The pathogenesis of obstructive sleep apnea

Luu V. Pham, Alan R. Schwartz

Division of Pulmonary and Critical Care Medicine, Department of Medicine, Johns Hopkins Sleep Disorders Center, Johns Hopkins University, Baltimore, Maryland, USA

Contributions: (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study materials or patients: None; (IV) Collection and assembly of data: None; (V) Data analysis and interpretation: None; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Luu V. Pham. 5501 Hopkins Bayview Circle, Baltimore, Maryland 21224, USA. Email: lpham1@jhmi.edu.

Abstract: Obstructive sleep apnea (OSA) is a major source of cardiovascular morbidity and mortality, and represents an increasing burden on health care resources. Understanding underlying pathogenic mechanisms of OSA will ultimately allow for the development of rational therapeutic strategies. In this article, we review current concepts about the pathogenesis of OSA. Specifically, we consider the evidence that the upper airway plays a primary role in OSA pathogenesis and provide a framework for modelling its biomechanical properties and propensity to collapse during sleep. Anatomical and neuromuscular factors that modulate upper airway obstruction are also discussed. Finally, we consider models of periodic breathing, and elaborate generalizable mechanisms by which upper airway obstruction destabilizes respiratory patterns during sleep. In our model, upper airway obstruction triggers a mismatch between ventilatory supply and demand. In this model, trade-offs between maintaining sleep stability or ventilation can account for a full range of OSA disease severity and expression. Recurrent arousals and transient increases in airway patency may restore ventilation between periods of sleep, while alterations in neuromuscular and arousal responses to upper airway obstruction may improve sleep stability at still suboptimal levels of ventilation.

Keywords: Obstructive sleep apnea (OSA); upper airway; inspiratory flow limitation; Starling resistor

Submitted Feb 13, 2015. Accepted for publication Jul 17, 2015.
doi: 10.3978/j.issn.2072-1439.2015.07.28
View this article at: http://dx.doi.org/10.3978/j.issn.2072-1439.2015.07.28

Introduction

Obstructive sleep apnea (OSA) is a highly prevalent disease, characterized by upper airway collapse during sleep resulting in recurring arousals and desaturations. Estimates of disease prevalence range between 3% and 10% of the population (1,2). Prevalence has risen with escalating rates of obesity, a major risk factor for OSA (2,3). Significant clinical consequences of the disorder cover a wide spectrum and include daytime hypersomnolence, neurocognitive dysfunction, cardiovascular disease, metabolic dysfunction, respiratory failure, and cor pulmonale (4-16). As a result, OSA represents an increasing burden on health care resources. Understanding underlying pathogenic mechanisms of OSA will ultimately allow for the development of rational therapeutic strategies in an era of personalized medicine.

In this article, we will review current concepts about the pathogenesis of OSA. Specifically, we will consider factors that initiate upper airway obstruction during sleep and examine responses to airway obstruction over the course of the ensuing event. In considering factors responsible for the initiation of an obstructive apnea, we will model the alterations in pharyngeal biomechanics that to lead airway obstruction during sleep. When the pharynx collapses, airflow obstruction elicits neuromuscular responses that can mitigate the obstruction and restore airway patency and ventilation. If these neuromuscular mechanisms are inadequate, additional factors contribute to the development of recurrent periods of airway obstruction and arousals from sleep. In this context, models that predict the likelihood of developing recurrent sleep disordered breathing episodes
will be considered. In modelling recurrent sleep disordered breathing episodes, we will elaborate generalizable mechanisms by which upper airway obstruction destabilizes respiratory and sleep-wake patterns.

**Upper airway biomechanics**

OSA is characterized by recurrent periods of upper airway occlusion during sleep (17). In modeling the biomechanics of pharyngeal airflow obstruction, we consider the fact that the upper airway collapses dynamically during sleep and reopens during wakefulness. Investigators have previously modelled dynamic alterations in patency as a function of transmural pressure across collapsible segments in biologic conduits in the cardiovascular, gastrointestinal, and genitourinary systems (18-24) (Figure 1). In the case of the upper airway, the collapsible segment is bordered by two rigid segments upstream (nasal passages) and downstream (trachea) (Figure 1A) (26). The segments upstream and downstream to the collapsible site have fixed diameters and resistances, RUS and RDS, respectively, and the pressures upstream and downstream are PUS and PDS, respectively.

Several important features of this model, known as the Starling resistor, are worth emphasizing. When PUS and PDS are less than the critical pressure surrounding the collapsible segment (Pcrit), the transmural pressure is negative, the airway closes and airflow ceases (Figure 1B). Flow can be re-established by raising PUS above Pcrit. If both PUS and PDS are greater than Pcrit however, transmural pressure remains positive (Figure 1D). Under these conditions, flow through the upper airway is proportional to the pressure gradient across the entire airway, and can be described by the voltage-current relationship in Ohm's law:

\[ V_I = \frac{P_{US} - P_{DS}}{R_{US} + R_{DS}} \]

In contrast, when PUS is greater than Pcrit and PDS is less than Pcrit, however, the airway enters a flow-limited condition (Figure 1C). To illustrate the mechanism by which inspiratory flow limitation arises, consider the effects of lowering PDS progressively during inspiration. At the start of inspiration, flow commences upon lowering PDS. As inspiration progresses and PDS falls below Pcrit, the airway begins to collapse. If the upper airway were to occlude, flow would cease transiently. As the upper airway occludes, the pressure immediately upstream of the occlusion would equilibrate with PUS and rise above Pcrit. This increase in pressure would inevitably lead to reopening of the airway. As the airway cycles rapidly between an open and closed state, the pressure at the collapsible segment remains
nearly constant at $P_{CRIT}$. Because pressure in the collapsible segment is constant, airflow also remains constant. Under these circumstances, flow becomes independent of $P_{DS}$ and plateaus at a maximal level ($V_{Imax}$) as $P_{CRIT}$ replaces $P_{DS}$ and becomes the effective back-pressure to inspiratory flow. The level of $V_{Imax}$ is therefore governed by the pressure gradient between $P_{US}$ and $P_{CRIT}$ and the resistance across the upstream segment, according to the following equation:

$$V_{max} = \frac{P_{US} - P_{CRIT}}{R_{US}} \tag{2}$$

In this model, decreasing $P_{DS}$ does not cause the upper airway to occlude and cannot account for the development of obstructive apneas during sleep.

Now consider the effects of altering $P_{US}$ on inspiratory flow in Figure 2. The effects of increasing levels of CPAP on upper airway pressure-flow dynamics is represented. In this figure, CPAP is applied at pressures of 4, 6, 9 and 12 cmH2O. At a low nasal pressure (4 cmH2O), both $P_{US}$ and $P_{DS}$ are less than $P_{CRIT}$ the airway occludes and no flow occurs. At an intermediate pressure of 6 cmH2O, $P_{US}$ exceeds $P_{CRIT}$ and flow is re-established. Nevertheless, $P_{DS}$ still falls below $P_{CRIT}$ over the course of inspiration, resulting in a plateau of mid-inspiratory airflow (flow-limitation) at $V_{max}$ (see arrows). With further increases in nasal pressure from 6 to 9 cmH2O, inspiratory flow plateaus at a proportionately higher level, as described by Eq. [2], but flow still remains limited. Finally, at nasal pressure of 12 cmH2O, both $P_{US}$ and $P_{DS}$ have increased sufficiently such that both pressures remain above $P_{CRIT}$ throughout inspiration and flow limitation is abolished. Conversely, decreases in $P_{US}$ elicit flow limitation and progressive decreases in $V_{max}$ until flow ceases (the upper airway occludes) when $P_{US}$ falls below $P_{CRIT}$. A simple pressure-flow plot (Figure 3) describes this linear relationship between $V_{max}$ and $P_{US}$ (Eq. [2]). This relationship can be used to define $P_{CRIT}$ at the zero flow intercept and $R_{US}$ as the reciprocal of the slope of this line (27,28).

It is important to note that inspiratory airflow limitation exerts two distinct loads on the respiratory system. First, airway resistance increases markedly in the flow limited compared to the non-flow limited state. During non-flow limited breathing (in the absence of upper airway collapse), the combined resistances of the upstream and downstream segments (Eq. [1]) is approximately 1 to 2 cmH2O/L/s, which accounts for approximately half of the total resistance of the respiratory system during tidal breathing. In contrast, resistance of the upstream segment alone (Eq. [2]) during periods of inspiratory airflow limitation increases into the range of 20 to 40 cmH2O/L/s. Second, additional load is imposed on the respiratory system during periods of flow limitation by virtue of the fact that patients continue to exert ever-increasing effort without increasing inspiratory airflow. In essence, a large portion of the pressure generated by the respiratory pump muscles is wasted by dynamic collapse of the upper airway without augmenting ventilation. Thus, increases in airway resistance and dynamic collapse of the upper airway augment work of breathing during periods of inspiratory airflow limitation and/or complete upper airway obstruction.
The role of upper airway obstruction in OSA pathogenesis

Current evidence suggests that disturbances in P\text{CRIT} play a primary role in OSA pathogenesis. The role of pharyngeal obstruction in OSA pathogenesis can be considered in light of Koch’s postulates (29), which establish criteria for demonstrating a causal relationship between pathogenic factors that promote upper airway obstruction and the overt polysomnographic manifestation of the disease. These principles require first and foremost that pathogenic factors causing upper airway collapse are associated with OSA. Investigators have examined the association between pharyngeal collapsibility and the clinical manifestations of OSA in several observational studies (25,28,30-35). Elevations in P\text{CRIT} have been demonstrated in OSA patients compared to age, sex and body mass index (BMI) matched controls under general anesthesia and neuromuscular blockade (35) as well as during sleep (25).

Further strength for the association between upper airway obstruction and OSA pathogenesis is evidenced by studies demonstrating that pharyngeal collapsibility (P\text{CRIT}) is both a sensitive (a large proportion of persons with OSA have collapsible upper airways) and specific (a large proportion of normal persons do not have collapsible airways) finding in OSA. High levels of sensitivity and specificity of P\text{CRIT} can be inferred from numerous studies that have demonstrated quantitative differences in P\text{CRIT} between health and disease (27,28,30-34,36-41). In the aggregate, these studies demonstrated that nearly all persons with OSA have a P\text{CRIT} greater than −5 cmH\text{2}O, indicating that upper airway collapsibility is a sensitive marker for OSA (Figure 4). In contrast, elevations in P\text{CRIT} were nearly absent in normal controls, suggesting that increased pharyngeal collapsibility is also highly specific to OSA.

\textbf{Figure 3} Relationship between maximal inspiratory flow (V_{\text{max}} ± standard deviation) and upstream nasal pressure (P_{\text{US}}) during NREM sleep. Inspiratory flow limitation in a non-apneic normal individual was induced by sequentially lowering P_{\text{US}}. Under these conditions, V_{\text{max}} is linearly dependent on P_{\text{US}}. The reciprocal of the line slope represents the upstream resistance (R_{\text{US}}). The critical pressure (P_{\text{CRIT}}) is the zero-flow intercept, −21 cmH\text{2}O (indicated by arrow), indicating total upper airway occlusion. Adapted from Schwartz \textit{et al.} 1988 (27).

\textbf{Figure 4} Summary of studies examining upper airway collapsibility (critical closing pressure, P_{\text{CRIT}}) in normal individuals and obstructive sleep apnea (OSA) patients. Mean P_{\text{CRIT}} (± standard deviation) in non-apneic controls (open squares) and apneic subjects (closed squares) are plotted. P_{\text{CRIT}} is elevated in sleep apnea patients compared to normal controls. Furthermore, these studies demonstrate minimal overlap of P_{\text{CRIT}} between disease and health, suggesting that an elevated P_{\text{CRIT}} is both sensitive and specific for OSA. Source studies, as indicated by enclosed numbers, are: 1. Meurice \textit{et al.}, 1996 (37); 2. Meurice \textit{et al.}, 1996 (36); 3. Issa & Sullivan, 1984 (30); 4. Gold \textit{et al.}, 2002 (33); 5. Philip-Joet \textit{et al.}, 1996 (38); 6. Schwartz \textit{et al.}, 1988 (27); 7. Smith \textit{et al.}, 1988 (28); 8. Ng \textit{et al.}, 2003 (39); 9. Kirkness \textit{et al.}, 2003 (34); 10. Farre \textit{et al.}, 2003 (40); 11. Series \textit{et al.}, 1996 (42); 12. Gold \textit{et al.}, 2002 (33); 13. Sforza \textit{et al.}, 2000 (32); 14. Sforza \textit{et al.}, 1999 (31); 15. Issa & Sullivan, 1984 (30). Adapted from Pack, 2009 (41).
Additional evidence for the primacy of upper airway collapse in OSA pathogenesis is provided by studies demonstrating a dose-response relationship between pharyngeal collapsibility and severity of OSA (Figure 5). As $P_{\text{CRIT}}$ increases, the severity of upper airway obstruction also increases progressively, suggesting a dose-response relationship between upper airway collapsibility and worsening obstruction in asymptomatic snorers and patients with upper airway resistance syndrome (UARS), obstructive hypopneas, and obstructive apneas. Adapted from Schwartz et al. 1988 and Gleadhill et al. 1991 (25,27).

Figure 5 Relationship between $P_{\text{CRIT}}$ and clinical disorders characterized by upper airway obstruction during sleep. As $P_{\text{CRIT}}$ increases, the severity of upper airway obstruction also increases progressively, suggesting a dose-response relationship between upper airway collapsibility and worsening obstruction in asymptomatic snorers and patients with upper airway resistance syndrome (UARS), obstructive hypopneas, and obstructive apneas. Adapted from Schwartz et al. 1988 and Gleadhill et al. 1991 (25,27).

Studies inducing experimental upper airway collapse during sleep also implicate pharyngeal obstruction in OSA pathogenesis. Indeed, manipulating nasal pressure recapitulates the entire OSA disease spectrum. With the application of subatmospheric nasal pressure during sleep, stable flow limited breathing and snoring were observed in healthy test subjects (27). Further reductions in nasal pressure resulted in recurrent obstructive hypopneas and apneas, which occurred at a rate of ~20-40 episodes per hour and were associated with oxyhemoglobin desaturations and arousals. Continuous application of subatmospheric nasal pressure during sleep also caused alterations in sleep architecture, with increases in stage 1 and stage 2 sleep, and decreases in stage 3/4 and REM sleep compared to baseline (43). Moreover, when study participants were subjected to two consecutive nights of experimentally induced OSA, multiple daytime sleep latency times fell markedly, indicating that excessive daytime somnolence had developed. Thus, experimental evidence suggests that airway collapse alone is sufficient to cause OSA.

Conversely, OSA can be treated with interventions designed to restore upper airway patency, further fulfilling Koch’s postulate that upper airway collapse is necessary for disease pathogenesis. In fact, treatments that decrease $P_{\text{CRIT}}$ (e.g., weight loss or uvulopalatopharyngoplasty) lead to improvements in OSA and to resolution of disease when $P_{\text{CRIT}}$ falls below $-5$ cmH$_2$O (27,44). At this level of $P_{\text{CRIT}}$, a transmural pressure of 5 cmH$_2$O provides adequate airflow to stabilize breathing patterns during sleep. Similarly, a positive transmural pressure can be induced by increasing $P_{\text{US}}$, leading to resolution of upper airway obstruction. With application of progressively increasing nasal pressure during CPAP titration, upper airway obstruction and recurrent obstructive apneas and hypopneas are reversed. As nasal pressure increases, episodic obstructive apneas give way to hypopneas when $P_{\text{US}}$ rises above $P_{\text{CRIT}}$. Further increases in nasal pressure stabilize respiratory patterns, leading to regular snoring and ultimately to the resolution of flow-limitation altogether as CPAP pressures rise to therapeutic levels.

Finally, the relationship between upper airway collapsibility and OSA is biologically plausible, as exemplified by several approaches to modelling upper airway obstruction. First, airway obstruction could be due to increasing airway resistance that is produced by narrowing of an otherwise rigid structure. In this model, no matter how narrow the tube, flow still remains dependent on downstream pressure. Although flow is reduced when resistance is high, this model cannot account for the development of inspiratory flow limitation (or snoring) in which airflow becomes independent of downstream pressure. Second, investigators have postulated that...
which flow limitation develops because surrounding tissue pressures produce a constant back pressure to airflow.

**Determinants of upper airway collapsibility**

Elevations in $P_{CRIT}$ can be attributed to passive structural defects in the upper airway and disturbances in neuromuscular control (45,46). Utilizing specialized physiologic techniques, investigators have separated structural from neuromuscular components by measuring $P_{CRIT}$ during sleep under conditions of reduced (passive) and elevated (active) neuromuscular activity, respectively (35,47,48). They demonstrated that airway collapsibility was elevated in OSA patients under passive conditions, suggesting underlying anatomic defects in OSA patients compared to age, sex and BMI matched normal controls (see Figure 6A). These OSA patients also exhibited blunted active responses to airway obstruction compared to controls, indicating concomitant deficits in pharyngeal neuromuscular control (48). Disturbances in neuromuscular control remained even in those OSA patients whose structural loads were comparable to those of normal controls (see Figure 6B). These findings suggest that elevations in $P_{CRIT}$ in OSA patients are due to defects in both upper airway structural and neuromuscular control (Figure 7), and those disturbances in both play a pivotal role in OSA pathogenesis. In fact, OSA can only develop when neuromuscular responses do not adequately mitigate the obstruction caused by excess pharyngeal mechanical loads.

**Anatomic alterations**

Investigators have identified a variety of anatomic factors that contribute to increases in airway collapsibility. Various craniofacial features related to either skeletal morphology or pharyngeal soft tissue may predispose to upper airway collapse. Mandibular size, maxillary height, and hyoid position have been associated with risk for OSA (32,35,49-55). Decreased velopharyngeal area, and tonsillar hypertrophy are soft tissue features that have been associated with increased upper airway collapsibility (35,56). In general, these anatomical variants are thought to increase $P_{CRIT}$ by restricting the size of the boney enclosure around the pharynx and/or increasing the amount of soft tissue contained therein (57).

Obesity and especially abdominal adiposity are also important anatomical risk factors for upper airway obstruction during sleep. The upper airways of obese
individuals are more susceptible to collapse (44) and $P_{\text{CRIT}}$ increases 1.0 and 1.7 cmH$_2$O per 10 kg/m$^2$ BMI increase in women and men, respectively (58). Increased fat deposition around the pharynx and airway narrowing (59-61) may increase the extraluminal tissue pressure and augment upper airway collapsibility (62). In addition, lung volumes are decreased in obese persons, leading to decreased caudal traction on the upper airway and an increased critical closing pressure (63-67). These reductions in caudal traction are most pronounced in patients with abdominal adiposity, which can decrease lung volume nearly to the level of residual volume (68-71). Conversely, improvements in OSA with weight loss are likely due to reductions in surrounding tissue pressure and increases in caudal traction, both of which decrease $P_{\text{CRIT}}$ (44).

**Disturbances in neuromuscular control**

Although structural defects play a clear role in the pathogenesis of OSA, these defects may only account for one-third of the variability in OSA severity (72), leaving neuromuscular responses accounting for much of the balance of OSA variability. OSA patients appear to be especially dependent on neuromuscular activity to maintain airway patency and ventilation during sleep (73). This activity varies markedly with reduction in pharyngeal dilator tone at sleep onset that predispose to airway obstruction (74). Reductions in neuromuscular tone are also suspected to contribute to increased OSA severity during REM compared to NREM sleep in selected patients, and particularly in women and children (75-78). Neuromuscular responses are also influenced by pharmacologic modulators of sleep-wake state (79). Alcohol, sedative medications and hypnotics may decrease active responses to upper airway occlusion, and contribute to upper airway obstruction during sleep. Benzodiazepines have been demonstrated to prolong obstructive apneas and hypopneas (80). The effects of opiate medications on upper airway collapsibility have not been well studied. Nevertheless, blockade of opioid receptors has been demonstrated to decrease $P_{\text{CRIT}}$, which suggests that narcotic medications may increase susceptibility to pharyngeal occlusion (37).

Current evidence also suggests that endogenous neurohumoral agents can modulate upper airway neuromuscular responses. Neurohormonal modulation of pharyngeal neuromuscular activation may in part account for differences in prevalence and severity of OSA between men and women. When matched by BMI, age, and passive mechanical loads, women demonstrated increased neuromuscular compensation and lower disease burden during NREM sleep compared to men (81). Sex differences in neuromuscular control may well be due elevations in circulating leptin levels in women compared to men (82,83). With weight loss, $P_{\text{CRIT}}$ falls by 6.2 cmH$_2$O per 10 kg/m$^2$
decrease in BMI in apneic men (44), which is greater than the above noted 1.7 cmH2O increase in passive Pcrit per 10 kg/m² increase in BMI attributable to weight gain (58). These observations suggest that obesity may increase upper airway collapsibility through alterations in pharyngeal neuromuscular responses in addition to imposing increased anatomical loads. Investigations have demonstrated elevations in circulating levels of somnogenic inflammatory chemokines, specifically TNF-α, TNF-α receptor I, IL-6, and IL-1β in association with obesity, which may account for the decreases in neuromuscular activity in obesity and OSA patients (84-97).

Pharyngeal neuromuscular activity is also controlled by chemical and mechanical reflexes. Hypercapnia is also a potent stimulator of upper airway neuromuscular activity, which decreases Pcrit (98-100). Hypocapnia, on the other hand, produces a relatively passive state, and is associated with elevations in Pcrit. The upper airway demonstrates decreased collapsibility during expiration compared to inspiration due to phasic activation of pharyngeal muscles (101). Phasic volume feedback, which is mediated by pulmonary stretch receptors, can also inhibit upper airway neuromuscular activity and increase Pcrit (102,103). Pharyngeal sensory afferents can detect intraluminal negative pressure swings during airflow obstruction, and recruit neuromuscular activity (104). Pharyngeal sensory inhibition with topical analgesics has been demonstrated to decrease these neuromuscular responses to upper airway obstruction (105,106). Similarly, mucosal inflammation may blunt local afferents and neuromuscular responses to upper airway obstruction, leading to worsening upper airway obstruction during sleep (107).

Responses in respiratory timing can further augment reflex responses to upper airway obstruction and stabilize ventilation during sleep. The inspiratory duty cycle (IDC), which is the ratio of inspiratory time to total respiratory cycle time, is the most significant determinant of ventilation during periods of inspiratory flow limitation (81,108,109) (see Figure 8). In response to pharyngeal obstruction during sleep, IDC increases nearly immediately with resultant increases in ventilation (108,110,111). In contrast, respiratory rate does not significantly alter minute ventilation in the flow-limited state. Thus, increases in pharyngeal neuromuscular activation and IDC can help maintain ventilation during periods of inspiratory flow limitation and will stabilize respiratory patterns accordingly (see below).

**Modeling the oscillatory patterns in OSA**

Thus far, we have examined factors that promote airway collapse at the onset of obstructive sleep disordered breathing events. To explain the repetitive nature of OSA events, investigators have cast the respiratory system as a closed-loop control system and have elucidated fundamental

---

**Figure 8** Flow profiles during inspiratory flow limitation. Positive flow represents expiration [E], whereas negative flow represents inspiration [I]. (A) During inspiratory flow limitation, minute ventilation (dark shaded area) can be approximated by the product of maximal inspiratory flow (Vim, dashed line) and inspiratory time (tI). Minute ventilation can be increased by increasing inspiratory duty cycle (IDC) (tI/ttot), as demonstrated by the arrow and dotted line; (B) in contrast, increasing the respiratory rate does not increase minute ventilation (the sum of dark shaded areas in B is equal to A) when tI/ttot remains constant. Adapted from Schneider et al. 2009 (108).
Younes and colleagues manipulated loop gain experimentally with proportional assist ventilators, and demonstrated that elevations in loop gain can cause periodic breathing (117). In cross-sectional studies by Wellman and colleagues, loop gain was identified as one of several possible determinants of sleep apnea pathogenesis (118,119). Studies by these investigators have also suggested that pharmacological manipulation of loop gain and arousal threshold can ameliorate OSA (120-122).

Two key assumptions underlie traditional approaches to mathematical models of periodic breathing patterns. The first is that ventilatory responses change in proportion to alterations in ventilatory demand. For instance, hypoventilation, which results in elevations in CO₂, would lead to proportionate increases in ventilatory drive. The second is that the mechanical components of the respiratory system and the control of ventilation remain relatively constant across cycles of periodic breathing. These two principles—linearity and time-invariance (123)—define linear control systems, which can be described by mathematical models that accurately predict smooth sinusoidal oscillations as those observed in Cheyne-Stokes respirations.

Rather than smooth oscillations in airflow, OSA appears to be characterized by abrupt transitions in upper airway patency and ventilation, which suggest a striking departure from the principles of linearity and time-invariance (see Figure 9). Remmers et al. demonstrated that obstructive apneas are characterized by the development of dynamic pharyngeal obstruction during sleep with prompt reopening upon arousal, suggesting marked state-dependence in pharyngeal neuromuscular activity and upper airway collapsibility (17). In fact, upper airway patency is suddenly restored at apnea termination as it shifts from a flow-limited state to a non-flow-limited state. The abrupt termination of these sleep disordered breathing events can be modelled as a switch in an electrical analog of sleep disordered breathing (see Figure 10). When this shift occurs, ventilation can once again track ventilatory drive. Ventilatory drive, in turn, responds to changes in mechanical and chemical afferent inputs, which detect differences between ventilatory supply and demand (left and right sides of Figure 10). At sleep onset, however, the upper airway transitions to a flow limited state. Under these circumstances, ventilation is determined by the degree of upper airway patency rather than ventilatory drive. If ventilatory supply no longer matches demand, ventilatory drive progressively increases (see marker ⊙ in Figure 11). These increases can be associated with alterations in upper airway patency and ventilatory timing that help mitigate the obstruction and/or restore ventilation.
If neuromuscular and ventilatory timing mechanisms do not adequately restore ventilation, ventilatory supply-demand mismatch continues, leading to progressive increases in ventilatory drive. Once drive exceeds a threshold, arousal is triggered (marker l in Figure 11), relieving pent-up ventilatory demand as the airway transitions to a non-flow-limited state (marker m in Figure 11). In fact, repetitive transitions have been demonstrated when ventilation decreases by more than 12-20% of baseline or \( V_{\text{t}} \max \) falls below 250 mL/s (81,124). The aforementioned model suggests that state-dependent alterations in upper airway patency and resulting mismatch between ventilatory supply and demand play pivotal roles in OSA pathogenesis.

In this model, several factors can influence the overall evolution and severity of obstructive breathing episodes. In Figure 11, we plot the cumulative difference between ventilatory supply and demand over time during an obstructive sleep disordered breathing event. In general, the duration of sleep disordered breathing events is governed by the time required for cumulative supply-demand mismatch to reach the arousal threshold. At the onset of sleep disordered breathing episodes, the initial shortfall in ventilation is determined by the severity of upper airway obstruction. As ventilation falls, \( CO_2 \) rises in proportion to the decrease in alveolar ventilation and the metabolic rate (\( CO_2 \) production). Metabolic rate, in turn, is determined by body mass and composition, food sources (e.g., respiratory quotient), basal metabolic rate, sex and work of breathing at rest. Inspiratory flow limitation can further increase the work of breathing and thereby widen ventilatory supply-demand mismatch. Under these circumstances, minute ventilation must increase well above that in the non-flow limit state to satisfy the additional ventilatory demand and stabilize breathing patterns. Supply-demand dynamics can also be affected by ventilatory efficiency. Decreases in ventilatory efficiency due to increased dead space and cardiopulmonary disease will accelerate the development of supply-demand differences. Along with the severity of upper airway obstruction, excess metabolic demand and underlying cardiopulmonary disease can increase the overall severity of sleep disordered breathing. In contrast, sleep disordered breathing events may be prolonged by factors

**Figure 10** Upper airway transition model for obstructive sleep apnea (OSA). In OSA, the upper airway alternates between states of flow limitation during sleep, and non-flow-limited breathing during wakefulness. During sleep, a flow limited upper airway caps ventilatory supply, which dissociates from ventilatory drive (see dashed line). When ventilatory supply does not meet demand, differences (\( \Delta \)) are detected by mechanical and chemical receptors and respiratory control centers, leading to increases in ventilatory drive. Increases in ventilatory drive during periods of flow limitation trigger arousal from sleep with restoration of upper airway patency.

**Figure 11** Ventilatory supply-demand dynamics. 1. At the start of an obstructive event, cumulative supply-demand difference (see \( \Delta \) in Figure 10) is initially determined by the degree of upper airway obstruction (supply) and \( CO_2 \) production (demand); 2. rate of accumulation may be decreased by neuromuscular responses and increase in ventilatory efficiency as \( CO_2 \) rises; 3. when supply-demand difference meets a threshold, arousal occurs and terminates the apnea; 4. pent-up ventilatory demand is relieved due to improved upper airway patency during the arousal.
that slow the development of supply-demand differences, including compensatory neuromuscular mechanisms that mitigate upper airway obstruction and augment ventilation. Sleep disordered breathing events can also be prolonged by raising the arousal threshold with sleep deprivation (125) or hypnotic agents (121,122). Because upper airway neuromuscular control is highly dependent upon sleep wake state, increasing arousal threshold will tend to increase the degree of hypoventilation and oxyhemoglobin desaturation during sleep, as well (79). On the other hand, lowering the arousal threshold may shorten events and increase sleep disruption while minimizing alterations in gas exchange. Thus, the arousal threshold and ventilatory supply-demand dynamics can interact to modulate the overall polysomnographic expression of OSA.

Conclusions

Upper airway obstruction is essential in the pathogenesis of OSA. OSA is largely absent in those individuals without an inherently collapsible upper airway on a structural basis. When $P_{CRIT}$ exceeds $-5$ cmH$_2$O, the risk for OSA markedly increases. The appearance of OSA features parallels the rise in $P_{CRIT}$, increasing from simple snoring, to cyclic hypopneas, and then to fully occlusive apneas. These features are recapitulated in normal persons when upper airway obstruction is induced, and are abolished in OSA patients when airway patency is restored. Therefore, upper airway obstruction alone constitutes both a necessary and sufficient condition for the development of OSA.

Once the airway has collapsed, several factors modify the response to airway obstruction, and affect the ultimate expression of sleep disordered breathing. Neuromuscular responses preserve ventilation and protect against the development of OSA. When neuromuscular compensatory mechanisms are insufficient for a given structural load, ventilatory demand and ventilation dissociate and repeated sleep disordered breathing events ensue. Trade-offs between sleep stability and ventilation can result in a full range of OSA severity and expression. Recurrent arousals and transient increases in airway patency may restore ventilation between periods of sleep, while alterations in neuromuscular responses to upper airway obstruction may improve sleep stability at still suboptimal levels of ventilation.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

71. Isono S. Obesity and obstructive sleep apnoea: mechanisms for increased collapsibility of the passive pharyngeal


90. Punjabi NM, Beamer BA. C-reactive protein is associated with sleep disordered breathing independent of adiposity. Sleep 2007;30:29-34.


