Critical Review Form
Clinical Prediction or Decision Rule

STOP Questionnaire - A Tool to Screen Patients for Obstructive Sleep Apnea
Anesthesiology 2008; 108:812-821

Objective: **"To develop and validate a concise and easy-to-use questionnaire for OSA screening in surgical patients".** (p. 812)

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, or expected abnormal EEG (brain tumor, epilepsy, brain stimulator). Sleep specialists and anesthesiologists identified four-questions based upon their previous work with the Berlin questionnaire (see PGY-III article) and a literature review. They then derived the STOP-questionnaire; a 4-item self-administered yes/no form written at the 5th grade Flesch-Kincaid reading level by performing a factor analysis of responses from 254 subjects who completed the STOP and 10-item from the Berlin questionnaire. All 4-items had sufficiently high factor loadings (all > 0.59) to be retained in the STOP model.**

Next, investigators validated the STOP-Questionnaire on 1,875 subjects, all of whom were invited to undergo an overnight polysomnographic study. Additional parameters were collected including BMI, age, gender, and neck circumference which were ultimately used to augment the STOP tool into an instrument with greater diagnostic accuracy called the STOP-Bang (see below).

One certified polysomnographic technologist scored all the sleep studies supervised by a single sleep physician and blinded to all STOP data and clinical information. Ten randomly selected subjects were re-scored by another technologist with nearly identical score results ($r = 0.984$, $p < 0.0001$). Sleep studies included EEG, electrooculogram, submental electromyogram, ECG, thoracoabdominal excursion, respiratory inductive plethysmography, and oronasal airflow measures with continues pulse oxymetry. OSA was defined by an Apnea-Hypopnea Index > 5. The study was powered using diagnostic accuracy method of Obuchowski for sensitivity 88% (± 9%), specificity 80% (± 9%), OSA prevalence 55%, power 80% and $\alpha = 0.05$ which would require 170 subjects to be enrolled.

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 – validation occurred prospectively on a unique group of patients.
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?	Probably since this work was developed by experienced sleep physicians and Anesthesiologists experienced with other OSA screening tools.
1b	Were all important predictors present in significant proportion of the study population?	Uncertain. Although the authors provided mean values for some of the measures stratified by the AHI (Table 3-4, p.817), they fail to provide the prevalence of variables in the study cohort.
1c	Does the rule make clinical sense?	Yes, STOP and STOP-Bang are straight-forward yes/no questionnaires with face-validity and user-friendly applicability.
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)?	Validation was on a single two-hospital population of preoperative patients in Toronto. Therefore, this is a Level III CDR only ready for general use in similar Toronto patient 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 depicted in Table 3 (p. 817) the patients had a variety of ages, body habiti, and ASA scores.
3b	Was there a blinded assessment of the gold standard?	Yes. "The certified technologist was blinded to the results of the STOP questionnaire (i.e., whether patients were at high or low risk of having OSA) and clinical information of the patient. (p. 814)

3c	Was there an explicit and accurate interpretation of the predictor variables & the actual rule without knowledge of the outcome?	Yes. "55 patients answered the STOP questionnaire twice at a time interval of 1 – 27 days (median: 8 days); 53 (96.4%) patients were found to have the same score upon re-testing with a κ coefficient of 0.923 (CI, 0.82 – 1.00)". (p. 815)
3d	Did the results of the assessment of the variables or of the rule influence the decision to perform the gold standard?	No. "All patients regardless of their score on the STOP questionnaire were invited to undergo an overnight polysomnographic study". (p. 813)
4	How powerful is the rule (in terms of sensitivity & specificity; likelihood ratios; proportions with alternative outcomes; or relative risks or absolute outcome rates)?	• Over the enrollment period 2,974 patients were willing to complete the questionnaire and 2,721 (91.5%) did so completely. • 254/2,721, were used for the factor analysis described above leaving 2,467 for derivation and validation of the rule. Unfortunately, only 416 (17%) consented to polysomnography and only 211 actually showed up for the sleep study. • Of 211 with a sleep study, 34 were included in the derivation phase and 177 were used to validate the tool. Among those agreeing to a sleep study, patients were older with greater BMI. Smokers tended to consent and then not show up for their sleep study. • As defined by AHI > 5, OSA prevalence was 69% in the validation cohort. Those with OSA had significantly more HTN (59% vs. 49%), GERD (42% vs. 17%), and DM (23% vs. 12%, NS. On average, they were also older (58 vs. 49 years) with greater neck circumference (40 vs. 36 cm) and more co-morbidity (ASA Class III 44% vs. 16%). • STOP-questionnaire sensitivity 66%, specificity 60%, LR+ 1.64 (95% CI, 1.17 – 2.39) and LR- 0.57 with AUC 0.703 for AHI > 5 (mild OSA). The diagnostic accuracy of the tool improves for severe OSA (LR+ 1.55, LR- 0.42, AUC 0.769). • STOP-Bang (see below) is slightly better for severe OSA LR+ 1.59 (1.28 – 1.84), LR- 0 (0 – 0.31) with AUC 0.822.

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:**	
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)?	Potential *selection bias* as only 17% accepted polysomnography (and only half of those followed through). Although criterion standard testing was likely biased by those with disease (as reflected by increased age and BMI among compliant subjects), such bias is largely out of the control of the investigators.
2	What was the impact on clinician behavior and patient-important outcomes?	No! This is the major flaw of this research. Even if the tool can reliably and accurately identify a subset of pre-op patients at high risk for OSA – so what? Will early identification alter surgical or Anesthesia plans? Can identification of occult OSA in pre-op, office, or ED alter OSA-related morbidity and mortality? Will the risks of a positive STOP screen (patient angst, expensive, confirmatory polysomnography testing) outweigh the potential benefits? Would an EM physician use this information accurately in the ED?

Limitations

1) **Incomplete polysomnography testing of all subjects likely increasing the prevalence of OSA and distorting the diagnostic test characteristics of the STOP and STOP-Bang instruments. Unfortunately, this bias is out of the realm of control of investigators although future researchers should devise strategies to increase criterion standard compliance to minimize bias.**

2) **Insufficient reporting of the prevalence of various candidate variables.**

3) **Lack of external validation outside the pre-op setting or outside of Toronto.**

4) **Lack of patient, clinician, or funding organization important outcomes. Identification of a high-risk subset is encouraging. Now what can we do about it to reduce OSA related morbidity.**

Bottom Line

In Toronto pre-op patients, the STOP-Bang (below) questionnaire can reliably identify a low risk subset of patients who are unlikely to have AHI-defined OSA. Before widespread application of this CDR, external validation should confirm the diagnostic accuracy of STOP-Bang in other patient populations. Future research should also examine the implementation science for this CDR. Specifically, how will recognition for a non-low risk subset alter management to reduce OSA-related morbidity?

STOP-Bang Rule

Answer each of the following yes or no:

1. Do you snore loudly (louder than talking or loud enough to be heard through closed doors)?
2. Do you often feel tired, fatigued, or sleepy during daytime?
3. Has anyone observed you stop breathing during your sleep?
4. Do you have or are you being treated for high blood pressure?
5. BMI more than 35kg/mm^2?
6. Age over 50 years old?
7. Neck circumference > 40 cm?
8. Male gender?

High-risk for OSA 3 ≥ yes answers ($LR^+ = 1.55$)
Low-risk for OSA 3 < yes answers ($LR^- = 0$)