Title page

Mortality-risk–based apnea-hypopnea index thresholds for diagnostics of obstructive sleep apnea

Korkalainen H M.Sc.[1,2], Töyräs J Ph.D.[1,2,3], Nikkonen S M.Sc.[1,2], Leppänen T Ph.D.[1,2]

[1]Department of Applied Physics, University of Eastern Finland, Kuopio, Finland

[2]Department of Clinical Neurophysiology, Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland

[3] School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia

Corresponding author:

Henri Korkalainen, M.Sc.

Department of Clinical Neurophysiology, Diagnostic Imaging Center,

Kuopio University Hospital, P.O Box 100, 70029 Kuopio, Finland

henri.korkalainen@kuh.fi

Tel: +358405344710

Fax: +35817173187

Short title: Mortality-risk–based AHI thresholds

Number of words: 4361

Number of references: 25

Author contributions:

J.T. and T.L. devised the project and the main conceptual ideas for the experiments. H.K., S.N., T.L. and J.T. carried out the data preparation and the experiments. H.K. drafted the manuscript and prepared the figures and tables. All the authors have revised the manuscript critically and approved the version submitted for publication.

Conflicts of interests:

This work was supported by the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (projects 5041767 and 5041768), by the Academy of Finland (Decision number 313697), by the Research Foundation of the Pulmonary Diseases, by Finnish-Norwegian Medical Foundation, by the Respiratory Foundation of Kuopio Region, by the Päivikki and Sakari Sohlberg Foundation, by Orion Research Foundation, and by Innovaatiorahoituskeskus Business Finland (decision number 5133/31/2018). The authors declare that they have no conflicts of interests.

Abstract

The severity of obstructive sleep apnea (OSA) is clinically assessed mainly using the apnea-hypopnea index (AHI). Based on the AHI, patients are classified into four severity groups: non-OSA (AHI<5), mild (5≤AHI<15), moderate (15≤AHI<30) and severe OSA (AHI≥30). However, these thresholds lack solid clinical and scientific evidence. We hypothesize that the current AHI thresholds are not optimal despite their global use and aim to assess this clinical shortcoming by optimizing the thresholds with respect to the risk of all-cause mortality. We analysed ambulatory polygraphic recordings of 1783 patients with suspected OSA (mean follow-up 18.3 years). We simulated 79079 different threshold combinations in 100 randomized subgroups of the population and studied the relative risk of all-cause mortality corresponding to each combination and randomization. The optimal thresholds were chosen according to three criteria: 1) the hazard ratios increase linearly between severity groups towards more severe OSA 2) each group includes at least 15% of the study population, 3) group sizes decrease with increasing OSA severity. The risk of all-cause mortality varied greatly across simulations; the threshold defining non-OSA group having the largest effect on the hazard ratios. The AHI threshold combination of 3-9-24 was optimal in most of the subgroups. In conclusion, the assessment of OSA severity based on the current AHI thresholds is not optimal as several patients with elevated risk of all-cause mortality are diagnosed to have no OSA. Our novel approach provides methods for optimizing AHI-based severity classification and the revised thresholds enhance the prediction of risk of all-cause mortality.

Word count: 250 words

Keywords: Obstructive sleep apnea; survival analysis; all-cause mortality; apnea-hypopnea index; severity classification

Introduction

Obstructive sleep apnea (OSA) is a highly prevalent nocturnal breathing disorder affecting approximately up to 38% of the adult population with the prevalence being higher in men and increasing with age (Senaratna et al. 2017). OSA is characterized by repetitive obstructions of the upper airways, resulting in complete or partial breathing cessations (AASM 1999). These cessations, called apneas and hypopneas respectively, could result in blood oxygen desaturations and arousals from sleep (AASM 1999). Furthermore, OSA is associated with an increased likelihood of severe health consequences such as cardiovascular disease, stroke, and daytime sleepiness (Marin et al. 2005; Somers et al. 2008; Young et al. 2002). Moreover, especially untreated OSA is associated with an elevated risk of all-cause mortality (Marshall et al. 2014; Punjabi et al. 2009; Young et al. 2008).

The most common parameter used to estimate the severity of OSA and the necessity of treatment is the apnea-hypopnea index (AHI), which is simply the number of apnea and hypopnea events per hour of sleep (AASM 1999; Patil et al. 2007). Currently, OSA is diagnosed for patients with an AHI value larger than or equal to 5 events/hour with associated daytime symptoms. Alternatively, an AHI value of ≥15 events/hour is independently sufficient for diagnosis according to the American Academy of Sleep Medicine (AASM) guidelines (AASM 1999). Based on the value of AHI, patients suffering from OSA are classified into four severity groups: non-OSA (AHI<5 events/hour), mild OSA (5≤AHI<15 events/hour), moderate OSA (15≤AHI<30 events/hour), and severe OSA (AHI≥30 events/hour) (AASM 1999).

Despite its widespread use, the AHI suffers from several shortcomings and inconsistencies (Kulkas et al. 2013; Otero et al. 2012; Punjabi 2016; Shahar 2014). First, the thresholds for OSA severity groups are arbitrary and lack strong clinical evidence (Penzel et al. 2015). Second, the definitions of hypopnea events counting towards the numerical value of AHI vary significantly affecting

diagnostics of OSA (Berry et al. 2012; Duce et al. 2015; Iber et al. 2007). Furthermore, the sensitivity of the diagnostic method used (e.g. inclusion of EEG recording) influences the number of obstructive events detected. Thus, the AHI of a patient differs depending on the used diagnostic method and event scoring criteria (Duce et al. 2015; Rapoport 2016). Nevertheless, the AHI thresholds used in the severity assessment of OSA remain constant across different measurement techniques and hypopnea definitions.

The current AHI thresholds (5-15-30 events/hour) used for the severity classification of OSA lack thorough examination and strong clinical evidence; they are simply suggestions that have become the clinical standard over decades (AASM 1999; Patil et al. 2007; Penzel et al. 2015). Despite the common use of the AHI thresholds, no comprehensive studies exist on whether the present classification is optimal for differentiating patients with the highest risk of OSA-related severe health consequences, such as an increased risk of mortality. Still, treatment decisions are based on these possibly outdated AHI thresholds and in some healthcare systems, only patients with moderate or severe OSA receive health insurance- or government-subsidized treatment.

We hypothesize that the current severity classification may not sufficiently reflect the true severity of OSA. To assess this hypothesis, we simulated numerous combinations (n=79079) of severity classification thresholds in a large pool of patients (n=1783) with a long-term follow-up (mean follow-up time of 18.3 years) and studied which threshold combination enables the optimal classification of disease severity with respect to all-cause mortality.

Methods

A follow-up dataset of 1989 suspected OSA patients collected at Kuopio University Hospital was used in this study. Based on clinical practice in Kuopio University Hospital during the years 1993-2003, patients had undergone polygraphic recordings conducted with a custom-made ambulatory

device recording four channels: airflow (thermistor), abdominal respiratory movements, oxygen saturation, and sleeping position (Kulkas et al. 2013; Leppänen et al. 2017). The recordings were manually reanalysed during the years 2012–2015 (Leppänen et al. 2017). In the analyses, hypopneas were scored using a desaturation threshold of 4% (2007 AASM rule 4A (Iber et al. 2007)) Patient information was collected from patients' medical records and the causes of death were obtained from Statistics Finland (Helsinki, Finland). The patients' medical records included information on body mass index (BMI), smoking status, age, gender, continuous positive airway pressure (CPAP) treatment, and comorbidities. Out of the 1989 patients, information on BMI and smoking was missing from 77 and 174 patients respectively, leading to a total of 206 patients being excluded from further analysis. The mean follow-up time for the included 1783 patients was 18.3 years. Ethical permissions were obtained from The Research Ethics Committee of the Hospital District of Northern Savo, Kuopio, Finland (decision numbers 127/2004 and 24/2013).

In addition to the traditional 5, 15 and 30 events/hour thresholds for mild, moderate and severe OSA, respectively, patients were reclassified into OSA severity groups using numerous combinations of simulated thresholds. The simulated thresholds were let to vary from a minimum of 1 event/hour for mild OSA to a maximum of 80 events/hour for severe OSA. The minimum separation between threshold values for OSA severity groups was chosen to be 1 event/hour, and all thresholds with 1 event/hour increments were examined leading to a total of 79079 different threshold combinations.

The relationship between all-cause mortality and OSA severity group was assessed using Cox proportional hazards model adjusted for smoking, BMI, age, gender, CPAP treatment, cardiovascular disease, diabetes, and occurrence of acute myocardial infarction (AMI). For each combination of thresholds, non-OSA patients corresponding to the current iteration were used as the comparison group. Hazard ratios, *p*-values, and confidence intervals were computed for all the

severity groups corresponding to each combination of the thresholds. The statistical analyses were conducted using MATLAB 2018b (The MathWorks Inc., Natick, Massachusetts, United States).

An optimization group was used to optimize the threshold combinations used in severity classification. Afterwards, a separate validation group was used to test the optimized thresholds. The validation was done to avoid overfitting and to better generalize the results. The patient population was split randomly into the two groups, optimization group (*n*=892) and validation group (*n*=891), as follows. First, female patients and patients who were treated with CPAP, had an existing diagnosis of diabetes or cardiovascular disease, or had suffered an acute myocardial infarction were grouped into a single group (*n*=613). These patients were further divided randomly and equally moved into the optimization and validation groups. This was done to increase the likelihood that both the optimization and validation groups contained patients with all the adjusting parameters used in the proportional hazards model. Second, the remaining patients were randomly divided into the optimization and validation groups such that both groups contained half of the whole population. After the group division, all the analyses were conducted. This protocol was then repeated 100 times. Therefore, all the 79079 combinations of thresholds were assessed in 100 different optimization and validation groups leading to the proportional hazards model being formulated a total of 7907900 times.

For every randomization protocol (*n*=100), out of all the 79079 combinations of OSA severity thresholds, a threshold combination was chosen to be optimal if it fulfilled the following criteria:

1) The differences in hazard ratios between mild and moderate and between moderate and severe OSA are similar (with a margin of ±0.02) i.e. the hazard ratios increase linearly towards more severe OSA. Furthermore, this criterion ensures that the diagnosis of severe OSA reflects the highest risk of all-cause mortality.

2) Each OSA severity group includes a minimum of 15% of the whole patient population.
3) The sizes of OSA severity groups decrease towards more severe disease i.e. for the group sizes: non-OSA>mild OSA>moderate OSA>severe OSA.

Results

The study population (*n*=1783) consisted of 1361 male and 422 female patients. The mean follow-up time was 18.3 years, during which 353 patients had died. Out of the population, 766 patients did not smoke, 480 had quit smoking and 537 were smokers. A total of 366 patients were treated with CPAP and 102 had suffered an acute myocardial infarction. Cardiovascular disease was diagnosed for 231 patients and diabetes for 363 patients. The median(range) was 5.7(0.0–148.7) events/hour for the AHI, 28.4(17.5–63.3) kg/m^2 for the BMI and 48.2(18.3–81.1) years for the age of patients.

The hazard ratios related to the risk of all-cause mortality of severity groups varied greatly between different combinations of the thresholds. This effect can be seen in Figure 1, which illustrates the iterations of different combinations of thresholds and the corresponding hazard ratios and group sizes in a single optimization group. The main factor determining the magnitude of hazard ratios for all severity groups was the threshold separating the non-OSA and mild OSA groups. The significance of this threshold can be seen in Figure 1, as all the numerical values of the hazard ratios increase significantly as the threshold of mild OSA decreases from the current clinical threshold of 5 events/hour. However, Figure 1 also illustrates that the hazard ratios decrease after decreasing the thresholds for mild OSA further from 3 events/hour.

Please insert Figure 1 approximately here.

Among the 100 different splits of random optimization and validation sets, the threshold combination of 3-9-24 events/hour met all the optimization criteria most often (37 times).

Therefore, the thresholds of 3, 9 and 24 events/hour for mild, moderate, and severe OSA groups, respectively, were considered optimal. The second most common threshold combinations were 3-9-23, 3-9-25, and 3-9-26 events/hour all of which appeared 34 times among thresholds fulfilling all three optimization criteria. For the remainder of the results, the threshold combination of 3-9-24 events/hour is referred to as the new optimized threshold combination. Figure 2 illustrates the thresholds amongst the 79079 combinations that fulfilled all the three criteria in a single random optimization group.

Please insert Figure 2 approximately here.

Tables 1 and 2 show the medians of the hazard ratios amongst all 100 validation and optimization sets. The median of the hazard ratios increased when using the optimized thresholds. In the validation sets with the current clinical thresholds, the medians of the hazard ratios for mild, moderate and severe OSA were 1.11 (p=0.50), 1.61 (p=0.05), and 1.64 (p=0.06) respectively. In contrast, the medians increased to 1.41 (p=0.15), 1.66 (p=0.05) and 1.82 (p=0.03) with the new thresholds for mild, moderate and severe OSA.

Please insert Table 1, and Table 2 approximately here.

Based on the current clinical diagnostic thresholds (5-15-30 events/hour), 838 patients had no OSA, 469 had mild OSA, 232 had moderate OSA, and 244 patients were diagnosed with severe OSA. With the optimized threshold combination of 3-9-24 events/hour, 630 patients had no OSA, 459 had mild OSA, 377 had moderate OSA, and 317 patients had severe OSA.

Discussion

In this study, we simulated 79079 different combinations of OSA severity classification thresholds in randomized subgroups (*n*=100) of the whole study population and assessed the potential of each combination to differentiate patients having elevated risk of all-cause mortality. The results indicate that the current clinical OSA severity classification (non-OSA, mild OSA, moderate OSA, and severe OSA) using the AHI threshold combination of 5-15-30 events/hour is not optimal with respect to all-cause mortality in our cohort i.e. ambulatory polygraphic recordings of patients with suspected OSA. Based on the results, using the AHI thresholds of 3-9-24 events/hour could improve the severity classification. These thresholds better differentiate patients into OSA severity groups ensuring that increase in OSA severity corresponds to an increase in the risk of all-cause mortality.

The present results suggest that the threshold dividing non-OSA from mild OSA should be lowered. The hazard ratios of all severity groups increased significantly when decreasing this threshold, suggesting that patients with AHI between 3 and 5 events/hour also suffer from an elevated risk of all-cause mortality. The increase in hazard ratios is due to the fact that by lowering the threshold from 5 to 3 events/hour, patients with an elevated risk of all-cause mortality were classified into OSA groups instead of the comparison group (i.e. non-OSA group). This, in turn, decreased the overall risk of all-cause mortality in the comparison group while increasing the relative risk in the mild, moderate, and severe OSA groups. However, decreasing the threshold value to lower than 3 events/hour decreased the hazard ratio for the mild OSA group. This indicates that patients with a low risk of all-cause mortality were classified as having mild OSA when the threshold was lower than 3 events/hour. Therefore, in our cohort, the optimal discrimination of non-OSA patients from OSA patients with respect to all-cause mortality is obtained with the AHI threshold value of 3 events/hour. Furthermore, our results suggest that the thresholds separating mild and moderate

OSA, and moderate and severe OSA, should also be lowered to achieve a more linear increase in the risk of all-cause mortality with increasing OSA severity.

Implementing the optimized OSA severity thresholds could lead to a significant change in OSA diagnostics and treatment decisions. Currently, the severity thresholds are used to guide treatment decisions and to convey information to the patients. The thresholds are needed to achieve rigorous treatment decision guidelines to convince payers of subsidized therapy in healthcare systems and insurance companies. The use of the proposed threshold combination of 3-9-24 events/hour increases the number of patients who fall into the moderate-to-severe OSA category. For example, in our population the number of patients in the moderate-to-severe OSA category increased from 476 (27% of the population) to 694 (39% of the population). This increase would then affect the treatment decisions as only moderate and severe OSA patients receive health insurance- or government-subsidized therapy in some healthcare systems. Thus, the number of patients receiving treatment might increase elevating treatment costs. However, it has been shown that health care costs of patients with OSA, even before diagnosis, are higher than those of matched individuals without OSA (Ronald et al. 1999). These elevated costs generally reduce to the level of the healthy population after treatment of OSA (Albarrak et al. 2005). Thus, lowering the severity thresholds of OSA could reduce total long-term costs for healthcare systems by reducing the prevalence of various OSA related comorbidities while improving the overall quality of life of OSA patients. However, further studies are needed to assess the effect of the lowered diagnostic thresholds on healthcare utilization and costs.

Across threshold simulations, the hazard ratios for all-cause mortality were strongly dependent on the threshold values and experienced large fluctuations with different combinations although we used a large dataset with a long follow-up time. This instability of the hazard ratios for all the severity

groups across simulations suggests that the AHI may not be the sole factor determining the OSA severity. Previous studies have shown that patients with a similar numerical AHI value can have very different severity of individual obstructive events and, therefore, have a completely different phenotype of OSA altogether (Kulkas et al. 2013; Zinchuk et al. 2018). The AHI denotes only the rate of obstructive events per hour of sleep and fails to consider the temporal distribution of events, their duration, or the physiological effects that obstructive events cause. The present results indicate that the evaluation of OSA severity based solely on the AHI might not be the best approach. As OSA is a largely heterogeneous disorder, perhaps an individual assessment of patients' risk factors, comorbidities, and the characteristics of the disorder alongside the AHI-based classification might provide a better way to determine the true severity of OSA. Individual risk-based assessment of patients might enable reliable identification of those patients with an elevated risk of OSA-related severe health consequences and all-cause mortality as well as those who would benefit the most from treatment.

The present study has certain limitations. The analysed ambulatory polygraphic data did not include EEG recording; thus, scoring of arousal-related hypopneas was not possible. The inclusion of EEG recording increases the number of hypopnea events detected and elevates the AHI values. It could thus be argued that the optimal thresholds, when EEG is included, would be higher than the presently proposed thresholds of 3-9-24 events/hour. However, it needs to be noted that when comparing the AHI values between ambulatory measurements and full polysomnography with EEG recording, patients with a similar AHI could have a completely different phenotype of OSA (e.g. different level of hypoxemia or degree of sleep fragmentation). Furthermore, the inclusion of EEG recording would additionally enable more reliable determination of the total sleep time and thus, affect the determined AHI values. Further studies are warranted to determine the optimal

thresholds in full polysomnography including a recording of EEG. Most likely, the optimal thresholds will differ between ambulatory recording and full polysomnography.

Furthermore, all polygraphic recordings were scored using a 4% desaturation threshold for hypopneas (AASM 2007 rule 4A (Iber et al. 2007)) instead of a 3% threshold proposed by AASM in 2012 (Berry et al. 2012). It has been shown that using a 3% desaturation threshold for hypopnea scoring significantly increases the number of scored hypopnea events leading to an increase in the AHI (Duce et al. 2015; Myllymaa et al. 2015). Therefore, it is reasonable to assume that the optimal thresholds, when using the 3% desaturation criterion for hypopnea scoring, would be higher than the presently suggested combination of 3-9-24 events/hour. Furthermore, it has been suggested that the OSA severity thresholds should not remain the same between different measurement techniques and scoring criteria (Hudgel 2016; Rapoport 2016). Therefore, the thresholds suggested in the present study are valid as such only for patients measured and scored with similar techniques and criteria as in the present study population. Further studies with a similar simulation approach are warranted to reliably define the optimal thresholds with different scoring criteria. Additional studies are also warranted to assess the classification with regards to different outcomes alongside the risk of all-cause mortality.

In conclusion, the current threshold combination of 5-15-30 events/hour used for OSA severity classification suffers from major shortcomings in predicting OSA related all-cause mortality. These thresholds should either be optimized for each scoring criteria and measurement technique or the classification based solely on the AHI should be complemented with a more comprehensive patient-specific assessment of the severity of OSA and the related individual risk of comorbidities and mortality.

References

AASM. Sleep–Related Breathing Disorders in Adults: Recommendations for Syndrome Definition and Measurement Techniques in Clinical Research. *Sleep*, 1999.

Albarrak, M., Banno, K., Sabbagh, A. A., et al. Utilization of Healthcare Resources in Obstructive Sleep Apnea Syndrome: a 5-Year Follow-Up Study in Men Using CPAP. Sleep, 2005, 28: 1306-1311.

Berry, R. B., Budhiraja, R., Gottlieb, D. J., et al. Rules for Scoring Respiratory Events in Sleep: Update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. *J. Clin. Sleep Med.*, 2012.

Duce, B., Milosavljevic, J. and Hukins, C. The 2012 AASM Respiratory Event Criteria Increase the Incidence of Hypopneas in an Adult Sleep Center Population. *J. Clin. Sleep Med.*, 2015, 11: 1425-1431.

Hudgel, D. W. Sleep Apnea Severity Classification - Revisited. *Sleep*, 2016, 39: 1165-1166.

Iber, C., Ancoli-Israel, S., Chesson, A. and Quan, S. F. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications*. Westchester, IL: American Academy of Sleep Medicine, 2007 (1st edition).

Kulkas, A., Tiihonen, P., Eskola, K., Julkunen, P., Mervaala, E. and Töyräs, J. Novel parameters for evaluating severity of sleep disordered breathing and for supporting diagnosis of sleep apnea-hypopnea syndrome. *J. Med. Eng. Technol.*, 2013, 37: 135-143.

Kulkas, A., Tiihonen, P., Julkunen, P., Mervaala, E. and Töyräs, J. Novel parameters indicate significant differences in severity of obstructive sleep apnea with patients having similar apnea–hypopnea index. *Med Biol Eng Comput*, 2013, 51: 697-708.

Leppänen, T., Töyräs, J., Mervaala, E., Penzel, T. and Kulkas, A. Severity of individual obstruction events increases with age in patients with obstructive sleep apnea. *Sleep Med.*, 2017, 37: 32-37.

Marin, J. M., Carrizo, S. J., Vicente, E. and Agusti, A. G. Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study. *Lancet*, 2005, 365: 1046-1053.

Marshall, N. S., Wong, K. K. H., Cullen, S. R. J., Knuiman, M. W. and Grunstein, R. R. Sleep apnea and 20-year follow-up for all-cause mortality, stroke, and cancer incidence and mortality in the Busselton Health Study cohort. *J. Clin. Sleep Med.*, 2014, 10: 355.

Myllymaa, S., Myllymaa, K., Kupari, S., et al. Effect of different oxygen desaturation threshold levels on hypopnea scoring and classification of severity of sleep apnea. *Sleep and Breathing*, 2015, 19: 947-954.

Otero, A., Félix, P., Presedo, J. and Zamarrón, C. An Evaluation of Indexes as Support Tools in the Diagnosis of Sleep Apnea. *Ann. Biomed. Eng.*, 2012, 40: 1825-1834.

Patil, S. P., Schneider, H., Schwartz, A. R. and Smith, P. L. Adult obstructive sleep apnea: pathophysiology and diagnosis. *Chest*, 2007, 132: 325.

Penzel, T., Schöbel, C. and Fietze, I. Revise Respiratory Event Criteria or Revise Severity Thresholds for Sleep Apnea Definition?. *J. Clin. Sleep Med.*, 2015, 11: 1357-1359.

Punjabi, N. M. COUNTERPOINT: Is the Apnea-Hypopnea Index the Best Way to Quantify the Severity of Sleep-Disordered Breathing? No. *Chest*, 2016, 149: 16-19.

Punjabi, N. M., Caffo, B. S., Goodwin, J. L., et al. Sleep-disordered breathing and mortality: a prospective cohort study. *PLoS Med.*, 2009, 6: e1000132.

Rapoport, D. M. POINT: Is the Apnea-Hypopnea Index the Best Way to Quantify the Severity of Sleep-Disordered Breathing? Yes. *Chest*, 2016, 149: 14-16.

Ronald, J., Delaive, K., Roos, L., Manfreda, J., Bahammam, A. and Kryger, M. H. Health Care Utilization in the 10 Years Prior to Diagnosis in Obstructive Sleep Apnea Syndrome Patients. Sleep, 1999, 22: 225-229.

Senaratna, C. V., Perret, J. L., Lodge, C., et al. Prevalence of Obstructive Sleep Apnea in the general population: A systematic review. *Sleep Med. Rev.*, 2017, 34: 70-81.

Shahar, E. Apnea-hypopnea index: time to wake up. *Nat. Sci. Sleep*, 2014, 6: 51-56.

Somers, V. K., White, D. P., Amin, R., et al. Sleep apnea and cardiovascular disease: an American Heart Association/American College of Cardiology Foundation Scientific Statement from the American Heart Association Council for High Blood Pressure Research Professional Education Committee, Council on Clinical Cardiology, Stroke Council, and Council on Cardiovascular Nursing. *J. Am. Coll. Cardiol.*, 2008, 52: 686.

Young, T., Finn, L., Peppard, P. E., et al. Sleep disordered breathing and mortality: eighteen-year follow-up of the Wisconsin sleep cohort. *Sleep*, 2008, 31: 1071.

Young, T., Peppard, P. E. and Gottlieb, D. J. Epidemiology of Obstructive Sleep Apnea: A Population Health Perspective. *Am. J. Respir. Crit. Care Med.*, 2002, 165: 1217-1239.

Zinchuk, A. V., Jeon, S., Koo, B. B., et al. Polysomnographic phenotypes and their cardiovascular implications in obstructive sleep apnoea. *Thorax*, 2018, 73: 472.

Figure captions

Figure 1. Illustration of the hazard ratios for all-cause mortality and the number of patients in each OSA severity group corresponding to different combinations of AHI thresholds for diagnosing OSA severity in a single optimization group. Out of the 79079 combinations, only a limited number of combinations (n=700) are shown for clarity. The current clinical thresholds (5-15-30) are indicated with a vertical dotted line and the optimized thresholds (3-9-24) with a vertical dashed line. The optimal region for the threshold of mild OSA between 3 and 5 events/hour is illustrated with a shaded area.

Figure 2. Illustration of the combinations of OSA severity thresholds fulfilling the three criteria for optimal threshold selection in a single optimization group. Corresponding hazard ratios for all-cause mortality, group sizes and statistical significances of hazard ratios are presented for each OSA severity group. The suggested combination of thresholds (3-9-24) is indicated with a vertical dashed line. The optimization was repeated similarly in 100 different randomized optimization groups.

Table 1. Medians of the hazard ratios and *p*-values in the optimization datasets.

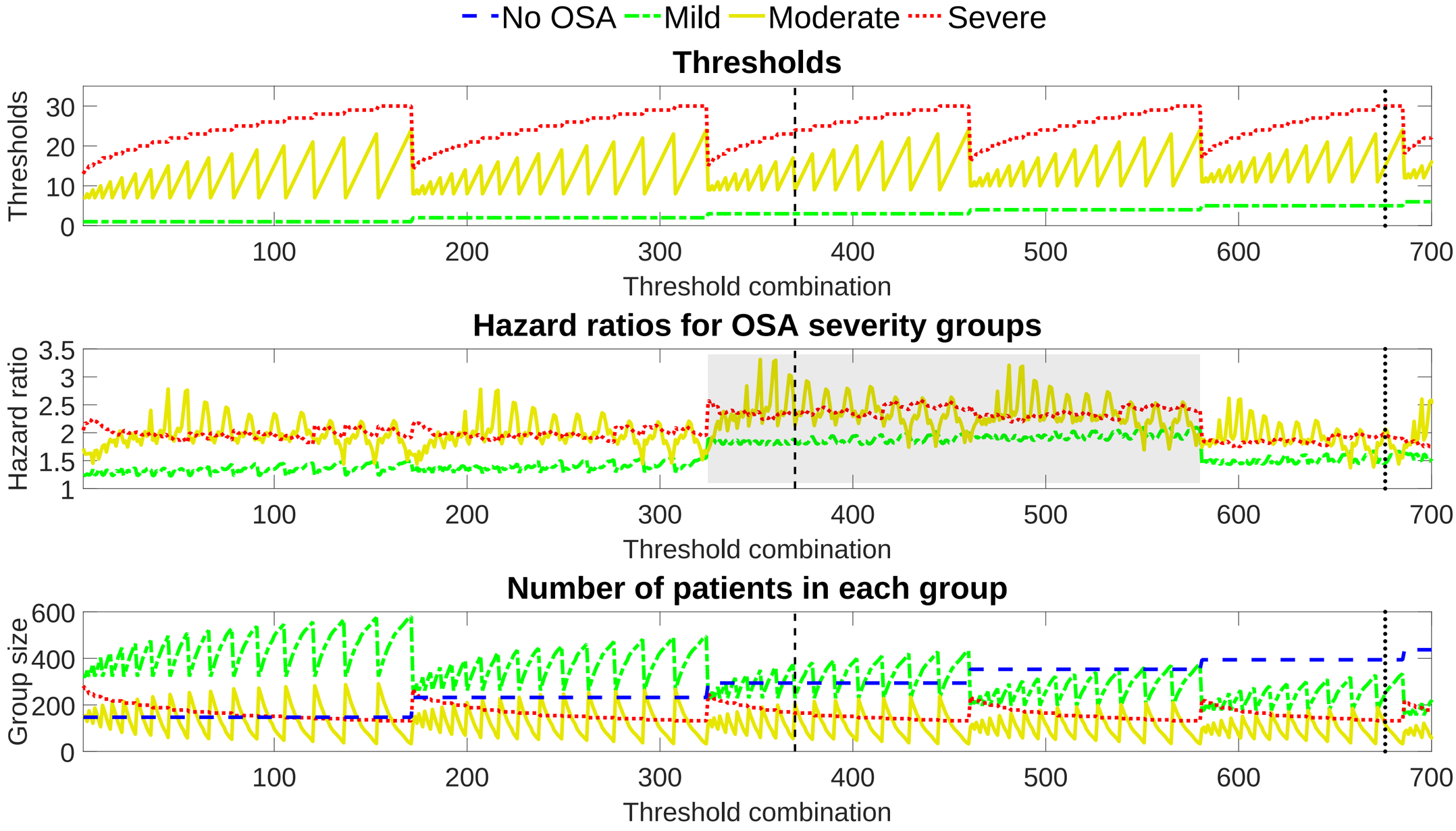
	5-15-30				3-9-24			
	Hazard ratios		*p*-values		Hazard ratios		*p-values*	
	Med	Range	Med	Range	Med	Range	Med	Range
Mild OSA	1.13	0.82–1.79	0.56	0.01–0.99	1.43	1.01–2.34	0.14	<0.01–0.95
Moderate OSA	1.64	1.12–2.44	0.03	<0.01–0.64	1.64	1.16–2.50	0.04	<0.01–0.55
Severe OSA	1.72	1.21–3.01	0.03	<0.01–0.48	1.76	1.31–3.72	0.03	<0.01–0.31
Age (risk/year)	1.08	1.07–1.10	<0.01	<0.01–<0.01	1.08	1.07–1.10	<0.01	<0.01–<0.01
BMI (risk/kgm^{-2})	1.03	1.01–1.06	0.02	<0.01–0.62	1.03	1.01–1.05	0.01	<0.01–0.54
Smoker	2.27	1.71–3.42	<0.01	<0.01–<0.01	2.29	1.71–3.52	<0.01	<0.01–<0.01
Former smoker	1.09	0.84–1.45	0.58	0.05–0.99	1.10	0.84–1.46	0.57	0.05–1.00
Male gender	1.49	0.96–2.34	0.07	<0.01–0.82	1.44	0.93–2.37	0.08	<0.01–0.73
CPAP	0.71	0.54–0.98	0.07	<0.01–0.9	0.75	0.54–1.02	0.12	<0.01–0.95
CVD	1.48	1.10–2.00	0.03	<0.01–0.62	1.49	1.10–2.02	0.03	<0.01–0.64
Diabetes	1.20	0.93–1.61	0.27	<0.01–0.96	1.21	0.92–1.59	0.28	<0.01–0.90
AMI	1.56	1.16–2.41	0.05	<0.01–0.54	1.55	1.12–2.38	0.05	<0.01–0.63

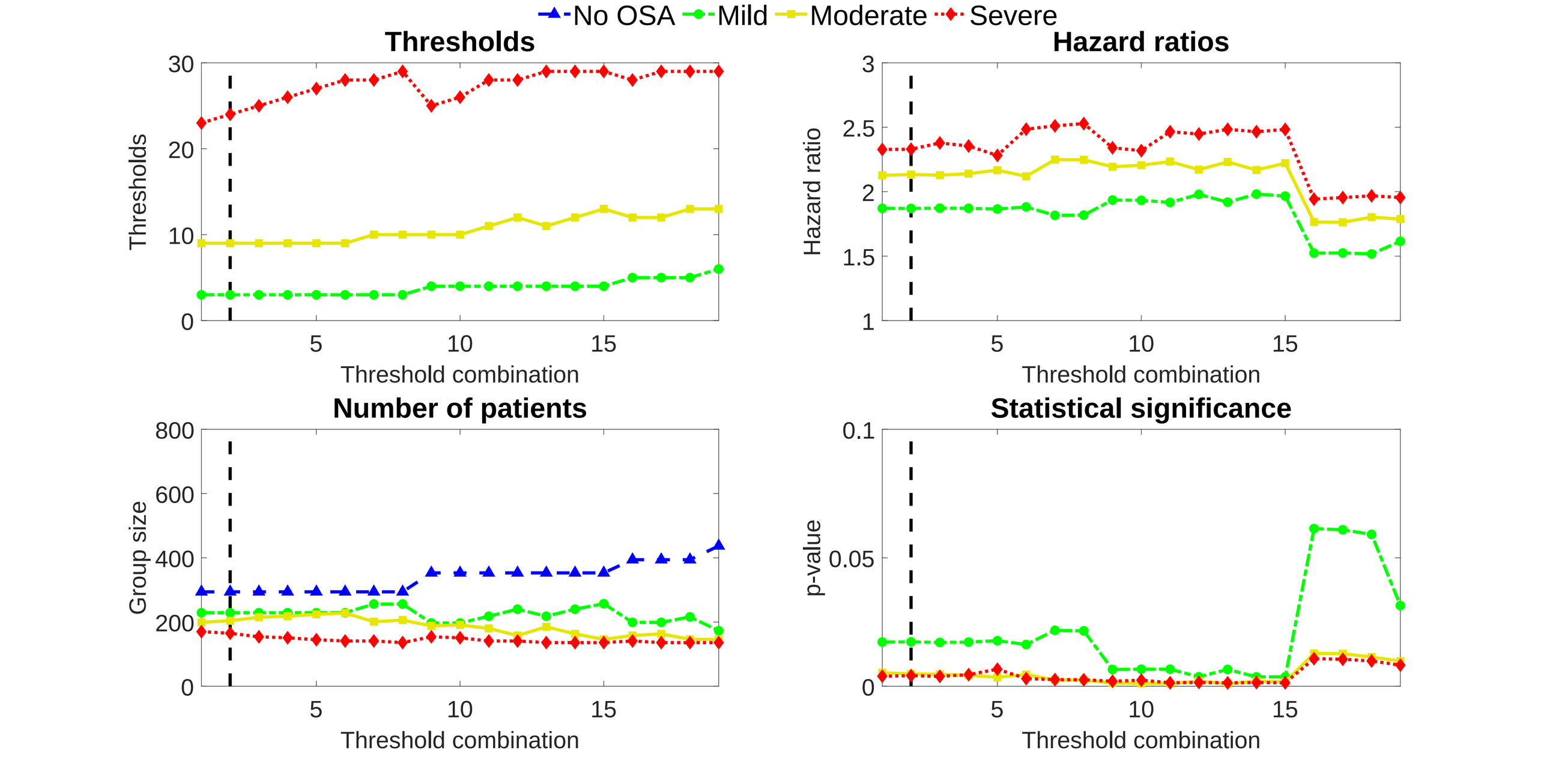
The hazard ratios with the current clinical thresholds (5-15-30) and the new optimized thresholds (3-9-24) were calculated using the Cox proportional hazards model and the medians of hazard ratios and p-values were calculated from all the randomized optimization sets (*n*=100). The proportional hazards model was adjusted for age, body mass index (BMI), smoking status, gender, continuous positive airway pressure (CPAP) treatment, cardiovascular disease (CVD), diabetes, and occurrence of acute myocardial infarction (AMI). Age and BMI were used as continuous variables.

Table 2. Medians of the hazard ratios and p-values in the validation datasets.

	5-15-30				3-9-24			
	Hazard ratios		*p*-values		Hazard ratios		*p-values*	
	Med	Range	Med	Range	Med	Range	Med	Range
Mild OSA	1.11	0.71–1.54	0.50	0.04–0.96	1.41	0.85–2.13	0.15	<0.01–0.88
Moderate OSA	1.61	0.99–2.47	0.05	<0.01–0.97	1.66	1.11–2.40	0.05	<0.01–0.68
Severe OSA	1.64	1.00–2.34	0.06	<0.01–0.99	1.82	0.91–2.62	0.03	<0.01–0.97
Age (risk/year)	1.08	1.07–1.10	<0.01	<0.01–<0.01	1.08	1.06–1.10	<0.01	<0.01–<0.01
BMI (risk/kgm^{-2})	1.03	1.01–1.06	0.02	<0.01–0.56	1.03	1.01–1.06	0.02	<0.01–0.36
Smoker	2.33	1.55–3.20	<0.01	<0.01–0.02	2.34	1.55–3.23	<0.01	<0.01–0.02
Former smoker	1.10	0.80–1.53	0.53	0.04–0.99	1.10	0.81–1.51	0.53	0.05–0.99
Male gender	1.39	0.92–2.29	0.12	<0.01–0.97	1.39	0.89–2.26	0.14	<0.01–0.94
CPAP	0.73	0.51–1.06	0.14	<0.01–0.97	0.78	0.54–1.11	0.21	0.01–0.99
CVD	1.49	0.99–2.05	0.05	<0.01–0.97	1.49	1.00–2.05	0.04	<0.01–0.98
Diabetes	1.22	0.87–1.60	0.28	0.01–0.98	1.19	0.83–1.61	0.34	0.01–0.99
AMI	1.53	0.92–2.58	0.09	<0.01–0.99	1.53	0.89–2.68	0.10	<0.01–0.95

The hazard ratios with the current clinical thresholds (5-15-30) and the new optimized thresholds (3-9-24) were calculated using the Cox proportional hazards model and the medians of hazard ratios and p-values were calculated from all the randomized validation sets (n=100). The proportional hazards model was adjusted for age, body mass index (BMI), smoking status, gender, continuous positive airway pressure (CPAP) treatment, cardiovascular disease (CVD), diabetes, and occurrence of acute myocardial infarction (AMI). Age and BMI were used as continuous variables.


No OSA
Mild
Moderate
Severe
Thresholds
Thresholds
30
20
10
0
100
200
300
400
500
600
700
Threshold combination
Hazard ratios for OSA severity groups
Hazard ratio
3.5
3
2.5
2
1.5
1
100
200
300
400
500
600
700
Threshold combination
Number of patients in each group
Group size
600
400
200
0
100
200
300
400
500
600
700
Threshold combination


No OSA
Mild
Moderate
Severe
Thresholds
Thresholds
Threshold combination
Hazard ratios
Hazard ratio
Threshold combination
Number of patients
Group size
Threshold combination
Statistical significance
p-value
Threshold combination