Development of the PSG PLUS PHYSIOLOGY model

Step (1) Picked possible surrogates of the physiological traits (See Table 1 in the main text).

Step (2) Looked for correlations between baseline surrogates and postoperative AHI.

Parameters were further considered if they showed correlation with postoperative AHI in either Pearson or Spearman rank correlation test.

The “mean duration of the respiratory events” was excluded from the model since it didn’t show any correlation with postoperative AHI (both Pearson and Spearman correlation \( P > .05 \)).

Step (3) looked for cutoffs that best discriminated between patients with different surgical outcomes.

The cutoff points of each parameter above were decided according to either (1) previously reported physiological cutoffs, if available, or (2) the calculated receiver operating characteristic (ROC) for a residual AHI > 20 or > 10 events/h, with some rounding to whole numbers (see Table S1). Each candidate parameter that might contribute to a residual postoperative AHI would be categorized as 0 (good), 1 (intermediate), or 2 points (bad). The higher the points are, the higher the risk of the traits may add to residual sleep apnea.
Table S1—Rationale for the cutoffs of the surrogates.

<table>
<thead>
<tr>
<th>Surrogate markers of physiology traits</th>
<th>Rationale</th>
<th>How the cutoffs were chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anatomy:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AHIREM is correlated with collapsibility of upper airway (Pcrit)(^1)</td>
<td>AHIREM &lt; 20 indicates a relatively good anatomy (0 points) AHIREM (\geq) 50 events/h indicates a bad anatomy (Pcrit &gt; +2-3 cmH(_2)O), surgery less possible to correct it to a normal range (2 points)</td>
<td></td>
</tr>
<tr>
<td><strong>Arousals:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated arousal threshold</td>
<td>Estimated arousal threshold &lt; -15 cmH(_2)O indicate high arousal threshold(^2)</td>
<td></td>
</tr>
<tr>
<td>Fraction of events that are hypopneas</td>
<td>Cutoffs were decided according to ROC for a residual AHI &gt; 20 and/or 10 events/h.</td>
<td></td>
</tr>
<tr>
<td><strong>Control of breathing:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central apnea index and mixed apnea Index may suggest clinically relevant abnormal control of breathing</td>
<td>Central or mixed apnea index &lt; 5 events/h, normal (0 point) Central or mixed apnea index (\geq) 5 events/h, abnormal (2 points)</td>
<td></td>
</tr>
<tr>
<td>High loop gain maybe a risk factor for postoperative sleep apnea.</td>
<td>Cutoffs were decided according to ROC for a residual AHI &gt; 20 and/or 10 events/h with the estimated loop gain derived from PSG as the test variable.</td>
<td></td>
</tr>
<tr>
<td><strong>Muscle responsiveness:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REM predominant OSA (AHIREM &gt; AHINREM) suggests ability of muscles to overcome passive anatomy; more frequent sleep apnea in NREM sleep may be a sign of negative effort dependence (or inadequate muscular compensation)(^3,4)</td>
<td>Cutoffs were decided according to ROC for a residual AHI &gt; 20 and/or 10 events/h with Ratio of AHIREM vs AHINREM as the test variable.</td>
<td></td>
</tr>
</tbody>
</table>

AHI, apnea-hypopnea index; AHIREM, apnea-hypopnea index during rapid eye movement sleep; AHINREM, apnea-hypopnea index during non-rapid eye movement sleep
• **Fraction of events that are Hypopneas**

Fraction of events that are hypopneas of 20% was chosen as cutoffs because it had a high value of specificity in both ROC curves for a residual AHI > 20 and 10 events/h (specificity = 0.92 when fraction of events that are hypopneas = 18.34%, and specificity = 1 when fraction of events that are hypopneas = 18.34%, respectively). That is, a fraction of hypopneas of more than 20% was a good predictive value for a low (less than 20 or 10) postoperative AHI.

Fraction of events that are hypopneas of 10% was chosen as a cutoff because it had the highest value of sensitivity plus specificity (sensitivity = 1, specificity = 0.875 when fraction of events that are hypopneas = 9.78%) in ROC for a residual AHI > 20. Thus, a fraction of hypopneas < 10% tended to accurately predict a postoperative AHI > 20 events/h. As noted in the discussion of the main paper, given that more percentage of hypopneas were associated with a better outcome, it may be that this is another marker of anatomy, as opposed to a surrogate of the respiratory arousal threshold.

• **Loop gain**

Loop gain of 0.6 was chosen as cutoffs because it had the highest value of sensitivity plus specificity (sensitivity = 1, specificity = 0.471, when loop gain = 0.583) in ROC for a residual AHI > 20. And the point had the highest value of sensitivity plus specificity (sensitivity = 0.875, specificity = 0.714 when loop gain = 0.583) in ROC for a residual AHI > 10. Thus, loop gain > 0.6 were less likely to have good outcomes.
Loop gain of 1 was chosen as cutoffs because it had the highest value of specificity in both ROC curves for a residual AHI > 20 and 10 events/h (specificity = 1 when loop gain = 1.022, and specificity = 1 when loop gain = 0.942, respectively)—ie, no subject with a loop gain > 1 had a postoperative AHI less than 20 events/h.

- **Ratio of AHI\textsubscript{REM} vs AHI\textsubscript{NREM}**

Because we examined the ratio of AHI\textsubscript{REM} and AHI\textsubscript{NREM}, we log transformed the data when looking for cutoffs. Two cutoffs became apparent for those who had a good response and those who were nonresponders. Log\(_{10}\) (ratio of AHI\textsubscript{REM} vs AHI\textsubscript{NREM}) = \(-0.0912\) (which indicates AHI\textsubscript{NREM} vs AHI\textsubscript{REM} = 1.3) was sensitive (likely to happen, sensitivity = 1, specificity = 0.5) for an AHI > 20 after surgery, and log\(_{10}\) (ratio of AHI\textsubscript{REM} vs AHI\textsubscript{NREM}) = \(0.0786\) was specific (unlikely to happen, sensitivity = 0.857, specificity = 0.958) for an AHI < 20 after surgery. Therefore, 130% AHI\textsubscript{NREM} < AHI\textsubscript{REM} and AHI\textsubscript{REM} < 130% AHI\textsubscript{NREM} were used as cutoffs.

**Models incorporating at least one surrogate parameter for each trait**

Based on correlations above, there was more than one surrogate parameter of arousal threshold and loop gain. A series of regression analyses were run to examine a model including just one of the parameters would perform as well as the model with all representative parameters.

- For arousal threshold terms, the predictive value of (1) fraction of events that were hypopneas, (2) estimated arousal threshold based on a previously published formula and (3) model including both fraction of events that were hypopneas and estimated
arousal threshold were compared.

Fraction of events that were hypopneas alone had an $r^2 = .354$, $F = 15.925$, $P < .001$.

Fraction of events that were hypopneas alone as a predictor did perform as well as the model including both fraction of events that were hypopneas and estimated arousal threshold, $r^2$ change = -.002, $F = 0.074$, $P = .788$. Therefore, for simplicity, we included only one surrogate of arousal threshold.

- For control of breathing terms, the predictive value of (1) loop gain derived from baseline PSG (2) Central or mixed apnea index and (3) model including both estimated loop gain and Central or mixed apnea index were compared. Estimated loop gain alone had an $r^2 = .288$, $F = 11.706$, $P = .002$. Central or mixed apnea index alone had an $r^2 = .430$, $F = 21.841$, $P < .001$. Including both terms as surrogates of stability of breathing control significantly increased $r^2$ of the model ($r^2$ change = .241, $P = .001$ and $r^2$ change = .099, $P = .022$) compared to the model with estimated loop gain/central or mixed apnea index alone. Thus both surrogates were incorporated into the control of breathing term. It is unclear from this training set the reason why two terms for control of breathing improves the predictive value of the model. It may be that control of breathing is of relatively greater importance than other traits. Alternatively, it may be that our surrogate markers actually incorporate other information. For example, loop gain is affected by baseline OSA severity.
References


