using curves from patients. In each of these steps the results were assessed against one expert's opinion. Finally, an independent set of FS curves from 291 patients was used for validation purposes. The novel mathematical approach to characterize the FS curves led to appropriate FS classification with high specificity (95%) and sensitivity (96%). The results constitute the basis for a successful transfer of FS testing to non-specialized professionals in the community.

Introduction

Forced spirometry (FS) testing aims at a global assessment of lung and chest wall mechanics. Specifically, FS provides measurements of expiratory volume and flow during a maximal expiratory manoeuvre. It is a relevant test in the clinical setting useful to perform both diagnosis and assessment of functional reserve in various lung-related health disorders. The test is also used for pre-operative evaluation and assessment of disability/impairment. Moreover, there is evidence that key spirometric indices (FVC, forced vital capacity; and, FEV1, forced expiratory volume during the first second) predict survival in the general population. For all these reasons, it is forecasted that the role of FS testing will expand across healthcare tiers and beyond respiratory medicine.

As part of the FS testing procedure, the patient performs maximum expiratory maneuvers under the guidance of a healthcare professional who should: (1) aim for a proper cooperation of the patient; (2) assess the quality of different FS manoeuvres; and, (3) select the most suitable spirometric values using the ATS/ ERS recommendations [1].

The equipment for FS measurements and the recommendations for testing are highly standardized [2,3], as well as the quality assessment [1]. The current systems measure expired flow using different technologies [4] that generate two types of FS curves: (1) a volume-time curve (VT) representing volume (in liters, L) along the ordinate and time (in seconds, s) in the abscissa; and, (2) a flow-volume curve (FV) depicting expired flow (in liters per second) in the ordinate and expired volume (in liters) in the abscissa. Clinically useful spirometric indices (i.e FVC and FEV_1) are calculated from selected curves following the international recommendations [1–3]. FS testing requires a high degree of patient cooperation with the support of a health professional to ensure that the quality of the maneuvers follows the recommended standards [1]. The transfer of FS testing to non-specialized professionals in the community generates a challenge in terms of preserving the quality of the testing to preclude misdiagnosis due to poor quality of FS curves.

There is evidence that remote off-line support of quality testing shows both feasibility and cost-effectiveness, but requires supervision by a specialist [5]. Unfortunately, currently available equipment with functionalities for automatic assessment of quality of FS testing generates poor outcomes due to an inadequate application of the ATS/ERS recommendations on quality control [5,6]. We recently reported the high potential of an automatic algorithm for real-time assessment of quality testing paving the way for the transfer of FS testing to the community [5,7].

It is hypothesized that the regional deployment of a comprehensive program ensuring: *i)* reliable automatic quality assessment of forced spirometric testing; ii) off-line remote assistance to non-specialized health professionals; and, *iii)* accessibility to quality labeled forced spirometric information across healthcare tiers, may have a marked positive impact on quality of diagnosis, healthcare outcomes and may generate cost savings.

The current research reports the identification of new metrics based on a mathematical approach that describes the entire spirometric curve allowing a proper quality assessment of volume-time (VT); flow-time (FT); and, flow-volume (FV) curves. It is of note that the results are under review in the European Patent Office with the registration number (PCT/EP2013/068732).

Materials and Methods

Databases

Three databases were used for building and validating the algorithm: (1) 24 simulated curves recommended by the ATS [2, 3]; (2) 270 curves from 90 patients examined at the Hospital Clinic de Barcelona [5] (P1); and, (3) 778 curves from 291 patients (P2) from one of the Primary Care centers in Barcelona. Forced spirometry testing in P1 and P2 was performed with the same equipment (Sibel 120, SIBELMED, Barcelona Spain). The simulated curves permitted the elaboration of the initial version of the algorithm; whereas the two patient databases (P1 and P2) were considered for refinement and validation purposes, respectively. The study was approved by the Ethical Committee of the Hospital Clínic de Barcelona. All the participants signed informed consent.

Algorithm development

The 24 simulated ATS curves were used to perform a comprehensive characterization of the curve morphology to facilitate the application of the different quality criteria defined in the ATS/ERS recommendations [2, 3]. To this end, three different concepts were introduced as defined below:

Criterion

Specific feature of the spirometric testing that requires quality assessment (i.e. back extrapolation, end-of-curve, peak flow, etc…). The quality analysis of the different criteria considered by the algorithm will provide an overall quality assessment of the spirometric curve.

Metric

Mathematical description of a given criterion. Several criteria may require one or more metrics to be properly defined.

Threshold

Quantitative values of a given metric used to assess the quality of a criterion. It is of note that some metrics may have primary and secondary thresholds.

The ultimate aim of the algorithm was to integrate the new criteria to enhance current quality assessment [1] and to allow on-line quality control of testing.

Each criterion (C_n) defined with the 24 ATS curves used one (primary) or more (secondary) metrics (M_j) with the respective threshold(s). In each step of the

algorithm development, the results were compared with the criteria of one expert in the field of pulmonary function testing.

In order the perform a global assessment of each spirometric curve, five different zones were identified in the flow-volume graph, as indicated in Fig. 1 and described in detail in S1 File. The overall result of applying the new methodology was the identification of three different quality grades, namely: i) Grade 0 → curve to be rejected because of a bad morphology; ii) Grade 1 → curve with acceptable morphology; and, iii) Grade 2 → curve requiring specialized professional judgment for acceptability.

The refinement of the threshold values was performed in an iterative process to maximize the agreement between the human expert and the automatic classification in grades 0 and 1 and to minimize the number of curves automatically classified in grade 2. The first two categories, grades 0 and 1 allow proper real-time and automatic classification of FS; whereas grade 2 requires off-line expert assessment. The automatic grade assignment is made as described in the S1 File.

The algorithm issued from the analysis of the 24 ATS curves was subsequently evaluated using the P1 database following identical procedures.

The current algorithm incorporates the 4 traditional ATS/ERS criteria commonly used in commercially available equipment with automatic quality assessment and applies several other ATS/ERS criteria for quality control of FS as indicated below in the description of the corresponding zones and in Fig. 2. The metrics corresponding to the 4 traditionally used ATS/ERS recommendations in commercially available equipment are defined as follows:

- *BEV* refers to back extrapolated volume (BEV>0.15 L or BEV <5% of FVC), which is the volume value for t=T_{zer} (T_{zero} refers to the back extrapolated time, which is the time at which the volume curve tangent with maximum slope crosses the horizontal time axis).
- *EOTV* refers to the difference between maximum and minimum volume in the last 1 second of exhalation. (T_{ex} refers to the time from T_{zero} to the time in which the VT curve reaches *EOTV* <0.025 L or the end of exhalation, as depicted in Fig. 1A),
- *FET100* refers to the time from T_{zero} to the time in which the VT curve reaches *FVC*, as depicted in Fig. 1A (6 seconds in adult population).
- Repeatability criteria (three good maneuvers, two of them with differences in FVC and FEV_1 less than 0.15 L).

Five spirometric zones

Fig. 2 displays the rationale for the five spirometric zones considered in the current analysis. The first zone (Z1) encompasses the area from zero to peak flow; whereas Z2 relates to the profile or the peak expiratory flow (PEF). The decrease of flow rate after the peak is analyzed in Z3; the end-of-test area is examined in Z4 and, finally, Z5 considers the overall shape of the curve.
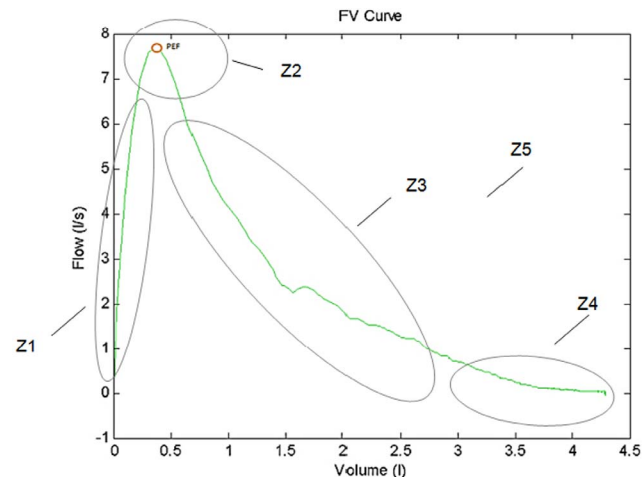
Fig. 1. Spirometer Zones. An example of FV curve with five zones described.

The Z1 criteria ensure that the slope of the curve is regular and free from fluctuations. The calculations are based on the first and second derivatives of the FV curve in zone Z1. Fig. 3 depicts an example of an FV curve that presents a bad morphology in zone Z1. The criterion C_1 detects that irregular concavity or convexity exists. A second criterion C_{2a} detects that the profile has an irregular slope. The criterion C_{2b} detects that the profile has an irregular concavity or convexity.

The Z2 criteria ensure that the PEF occurs at an early point in the maneuver, it has an appropriate height to width ratio and there are no secondary peaks present. Fig. 4 depicts example curves in which the corresponding tests are performed in the zone Z2. Criterion C_3, detects that the PEF point has occurred too late. Criterion C_4 detects that the PEF point is too early. The criteria C_5 analyze the peak. C_{5a} and C_{5b} detect a flat peak (Fig. 4B). The criterion C_{5c} detects a situation of bimodal peaks as depicted in Fig. 4A (multiple peaks). The criterion C_6 detects if the V value in the position of the PEF is lower than a fixed threshold (Fig. 4C).

The Z3 criteria ensure that the slope of the curve is regular and free from fluctuations and are based on the first derivative, and definite integrals of the FV curve in zone Z3. Fig. 5 depicts an example curve in which the corresponding tests are performed in zone Z3. Fig. 5A depicts criterions C_{7a}, C_{7b} and C_{7d}. Fig. 5B depicts criterions C_{7c} and C_{7d}. The criterion C_{7a} detects a situation of high slopes during FV curve descent. The criterion C_{7b} detects an excessive variation in the slope of the FV curve in zone Z3. The criterion C_{7c} detects an excessive variation in the slope calculated in a V segment of the FV curve in the zone Z3. The criterion C_{7d} detects an irregular slope.

The Z4 criteria ensure that the flow at the end of the curve is regular and free from fluctuations and are based on the first order derivative of the FT curve in zone Z4, and the difference between the maximum and minimum volume in the last second of exhalation. Fig. 6 depicts an example curve in which the
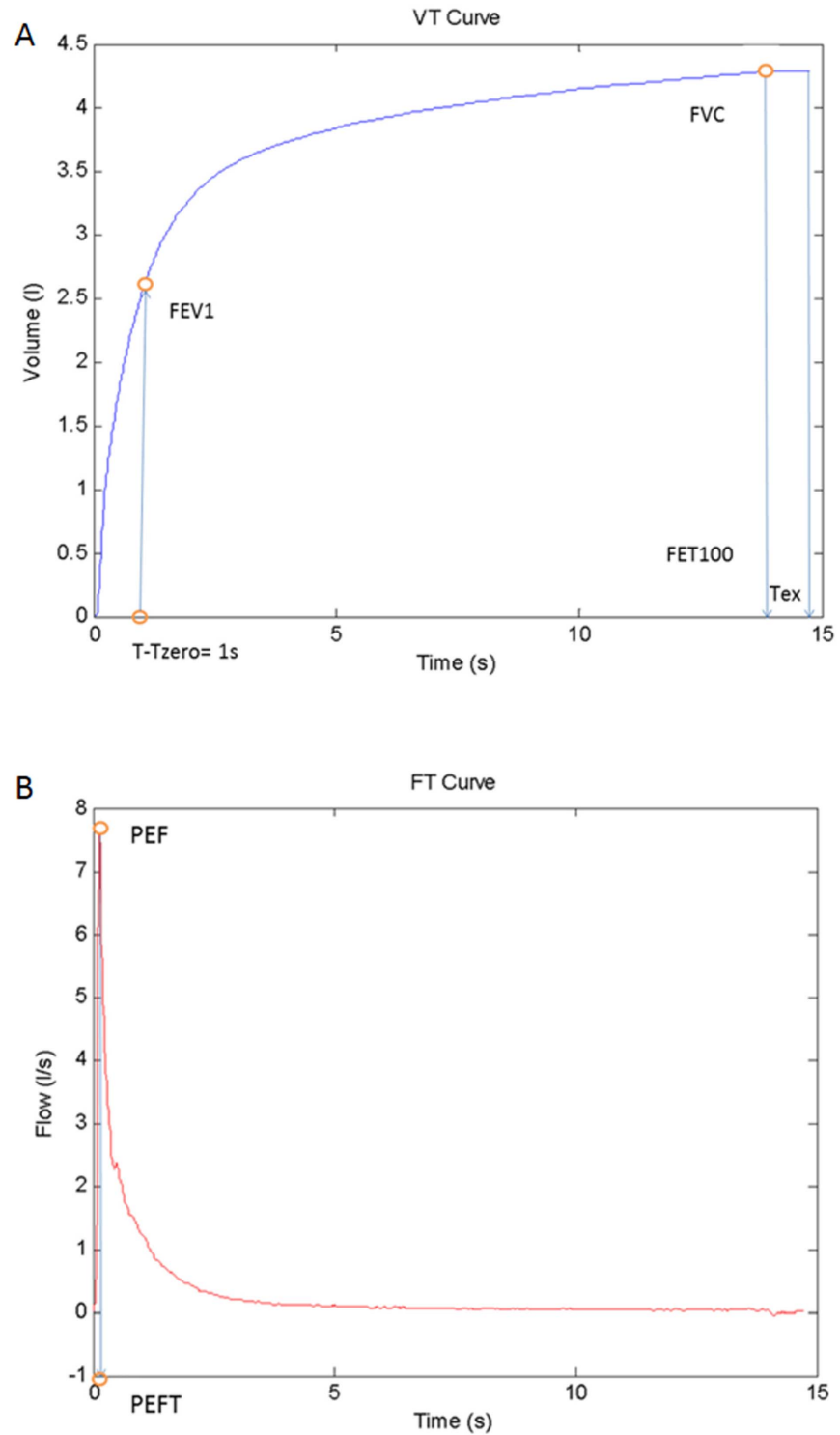
Fig. 2. Spirometer metrics. Metrics involved in the traditional criteria: (A) *FVC* and *FEV1*, (B) *PEFT*.
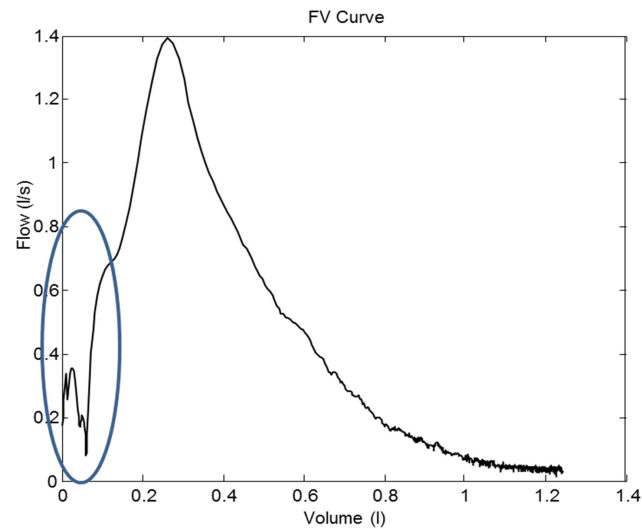
Fig. 3. Zone Z1 analysis. An example of a FV curve that presents irregularity on the ascent to the PEF.

corresponding tests are performed in the zone Z4. Fig. 6A and 6B depict criterion C_{11}. Criteria C_8 is the traditional *BEV* criteria (*BEV*>0.15 L or *BEV* <5% of *FVC*). Criteria C_9 is the traditional *EOTV* criteria (V<0.025 L in $t \geq 1$ s). Criteria C_{10} is a combination of 5 criteria explained in the following lines. C_{10a} detects if the *EOTV* and *Tex* both does not satisfy their traditional criteria. C_{10b} defines a new period to calculate *EOTV* if the *Tex* traditional criteria is satisfied. C_{10c} defines a new threshold for *EOTV* if the *Tex* traditional criterion is satisfied. C_{10d} uses threshold to define the *EOTV(Tex)* and compare with the traditional threshold. C_{10e} defines *EOTV* calculated as a function of *Tex*. Criterion C_{11} detects irregularity or oscillation at the end part of FT curve.

The Z5 criteria ensure that there only exists one local maximum (the *PEF* point) and they are based on the derivative of the FV curve. Fig. 7 depicts criterions C_{12a} and C_{12b}. The criterion C_{12a} detects a situation of multiple peaks that typically occurs when the subject coughs. The criterion C_{12b} detects a situation of multiple peaks for values of V adjacent to FEV_1.

For more details, see S1 File.

Algorithm evaluation

The quality grades (0 to 2) generated by the algorithm using the P2 dataset were compared with those generated by the expert evaluator. Sensitivity and specificity of the algorithm were calculated for all curves classified as classes 0 or 1 in order to quantify the agreement between the algorithm and the evaluator and between the 4 traditional ATS criteria and the evaluator. Sensitivity is defined as the number of curves classified as class 0 by both the algorithm (or the 4 traditional ATS/ERS criteria) and the evaluator divided by the total of the curve evaluated in grade 0 by

Fig. 4. Zone Z2 analysis. Examples of FV curves that present (A) bimodal peak; (B) flat peak and (C) slow peak.

the evaluator, while specificity is defined as the number of curves classified as class 1 by both the algorithm (or the 4 traditional ATS/ERS criteria) and the evaluator divided by the total of the curve evaluated in grade 1 by the evaluator.
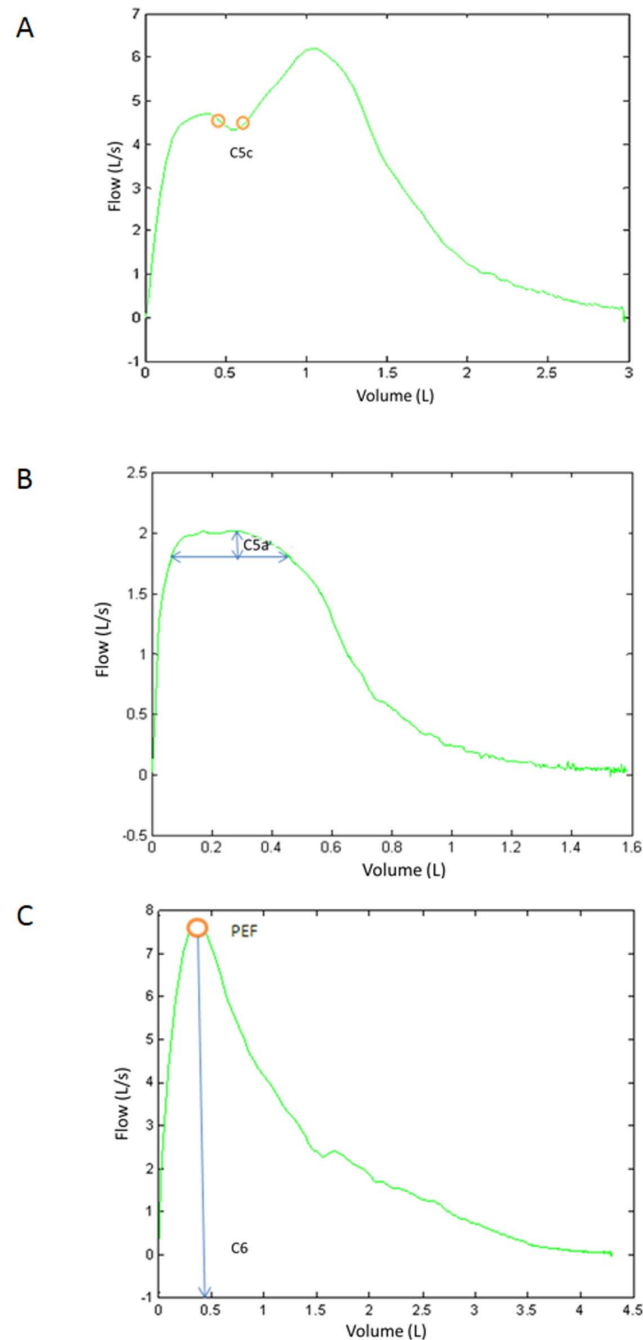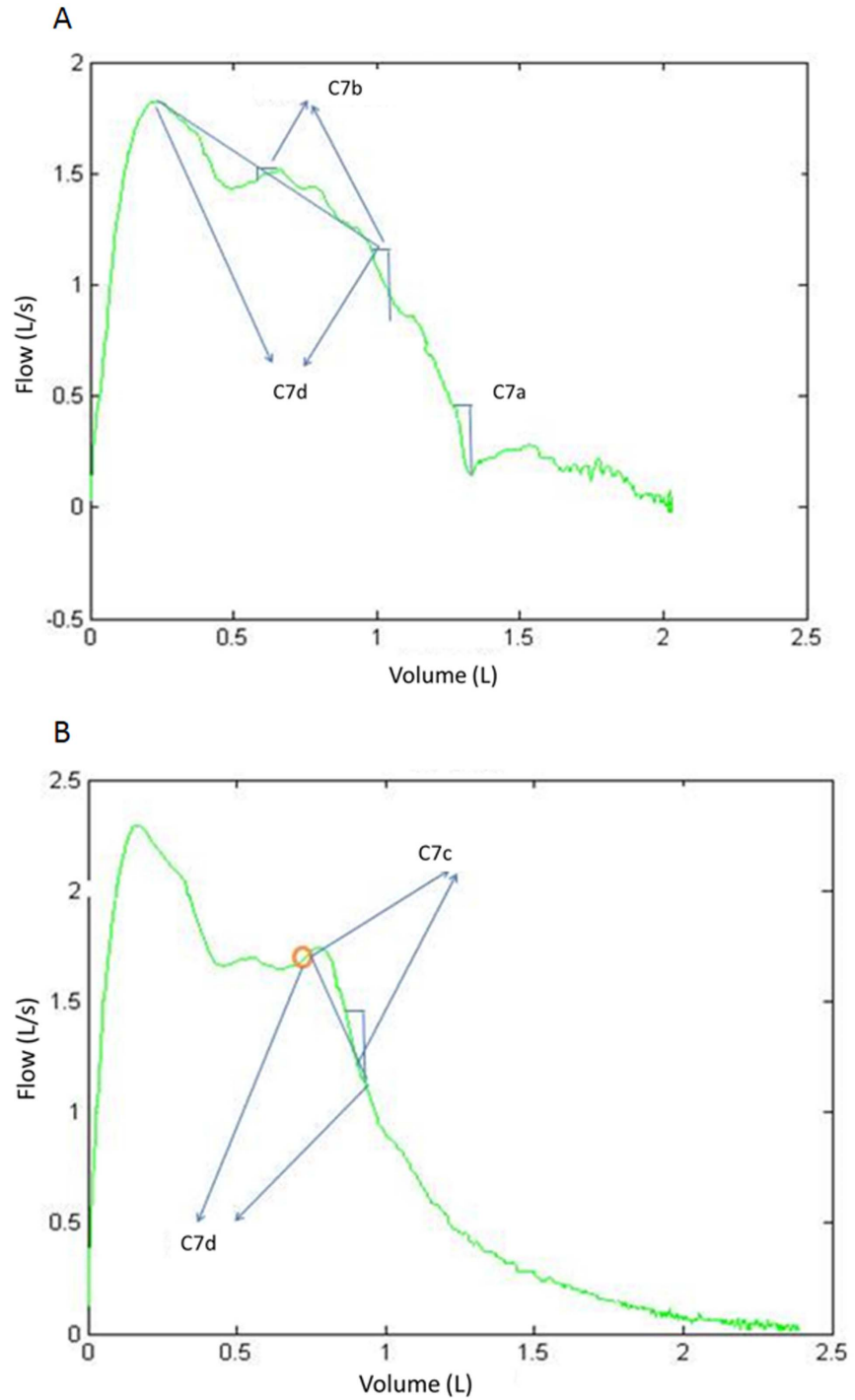
Fig. 5. Zone Z3 analysis. Examples of FV curves that present irregularity in the descent from PEF.
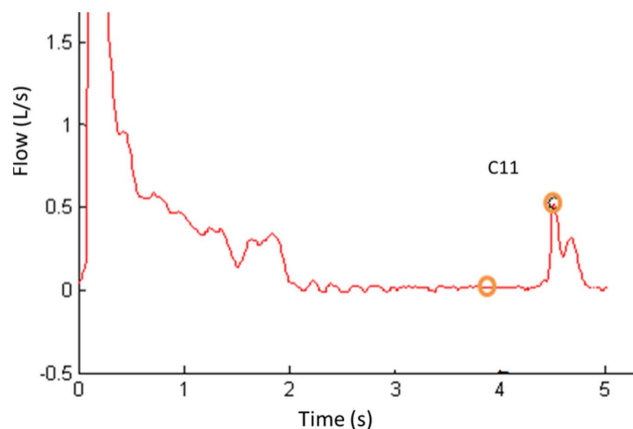
Fig. 6. Zone Z4 analysis. Example of FT curve that present irregularity in the final part.

Results and Discussion

Table 1 summarizes the evaluation of the new algorithm showing that a reasonable percentage of FS curves (88%), could be automatically assessed as either acceptable (grade 1) or unacceptable (grade 0) in concordance to an expert evaluation. Twelve percent of the curves (n=93) were automatically classified as grade 2 requiring an expert opinion. It is of note that 43% of these grade 2 curves were evaluated as grade 0 and 57% as grade 1 by the expert evaluator. The table also compares the results of the current research against those obtained only using the four traditional ATS/ERS criteria for quality assessment. Think

Several alternative technological approaches [8–14] for automatic quality assessment of FS testing were considered during the current study design. But, we
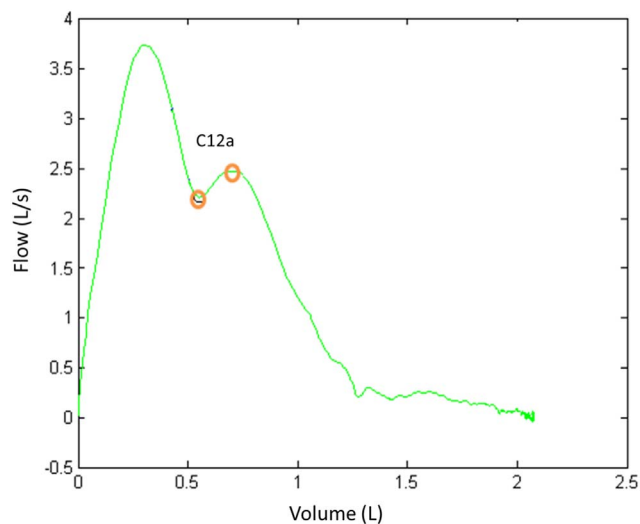

Fig. 7. Zone Z5 analysis. Examples of FV curve with peak and valley.

Table 1. Computed sensitivity (Sen) and specificity (Spe) using the current automatic classification algorithm and using only the four traditional ATS/ERS quality criteria applied to P2.

Automatic Classification Algorithm	Sen: 96.1%
	Spe: 94.9%
Number of Curves Detected in each Grade	*Grade 0*: 266
	Grade 1: 419
	Grade 2: 93
Four traditional ATS/ERS Criteria	Sen: 67.7%
	Spe: 75.0%
Number of Curves Detected in each Grade	*Grade 0*: 320
	Grade 1: 458
TOTAL CURVES ANALYZED: 778	

doi:10.1371/journal.pone.0116238.t001

consider that the method reported allows a comprehensive and efficient application of the different quality control ATS/ERS criteria [1], it does not show limitations in terms of computer requirements, or unnecessary delays using regular computers used in the clinical setting, and it was well accepted by health professionals as on-line clinical decision support systems support during performance of FS testing.

We acknowledge as a limitation of the current study that the algorithm has been developed with the feedback of only one expert. Consequently, despite the positive results reported, there is a need for a formal assessment of the variability among various expert observers. Despite that internal interim data indicates that interobserver variability is not a relevant factor, we are planning its evaluation as part of a large prospective future trial analyzing both clinical and cost saving impact of the regional deployment of the high quality FS program, as described below.

The need for an external, likely centralized, quality control of FS testing [15–16] is widely accepted if based on well-established objective criteria. It is of note, however, that low specificity of any combination of the computer-based quality control criteria using only the four traditional ATS/ERS [3] has been reported [5, 6] such that automatic quality assessment using algorithms incorporated in commercially available equipment cannot replace the visual inspection by an expert. In contrast, our proposed algorithm shows two advantages: it enhances quality control of FS testing and allows on-line assessment of the testing.

Previous reports have indicated the potential of telemedicine to enhance both quality and diagnostic potential of FS testing carried out by non-expert professionals [5, 17–19], but the studies are based on off-line analyses by specialists [20–22]. The findings of the current research suggest that a vast majority of FS testing carried out by non-specialized professionals in primary care could be reliably assessed in real-time. Consequently, the results of the current study refine previous achievements [5] and open the way to explore extensive and efficient adoption of this type of high quality FS programs.

Several factors limiting regional deployment of a quality control program of FS using the current algorithm in the clinical practice are acknowledged, namely: (i) implementation of standardized raw spirometric data transfer through a clinical document architecture (CDA) [23]; (ii) an ICT architecture providing interoperability across healthcare tiers; (iii) design of an educational program for professionals; and, (iv) implementation of incentives fostering professional engagement. The region of Catalonia will be ready in 2015 for the regional deployment of a high quality FS program overcoming the barriers alluded to above. Such a comprehensive program: *i)* will likely have a positive clinical impact on the quality of diagnosis of patients with respiratory disorders, *ii)* should prevent unnecessary duplication of FS testing; *iii)* will likely enhance longitudinal follow-up of patients and support cost-effective preventive strategies aiming at modulating disease progress; *iv)* will pave the way to generate novel approaches to assess abnormal biological variability of FS testing; and, *v)* may likely produce cost savings. No doubt that such a program will require a proper evaluation on a longitudinal basis to assess its potential for generation of healthcare value.

Conclusion

The results of the current study provides a tool that makes operational a comprehensive application of the ATS/ERS recommendations for automatic quality control of FS testing. It constitutes a pivotal element facilitating the design and future deployment of a high quality FS program based on remote automatic evaluation of the testing.

Supporting Information

S1 File. Metrics and flow-chart. This file provides a detailed description of the metrics and the decision process used in each of the zone (Z1, Z2, Z3, Z4 and Z5) in order to automatically evaluate the FS curves.
doi:10.1371/journal.pone.0116238.s001 (PDF)

Aknowledgments

Author Contributions

Conceived and designed the experiments: UM FB MV PC JR. Performed the experiments: UM FB. Analyzed the data: UM FB MV PC FV ML JR. Contributed reagents/materials/analysis tools: UM FB MV PC FV ML JR. Wrote the paper: UM FB MV PC FV ML JR.

References

1. **Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, et al.** (2005) Standardisation of spirometry. Eur Respir J. 26: 319–338.

2. **Gardner RM, Hankinson JL** (1988) Standardization of spirometry ATS update American Thoracic Society, J Occup Med March. 30 (3): 272–3.

3. Medical Section of the American Lung Association, Standardization of spirometry: 1994 update (1994), Am Rev Respir Dis. 152: 1107–1136.

4. **Burgos F, Roca J** (2013) Is Plug-and-Play Spirometry the Future? Buyers' Guide To Respiratory Care Products. 14: 11–22.

5. **Burgos F, Disdier C, de Santamaria EL, Galdiz B, Roger N, et al.** (2012) Telemedicine enhances quality of forced spirometry in primary care. Eur Respir J. 39 (6): 1313–8.

6. **Müller-Brandes U, Krämer M, Gappa G, Seitner-Sorge A, Hüls A, et al.** (2014) LUNOKID: can numeric ATS/ERS quality criteria replace visual inspection of spirometry? Eur Respir J. 43: 1347–1356.

7. **Burgos F** (2013) Impact of Information and Communication Technologies on Remote Testing: Forced Spirometry as a User Case. PhD Thesis, University of Barcelona 2013. http://hdl.handle.net/10803/284587.

8. **Sankar AB, Selvi JAV, Kumar D, Lakshmi KS** (2013) Effective enhancement of classification of respiratory states using feed forward back propagation neural networks. Sadhana, 38(3): 377–395.

9. **de Oliveira Camargo-Brunetto MA, Gonçalves AR** (2013) Diagnosing Chronic Obstructive Pulmonary Disease with Artificial Neural Networks using Health Expert Guidelines. In HEALTHINF 1: 207–214.

10. **Devi GM, Venkatesan P** (2013) Least Squares Support Vector Regression for Spirometric Forced Expiratory Volume (FEV1) Values. International Journal of Science and Technology 3: 1.

11. **Polak AG** (1998) A forward model for maximum expiration. Comp. Biol. Med. 28: 613–625.

12. **Pardaens J, van de Woestijne KP, Clément J** (1975) Simulation of regional lung emptying during slow and forced expirations. J. Appl. Physiol. 39: 191–198.

13. **Lambert RK** (1989) A new computational model for expiratory flow from nonhomogeneous human lungs. ASME Trans. J. Biomech. Eng. 111: 200–205.

14. **Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, et al.** (2012) Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. European Respiratory Journal 40(6): 1324–1343.

15. **Enright PL** (2003) How to make sure your spirometry tests are of good quality, Respir Care. 48(8): 773–776.

16. **Enright PL, Skloot GS, Cox-Ganser JM, Udasin IG, Herbert R** (2010) Quality of spirometry performed by 13,599 participants in the World Trade Center Worker and Volunteer Medical Screening Program. Respir Care 55(3): 303–309.

17. **Masa JF, Gonzalez MT, Pereira R, Mota M, Riesco JA, et al.** (2011) Validity of spirometry performed online. Eur Respir J. 37 (4): 911–918.

18. **Bellia V, Pistelli R, Catalano F, Antonelli I, Grassi V, et al.** (2000) Quality control of spirometry in the elderly, Respiratory Health in the Elderly. Am J Respir Crit Care Med. 161: 1094–100.

19. **Spirometry Monitoring Technology - SPIROLA.** NIOSH - Spirometry in Occupational Medicine. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health. www.cdc.gov/niosh/spirometry/spirola.htlm.

20. **de Montes OM, Talamo C, Perez-Padilla R, Jardim JR, Muino A, et al.** (2008) Chronic obstructive pulmonary disease and body mass index in five Latin America cities: the PLATINO study, Respir Med. 102 (5): 642–50.

21. **Janssens W, Liu Y, Liu D, Kesten S, Tashkin DP, et al.** (2013) Quality and reproducibility of spirometry in COPD patients in a randomized trial (UPLIFT), Respir Med. 25.

22. **Perez-Padilla R, Vazquez-Garcia JC, Marquez MN, Menezes AM** (2008) Spirometry quality-control strategies in a multinational study of the prevalence of chronic obstructive pulmonary disease, Respir Care, 53 (8): 1019–26.

23. **Sala T, Rubies C, Gallego C, Muñoz P, Burgos F, et al.** (2011) Technical Requirements of Spirometers in the Strategy for Guaranteeing the Access to Quality Spirometry. Arch Bronconeumol. 47(9): 466–469.