Standardizing drug-induced sleep endoscopy scoring by an expert review panel : our experience in 81 patients

Standardizing Drug-Induced Sleep Endoscopy Scoring by an Expert Review Panel: Our Experience in 81 Patients.

Short running title: Expert Review Panel for DISE

Authors: Eli Van de Perck [*1,2], Annelies E. Verbruggen [2], Anneclaire V. Vroegop [1,2,4], Marijke Dieltjens [1,2,5], Evert Hamans [3], Marc J. Braem [1,5], Paul H. Van de Heyning [2], Olivier M. Vanderveken [1,2,4]

[1]Faculty of Medicine and Health Sciences, University of Antwerp, Wilrijk, Belgium; [2]Department of Otorhinolaryngology, Head and Neck Surgery, Antwerp University Hospital, Edegem, Belgium; [3]Department of Otorhinolaryngology, Head and Neck Surgery, Jan Palfijn Hospital, Antwerp, Belgium; [4]Multidisciplinary Sleep Disorders Centre, Antwerp University Hospital, Edegem, Belgium; [5]Special Dentistry Care, Antwerp University Hospital, Edegem, Belgium.

* Corresponding author
Eli Van de Perck, MD
Antwerp University Hospital
ENT, Head and Neck Surgery department
Wilrijkstraat 10
2650 Edegem
BELGIUM
E-mail: eli.vandeperck@uza.be
Telephone number: +32 3 821 42 44

Author contribution statement:

Dr Van de Perck drafted the manuscript and performed statistical analyses. The data were collected by Drs Verbruggen, Vroegop, Hamans, and Vanderveken. Drs Braem, Van de Heyning, and Vanderveken conceived the study and acquired funding. Drs Verbruggen, Vroegop, and Dieltjens coordinated the project. Dr Vanderveken was the primary supervisor. All authors revised and approved the final version of the manuscript.

Conflict of Interest:

Dr Dieltjens holds a Senior Postdoctoral Fellowship from the Research Foundation Flanders (FWO: 12H4520N). Dr Hamans reports other from Nyxoah outside the submitted work. Dr Vanderveken holds a Senior Clinical Investigator Fellowship from the Research Foundation Flanders (FWO: 1833517N), and reports grants and personal fees from Somnomed, grants and non-financial support from Philips, personal fees and other from Inspire, and other from Nyxoah, outside the submitted work. All other authors declare that they have no competing interests.

Funding Statement:

This study was funded by a 3-year grant of the Flemish government agency for Innovation by Science and Technology (IWT-0908864).

MR ELI VAN DE PERCK (Orcid ID : 0000-0002-7180-0473)

Article type : Correspondence: Our Experience

Standardizing Drug-Induced Sleep Endoscopy Scoring by an Expert Review Panel: Our Experience in 81 Patients.

KEY POINTS

- Assessment of pharyngeal collapse patterns by drug-induced sleep endoscopy (DISE) is often paramount to guide clinical decision-making in patients with obstructive sleep apnoea.
- Being a subjective investigation, DISE findings are prone to interindividual variation.
- The present study introduces the concept of an expert review panel to standardize DISE scoring.
- Comparing the ratings of the review panel with the original ratings revealed a substantial concordance at the level of the oropharynx and a fair to moderate concordance at the levels of the soft palate, tongue base, hypopharynx, and epiglottis.
- Several recommendations were made to address the observed discrepancies in DISE ratings.

KEY WORDS

DISE, endoscopy, expert review, interobserver agreement, obstructive sleep apnoea, quality control

WORD COUNT

1655 / 1500

TABLES & FIGURES

4 / 4

INTRODUCTION

Since its introduction in 1991,[1] drug-induced sleep endoscopy (DISE) has gained great popularity for assessing the site(s) of upper airway (UA) obstruction in patients with obstructive sleep apnoea (OSA). By providing a three-dimensional snapshot of the UA anatomy during a sleep-mimicking state, DISE is often used to guide therapeutic decision-making in patients who are seeking alternatives to positive airway pressure therapy.[2] The validity and reliability of the procedure have been demonstrated in previous research[3,4]; however, as the assessment is subjective, the results may vary based on personal experience.[5] Furthermore, due to the complex anatomy of the UA, a plethora of classification systems for DISE findings currently exists.[2] These shortcomings may limit the generalisability of DISE findings in both research settings and clinical practice.

In our high-volume centre,[6] DISE is performed by several experienced ear, nose and throat (ENT) surgeons in a standardized manner. Nevertheless, despite long-term use of the same classification system, comparison of DISE findings revealed several incongruities among the examiners. Hence, a review committee was installed to discuss these incongruities and to further standardize the DISE assessment. The results and experiences of this expert review process are presented and discussed in the current paper.

METHODS

Ethical Considerations

The current data are part of a prospective trial (PROMAD)[7] that was registered at clinicaltrials.gov (NCT01532050) and approved by the local ethics committee at the Antwerp University Hospital and University of Antwerp (B300201212961). All patients gave written informed consent.

Patient population

Patients with previous diagnosis of OSA were prospectively recruited for DISE at the ENT department of the Antwerp University Hospital. Key exclusion criteria were body mass index (BMI) > 35 kg/m², previous pharyngeal surgery other than tonsillectomy, syndromic craniofacial or UA anomalies, and history of psychiatric disorders. Type 1 polysomnography was repeated to establish a recent diagnosis prior to performing DISE.

DISE procedure: original scoring

DISE was performed by an experienced ENT surgeon (AEV) in a semi-dark and silent operating theatre with the patients in supine position. Glycopyrrolate (0.2 mg) was administered intravenously to avoid mucosal hypersalivation. Sedation was induced by intravenous bolus injection of midazolam (1.5 mg) and maintained by target-controlled infusion of propofol (2.0–3.0 µg/ml). The sedation level was continuously assessed by bispectral index monitoring (pursued value 50 to 70). Sound and oximetry measurements were integrated in the DISE recordings using specialised software. At the end of each procedure, the presence of UA collapse was noted using a standard scoring system, specifying the level (soft palate, oropharynx, tongue base, epiglottis, or hypopharynx), degree (none, partial, or complete), and direction (anteroposterior, latero-lateral, or concentric).[5]

Review panel: consensus scoring

DISE recordings of 81 patients (Table 1) were reviewed by a panel formed by all three experts in our centre (AVV, EH, OMV). The same scoring system for UA collapse was used. Specific definitions of UA collapse (see results) were established to streamline the review process. All members were blinded for patient characteristics and original DISE scoring. Scoring discrepancies were deliberated until global agreement was reached. A neutral attendant oversaw the discussions and ensured that each panel member was actively involved.

Statistical analysis

The original DISE scoring was compared to the consensus-based DISE scoring by calculating the percentage agreement and Krippendorff's alpha coefficient (α) for the degree and pattern of UA collapse. The pattern involved both the degree and direction of collapse (*e.g.* partial concentric collapse of the soft palate or complete anteroposterior tongue base collapse). 95% confidence intervals were produced using bootstrap with 1000 iterations. All analyses were performed using SPSS statistics 26.0 (IBM corporation, Armonk, NY, USA) with an amended macro designed for Krippendorff's alpha.

RESULTS

Table 2 summarizes the original and consensus-based DISE ratings. The agreement between both ratings is presented in Table 3. Overall, interobserver agreement was highest for oropharyngeal collapse (α_{degree} = 0.63; $\alpha_{pattern}$ = 0.47) and lowest for tongue base collapse (α_{degree} = 0.42; $\alpha_{pattern}$ = 0.27). The agreement for observations at other UA levels ranged from fair to moderate.

In order to further standardize the DISE assessment, discrepancies among the panel members were extensively discussed and reconciled. Global agreement was reached in all but two patients. This consensus-based approach was the foundation for the following recommendations (Figure).

- In three patients (4%), the uvula stuck to the posterior pharyngeal wall without obstructing the lumen. This eight-shaped configuration was considered as no collapse instead of partial collapse.
- Complete palatal collapse was defined as a complete obstruction sustaining at least three seconds. Complete obstructions lasting less than three seconds were classified as partial collapse. Review of the DISE footage revealed 12 complete obstructions (15%) classified as partial collapse and 44 complete obstructions (54%) classified as complete collapse according to this definition.
- Three patients (4%) demonstrated combined anteroposterior and lateral collapse of the soft palate. This specific rectangular or quadrangular configuration was distinguished from concentric collapse of the soft palate (32%) and was referred to as anteroposterior-laterolateral (AP-LL) collapse.
- Tongue base collapse was considered to be partial whenever the vallecula was completely obscured. This was present in 35 patients (43%).
- Lingual tonsil hypertrophy, as observed in 26 patients (32%), was systematically noted in case of tongue base collapse. We advocate to evaluate this aspect as part of the general DISE assessment.
- Complete and partial epiglottic collapse were classified using the same definition as that of palatal collapse. The reviewed recordings demonstrated 8 patients (10%) with partial collapse (*i.e.* < 3 seconds) and 8 patients (10%) with complete collapse (*i.e.* ≥ 3 seconds).

DISCUSSION

Synopsis of key findings

The current study introduced the concept of an expert review panel for DISE scoring. Comparing the consensus-based assessment with the original assessment revealed substantial agreement for ratings of the oropharynx and fair to moderate agreement for ratings of other UA levels. These findings emphasize the importance of specific definitions for certain collapse types as reported in this study.

Clinical applicability

There is a high need for development of a universally accepted approach to classify DISE findings.[2] However, until such classification system exists, it is imperative to attain a uniform and consistent scoring method among examiners in one centre. Our study highlights the role of an internal review panel for such purpose.

By discussing a limited numbers of samples, review panels allow to identify important discrepancies between examiners, thus standardizing DISE scoring. In this study, DISE review mainly changed the scoring at the levels of the soft palate, tongue base, and epiglottis; oropharyngeal findings were fairly consistent. According to these findings, both experienced and inexperienced clinicians may benefit from DISE review. Obviously, however, it is not feasible, both practically and financially, to put every single DISE case through panel discussion. Nevertheless, review panels can prove useful for specific indications, such as (1) patients who need multimodal treatment, (2) patients with complex or rare types of UA collapse, (3) candidates for hypoglossal nerve stimulation (to exclude the risk of complete concentric collapse of the soft palate),[8] and (4) for training purposes.

Expert review panels can be organised in different shapes and sizes depending on the available resources and number of members. This can range from informal case discussions to periodic (multidisciplinary) meetings, or even regional (online) platforms, which might facilitate patient referrals. In any case, DISE review panels might streamline therapeutic decision-making and, accordingly, contribute to an increased quality of patient care.

Comparison with other studies

To the best of our knowledge, this is the first report on the implementation of an internal review panel for DISE findings. Studies in other, mainly oncological, fields have emphasized the need of expert panels and harmonisation processes for diagnostic examinations, demonstrating significant

improvements in interobserver agreement between experts.[9] As DISE is commonly recognized as the cornerstone of anatomical and clinical phenotyping in patients with OSA, review platforms are essential to maintain and increase the quality of this examination.

This study compared DISE ratings of an expert review panel with the original ratings. Notably, both the original and panel assessment were performed by experienced clinicians. Despite differences in study design, the present interobserver agreement rates are in accordance with previous research on DISE.[3-5] *Vroegop et al.* investigated variations in interobserver agreement between experienced and inexperienced examiners.[5] The authors found the highest concordance among experienced examiners at the level of the oropharynx (significantly higher than in the inexperienced group). Lower levels of agreement, however, were reported for the degree of palatal and tongue base collapse, which is consistent with the current findings.

Some of the current recommendations are in line with previously proposed scoring systems for DISE. *Gillespie et al.* systematically assessed lingual tonsil hypertrophy during DISE and specified the occurrence of tongue base collapse according to this anatomical feature.[10] Moreover, *Victores et al.* defined epiglottis collapse as intermittent or sustained,[11] which resembles our three seconds rule of complete collapse.

Strengths and limitations

The main strength of this study relates to the large dataset of DISE samples (n = 81) that was reviewed by the expert panel. This allowed us to include a wide biodiversity of collapse patterns. Additionally, all review members were blinded for patient characteristics and original DISE scoring.

We also acknowledge a number of potential limitations. Ideally, an expert review panel consists of more members, preferably from different centres, to coalesce various experiences and opinions. However, first and foremost, all clinicians who regularly perform DISE in one or more centres should be included. Furthermore, panel review of large numbers of cases may not be cost-effective or practical. Finally, in this study, DISE samples were not independently assessed by the panel members. Although this approach could have revealed more discrepancies, we opted for an exhaustive panel discussion to promote a consensus-based review process.

CONCLUSION

DISE findings may vary significantly depending on personal experience and judgment. The current study demonstrates the benefits of an expert review panel to overcome this issue. Several scoring discrepancies were identified and translated into specific future recommendations. Such endeavours are paramount to standardize DISE scoring and might expedite the development of a universal classification system.

DATA AVAILIBILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Pringle MB, Croft CB. A grading system for patients with obstructive sleep apnoea--based on sleep nasendoscopy. *Clin Otolaryngol Allied Sci.* 1993;18(6):480-484.
2. De Vito A, Carrasco Llatas M, Ravesloot MJ, et al. European position paper on drug-induced sleep endoscopy: 2017 Update. *Clin Otolaryngol.* 2018;43(6):1541-1552.
3. Kezirian EJ, White DP, Malhotra A, Ma W, McCulloch CE, Goldberg AN. Interrater reliability of drug-induced sleep endoscopy. *Arch Otolaryngol Head Neck Surg.* 2010;136(4):393-397.
4. Carrasco-Llatas M, Zerpa-Zerpa V, Dalmau-Galofre J. Reliability of drug-induced sedation endoscopy: interobserver agreement. *Sleep Breath.* 2017;21(1):173-179.
5. Vroegop AV, Vanderveken OM, Wouters K, et al. Observer variation in drug-induced sleep endoscopy: experienced versus nonexperienced ear, nose, and throat surgeons. *Sleep.* 2013;36(6):947-953.
6. Vroegop AV, Vanderveken OM, Boudewyns AN, et al. Drug-induced sleep endoscopy in sleep-disordered breathing: report on 1,249 cases. *Laryngoscope.* 2014;124(3):797-802.
7. Verbruggen AE, Vroegop AV, Dieltjens M, et al. Predicting therapeutic outcome of mandibular advancement device treatment in obstructive sleep apnoea (PROMAD): study design and baseline characteristics. *J Dent Sleep Med.* 2016;3(4):119-138.
8. Vanderveken OM, Maurer JT, Hohenhorst W, et al. Evaluation of drug-induced sleep endoscopy as a patient selection tool for implanted upper airway stimulation for obstructive sleep apnea. *J Clin Sleep Med.* 2013;9(5):433-438.
9. Nestle U, Rischke HC, Eschmann SM, et al. Improved inter-observer agreement of an expert review panel in an oncology treatment trial – Insights from a structured interventional process. *Eur J Cancer.* 2015;51(17):2525-2533.
10. Gillespie MB, Reddy RP, White DR, Discolo CM, Overdyk FJ, Nguyen SA. A trial of drug-induced sleep endoscopy in the surgical management of sleep-disordered breathing. *Laryngoscope.* 2013;123(1):277-282.
11. Victores AJ, Takashima M. Effects of nasal surgery on the upper airway: a drug-induced sleep endoscopy study. *Laryngoscope.* 2012;122(11):2606-2610.

TABLES

Variables	Patients (n = 81)
Gender (% male)	81.5
Age (years)	47.8 ± 9.8
BMI (kg/m²)	27.5 ± 3.3
Neck circumference (cm)	39.0 (37.5–42.0)
AHI (events/h)	14.6 (9.4–23.3)
Supine AHI (events/h)	28.8 (12.7–52.2)
ODI (events/h)	3.5 (1.9–10.7)
Mean O_2 saturation (%)	95.2 (94.1–96.1)
Minimal O_2 saturation (%)	87.0 (84.0–90.0)

Table 1. Patient characteristics.

Data are presented as mean ± standard deviation or median (25^{th}–75^{th} percentile). AHI = apnoea-hypopnoea index; BMI = body mass index; ODI = oxygen desaturation index.

Upper airway collapse	Original scoring				Consensus scoring			
	NA	Absent	Partial	Complete	NA	Absent	Partial	Complete
Soft palate								
AP	0	2	11	32	0	5	22	25
LL			0	2			0	0
Concentric			12	22			9	17
AP-LL			0	0			1	2
Oropharynx								
AP								
LL	6	47	22	6	0	56	18	7
Concentric								
Tongue base								
AP	0	47	25	7	0	37	34	9
LL			2	0			1	0
Concentric								
Epiglottis								
AP	3	42	18	13	1	64	7	8
LL			3	2			1	0
Concentric								
Hypopharynx								
AP								
LL	2	47	28	1	2	61	18	0
Concentric			2	1			0	0

Table 2. Number of patients per collapse pattern.

Shaded areas represent collapse patterns that did not occur. AP = anteroposterior; LL = latero-lateral; NA = not assigned (due to inconclusive findings or panel disagreement).

	Agreement (95%CI)	α-value (95%CI)
Soft palate		
Degree	0.64 (0.53–0.75)	0.39 (0.20–0.60)
Pattern	0.51 (0.39–0.62)	0.40 (0.26–0.55)
Oropharynx		
Degree	0.72 (0.63–0.82)	0.63 (0.43–0.80)
Pattern	0.72 (0.63–0.82)	0.47 (0.29–0.66)
Tongue base		
Degree	0.57 (0.46–0.68)	0.42 (0.24–0.57)
Pattern	0.56 (0.45–0.68)	0.27 (0.08–0.46)
Epiglottis		
Degree	0.63 (0.53–0.73)	0.46 (0.25–0.65)
Pattern	0.63 (0.52–0.75)	0.28 (0.08–0.48)
Hypopharynx		
Degree	0.72 (0.61–0.82)	0.41 (0.17–0.62)
Pattern	0.70 (0.59–0.80)	0.36 (0.12–0.59)

Table 3. Agreement between original and consensus-based ratings.

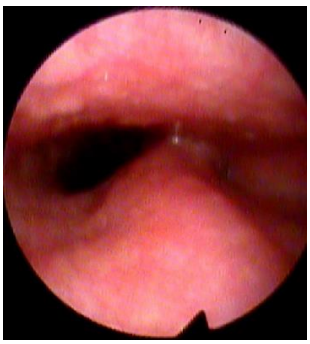
α-value = Krippendorff's alpha coefficient; 95%CI = 95% confidence interval.

FIGURES

Figure 1. Examples of upper airway collapse.

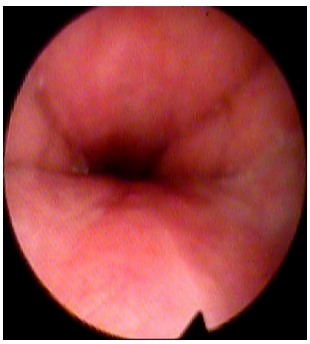
AP–LL = anteroposterior–laterolateral.

A. No collapse

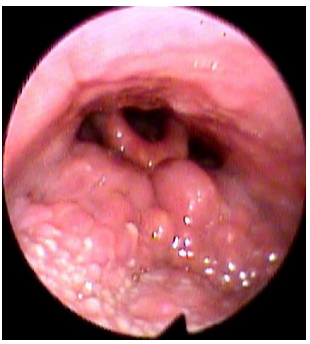


Uvula sticking to the posterior pharyngeal wall

B. Partial collapse

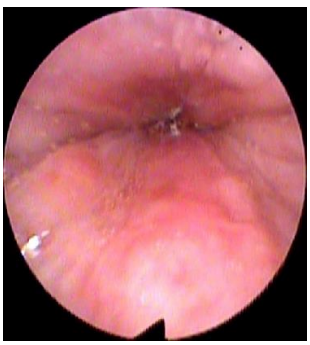


AP–LL collapse of the soft palate

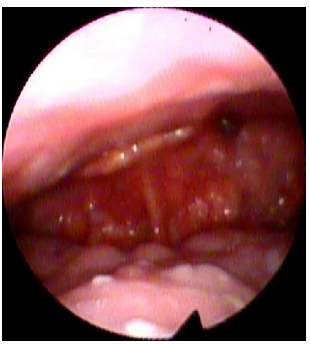


Tongue base collapse due to lingual tonsil hypertrophy

C. Complete collapse (≥ 3 seconds)



Concentric collapse of the soft palate



Anteroposterior epiglottis collapse