Interrater Reliability of Drug-Induced Sleep Endoscopy

Eric J. Kezirian, MD, MPH; David P. White, MD; Atul Malhotra, MD; Wendy Ma, BA;
Charles E. McCulloch, PhD; Andrew N. Goldberg, MD, MSCE

Objective: To determine the interrater reliability of drug-induced sleep endoscopy (DISE).

Design: Prospective cohort; blinded comparison.

Setting: Academic referral center.

Participants: Subjects with obstructive sleep apnea unable to tolerate positive airway pressure therapy.

Interventions: Drug-induced sleep endoscopy was performed with intravenous propofol infusion to achieve sedation, and the videoendoscopy recording was evaluated by 2 independent reviewers.

Main Outcome Measures: The following outcomes were measured: a global assessment of obstruction at the palate and/or hypopharynx; the degree of obstruction at the palate and hypopharynx; and the contribution of individual structures (palate, tonsils, tongue, epiglottis, and lateral pharyngeal walls) to obstruction.

Results: A total of 108 subjects underwent DISE examination. Diagnostic sleep studies demonstrated a mean (SD) apnea-hypopnea index of 39.6 (24.0). Three-quarters of the subjects demonstrated multilevel airway obstruction at the palate and hypopharynx, with a diversity of individual structures contributing to obstruction. The interrater reliability for the presence of obstruction at the palate and hypopharynx (κ values, 0.76 and 0.79, respectively) was higher than for the degree of obstruction (weighted κ values, 0.60 and 0.44). The interrater reliability for the assessment of primary structures contributing to obstruction at the palate and hypopharynx (0.70 and 0.86) was higher than for the contributions of individual structures (κ values, 0.42-0.71). The interrater reliability for evaluation of the hypopharyngeal structures was higher than for those of the palate region.

Conclusion: The interrater reliability of DISE is moderate to substantial.

Trial Registration: clinicaltrials.gov Identifier: NCT00695214

AIRWAY OBSTRUCTION IN obstructive sleep apnea (OSA) can occur at many levels, and the principal regions of dynamic obstruction are the palate and hypopharynx (actually corresponding to the hypopharynx and the retrolingual portion of the oropharynx). Surgical procedures are inherently directed at specific regions of the upper airway, and by addressing airway obstruction in a targeted fashion, it may be possible to tailor surgical treatment to a patient's specific pattern of obstruction—improving surgical results and/or minimizing the scope of surgical intervention. A major goal of surgical assessment is determining the pattern of obstruction, but upper airway anatomical assessment is limited by the fact that evaluation is often static and performed during wakefulness, which may not represent dynamic upper airway behavior during sleep. Drug-induced sleep endoscopy (DISE) differs and may provide a useful upper airway examination. First described as sleep nasendoscopy in 1991,[1] the technique requires pharmacologic sedation and fiberoptic visualization of the upper airway to observe directly and characterize the upper airway collapse that occurs during sedation.[2] Drug-induced sleep endoscopy has been shown to be a safe, feasible, and valid assessment of the upper airway,[3-5] and we have demonstrated moderate to substantial test-retest reliability.[6] The objective of this study was to examine DISE interrater reliability.

METHODS

This prospective cohort study included consecutive subjects seen by the lead author (E.J.K.) in the University of California, San

Author Affiliations:
Departments of Otolaryngology–Head and Neck Surgery (Drs Kezirian and Goldberg and Ms Ma) and Epidemiology and Biostatistics (Dr McCulloch), University of California, San Francisco; and Department of Medicine, Harvard Medical School, Boston, Massachusetts (Drs White and Malhotra).

Francisco (UCSF), Department of Otolaryngology–Head and Neck Surgery. Inclusion criteria included age older than 18 years, apnea-hypopnea index (AHI) higher than 5/h on sleep study, and inability to tolerate positive airway pressure therapy. Exclusion criteria included pregnancy and allergy to propofol or to components of propofol, such as egg lecithin or soybean oil. This study was approved by the UCSF institutional review board, and all subjects provided written informed consent.

All subjects underwent DISE in the operating room. The DISE technique has been described previously.[2] A continuous intravenous infusion of propofol was used as the sole agent to achieve sedation, with the target level of sedation being arousal to loud verbal stimulation, similar to a Modified Ramsay score of 5 or Observer's Assessment of Alertness/Sedation score of 4. The initial infusion rate of propofol was 50 to 75 µg/kg/min, and the rate was adjusted to meet this target level of sedation. The lead author (E.J.K.) performed all DISE examinations, and the digitally recorded video images were later reviewed concurrently but independently by 2 surgeons. The unblinded surgeon (E.J.K.) was aware of subject identity throughout; the blinded surgeon (A.N.G.) was informed only of whether the subject had previously undergone tonsillectomy and had no knowledge of history or physical examination findings, sleep study results, or planned procedures.

The DISE findings were summarized with 3 analyses. Analysis 1 was a global dichotomous (yes or no) assessment of obstruction at each of 2 levels: the palate and the hypopharynx. Analysis 2 reflected the degree of palatal and hypopharyngeal obstruction. This was graded separately for each region subjectively and categorized in an ordinal fashion as less than 50%, 50% to 75%, and more than 75% obstruction; these were not quantitative but were a qualitative assessment of no or mild, moderate, and severe obstruction, respectively. Analysis 3 evaluated specific structures with a determination of which structure at the level of the palate and hypopharynx was the primary factor in airway obstruction, if present, and a dichotomous evaluation of whether each of the individual structures contributed to airway obstruction. Structures were grouped as those at the level of the palate (palate, tonsils when present, and lateral pharyngeal walls at the velopharynx) and the hypopharynx (tongue, epiglottis, and lateral pharyngeal walls at the hypopharynx).

Descriptive statistics were calculated for baseline subject characteristics, and results are reported with means (SDs). Summary statistics for the DISE findings were also calculated, with the McNemar test for paired proportions to evaluate differences between the unblinded and blinded reviewer ratings. The percentage of agreement between reviewer ratings was calculated, and Cohen κ (for Analyses 1 and 3) and weighted κ using linear weights (for Analysis 2) statistics were calculated to assess interrater reliability. Statistical analyses were conducted using Stata software (version 10.0; StataCorp LP, College Station, Texas).

RESULTS

A total of 108 subjects underwent DISE examinations from 2004 through 2008. The mean (SD) age was 43.7 (10.2) years (range, 20-68 years), and 14 of 108 (13%) were female. Most (85 of 108 [79%]) were non-Hispanic white, per subject report. On diagnostic sleep studies, the mean (SD) AHI was 39.6 (24.0), with the following distribution across commonly used AHI cutpoints: 11 of 108 (10%) with an AHI of 5 to less than 15, 36 of 108 (33%) with an AHI of 15 to 30, and 61 of 108 (56%) with an AHI higher than 30. The lowest oxygen satura-

tion during sleep was 80.6% (12.1%), and the subjects with oxygen desaturation level below 90% during sleep (n=74) spent 13.2% (19.0%) of sleep time with an oxygen saturation level below 90%. Twenty-four subjects (22%) had prior tonsillectomy. The mean propofol infusion rate required to achieve sedation was 110 (25) µg/kg/min (range, 50-175 µg/kg/min). The total propofol dose was variable, as the DISE evaluation time varied widely. A subset of subjects (n=22) underwent clinical evaluation of the depth of sedation using the Modified Ramsay score (mean [SD], 4.8 [0.8]; range, 4-6) and the Observer's Assessment of Alertness/Sedation score (2.8 [1.0], range, 1-3).

A complete DISE examination was performed in all cases, and all subjects demonstrated airway obstruction. The DISE findings are presented in **Table 1**. Almost all subjects demonstrated evidence of palatal obstruction, and most also demonstrated hypopharyngeal obstruction (Analyses 1 and 2, reviewer ratings). Both reviewers determined that most subjects (81 of 108 [75%] for the unblinded reviewer and 85 of 108 [79%] for the blinded reviewer) demonstrated obstruction at the levels of both the palate and hypopharynx. Although multilevel obstruction was common, there was diversity in the structures that contributed to obstruction, both in the primary structure and the contribution of individual structures (Analysis 3, reviewer ratings). **Table 2** presents the combinations of individual structures contributing to palatal and hypopharyngeal obstruction, reflecting the multiple observed combinations of involved structures.

The reviewer ratings differed statistically, but the overall distribution of findings was largely similar. Percent agreement and interrater reliability results (κ statistics) are also presented in Table 1. The reliability of the global assessment of obstruction (0.79 and 0.76 in Analysis 1) was somewhat higher than for the degree of obstruction (0.60 and 0.44 in Analysis 2); this was particularly true for the hypopharynx. Analysis 3 results showed greater interrater reliability for the evaluation of the primary structure contributing to airway obstruction (0.70-0.86) than for individual structures (0.42-0.71). The assessments of the palate, tongue, and epiglottis contributions had greater reliability than for other structures.

COMMENT

Drug-induced sleep endoscopy has moderate to substantial interrater reliability. The interpretation of κ values is controversial, but these descriptive terms come from the framework proposed by Landis and Koch.[7] Useful diagnostic tests must demonstrate important characteristics such as safety, validity, and reliability, and this study complements the previous work of others and our own research on test-retest reliability.

Drug-induced sleep endoscopy offers a unique structure-based assessment of the airway compared to other commonly used evaluation techniques. Many DISE classification schemes have been proposed,[3,4,8-17] and we developed our own to balance completeness and simplicity. Our region-based (Analyses 1 and 2) and structure-

Table 1. Summary of DISE Examination Ratings

Analysis	No. (%)		% Agreement	κ Value
	Reviewer 1, Unblinded (n=108)	Reviewer 2, Blinded (n=108)		
Analysis 1: Global assessment of level(s) of obstruction				
Palate	99 (92)	102 (94)	97	0.79
Hypopharynx	90 (83)	91 (84)	94	0.76
Analysis 2: Degree of obstruction				
Palate,%				
<50	9 (8)	7 (6)	81	0.60[d]
50-75	15 (14)	18 (17)		
>75	84 (78)	83 (77)		
Hypopharynx, %[c]				
<50	17 (16)	15 (14)	90	0.44[d]
50-75	29 (27)	47 (44)		
>75	62 (57)	46 (43)		
Analysis 3: Specific structures				
Primary structure contributing to obstruction at level of the palate[c]				
None	9 (8)	7 (6)	83	0.70
Palate	59 (55)	71 (66)		
Tonsils[b]	32 (30)	25 (23)		
LPW-velopharynx	8 (7)	5 (4)		
Primary structure contributing to obstruction at the level of the hypopharynx[c]				
None	16 (15)	14 (13)	92	0.86
Tongue	60 (56)	63 (59)		
Epiglottis	16 (15)	14 (13)		
LPW-hypopharynx	16 (15)	17 (16)		
Specific structures contributing to obstruction[a]				
Palate	91 (84)	95 (88)	93	0.69
Tonsils[b,c]	73 (68)	64 (59)	82	0.42
LPW-velopharynx	62 (57)	64 (59)	74	0.47
Tongue[c]	77 (71)	84 (78)	86	0.64
Epiglottis	31 (29)	32 (30)	88	0.71
LPW-hypopharynx[c]	27 (25)	47 (44)	78	0.53

Abbreviations: DISE, drug-induced sleep endoscopy; LPW, lateral pharyngeal wall.
[a] Each structure was considered separately, with percentages expressed as a fraction of total DISE evaluations. Percentages sum to greater than 100% because it was possible for a subject to have more than 1 structure contributing to airway obstruction.
[b] Twenty-four subjects had undergone previous tonsillectomy.
[c] $P<.05$ (McNemar test on paired proportions comparing unblinded and blinded reviewer ratings).
[d] Reported statistic is for weighted κ.

based (Analysis 3) method serves 2 major purposes: characterizing the pattern of obstruction and selecting among treatment options. This scheme uniquely focuses attention on the primary structures contributing to obstruction in each region; we posit that treatment of these primary structures may be required, at a minimum, to eliminate upper airway obstruction. The interrater reliability—like the test-retest reliability[6]—is higher for the identification of primary structures than for the involvement of individual structures.

According to both reviewers, three-quarters of the subjects in this study demonstrated multilevel obstruction during DISE. Although the upper airway does not consist of 2 regions (palate and hypopharynx) in isolation, if DISE provides an accurate airway assessment, single pharyngeal procedures may be less likely to treat OSA successfully than combinations—or at least single procedures that treat both the palate and hypopharynx. Because surgical procedures are ultimately directed at specific structures, DISE may improve procedure selection and outcomes. This may not be as important for palatal obstruction, for which the most common surgi-

cal treatment is uvulopalatopharyngoplasty with possible tonsillectomy, regardless of specific contribution of the lateral pharyngeal walls at the velopharynx. Although there are alternative palate procedures, it remains unclear how to identify specific subgroups of patients who obtain better or worse outcomes, compared to uvulopalatopharyngoplasty.

However, the involvement of specific structures may be critical for the hypopharynx, where DISE may inform decisions if the multiple treatment options exert differential effects on the tongue, epiglottis, and/or lateral pharyngeal walls. The 3 structures that most commonly contribute to hypopharyngeal airway obstruction are the tongue, epiglottis, and lateral pharyngeal walls, and the results for Analysis 3 indicate that there seems to be an important diversity in the patterns of hypopharyngeal obstruction, a diversity that is evaluated with moderate to substantial interrater reliability. This may prove invaluable if the array of surgical and nonsurgical treatment options to treat the hypopharyngeal airway truly exert differential effects on these various structures. For example, the genioglossus advancement and tongue radiofre-

Table 2. Combinations of Individual Structures Contributing to Airway Obstruction at the Levels of the Palate and Hypopharynx

Type of Obstruction	Reviewer Rating	
	Unblinded	Blinded
Palatal obstruction, %		
None	8	6
Palate	11	20
Tonsils	5	4
LPW-velopharynx	0	1
Palate + tonsils	19	10
Palate + LPW-velopharynx	13	13
Tonsils + LPW-velopharynx	3	1
Palate + tonsils + LPW-velopharynx	42	44
Hypopharyngeal obstruction, %		
None	14	15
Tongue	37	26
Epiglottis	6	5
LPW-hypopharynx	7	1
Tongue + epiglottis	18	10
Tongue + LPW-hypopharynx	13	29
Epiglottis + LPW-hypopharynx	1	1
Tongue + epiglottis + LPW-hypopharynx	4	13

Abbreviation: LPW, lateral pharyngeal wall.

quency procedures likely produce greater changes in tongue position during sleep than in the lateral pharyngeal walls. The hyoid suspension may have less effect on tongue position but may alter the behavior of the epiglottis and/or lateral pharyngeal walls during sleep.

The most common surgical treatment for palatal obstruction is uvulopalatopharyngoplasty, with tonsillectomy in most patients without previous tonsillectomy. Because a similar surgical approach is used for patients regardless of whether the soft palate or velopharynx lateral pharyngeal walls contribute more to obstruction, the question of whether a patient has palate-level obstruction (as in Analysis 1) may be more important than determining whether specific structures contribute to collapse (Analysis 3). Because almost all subjects in this study demonstrated palatal obstruction, the significance of differentiating palate vs velopharynx-level lateral pharyngeal wall obstruction based on DISE (Analysis 3) is unclear. With the adoption of a wider variety of first-line palate procedures, this may prove more important.

The precise relationship between natural sleep and propofol sedation is unclear. Propofol has dose-dependent effects on muscle tone and airway collapsibility, and it is unlikely that propofol sedation is a perfect simulation of natural sleep with precisely the same effects on upper airway dilator muscle activity. Hillman et al[18] demonstrated that propofol sedation, compared to wakefulness, is associated with decreases in genioglossus neuromuscular activity and increases in airway collapsibility similar to that observed in stable non–rapid eye movement sleep. Our target level of sedation (arousal only to loud verbal stimulation) was based on previous research showing more pronounced changes in genioglossus activity and Pcrit during propofol anesthesia (ie, deeper than unconscious sedation),[19] and our overriding concern in this study was minimizing propofol in-

fusion rates to avoid oversedation. We monitored the depth of sedation clinically and in the latter stages incorporated 2 clinical assessments: the Modified Ramsay score and the Observer's Assessment of Alertness/ Sedation score. Although the depth of sedation (with spontaneous respiration and the ability to tolerate fiberoptic endoscopy) remained within a defined range, some authors have proposed standardizing propofol dose with use of target-controlled infusion, a proprietary technology (Diprifusor; Astra-Zeneca Inc, London, England) that calculates effect site (brain) concentration using a 3-compartment pharmacokinetic model.[19,20] This approach is well suited to demonstrating the effects of varying propofol doses, but our objective is based on the target depth of anesthesia. Instead, our work and that of others[18] has shown that different subjects will achieve a defined level of sedation at different propofol doses, so we instead targeted the depth of sedation and adjusted propofol infusion rates accordingly. Fortunately, previous research has demonstrated a linear relationship between propofol infusion rates and serum concentrations at rates of 50 to 200 µg/kg/min,[21] a range that captures the doses used in this study; this suggests that there is little difference between our adjustment of infusion rates within this range and the target-controlled infusion method. As an objective measure of the depth of anesthesia, future investigations should incorporate bispectral index monitoring or some other, more objective measure of the depth of sedation.

Any useful diagnostic evaluation must demonstrate validity and reliability, among other qualities. Other studies have also supported the validity of DISE. Berry et al[5] showed that no subjects without a history of snoring or witnessed apneas (0 of 54) developed snoring or airway obstruction with escalating doses of propofol using target-controlled infusion to a maximum level, whereas all subjects with snoring at baseline (53 of 53) developed snoring and/or airway obstruction. Another study compared 207 primary snorers without OSA with 117 subjects with OSA after receiving bolus doses of propofol and found a higher degree of collapsibility in the latter group, with a correlation between the AHI during natural sleep and the degree of hypopharyngeal obstruction during DISE.[4] Another group[3] administered diazepam (10 mg, with additional doses as needed) to 50 subjects (30 with OSA and 20 with primary snoring but not OSA) and performed polysomnography measurement during DISE for a mean period of over 2 hours; there were no differences between natural sleep and diazepam-induced sedation in the AHI, apnea index, or measures of oxygen saturation. The diversity of patterns on DISE in this study is also reassuring because it likely reflects underlying variation in anatomy and patterns of obstruction during natural sleep rather than an artifact of propofol infusion that might produce identical airway obstruction patterns. Although the gold standard to establish validity would be natural sleep endoscopy, previous researchers have shown that this is challenging.[22,23]

Drug-induced sleep endoscopy as a diagnostic procedure has important theoretical and logistical limitations. Both reviewers are experienced sleep surgeons who have worked together to develop a novel DISE

scoring method. The generalizability of the findings can be explored with larger studies that include more reviewers. Finally, DISE has associated costs and risks (albeit relatively low) that must be balanced against the benefits of the procedure. These benefits—its role in procedure selection and improving outcomes—have been examined in one study in which subjects with isolated palatal obstruction on DISE achieved better outcomes after uvulopalatopharyngoplasty than those with combined palatal and hypopharyngeal obstruction.[13] Future research will be invaluable in making risk-benefit determinations.

Correspondence: Eric J. Kezirian, MD, MPH, Department of Otolaryngology–Head and Neck Surgery, University of California, San Francisco, 2233 Post St, Third Floor, Campus Box 1225, San Francisco, CA 94115 (ekezirian @ohns.ucsf.edu).

REFERENCES

1. Croft CB, Pringle M. Sleep nasendoscopy: a technique of assessment in snoring and obstructive sleep apnoea. *Clin Otolaryngol Allied Sci*. 1991;16(5):504-509.
2. Kezirian EJ. Drug-induced sleep endoscopy. *Op Tech Otolaryngol*. 2006;17(4):230-232.
3. Sadaoka T, Kakitsuba N, Fujiwara Y, Kanai R, Takahashi H. The value of sleep nasendoscopy in the evaluation of patients with suspected sleep-related breathing disorders. *Clin Otolaryngol Allied Sci*. 1996;21(6):485-489.
4. Steinhart H, Kuhn-Lohmann J, Gewalt K, Constantinidis J, Mertzlufft F, Iro H. Upper airway collapsibility in habitual snorers and sleep apneics: evaluation with drug-induced sleep endoscopy. *Acta Otolaryngol*. 2000;120(8):990-994.
5. Berry S, Roblin G, Williams A, Watkins A, Whittet HB. Validity of sleep nasendoscopy in the investigation of sleep related breathing disorders. *Laryngoscope*. 2005;115(3):538-540.
6. Rodriguez-Bruno K, Goldberg AN, McCulloch CE, Kezirian EJ. Test-retest reliability of drug-induced sleep endoscopy. *Otolaryngol Head Neck Surg*. 2009;140(5):646-651.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
8. Abdullah VJ, Wing YK, van Hasselt CA. Video sleep nasendoscopy: the Hong Kong experience. *Otolaryngol Clin North Am*. 2003;36(3):461-471, vi.
9. Camilleri AE, Ramamurthy L, Jones PH. Sleep nasendoscopy: what benefit to the management of snorers? *J Laryngol Otol*. 1995;109(12):1163-1165.
10. den Herder C, van Tinteren H, de Vries N. Sleep endoscopy versus modified Mallampati score in sleep apnea and snoring. *Laryngoscope*. 2005;115(4):735-739.
11. El Badawey MR, McKee G, Heggie N, Marshall H, Wilson JA. Predictive value of sleep nasendoscopy in the management of habitual snorers. *Ann Otol Rhinol Laryngol*. 2003;112(1):40-44.
12. Hessel NS, de Vries N. Diagnostic work-up of socially unacceptable snoring, II: sleep endoscopy. *Eur Arch Otorhinolaryngol*. 2002;259(3):158-161.
13. Iwanaga K, Hasegawa K, Shibata N, et al. Endoscopic examination of obstructive sleep apnea syndrome patients during drug-induced sleep. *Acta Otolaryngol Suppl*. 2003;(550):36-40.
14. Johal A, Battagel JM, Kotecha BT. Sleep nasendoscopy: a diagnostic tool for predicting treatment success with mandibular advancement splints in obstructive sleep apnoea. *Eur J Orthod*. 2005;27(6):607-614.
15. Marais J. The value of sedation nasendoscopy: a comparison between snoring and non-snoring patients. *Clin Otolaryngol Allied Sci*. 1998;23(1):74-76.
16. 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.
17. Quinn SJ, Daly N, Ellis PD. Observation of the mechanism of snoring using sleep nasendoscopy. *Clin Otolaryngol Allied Sci*. 1995;20(4):360-364.
18. Hillman DR, Walsh JH, Maddison KJ, et al. Evolution of changes in upper airway collapsibility during slow induction of anesthesia with propofol. *Anesthesiology*. 2009;111(1):63-71.
19. Eastwood PR, Platt PR, Shepherd K, Maddison K, Hillman DR. Collapsibility of the upper airway at different concentrations of propofol anesthesia. *Anesthesiology*. 2005;103(3):470-477.
20. Roblin G, Williams AR, Whittet H. Target-controlled infusion in sleep endoscopy. *Laryngoscope*. 2001;111(1):175-176.
21. Schnider TW, Minto CF, Gambus PL, et al. The influence of method of administration and covariates on the pharmacokinetics of propofol in adult volunteers. *Anesthesiology*. 1998;88(5):1170-1182.
22. Borowiecki B, Pollak CP, Weitzman ED, Rakoff S, Imperato J. Fibro-optic study of pharyngeal airway during sleep in patients with hypersomnia obstructive sleep-apnea syndrome. *Laryngoscope*. 1978;88(8, pt 1):1310-1313.
23. Rojewski TE, Schuller DE, Clark RW, Schmidt HS, Potts RE. Synchronous video recording of the pharyngeal airway and polysomnograph in patients with obstructive sleep apnea. *Laryngoscope*. 1982;92(3):246-250.