Critical Review Form
Clinical Prediction or Decision Rule

Validation of the Berlin Questionnaire and American Society of Anesthesiologists Checklist as Screening Tools for Obstructive Sleep Apnea in Surgical Patients, *Anesthesiology* 2008;108:822-830

Objectives: **"To validate the Berlin questionnaire and the ASA checklist as screening tools for OSA in surgical patients and to compare them with the STOP questionnaire.** (p. 822)

Methods: **Over 16-months all patients scheduled for elective surgery at Toronto Western Hospital or Mount Sinai Hospital (Toronto, ON) were approached for consent. Excluding criteria included age < 18 years, inability to consent, previously diagnosed OSA, or other sleep disorder, expected abnormal EEG (brain tumor, epilepsy, brain stimulator). All patients who set inclusion criteria and consented were administered three OSA screening tools: ASA checklist, Berlin questionnaire, and the STOP questionnaire. The latter two were administrated in random order followed by the ASA checklist questions by one of three researchers. All those who completed the surveys were invited to an overnight polysomnographic study which included EEG, electrooculogram, submental electromyogram, ECG, thoracoabdominal excursion, respiratory inductive plethysmography, and oronasal airflow measurements with continuous pulse oxymetry. OSA was defined by an Apnea-Hyponea Index > 5. The study was powered using diagnostic accuracy methods of Obuchowski with OSA prevalence 55%, power 80%, and α= 0.05 which would require enrollment of 170 subjects. To assess reliability the Fleiss κ coefficient was calculated on 24 subjects upon whom all three researchers administered the ASA, while the Cohen κ coefficient on 54 and 55 subjects respectively who were re-tested on the Berlin and STOP questionnaires. Post-operative complications were assessed through blinded retrospective chart reviews.**

Guide		Comments
I.	***Is this a newly derived instrument*** **(Level IV)?**	
A.	Was validation restricted to the retrospective use of statistical techniques on the original database? (If so, this is a Level IV rule & is not ready for clinical application).	No – this validation is prospective and on a population distinct from that upon which it was derived so at least a Level III CDR.

II.	**Has the instrument been validated? (Level II or III). If so, consider the following:**	
1a	Were all important predictors included in the derivation process?	ASA checklist and Berlin questionnaire were never derived or previously validated. They were simply consensus statements.
1b	Were all important predictors present in significant proportion of the study population?	Unknown since the authors fail to report the prevalence of the 11 Berlin variables, 14 ASA checklist variables, or four STOP questions.
1c	Does the rule make clinical sense?	Yes, all three rules have face- and content-validity. However, for EM purposes, the Berlin and ASA guides are likely too long with cumbersome scoring systems.
2	Did validation include prospective studies on several different populations from that used to derive it (II) or was it restricted to a single population (III)?	Validated was on a single two-hospital population of Toronto pre-op patients. Therefore, this is Level III CDR and ready for general use in similar Toronto populations.
3	*How well did the validation study meet the following criteria?*	
3a	Did the patients represent a wide spectrum of severity of disease?	Yes, as described in Table 1 (p. 823) the 177/2467 who completed the screening tools and overnight polysomnography had a wide distribution of age, gender, body habitus, and co-morbidities. Baseline functional status and socioeconomic strata are not reported.
3b	Was there a blinded assessment of the gold standard?	Yes, as reported in the PGY-II paper "the certified technologist was blinded to the results of the STOP questionnaire….. and clinical information of the patients".(p. 814).
3c	Was there an explicit and accurate interpretation of the predictor variables & the actual rule without knowledge of the outcome?	Yes. "the perioperative complications of patients were obtained through chart review by a research anesthesiologist who was blinded to the results of the three questionnaires and polysomnography". (p. 823) The investigators also assessed reliability of the three screening tools with the following results ASA Fleiss $\kappa = 0.7460$ ($p<0.001$) Berlin Cohen $\kappa = 0.9168$ (95% CI 0.804-1.000) STOP Cohen $\kappa = 0.923$ (95% CI 0.818-1.000)

<table>
<tr><td>3d</td><td>Did the results of the assessment of the variables or of the rule influence the decision to perform the gold standard?</td><td>No, all subjects who competed the screening tools were invited for overnight sleep study – most declined.</td></tr>
<tr><td>4</td><td>How powerful is the rule (in terms of sensitivity & specificity; likelihood ratios; proportions with alternative outcomes; or relative risks or absolute outcome rates)?</td><td>

2974 eligible subjects were willing to complete the surveys, 2721 (91.5%) did so completely with 254 used for a factor analysis to construct the STOP questionnaire leaving 2467. Unfortunately only 416 (17%) consented to polysomnography and only 211 showed up for the overnight test. Of these, 34 were used to derive the STOP leaving 177 to validate the Berlin and ASA checklist.
Among the 2467, high-risk subjects were identified in 33% by Berlin, 27% by ASA and 28% by STOP. All three screening tools identified subjects with higher AHI but only Berlin (85 vs. 80%) and STOP (84 vs. 80%) identified those with lower overnight oxygen saturations among high risk.
All three rules had reasonable sensitivity but poor specificity for AHI > 5 or AHI > 30 (Table 2, p. 824)

<table>
<tr><th>AHI>5</th><th>sen</th><th>spec</th><th>LR+</th><th>LR-</th><th>AUC</th></tr>
<tr><td>STOP</td><td>66</td><td>60</td><td>1.6</td><td>0.57</td><td>0.703</td></tr>
<tr><td>Berlin</td><td>69</td><td>56</td><td>1.6</td><td>0.55</td><td>0.690</td></tr>
<tr><td>ASA</td><td>72</td><td>38</td><td>1.2</td><td>0.74</td><td>0.783</td></tr>
</table>
<table>
<tr><th>AHI>30</th><th>sen</th><th>spec</th><th>LR+</th><th>LR-</th><th>AUC</th></tr>
<tr><td>STOP</td><td>80</td><td>49</td><td>1.6</td><td>0.41</td><td>0.769</td></tr>
<tr><td>Berlin</td><td>87</td><td>46</td><td>1.6</td><td>0.28</td><td>0.668</td></tr>
<tr><td>ASA</td><td>87</td><td>36</td><td>1.4</td><td>0.36</td><td>0.617</td></tr>
</table>

As defined by AHI > 5, the prevalence of OSA was 69% in the validation cohort. Therefore, a positive STOP, Berlin, or ASA would increase the post-test probability of OSA AHI>30 to 78%, 78% or 73%, respectively. On the other-hand a negative result would decrease the post-test probability to 48%, 38%, or 44% respectively.
There were no deaths or life-threatening complications in either group of patients.
Patients who underwent polysomnography were more likely to suffer post-operative complications (22.8%) vs. 14.6%, p=0.034) but severe desaturation without obvious clinical sequelae (10.9% vs. 5.4%, p=0.039) was the leading complications.
All three screening tools identified increased patients at post-operative risk for respiratory complications including severe desaturation, but STOP was the superior instrument (4.7% vs.

</td></tr>
</table>

		• 15.1% for high-risk, p<0.05). Berlin (9.1 vs. 13.4%) and ASA (5.6% vs. 15%) were less impressive.

III.	**Has an impact analysis demonstrated change in clinical behavior or patient outcomes as a result of using the instrument? (Level I). If so, consider the following:**	No Level I impact analysis was performed.
1	How well did the study guard against bias in terms of differences at the start (concealed randomization, adjustment in analysis) or as the study proceeded (blinding, co-intervention, loss to follow-up)?	Given the patient population derived from those referred to a sleep clinic, likely *selection bias*. Also, since IRB forced management of those with AHI > 30 *co-intervention bias* also likely to have introduced non-random error to observations.
2	What was the impact on clinician behavior and patient-important outcomes?	"Ethics board required Anesthesiologists be notified if the patient's AHI was 30 or greater. In one of our study hospitals, we were required to admit all patients with an AHI of 30 or greater to the Intensive Care Unit.". (p. 828)

Limitations

1) **Selection bias – Incomplete polysomnography testing of all subjects likely increased the prevalence of OSA and distorted the diagnostic and prognostic test characteristics of these screening tools. Unfortunately, this bias is outside the control of researchers, but future investigators should devise strategies to increase compliance with criterion standard testing.**
2) **Insufficient reporting of the prevalence of variables in each screening tool.**
3) **Lack of external validation outside the pre-op setting in Toronto.**
4) **Lack of patient, clinician, or funding-organization important outcomes. Severe desaturation without clinical sequelae was the worst event – most patients and clinicians will never know this event occurred so who cares? Identification of a high-risk subset is encouraging, but how will short-term management be altered to reduce OSA-related morbidity?**
5) **Potential co-intervention bias by admitting AHI > 30.**
6) **No comparison to unaided clinical gestalt.**

Bottom Line

In Toronto pre-op patients the STOP questionnaire is superior to the ASA checklist in identifying a subset at high-risk for severe OSA (AHI > 30), but inferior to the Berlin questionnaire at ruling out severe OSA. None of these screening tools offer impressive diagnostic test characteristics for either mild or severe OSA with the best LR^- 0.28. Furthermore, no patient or clinician important outcomes are predicted by the early identification of OSA in these patients. Future trials will need to demonstrate a benefit for these patients before widespread acceptance can be anticipated.

Berlin Questionnaire

Height _____ m Weight _____ kg Age_____ Male/Female
Please choose the correct response to each question.

Category 1
1. Do you snore?
a. Yes
b. No
c. Don't know

If you snore:
2. Your snoring is:
a. Slightly louder than breathing
b. As loud as talking
c. Louder than talking
d. Very loud—can be heard in adjacent rooms

3. How often do you snore?
a. Nearly every day
b. 3–4 times a week
c. 1–2 times a week
d. 1–2 times a month
e. Never or nearly never

4. Has your snoring ever bothered other people?
a. Yes
b. No
c. Don't know

5. Has anyone noticed that you quit breathing during your sleep?
a. Nearly every day
b. 3–4 times a week
c. 1–2 times a week
d. 1–2 times a month
e. Never or nearly never

Category 2
6. How often do you feel tired or fatigued after your sleep?
a. Nearly every day
b. 3–4 times a week
c. 1–2 times a week
d. 1–2 times a month
e. Never or nearly never

7. During your waking time, do you feel tired, fatigued, or not up to par?
a. Nearly every day
b. 3–4 times a week

c. 1–2 times a week
d. 1–2 times a month
e. Never or nearly never

8. Have you ever nodded off or fallen asleep while driving a vehicle?
a. Yes
b. No

If yes:
9. How often does this occur?
a. Nearly every day
b. 3–4 times a week
c. 1–2 times a week
d. 1–2 times a month
e. Never or nearly never

Category 3
10. Do you have high blood pressure?
a. Yes
b. No
c. Don't know

Scoring Berlin Questionnaire
The questionnaire consists of three categories related to the risk of having OSA.
Categories and scoring:
Category 1: items 1, 2, 3, 4, and 5

Item 1: If *yes* is the response, assign 1 point.
Item 2: If *c* or *d* is the response, assign 1 point.
Item 3: If *a* or *b* is the response, assign 1 point.
Item 4: If *a* is the response, assign 1 point.
Item 5: If *a* or *b* is the response, assign 2 points.

Category 1 is positive if the total score is 2 or more points.
Category 2: items 6, 7, and 8 (item 9 should be noted separately)

Item 6: If *a* or *b* is the response, assign 1 point.
Item 7: If *a* or *b* is the response, assign 1 point.
Item 8: If *a* is the response, assign 1 point.

Category 2 is positive if the total score is 2 or more points.
Category 3 is positive if the answer to item 10 is yes *or* if the BMI of the patient is greater than 30 kg/m^2.

High risk of OSA: ***two or more categories scored as positive***
Low risk of OSA: ***only one or no category scored as positive***

ASA Checklist

Category 1: Predisposing Physical Characteristics
a. BMI _35 kg/m^2
b. Neck circumference _43 cm/17 inches (men) or 40 cm/16 inches (women)
c. Craniofacial abnormalities affecting the airway
d. Anatomical nasal obstruction
e. Tonsils nearly touching or touching the midline

Category 2: History of Apparent Airway Obstruction during Sleep
Two or more of the following are present (if patient lives alone or sleep is not observed by another person, then only one of the following need be present):

a. Snoring (loud enough to be heard through closed door)
b. Frequent snoring
c. Observed pauses in breathing during sleep
d. Awakens from sleep with choking sensation
e. Frequent arousals from sleep

Category 3: Somnolence
One or more of the following are present:

a. Frequent somnolence or fatigue despite adequate "sleep"
b. Falls asleep easily in a nonstimulating environment (*e.g.*, watching TV, reading, riding in or driving a car) despite adequate "sleep"
c. [Parent or teacher comments that child appears sleepy during the day, is easily distracted, is overly aggressive, or has difficulty concentrating]*
d. [Child often difficult to arouse at usual awakening time]*

Scoring:
If two or more items in category 1 are positive, category 1 is positive.
If two or more items in category 2 are positive, category 2 is positive.
If one or more items in category 3 are positive, category 3 is positive.

* Items in brackets refer to pediatric patients.

High risk of OSA: **two or more categories scored as positive**
Low risk of OSA: **only one or no category scored as positive**