LARYNGOLOGY

Correlations between bedside and instrumental endoscopic parameters in determining severity of dysphagia: an integrated clinical evaluation of safety and efficiency

Correlazioni tra parametri non strumentali e strumentali endoscopici nel determinare la severità della disfagia: una valutazione clinica integrata di sicurezza ed efficienza

Daniele Farneti[1], Elisabetta Genovese[2]

[1] Audiology and Phoniatrics Department - Romagna Health Service, Rimini Hospital, Rimini, Italy; [2] Audiology Service, University of Modena and Reggio Emilia, Policlinico Hospital of Modena, Italy

SUMMARY

Interaction between bedside and endoscopic parameters is of great interest in the management of patients with swallowing disorders. Our aim is to document if and how bedside parameters correlate with severity using endoscopic assessment. 556 consecutive patients (318 M/238 F, mean age 65.56 ± 10.36 years, range 18-91), were evaluated in our Swallowing Centre during 2008. All underwent bedside evaluation and fiberoptic endoscopic evaluation of swallowing (FEES), considering the pooling score (p-score) and the pooling sensation, collaboration and age score (p-SCA score) to express criteria of clinical severity of dysphagia. The correlation between the two tests (Spearman correlation coefficient) and their agreement to classify severity (Cohen's kappa) was defined. After dichotomisation (cut-off: no risk/any kind of risk of aspiration), values of sensitivity and specificity were obtained after comparison with FEES results (gold standard). A close and significant correlation between the p-score and p-SCA score was found (rho = 0.88; p < 0.001). The agreement among scores in attributing the categories of risk is moderate (Cohen's Kappa = 0.46; p < 0.001). The p-score had a sensitivity of 96% and specificity of 60%, while the p-SCA score has a sensitivity of 98% and specificity of 40%. Our results suggest that including even a few parameter from bedside evaluation to an endoscopic score, the level of severity expressed by the latter, decreases. The evaluation of patients with swallowing disorders should consider as many elements as possible, deriving from non-instrumental and instrumental evaluation (integrated clinical evaluation).

KEY WORDS: swallowing, deglutition disorders, aspiration, residue, FEES, p-score

Conflict of interest
The Authors declare no conflict of interest.

RIASSUNTO

L'interazione tra parametri bedside e parametri endoscopici è di grande interesse nella gestione di pazienti con disturbi di deglutizione. Il nostro obiettivo è documentare se e come i parametri bedside modificano la gravità espressa dalla valutazione endoscopica, per definire un criterio di gravità che meglio aderisca al reale contesto clinico. 556 pazienti consecutivi (318 M/238 F, età media 65,56 ± 10,36 anni, intervallo 18-91 anni), sono stati valutati nel nostro Centro Disfagie durante il 2008. Tutti sono stati sottoposti a una valutazione bedside e una valutazione endoscopica della deglutizione, considerando il punteggio del pooling score (p-score) e del pooling sensibilità, collaborazione ed età score (p-SCA ascore) per esprimere i criteri di gravità clinica della disfagia. È stata definita la correlazione tra i due test (coefficiente di correlazione di Spearman) e il loro accordo per classificare la gravità (kappa di Cohen). Dopo la dicotomizzazione (cut-off: nessun rischio / qualunque rischio di aspirazione) sono stati ottenuti valori di sensibilità, specificità, dal confronto con i risultati FEES (gold standard). È stata trovata una stretta correlazione tra il p-score e il p-SCA score (rho = 0,88): la correlazione è significativa (p < 0,001). L'accordo tra i punteggi nell'attribuzione delle categorie di rischio è moderato (Cohen's Kappa = 0,46; p < 0,001). Il p-score ha raggiunto valori di sensibilità del 96% e specificità del

60%, mentre il punteggio p-SCA ha raggiunto valori di sensibilità del 98% e specificità del 40%. I nostri risultati suggeriscono che, includendo anche pochi elementi dalla valutazione bedside a un punteggio endoscopico, il livello di gravità espresso da quest'ultimo diminuisce. La valutazione di pazienti con disturbi della deglutizione dovrebbe considerare il maggior numero possibile di elementi, derivanti dalla valutazione non strumentale e strumentale (valutazione clinica integrata).

PAROLE CHIAVE: *deglutizione, disturbi della degradazione, aspirazione, ristagni, FEES, p-score*

Introduction

Swallowing is a complex neuromuscular act, which requires sensory input and automated neuromuscular activities under real-time modulation guaranteed by the central nervous system in all its parts [1]. In optimal physiological and functional conditions, all the ingested bolus is transferred to the stomach, for digestive processing, without invasion of the airway (false route) or residue. In dysfunctional or pathological conditions, material pooling or residue can invade the respiratory tract, upwards and/or downwards, leading to respiratory or nutritional complications (ineffective / inefficient swallow or dysphagia), respectively [2]. Several conditions, acting as morbidity or comorbidities, can affect the quality of life (QOL) of dysphagic patients, offering a wide variety of events that are able to influence each other in guiding the clinical options adopted by the multidisciplinary team (MDT) that manages the patient [3].

During the processing of data leading to therapeutic planning, conditions of ineffective/inefficient swallowing (i.e. airway invasion/residue) must be documented with instrumental assessment. However, complete strategic planning cannot disregard other clinical not instrumental information, which is no less important in suggesting therapeutic strategies, even when apparently in contrast with instrumental evaluation. Data from the literature show that clinical non-instrumental evaluation alone tends to underestimate the risk of aspiration, while endoscopic evaluation tends to overestimate it [4]. A reasonable balance between non-instrumental and instrumental evaluation seems to be the best combination that can offer the optimal strategy to the MDT, in order to obtain the best compliance of patients and caregivers. The balance mentioned above also seems to be the best way to achieve a concrete and reasonable evaluation of severity of impaired swallowing and the risk of complications [5]. In addition, the real and logistical aspects of clinical practice must also be considered, as well as the availability of local resources [3].

In our Swallowing Center all patient referring swallowing disorders are routinely submitted to Bedside Swallowing Evaluation (BSE) [6] and endoscopic evaluation of the upper aerodigestive tract, performing a dynamic test with bolus for the study of swallowing (fiberoptic endoscopic evaluation of swallowing; FEES) [7]. The evaluation of the oral and oesophageal phases of swallowing (O-FEES and E-FEES, respectively) [5] is performed considering the patient's specific complaints or to answer to specific questions that the clinician poses. These approaches, only recently introduced in clinical practice, are not yet adequately supported by the international literature, but their value is purely local and far from any research intent. The pooling score (p-score) and the pooling Sensation, Collaboration and Age score (p-SCA score) [8] are simple scores used in our endoscopic practice to define a criterion of severity of impaired swallowing, considering the material pooling/residue and the risk of aspiration. Material pooling/residue is evaluated in the hypopharynx/laryngotracheal cavities, considering specific anatomical landmarks, but also considering its amount and management (Tab. I).

After these preliminary considerations, the aims of this study are to evaluate: 1) how the severity of dysphagia, endoscopically defined, changes when BSE parameters are considered together with FEES parameters; and 2) the weight of BSE parameters in determining severity, as mentioned above.

Materials and methods

In this prospective study a sample of 556 consecutive patients (318 M/238 F, mean age 65.56 ± 10.36 years, range 18-91 years) seen at our Swallowing Centre during 2008 was considered. All patients underwent BSE performed alternately by two speech-language pathologists (SLPs) and a phoniatrician. During BSE a preliminary collection of information on pathologies, interventions and drugs was made. Subsequently, cognitive and language skills were assessed, observing the patient's facies and carrying out manual exploration of the mouth and oropharynx. Finally, assessment of swallowing skills was performed with boluses of different consistency and volume, verifying the appearance of cough, throat clearing and modification of vocal quality.

After BSE, all patients were evaluated by a phoniatrician and submitted to FEES according to a protocol in use at our centre [3]. Endoscopic evaluation was performed with a Storz endoscope (model 11101RP2, 30 cm long, 3.5 mm in diameter) and recorded with a workstation (Richard Wolf GmbH, Knittlingen, Germany). The patients were given three trials of different consistencies: 5 cc pureed,

Table I. Anatomical landmarks and bedside parameters with relative values.

Pooling	Endoscopic landmarks	Bedside parameters		
		Sensation	Collaboration	Age (years)
Site	Vallecule: 1 Marginal zone: 1 Pyriform sinus: 2 Vestibule/vocal cords: 3 Lower vocal cords: 4			
Amount	Coating: 1 Minimum: 2 Maximum: 3	Presence = -1 Absence = +1	Presence = -1 Absence = +1	+1 (< 65) +2 (65-75) +3 (> 75)
Management	< 2 2 </> 5 > 5			
Score	p: 4-11	p-SCA: 3-16		

p: pooling; p-SCA: pooling-sensation-collaboration-age. P-score: 4-5 = minimum score, corresponding to no dysphagia; 6-7 = low score, corresponding to a mild dysphagia; 8-9 = middle score, corresponding to a moderate dysphagia; 10-11 = high score, corresponding to a severe dysphagia. P-SCA score: 3-4 = minimum score, corresponding to no dysphagia; 5-8 = low score, corresponding to a mild dysphagia; 9-12 = middle score, corresponding to a moderate dysphagia; 13-16 = high score, corresponding to a severe dysphagia.

5 cc liquid dyed with 5% methylene blue and 1/4 cracker. After each consistency the p-score and the p-SCA score were obtained (Tab. I). The p-score expresses a continuum of severity, summarised in a simple number ranging from 4 to 11, and clinically distributed over 4 levels of severity. The p-SCA score [8] is the p-score enriched with simple information achieved from BSE, namely sensation, collaboration and age: it ranges from 3 to 16, with the same clinical application on 4 levels of severity. The inter-rater and intra-rater reliability of the p-score was recently determined [8]. The p-score considers the risk of aspiration occurring over the time (before, during and after swallowing) as an interaction between material pooling and false route (ineffective/inefficient swallowing). The correlation between material pooling and false route, i.e. safety and efficiency, has been recently evaluated [9,10].

Statistical analysis

The correlation between the p-score and p-SCA was determined with Spearman's correlation coefficient. The agreement between the two scores was calculated (Cohen's Kappa) considering the categories of risk corresponding to the total scores (no dysphagia, mild, moderate, severe). The categories of risk individualised with the two scores was studied with the aim of underlining possible systematic divergences in the attribution of the severity to individual cases. Subsequently, the p-score and the p-SCA score were dichotomised, dividing patients without risk from those with middle and high risk of aspiration. By comparison of dichotomic scores with the results of FEES (considered as the gold standard), the values of sensitivity and specificity were obtained.

All patients were over 18 years of age and consenting, in accordance with the Declaration of Helsinki. The study was approved by the local Ethical Research Committee. Statistical analysis was performed with Intercooled STATA 8.0 for Windows software.

Results

A sample of 556 consecutive patients (318M/238F, mean age 65.56 ± 10.36 years, range 18-91 years) was evaluated. A close correlation between the p-score and p-SCA score in determining dysphagia severity (rho = 0.88) was found (Figs. 1A, B). The agreement among scores regarding the categories of risk showed moderate correlation (Cohen's Kappa = 0.46; p < 0.001). Table II shows how the two scores classified patients in different categories of risk. Table III summarises the percentage of patients classified by the two scores: the two scores are correlated (Wilcoxon signed-rank test, p < 0.001). Subsequently, the judgement expressed by the scores was dichotomised, setting the cut-off point between patients without risk and those with any risk, with the purpose of comparing the scores with the results of FEES in terms of "aspiration" and to obtain, for both, values of sensitivity and specificity. The values of sensitivity and specificity are reported in Table IV.

Discussion

In daily clinical practice the possibility of correlating signs and symptoms with residue/aspiration in patients with deglutition disorders due to different aetiologies is an important goal, and is needed to obtain better guidance for MDTs and prevent complications [10]. The interaction between signs (instrumentally documented) and symptoms is a complex and intriguing relationship, but nonetheless is likely the best way to offset the trend of instrumental
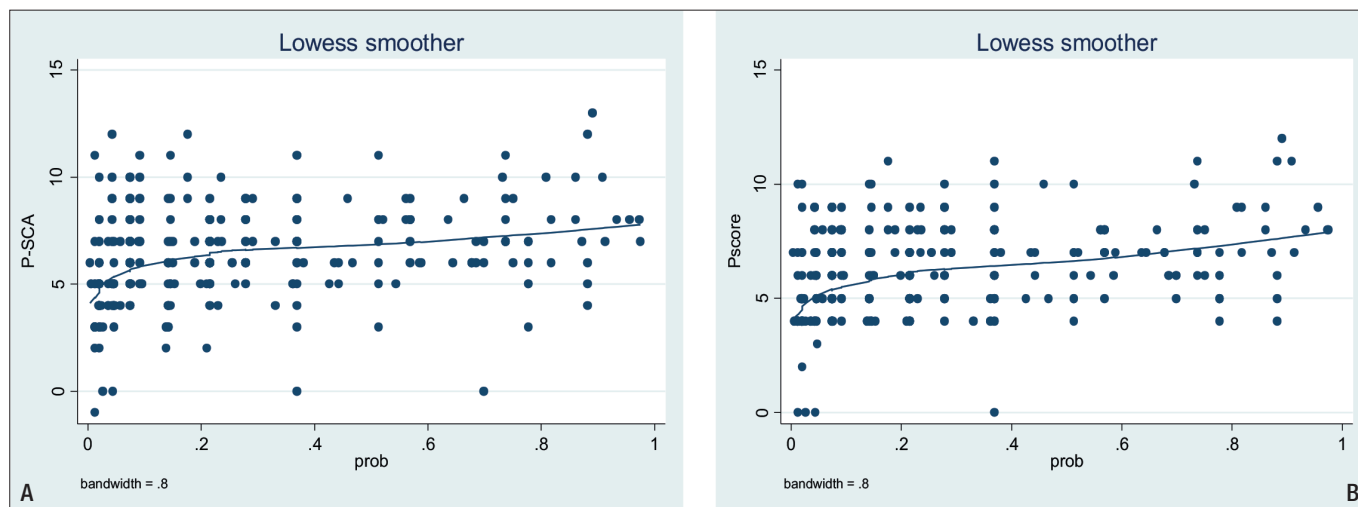
Figure 1. Scatterplot whit non-parametric regression curve (Lowess smoother): (**A**) p-score; (**B**) p-SCA score.

Table II. Comparison of P-score and P-SCA score and classification of risk.

P-score	P-SCA score				
	No	**Mild**	**Moderate**	**Severe**	**Total**
No	163	100	0	0	263
	61.98	38.02	0.00	0.00	100.00
Mild	11	139	0	0	150
	7.33	92.67	0.00	0.00	100.00
Moderate	0	43	35	0	78
	0.00	55.18	44.87	0.00	100.00
Severe	0	12	20	1	33
	00.0	36.36	60.61	3.03	100.00
Total	174	294	55	1	524
	33.21	55.61	10.50	0.19	100.00

Table III. P-score and P-SCA score and risk of aspiration.

	Frequency	**Percentage**
P-score		
No dysphagia	264	50.29
At risk	261	49.71
Total	525	100.00
P-SCA score		
No dysphagia	174	33.21
At risk	350	66.79
Total	524	100.00

Table IV. Screening properties of the scores considering no risk and any risk of aspiration.

	Sensitivity	**Specificity**	**Correctly classified**	**ROC Area [95% Conf. Interval]**
P-score	95.56%	59.77%	65.90%	0.7766 (0.73908 0.81205)
P-SCA score	97.78%	39.63%	49.62%	0.6870 (0.64539 0.72653)

The p-SCA score tends to increase severity in the category with lower risk, while in those with higher risk it tends to be more cautious, attributing a category with lower severity in comparison to p-score (Tab. II). Overall, the patients classified as at risk of aspiration by the p-score are 50%, while the p-SCA score considers 67% of patients to be at risk (Tab. III). After dichotomisation (cut-off: no risk/any risk class) and comparison with the gold standard (aspiration documented during FEES), the two scores still show a different trend, offering different levels of sensitivity and specificity (Tab. IV). It is as if the p-SCA score recognises more patients at risk, more sensitive than the p-score, but also less specific with a greater risk of identifying false positive patients. In other words, both have high sensitivity to identify patients with any risk of aspiration, from minimal to high, but the p-score is more specific, and better at recognising false positives and therefore more reliable in correctly classifying patient without dysphagia and those with a risk of swallowing disorders. The low specificity of the p-SCA score also expresses a low ability of the score to identify patients who do not have swallowing disorders or who have low swallowing concerns, despite the more precise contribution of endoscopy results.

assessments to overestimate risks. With the contribution of instrumental evaluation, the risk of lost episodes of silent aspiration/penetration at the bedside is less, but the risk in generalising pathological random or extraordinary airway invasion events is higher [11,12]. Considering our data, the p-score and p-SCA score are both useful tools to define severity, and both show the same statistical significant trend (Figs. 1A, B). However, considering the categories of risk attributed, the two scores seem to work in a different way.

It is worth noting, as previously mentioned, that BSE alone, even if well conducted, underestimates the risk of silent aspiration, while endoscopy leads to generalisation of occasional episodes of false routes and residues. In addition, aspiration, as a marker of impaired swallowing, it is the most significant but not the only one, capable of determining the clinical severity of dysphagia and suitable for guiding the treatment plan and the activities of the MDT.

Conclusions

The simplest conclusion of the current experience, in accordance with our aims, seems to be that, by including even a few factors from the BSE in an instrumental score, the level of severity expressed by the latter decreases. Namely, the simple evaluation of sensation, collaboration and age tends to mitigate the judgment of severity expressed by the p-score, putting patients back into categories with lower risk: not only, but the greater impact of this contamination affects the ability of a "hybrid" score to correctly identify false negative patients, with a decrease in specificity. This experience also statistically quantified the weight of BSE parameters in reducing the specificity of a FEES parameter, such as the p-score.

The most relevant clinical implication of our work seems to be that the evaluation of patients with swallowing disorders should consider as many elements as possible, deriving from both non-instrumental and instrumental clinical evaluation: this could be considered as integrated clinical evaluation [5].

The main limitation of this preliminary work is that the p-SCA score considers a limited number of BSE parameters, although the logic supporting it is strong [10].

References

[1] Jean A. Brain stem control of swallowing: neuronal network and cellular mechanisms. Physiol Rev 2001;81:929-69. https://doi.org/10.1152/physrev.2001.81.2.929

[2] Clavé P, Verdaguer A, Arreola V. Oropharyngeal dysphagia in the elderly. Med Clin (Barc) 2005;124:742-8. https://doi.org/10.1157/13075447

[3] Brewer K, Leonard L, Kendall DL. Dysphagia assessment and treatment planning - a team approach. Perspectives on Swallowing and Swallowing Disorders (Dysphagia) 2000;9:9-10. https://doi.org/10.1044/sasd9.1.9

[4] Leder BS, Espinosa JF. Aspiration risk after acute stroke: comparison of clinical examination and fiberoptic endoscopic evaluation of swallowing. Dysphagia 2002;17:214-8. https://doi.org/10.1007/s00455-002-0054-7

[5] Farneti D, Genovese E. Endoscopic criteria in assessing severity of swallowing disorders. In: Speyer R, editor. Seminars in Dysphagia. London: In-Tech; 2015. https://doi.org/10.5772/60836

[6] Murry T, Carrau RL. Clinical manual of swallowing disorders. Second Edition. San Diego, CA: Singular Thomson Learning; 2001.

[7] Langmore SE, Schatz K, Olsen N. Fiberoptic endoscopic examination of swallowing safety: a new procedure. Dysphagia 1988;2:216-9. https://doi.org/10.1007/bf02414429

[8] Farneti D, Fattori B, Nacci A, et al. The Pooling-score (P-score): inter- and intra-rater reliability in endoscopic assessment of the severity of dysphagia. Acta Otorhinolaryngol Ital 2014;34:105-10.

[9] Nordio S, Di Stadio A, Koch I, et al. The correlation between pharyngeal residue, penetration/aspiration and nutritional modality: a cross-sectional study in patients with neurogenic dysphagia. Acta Otorhinolaryngol Ital 2020;40:38-43. https://doi.org/10.14639/0392-100X-2136

[10] Farneti D, Turroni V, Genovese E. Aspiration: diagnostic contributions from bedside swallowing evaluation and endoscopy. Acta Otorhinolaryngol Ital 2018;38:511-6. https://doi.org/10.14639/0392-100X-1967

[11] Pilz W, Vanbelle S, Kremer B, et al. Observers' agreement on measurements in fiberoptic endoscopic evaluation of swallowing. Dysphagia 2016;31:180-7. https://doi.org/10.1007/s00455-015-9673-7

[12] Molfenter SM, Steele CM. The relationship between residue and aspiration on the subsequent swallow: an application of the normalized residue ratio scale. Dysphagia 2013;28:494-500. https://doi.org/10.1007/s00455-013-9459-8