Subjective assessment using still bronchoscopic images misclassifies airway narrowing in laryngotracheal stenosis

Permalink
https://escholarship.org/uc/item/0hx0d769

Journal
Interactive CardioVascular and Thoracic Surgery, 16(5)

ISSN
1569-9293 1569-9285

Authors
Murgu, S.
Colt, H.

Publication Date
2013-02-13

DOI
10.1093/icvts/ivt015

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed
Subjective assessment using still bronchoscopic images misclassifies airway narrowing in laryngotracheal stenosis

Septimiu Murgu* and Henri Colte

a Department of Medicine, Division of Pulmonary and Critical Care Medicine, University of Chicago Pritzker School of Medicine, Chicago, IL, USA
b Department of Medicine, Division of Pulmonary and Critical Care Medicine, University of California, Irvine, CA, USA
c Corresponding author. Department of Medicine, Division of Pulmonary and Critical Care Medicine, University of Chicago Pritzker School of Medicine, 5841 South Maryland, MC 6076, Chicago, IL 60637, USA. Tel: +1-773-7024773; fax: +1-773-7026500; e-mail: tim.murgu@gmail.com; smurgu@medicine.bsd.uchicago.edu (S. Murgu).

Received 22 August 2012; received in revised form 17 November 2012; accepted 27 November 2012.

Abstract

OBJECTIVES: Severity of airway narrowing is relevant to management decision-making processes in patients with laryngotracheal stenosis. Airway lumen is frequently assessed subjectively based on still images obtained during airway examinations or objectively using image analysis software applied to radiological or bronchoscopic images. The purpose of this study was to determine whether strictures classified as mild, moderate or severe degrees of airway narrowing based on subjective assessments by a group of experienced bronchoscopists using still images, matched the classifications derived from morphometric bronchoscopy measurements and whether the results of subjective assessments correlated with the level of bronchoscopic experience.

METHODS: Thirty-five bronchoscopic doublet still images of benign causes of laryngotracheal stenosis containing normal and abnormal airway cross-sectional areas were objectively analysed using morphometric bronchoscopy and classified as mild (<50%), moderate (50–70%) or severe (>70%). These images were then subjectively assessed by 42 experienced bronchoscopists participating in an interventional bronchoscopy course. Descriptive statistics were used to explore the accuracy of the participants’ classifications. Correlation coefficients were used to study the relationship between participants’ subjective assessments and bronchoscopy experience.

RESULTS: Only 47% of strictures were correctly classified by study participants (mean 16.48 ± 2.8). Of the 1447 responses included in this analysis, 755 were incorrect: 71 (9%) were over-classifications of strictures’ severity and 684 (91%) were under-classifications. There was no correlation between number of strictures correctly classified and number of lifetime bronchoscopies or number of strictures seen by bronchoscopists in an average month.

CONCLUSIONS: Experienced bronchoscopists often misclassify the degree of airway narrowing when using still bronchoscopic images to subjectively assess strictures of benign aetiology.

Keywords: Morphometric bronchoscopy • Central airway obstruction • Bronchoscopy • Tracheal stenosis • Laryngotracheal stenosis

INTRODUCTION

The severity of an airway stricture affects airflow, symptoms, functional status and management decisions before and after treatment [1]. The work of breathing depends on the airway pressure drop along the stenosis, which in addition to being affected by the degree of airway narrowing, is subject to the effects of a stricture’s morphology and flow velocity through the stenosis [2]. Airflow dynamic studies show, for example, that a <50% reduction in airway lumen cross-sectional area (CSA) results in a pressure drop of the same magnitude as the normal glottis opening, and is thus unlikely responsible for symptoms [3]. A reduction of more than 70%, on the other hand, results in a significant pressure drop even at low flow velocities, with a resultant impact on symptoms and functional status during minimal activities or at rest [2, 4]. These findings seem to justify classifying strictures as mild, moderate or severe based on the degree of airway narrowing and impact on airflow limitation.

When airway calibre is quantified using still images viewed directly during bronchoscopy or laryngoscopic examination, however, or afterwards using printed photos or digital images stored in electronic medical records, assessments are frequently subjective. Subjective assessments are dependent on airway examination conditions such as patient’s respiratory effort (tidal breathing vs forced respiration), body position (supine vs sitting), neck position (flexion vs extension), distance between the tip of the scope and the stricture, extent and precision of bronchoscopic or laryngoscopic examination and physician’s experience evaluating patients with laryngotracheal stenosis (LTS).

Spirometry is not feasible in very symptomatic, critically ill or uncooperative patients with LTS, does not identify the precise extent or morphology of the stricture within the airway and has low sensitivity for detecting mild to moderate reductions in airway calibre [5, 6]. Radiological studies such as multidetector computed tomography (MDCT) are used, but quantification of airway calibre is difficult in dyspnoeic or uncooperative patients.
experience. Objective assessments correlated with the level of bronchoscopic measurements, and whether the results of subjective assessments using still bronchoscopic images. The purpose of this study, therefore, was to determine whether strictures classified as mild, moderate, or severe degrees of airway narrowing based on subjective assessments made by a group of experienced bronchoscopists using still images, matched the classifications of benign causes of LTS derived from morphometric bronchoscopy measurements, and whether the results of subjective assessments correlated with the level of bronchoscopic experience.

**MATERIALS AND METHODS**

Still bronchoscopic images were obtained from abnormal and normal airway lumen in patients with benign causes of laryngotracheal stenosis. A series of 35 bronchoscopic doublet images consisting of the bronchoscopic image of a benign LTS and of the normal airway distal to the stenosis in the same patient were thus compiled from the authors’ records of 35 different patients (study protocol # 2005–4236 was approved by UC Irvine’s IRB committee). Images were captured using an Olympus BF 160 videobronchoscope positioned 1 cm proximal to the stricture or to a normal cartilaginous ring, respectively. The scope was kept centred so that the entire circumference of the airway lumen was visualized (Fig. 1). Images were saved on CD ROM using a Sony image-capturing device (Sony Digital Capture Unit UPA-P100 MD). Manual morphometric bronchoscopy was then performed by one of the authors (S.M.) on images from each of the 35 strictures. The CSA of the airway lumen was calculated using Image J analysis (software available free of charge at http://rsb.info.nih.gov/ij/). Using the manual method measurement technique, the area to be measured was selected after first opening the image using Image J [12]. This was followed by highlighting the area of interest using the Polygon Selection Tool. By selecting Analyse → Measure, the selected area was measured in pixels (Fig. 2). A Stenosis Index (SI) defined as the CSA of the obstructed area relative to that of normal airway lumen distal to the stenosis (SI = CSAabnormal/CSAnormal × 100%) was then calculated [12]; the greater the SI, the greater the degree of airway narrowing and more severe the obstruction. Based on the degree of airway narrowing, 63% (22 of 35) of strictures were classified as severe (>70% narrowing), 26% (9 of 35) were classified as moderate (50–70% narrowing) and 11% (4 of 35) were classified as mild (<50% narrowing). The aetiology, location, morphology, stenosis index and classification based on the degree of narrowing are presented in Table 1.

After a short instruction session conducted by one of the authors (H.C.), who remained blinded as to the calculated SI and classification based on the degree of narrowing, the 35 doublet images were subsequently shown to a group of 43 bronchoscopists participating in an interventional bronchoscopy course (Barcelona, Spain). Doublet images were projected on the auditorium screen one set at a time, with a move to a new image set.

![Figure 1: Bronchoscopic images of an idiopathic subglottic stricture (A) and the normal distal tracheal lumen (B). Images were obtained with the scope being centred and positioned 1 cm proximal from the regions of interest (arrows) which are clearly visualized in their entire circumference.](image-url)

[7], or in patients with cognitive dysfunction and those hospitalized in the intensive care unit. Imaging results are affected by airway secretions that mimic airway narrowing [8], and significant intrasubject variations in airway lumen measurements across sites and respiratory cycles have been documented [9]. The risks of radiation exposure make these tests poorly suited for repeat testing, especially in younger patients [10], and the need for costly computer hardware and an experienced interpreter of radiographical findings limits their application in daily practice [11]. The role of radiological studies in this condition, however, has to be individualized based on patient- and institution-related factors. For instance, uncooperative patients with stable respiratory status can be administered anxiolytics and/or sedatives to facilitate CT scanning. In addition, experienced radiologists could differentiate secretions from stenosis and the risk of radiation is not relevant in life-threatening situations. Morphometric bronchoscopy uses image software-processing techniques that enable analysis of still digital images in order to measure airway lumen calibre, providing an objective quantification of the degree of airway narrowing in patients with various forms of central airway obstruction [12]. While certain limitations such as lack of real-time image analysis at the time of airway examination, use of calibration markers for absolute airway calibre measurement or lack of familiarity with the technique seem to have prevented this technique from becoming a universally accepted gold standard for quantifying airway calibre, morphometric bronchoscopy has been shown to reliably assess airway calibre in paediatric and adult patients with normal airways and tracheobronchomalacia. Results of studies also demonstrate high inter- and intraobserver reliability and show that measurements obtained by morphometric bronchoscopy correlate well with measurements performed by MDCT [12].

While no studies identify a universally accepted reference standard for assessing airway calibre, it is safe to say that to our knowledge, there are no published data regarding the accuracy of subjective assessments using still bronchoscopic images. The purpose of this study, therefore, was to determine whether strictures classified as mild, moderate or severe degrees of airway narrowing based on subjective assessments made by a group of experienced bronchoscopists using still images, matched the classifications of benign causes of LTS derived from morphometric bronchoscopy measurements, and whether the results of subjective assessments correlated with the level of bronchoscopic experience.
only after all participants had completed their assessments. Based on comparison with the image of the normal airway, participants were asked to independently classify each of the 35 strictures as mild (<50% narrowing), moderate (50–70% narrowing) or severe (>70% narrowing). Participants were also asked to estimate as accurately as possible their total lifetime number of bronchoscopies, and the number of patients with LTS they examined in an average month. All scoring sheets were kept anonymous.

Statistical methods

Results were tabulated using a Microsoft Excel-2007 spreadsheet and analysed using SyStat-13 (2009). Classification results based on the subjective assessments from the participants were compared with those derived from the objective measurements made by morphometric bronchoscopy. Means, standard deviations and ranges were calculated for the total number of correct stricture classifications and for the mild, moderate and severe subsets (i.e. the number of strictures classified as mild, moderate or severe by study participants that matched the classifications derived from morphometric bronchoscopic measurements). For each stricture, the number of correct classifications was also determined. Results were separately plotted for individual participants and for individual strictures in order to distinguish distribution patterns.

Descriptive statistics were used to assess the percentage of erroneous (under- or over-)classifications based on subjective assessments. Non-parametric (Spearman’s Rank correlation coefficient) measures were used to analyse the relationship between the participants’ classification results and level of experience based on the number of lifetime bronchoscopies and number of LTS seen monthly.

RESULTS

All 43 bronchoscopists in this interventional bronchoscopy course participated in the study, but one scoring sheet had to be excluded because it was almost completely blank. The 42 remaining scoring sheets were almost all completed properly, with a single response per image: the rate of blank and double responses was only 1.5% (23 of 1470 images), leaving 1447 responses for analysis. Of the 42 participants who completed the scoring sheets, 69% (29 of 42) had performed more than 1000 lifetime bronchoscopies, 26% (11 of 42) had performed between 200 and 1000 and 5% (2 of 42) had done 200 or less. Seven participants were seeing more than two strictures in an average month.
Forty-seven percent of strictures (mean number of stricture 16.48 ± 2.8) classified as mild, moderate or severe by study participants matched the classifications derived from morphometric bronchoscopic measurements. The participants’ distribution plot was relatively tight, with 79% of participants (33 of 42) correctly classifying between 40 and 57% of strictures (Fig. 3: top panel). Ten of the 35 strictures (29%) were correctly classified by 88–98% of participants; another 10 strictures (29%) were correctly classified by 40–70% of participants; 13 of 35 strictures (37%) were correctly classified by 0–17% of participants and 9 strictures (26%) were correctly classified by ≤5% of participants (Fig. 3: bottom panel).

Misclassifications were sorted by whether participants had under- or over-classified a stricture’s degree of airway narrowing. Of the 1447 responses included in this analysis, 755 were incorrect: 71 (9%) were an over-classification of a stricture’s severity and 684 (91%) were an under-classification. Each over-classification was by one degree (from mild to moderate, or from moderate to severe). Of the 684 under-classifications, 154 (23%) were under-classified by two degrees (from severe to mild) and 530 (77%) by one degree (from severe to moderate or from moderate to mild). Almost all incorrect classifications were in the severe and moderate categories, whereas almost all (161 of 168 or 96%) mild strictures were classified correctly.

There was no correlation between number of strictures correctly classified and number of lifetime bronchoscopies (\(\rho = 0.17\)) or number of strictures seen (\(\rho = 0.09\)); there was minimal correlation between the number of lifetime bronchoscopies with the number of strictures seen per month (\(\rho = 0.23\)). The lack of correlation between percentage of correctly classified strictures and number of bronchoscopies performed or strictures seen was confirmed by Spearman Rank test (\(\rho = 0.21\) and \(\rho = 0.15\), respectively).

**DISCUSSION**

The results of this study demonstrate that subjective assessment using still bronchoscopic images more often than not underclassifies and occasionally over-classifies the degree of airway narrowing in patients with benign causes of LTS, when compared with an objective measurement such as morphometric bronchoscopy. Subjective assessments were accurate for classifying mild strictures, but an under-classification by two degrees from severe to mild airway narrowing was noted for almost a quarter of the total number of incorrect classifications. These findings are relevant to patient care because: (i) alternative causes may be sought for a patient’s shortness of breath if the stenosis is classified as mild, when in actuality it is moderate or severe, potentially delaying care, compromising safety and increasing costs;
ii) if strictures are unrecognized as severe (for example, a severe stricture is erroneously classified as mild or moderate), LTS can become critical in case of subsequent obstruction by mucus production or oedema from upper respiratory tract infection, which can also precipitate respiratory failure; (iii) over-classification (e.g. a mild or moderate stricture is erroneously classified as severe), may prompt referral for unnecessary invasive interventions despite a possibility of alternative explanations and management for their symptoms and (iv) accurate classification of the degree of airway narrowing allows physicians to accurately communicate with each other before and after therapeutic interventions, and more precisely employ a validated classification such as Myer-Cotton system when managing patients with LTS [1]. While we did not study this issue specifically, the findings of this study suggest the need for surgeons contemplating tracheal resection to be present at the time of bronchoscopy or to perform their own airway examination prior to surgical intervention.

Paradoxically, there was no correlation between bronchoscopist experience (based on the number of bronchoscopies and number of strictures seen per month) and the ability to accurately classify the degree of airway narrowing as mild, moderate or severe based on still bronchoscopic images. Other than suggesting a need for educational programmes pertaining to LTS, the impact of this finding on patient care is difficult to assess because treatment decisions in benign LTS are based not only on the degree of narrowing, but also on other factors such as extent, location, morphology of airway narrowing, vocal cord function, swallowing dysfunction, dyspnoea class and functional status [1, 13, 14]. The reasons pulmonologists underestimate the degree of narrowing are unclear and were not the purpose of this study. We did not find a correlation between operators’ experience and ability to accurately estimate the degree of stenosis. Furthermore, as the airway narrowing was assessed on still images, it is unlikely that the participants’ bronchoscopic technique contributed to these findings. One possible explanation for these findings is the lack of familiarity with the optics principle of image distortion partially caused by intrinsic lens properties but also by the location of the bronchoscope in relation to the stricture, factors which could significantly influence the airway size perception.

In order to accurately assess and classify airway strictures, several techniques have been proposed. Spirometry and flow-volume loops, however, do not precisely localize strictures, are insensitive diagnostic tests for mild to moderate narrowing and provide functional rather than accurate anatomical information [5, 6]. Results from ultrasonography correlate with bronchoscopic imaging with regard to airway lumen size estimations, but allow non-invasive measurements only of transverse cervical tracheal diameter [15]. MDCT, with or without 3D reconstruction, can be used to quantify the degree of airway narrowing [16, 17]. MDCT airway measurements correlate with results of objective bronchoscopic assessments [7, 18, 19]. Radiographical quantification of LTS, however, is not standardized, may be hindered by secretions, and scanning may be difficult or unreasonable in critically ill patients or in those who are uncooperative or have severe dyspnoea [7–9]. Morphometric bronchoscopy using image analysis software provides quantitative (using calibration bodies to report CSA in metric units) or semi-quantitative (reporting CSA as percentages of normal airway) measurements of degree of airway narrowing [12]. From a physiological standpoint, the latter parameter matters more than the absolute CSA or airway diameter in metric units. Given the airway size dependence on age, sex and body height, the percentage reduction in airway calibre is more relevant to flow limitation, symptoms and treatment decisions than the absolute CSA measured in metric units. Morphometric bronchoscopy measurements correlate well with clinical findings as well as with measurements made using MDCT in both experimental and clinical studies [18–22]. High intra- and interobserver reliabilities have been demonstrated [20–22]. Real-time morphometric bronchoscopic measurements can be obtained using new stereovision bronchoscopes [23], and in the near future, correlations between the degree of narrowing and pressure changes responsible for increased work of breathing will become possible intraoperatively [4].

This study has several limitations that impact the generalizability of our results. Because results were based on course participants’ responses, the problems of recall bias, personal motivation or indifference at the time of data gathering must be recognized. The results of our survey may not be applicable to all practitioners. All participants were pulmonologists, so our results may not apply to bronchoscopists in other specialties such as thoracic surgery, ENT or anaesthesiology. Thirdly, morphometric bronchoscopy was used as a reference standard for objectively assessing airway narrowing because there is no true gold standard. While shown to be accurate in paediatric patients with both normal and abnormal airway findings [20], the manual method of morphometric bronchoscopy used in this study has not been rigorously studied in adult patients with LTS. From an optics standpoint, the airway size does not affect the morphometric bronchoscopy methodology; there is no physics rationale for expecting different results when measuring paediatric or adult airways. Further studies of this technique should, however, attempt to correlate the airway measurements with lung function.
testing or radiological investigations. Therefore, the use of calculated stenosis index on the margins of categories could occasionally be misleading. For example, stricture #6 was a triangular stenosis with a calculated SI of 71%, and stricture #20 had an SI of 53% (Table 1). If morphometric bronchoscopy technique has an error of ±3%, both of these strictures would be correctly classified by the participants, and incorrectly classified by morphometric bronchoscopy. Finally, from an optics standpoint, interpretations of airway findings are affected by the presence of image distortion [24] because length, perimeter and area measurements depend on the target’s location within the field of view of the bronchoscope [25]. (It is noteworthy that airway targets move away from the bronchoscope lens during inspiration and towards the lens during expiration, causing an optical effect whereby the airway lumen appears greatest at end-expiration. While obtaining images for morphometric bronchoscopy, it is important, therefore, to keep the bronchoscope centred and at a fixed distance from the target regardless of the respiratory phase). This distortion, however, applies to both normal and abnormal airway abnormalities. Results from this study suggest, however, that experienced bronchoscopists often misclassify the degree of airway narrowing when subjectively assessing strictures of benign aetiology using still bronchoscopic images. Whether in the bronchoscopy suite or post-procedure, caution is warranted when classifying airway strictures based on printed or digital still images captured at the time of airway examination.

ACKNOWLEDGEMENTS

We thank Pablo Diaz-Jimenez for facilitating the performance of the study at his interventional bronchoscopy course, Mohsen Davoudi for his help with statistical analysis and Stephanie Langer for her assistance with data gathering.

Conflict of interest: none declared.

REFERENCES