Poor interexaminer agreement on Friedman tongue position

Joar Sundman, Johan Bring & Danielle Friberg

To cite this article: Joar Sundman, Johan Bring & Danielle Friberg (2017) Poor interexaminer agreement on Friedman tongue position, Acta Oto-Laryngologica, 137:5, 554-556, DOI: [10.1080/00016489.2016.1255776](https://doi.org/10.1080/00016489.2016.1255776)

To link to this article: <https://doi.org/10.1080/00016489.2016.1255776>

Published online: 23 Nov 2016.

Submit your article to this journal [↗](#)

Article views: 95

View Crossmark data [↗](#)

Citing articles: 1 View citing articles [↗](#)

ORIGINAL ARTICLE

Poor interexaminer agreement on Friedman tongue position

Joar Sundman^a , Johan Bring^b and Danielle Friberg^a

^aDepartment of Otorhinolaryngology, Karolinska University Hospital, Stockholm, Sweden; ^bStatisticon AB, Stockholm, Sweden

ABSTRACT

Conclusion: The Friedman tongue position demonstrated only a slight inter-examiner agreement among 15 medical doctors, indicating that the method is difficult to perform and could be an uncertain method to select patients for uvulopalatopharyngoplasty.

Objective: The Friedman staging system is a clinical tool for selecting patients with obstructive sleep apnea who are appropriate for uvulopalatopharyngoplasty. The objective of this study was to evaluate the staging system by determining the inter-examiner agreement of one of its key components, the Friedman tongue position.

Methods: Eleven residents and four specialists in Otorhinolaryngology were recruited from a course in surgical management of obstructive sleep apnea. They examined each other's tongue positions, resulting in a total of 210 evaluations. Cohen's kappa analysis was performed to assess the inter-examiner agreement. The range of kappa is usually between 0–1, where 0 is complete disagreement and 1 is complete agreement.

Results: The median kappa was 0.36 (1st and 3rd quartile, 0.23 and 0.42), corresponding to only a slight agreement.

ARTICLE HISTORY

Received 7 September 2016

Revised 12 October 2016

Accepted 17 October 2016

KEYWORDS

Uvulopalatopharyngoplasty;
obstructive sleep apnea;
patient selection; inter-rater
reliability; Cohen kappa

Introduction

Uvulopalatopharyngoplasty (UPPP) is the most common surgical treatment for obstructive sleep apnea (OSA) worldwide. However, there is no international consensus on how to appropriately select patients for surgery. The Friedman staging system was developed in 2002, and was originally based on four different tongue positions (Figure 1) and tonsillar size [1]. It was demonstrated that, in 134 patients with OSA, the Friedman staging system correlated to surgical success after UPPP. Stage I patients had a 81% success rate, defined as a post-operative Respiratory Distress Index (RDI) below 20 and, reduced by at least 50%, stage II patients had a success rate of 38%, and stage III patients of 8%. The method is inexpensive, requires no special equipment, and could facilitate surgeons to more accurately select patients for UPPP. At our clinic the Friedman staging system is important in the decision whether to recommend surgery or not, and we also used it as an inclusion criteria in our randomized controlled study, SKUP3, from 2013 [2].

As any clinical test the findings might be more or less subjective and may vary between different examiners. Ideally the variation is small or non-existent. Too much inter-examiner variation limits the value of the test and puts its reliability into questioning. The objective of this study was to evaluate the staging system by determining the inter-examiner agreement of one of its key components, the Friedman tongue position. A previous study by the inventor of the

system, Friedman et al. [3], found a high inter-examiner agreement with a kappa value of 0.83 using video clips of oropharyngeal examinations. Our hypothesis was that such agreement would be less pronounced when examining real life subjects.

Materials and methods

Eleven residents and four specialists in Otorhinolaryngology participated. They were recruited from a course in surgical management of OSA, which was held at the ORL-Clinic, Karolinska University Hospital in Stockholm, Sweden in October 2015. The study was performed on colleagues only and not on patients. They all gave consent to participate. Seven men and eight women participated. The median age was 37 (range 30–60) years; median BMI was 22.9 (range 19.7–30.5). They all knew about the Friedman tongue position, although the extent to which they had used it in clinical practice varied. All were given instructions both verbally and in writing. They received a paper with a picture of the four Friedman tongue positions (previously known as the Friedman palate position) (Figure 1). The instructions were given as in Friedman et al. [1]: Ask the test subjects to open their mouth widely without protruding their tongue and repeat this five times and note the most consistent value. A headlight was used and the examination was done with the rater and subject sitting in an upright position. Every participant then examined each other, resulting in a total of

Figure 1. The four different Friedman tongue positions. Courtesy of Professor Friedman and co-authors.

Rater		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
1		1	4	2	4	3	4	3	1	1	4	2	2	3	2	
2	4		3	1	4	3	2	2	1	3	4	2	1	4	3	
3	3	3		1	4	3	3	3	1	3	4	2	2	3	1	
4	3	2	2		4	2	3	4	1	2	4	3	2	3	2	
5	3	2	3	1		3	3	4	1	3	4	2	2	3	2	
6	4	4	4	1	4		4	3	1	3	4	3	2	3	1	
7	4	1	3	1	4	2		4	1	2	4	2	2	2	2	
8	4	2	4	1	4	4	3		2	2	4	2	2	2	2	
9	3	1	1	1	4	2	2	2		3	4	2	1	2	2	
10	4	4	4	1	4	4	3	4	1		4	3	2	2	1	
11	4	4	4	1	4	4	3	4	1	3		2	1	2	2	
12	4	4	1	1	4	3	3	2	1	1	4		1	3	2	
13	3	3	2	1	3	3	2	2	1	2	4	2		1	2	
14	4	4	4	1	4	4	1	4	2	2	3	2	2		1	
15	3	1	3	1	4	3	3	3	1	1	4	1	1	2		
		a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
		Study objects														

Figure 2. The results from all ratings; each rater on the x-axis; 1–15, and as study objects on the y-axis; a–o. The black boxes indicate that they cannot examine themselves.

210 evaluations during 60 min. The participants were instructed to note their findings on a paper and not to share them, so as not to influence each other's judgment. Data were then transferred to a computer by a single person and control samples were obtained by a second individual. No power analysis was performed. The data were analyzed according to Cohen kappa, which measures the amount of inter-examiner agreement that occurs beyond what would be explained by chance alone. The software R was used for statistical computing [4]. The range of kappa is usually between 0–1, where 0 is complete disagreement and 1 is complete agreement.

Each of the 15 participants rated 14 subjects. Since they did not rate themselves, there were, for each pair of raters, 13 subjects that both raters evaluated. Based on these 13 ratings, a Kappa coefficient was derived and 105 pairs of raters were combined. Therefore, 105 Kappa coefficients were calculated.

Results

All participants filled in all ratings and there were no missing values. The results from all ratings are presented in

Figure 2, and the individual kappa coefficients are presented as box plots in Figure 3.

The median Cohen's kappa coefficient for Friedman tongue position was 0.36, the 1st quartile 0.23, and the 3rd quartile 0.42, with a minimum of 0.03 and a maximum of 0.69. This corresponds to only a slight agreement according to Byrt [5].

Discussion

In the present study, Friedman tongue position did not achieve more than a slight inter-examiner agreement among 15 medical doctors. This indicates that the method is difficult to perform for non-experts and could be an uncertain method when selecting patients for UPPP.

The strength of this study is the 'live' examination, in contrast to the original study, which used video clips [3]. We also believe that the time pressure is more similar to that of everyday clinical practice. The main weaknesses are that our raters were non-experts in using the Friedman staging system, and that our study subjects were not patients

Figure 3. Box plots consisting of 13 kappa coefficients for each rater. The median kappa values are below 0.5 in all raters, and there is a large variability between different raters.

and did not have an OSA diagnosis. The participants may have also become fatigued to a certain degree after having performed and undergone 14 examinations during 60 min, possibly resulting in both impaired judgment and affecting how well they let themselves be examined. Furthermore, a Cohen's kappa analysis does not make a difference in how inaccurate an estimate is. This means that it is perceived as equally incorrect to grade 1 and 2 as it is to grade 1 and 4.

One could claim that, since Friedman tongue position is used in combination with tonsillar size to determine Friedman stage, the uncertainty is reduced when the two are combined. While this may be true, even small differences in the assessment of tonsillar size and tongue position are potentially a cause of major grading errors. Patients in the median range of both tonsillar size and tongue position are especially prone to this. Even a slight inter-examiner disagreement could result in a Friedman grade I (tonsil size 3 and Friedman tongue position 2) or Friedman grade III (tonsil size 2 and Friedman tongue position 3), causing a difference in the estimated success rate between 81% and 8%. Similar studies on OSA patients are needed, as well as renewed studies on the correlation between Friedman stage and surgical outcome.

In 2008, Friedman et al. [3] modified the tongue position with an added extra grade (2a and 2b). However, since this modification has not been validated against surgical outcome, we used the original system with four grades.

Conclusions

In this study, Friedman tongue position only achieved a slight inter-examiner agreement among 15 medical doctors,

indicating that the procedure is difficult to perform for non-experts and could be an uncertain method to select patients for UPPP. We conclude that educational efforts are required to improve the inter-examiner agreement.

Disclosure statement

The authors reports no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

ORCID

Joar Sundman <http://orcid.org/0000-0003-4475-603X>
 Johan Bring <http://orcid.org/0000-0002-8204-8432>
 Danielle Friberg <http://orcid.org/0000-0001-8327-1292>

References

- [1] Friedman M, Ibrahim H, Lee G. Clinical staging for sleep-disordered breathing: a guide to surgical treatment. *Oper Tech Otolaryngol Neck Surg* 2002;13:191–5.
- [2] Browaldh N, Nerfeldt P, Lysdahl M, Bring J, Friberg D. SKUP3 randomised controlled trial: polysomnographic results after uvulopalatopharyngoplasty in selected patients with obstructive sleep apnoea. *Thorax* 2013;68:846–53.
- [3] Friedman M, Soans R, Gursipinar B, Lin HC, Joseph NJ. Interexaminer agreement of Friedman tongue positions for staging of obstructive sleep apnea/hypopnea syndrome. *Otolaryngol Head Neck Surg* 2008;139:372–7.
- [4] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. URL <https://www.R-project.org/>.
- [5] Byrt T. How good is that agreement? *Epidemiology* 1996;7:561.