



# Performance of an Automated Polysomnography Scoring Using Noxturnal Program versus Manual Scoring

Su Jeong Jeong\*, Jong Yuap Seong\*, Jisu Kim, Jong Min Park, Chung Man Sung, Hyung Chae Yang, and Sang Chul Lim

Department of Otolaryngology-Head and Neck Surgery, Chonnam National University Hospital, Chonnam National University Medical School, Gwangju, Korea

레벨2 장비로 전용가능한 레벨1 장비의 수면다원검사 자동화 판독과 검사자 판독 간의 비교

정수정\* · 성종엽\* · 김지수 · 박종민 · 성충만 · 양형채 · 임상철

전남대학교 의과대학 전남대학교병원 이비인후과학교실

Received March 20, 2020

Revised June 16, 2020

Accepted June 22, 2020

Address for correspondence

Sang Chul Lim, MD, PhD  
Department of Otolaryngology-Head and Neck Surgery, Chonnam National University Hospital, Chonnam National University Medical School, 42 Jebong-ro, Dong-gu, Gwangju 61469, Korea  
Tel +82-62-220-6776  
Fax +82-62-228-7743  
E-mail limsc@jnu.ac.kr

\*These authors contributed equally to this work.

**Background and Objectives** A lack of investigators for polysomnography has risen due to increased demand since health insurance started to cover the cost of the test. We examined the reliability of the automated scoring of polysomnography, which has been deployed to improve this imbalance.

**Subjects and Method** We analyzed the data of 20 patients who underwent level 1 polysomnography from April 1 to July 27, 2019. The software from Noxturnal (Nox Medical) was used for the scoring of the Polysomnography data. Each of the polysomnographic data was scored both by the automated scoring system and by a skilled technician.

**Results** Twenty patients were analyzed. There was no significant difference between automated scoring and manual scoring in sleep latency, apnea index, and rapid eye movement sleep stage ratio. However, the concordance rate of the sleep stage by epoch was 83.32%, and there was a significant difference with regards to apnea-hypopnea index (AHI) and respiratory disturbance index (RDI). Two obvious errors were noted in the automated scoring that could be easily fixed; the failure to recognize wakefulness during sleep and the fragmentation of respiratory events. When two errors were corrected, many differences in polysomnography parameters, including AHI and RDI, were eliminated.

**Conclusion** It showed 80% coincidence of epoch in the sleep stage between the automated scoring and manual scoring. However, there was no difference in AHI and RDI when the fragmented respiratory events of the automated scoring were adjusted. Therefore, automated scoring is considered to be useful if only a little modification could be made.

Korean J Otorhinolaryngol-Head Neck Surg 2021;64(3):169-75

**Key Words** Artificial intelligence · Economics · Obstructive sleep apneas · Polysomnographies · Sleep monitoring.

## 서 론

폐쇄성 수면무호흡증은 수면 중 호흡 시에 부분적 또는 완

전한 상기도 기류 폐쇄가 발생하여 저호흡이나 무호흡이 나타나는 질환으로 보통 성인의 5~10%에서 발병하며 나이, 성별, 비만도 등과 밀접한 관련이 있고 특히 뇌졸중, 허혈성 심장 질환, 부정맥, 본태성 또는 약물 저항성 고혈압 등의 환자에서는 그 유병률이 상대적으로 더 높다.<sup>1-3)</sup> 이의 진단을 위해 이학적검사, 간이형수면검사, 약물유발수면내시경 등 다양한

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

방법이 제시되지만,<sup>4,5)</sup> 아직까지는 수면다원검사를 통한 수면 무호흡 지수의 확인이 필요하다.

폐쇄성 수면무호흡증 진단을 위해 필요한 수면다원검사는 2018년 7월 1일부터 고시 제 2018-135로 급여화되었다. 급여화 이후 수면다원검사 검사건수가 급격히 증가하였으며, 이와 더불어 수면검사실 및 검사자 및 판독을 위한 인력 부족을 야기하였다. 이런 현실에서 수면다원검사 자료의 자동화 판독은 검사의 지연 및 판독자 부족 문제를 해결할 수 있는 하나의 대안이 될 수 있다. 하지만 이는 자동화 판독의 신뢰성이 높다는 가정하에서만 가능하며, 최근 기술의 발전과 함께 자동화 판독 기술 역시 많은 진보를 이루었다.

하지만 이런 새로운 기술에 대한 검증된 연구는 많지 않다. 이에 본 연구에서는 최근에 나온 레벨 I 수면검사장비 중, 수면검사의 자동판독이 가능하며, 이와 함께 검사실 외에서도 레벨 II로 검사가 가능한 장비를 선택하였다. 이 장비를 이용해 수면검사를 시행한 후, 검사결과를 자동화 판독과 검사자 판독을 통해 분석하고 이들 간의 차이를 비교해 해당장비의 자동화검사 신뢰성을 검증해 보고자 하였다.

## 대상 및 방법

### 대 상

2019년 4월 1일~7월 27일까지 심한 코골이나 수면호흡장애 증상을 호소하여 Nox AI<sup>®</sup> PSG System(Nox Medical, Reykjavík, Iceland) 수면다원검사기기 및 프로그램을 통해 레벨 I 수면다원검사를 시행한 20명의 데이터를 분석하였다. 본 연구의 장비선택 기준은 레벨 I 수면검사 장비 중, 레벨 II 가정수면다원검사 기기로 사용이 가능하며, 장소 이동이 가능하고, 자동화 판독이 가능한 장비 중 해외 우수 논문에서 인용횟수가 높고, 해외 유명 수면 검사실에서 사용하고 있는 장비를 선택하였다.<sup>6-8)</sup>

환자의 키, 몸무게, 체질량 지수(body mass index), 평균산소포화도, 최저산소포화도, 총수면시간(total sleep time, TST), 총분석시간(total recording time, TRT), 수면 잠복기, rapid eye movement(REM)수면잠복기, 입면 후 각성시간(wake after sleep onset, WASO), 수면효율성, 호흡장애 지수, 뇌파에 따른 수면단계정도, 무호흡-저호흡 지수(apnea-hypopnea index, AHI), 호흡곤란 지수(respiratory disturbance index)를 자동화 판독과 검사자 판독에 따라 데이터를 분석했다. 본 연구는 전남대학교병원 기관생명윤리위원회의 연구 승인(IRB NO. CNUH-2020-045)을 받았다.

### 수면다원검사

미국수면학회 가이드라인에 따른 표준화된 수면다원검사 방법으로 독립된 수면검사실에서 검사를 진행하였다. 뇌파도(electroencephalogram), 안전위도(electroculogram), 심전도, 턱근전도(chin electromyogram), 기류를 포함하였고, 경피적 방법을 이용하여 산소포화도를 측정하였다. 호흡기류는 구강비강열전도센서와 비강압력센서, 이산화탄소 모니터를 이용하여 측정하였는데, 구강비강열전도센서를 이용하여 무호흡을 측정하였고, 비강압력센서를 이용하여 저호흡을 측정하였다. 체위와 다리 근전도를 포함하여 검사를 시행했다.<sup>9)</sup>

판독 역시 미국 수면학회에서 발간한 The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.4에 따라 시행하였다.<sup>10)</sup> 무호흡은 수면 중 10초 이상의 시간 동안 평소 호흡기류의 90% 이상의 감소를 보이는 것으로 정의하고 한시간 동안 무호흡이 일어나는 횟수를 무호흡 지수라 정의하였다. 저호흡은 수면 중 10초 이상의 시간 동안 평소 호흡기류의 30% 이상의 감소를 보이며 혈중 산소포화도가 3% 이상 감소를 동반하거나 각성을 동반하는 경우로 정의하였다. 폐쇄성 무호흡은 전 무호흡시간 동안 호흡노력이 보이는 것을 의미하고, 중추성 무호흡은 전체 무호흡시간 동안 호흡노력이 없는 것이며, 혼합성 무호흡은 처음에는 호흡노력이 없으나 후반부에 호흡노력이 나타나는 경우로 정의하였다.

### 수면다원검사의 판독

각 환자의 데이터는 2가지 방법으로 판독되었다. 먼저 20명의 데이터를 Noxturnal software system 5.1.3(Nox Medical) 프로그램을 사용하여 자동화 판독 결과값을 확인하였고 이후 경험이 많은 단일 판독자가 기존 자동화 판독 자료를 참고하지 않고, 검사자 분석을 하여 각각 결과값을 비교 분석하였다.

### 통계 분석

모든 통계학적 분석은 윈도우용 IBM SPSS 버전 25.0(IBM Corp, Armonk, NY, USA)을 이용하였으며, 총수면시간(TST), 총분석시간(TRT), 수면잠복기, REM수면잠복기, 입면 후 각성시간(WASO), 수면 효율성, 무호흡-저호흡 지수, 뇌파 단계를 대응표본과 Wilcoxon 부호순위검정을 이용하여 비교 분석하였다. 양측검정에서  $p$ 값이 0.05 미만을 보이는 변수에 대해 통계적으로 유의미한 것으로 판단하였다.

## 결 과

### Baseline characteristics

총 대상인원은 20명이었으며, 남성이 15명, 여성이 5명이었 다. 중간연령은 49세였다. 키의 중위수는 170.0 cm, 몸무게의 중위수 66.0 kg였으며, 체질량지수의 중위수는 24.9 kg/cm<sup>2</sup> 였다. 평균산소포화도의 중위수는 94.1%, 최소산소포화도의 중위수는 84.5%였다(Table 1).

### Comparison of Polysomnographic parameters between groups

총 수면시간(TST)의 경우 자동화 판독은 340.8분, 검사자 판독은 299.3분( $p=0.001$ )였다. 수면 잠복기의 경우 자동화 판독은 17.9분, 검사자 판독은 18.0분( $p=0.619$ )이었으며, REM 수면잠복기의 경우 자동화 판독은 75.8분, 검사자 판독은 139.0분( $p=0.001$ ) 소견을 보였다. 입면 후 각성시간(WASO)의 경우 자동화 판독은 67.3분, 검사자 판독은 104.3분( $p=0.001$ ), 수면 효율성의 경우 자동화 판독은 79.4%, 검사자 판독은

68.4%( $p=0.001$ )였다(Table 2).

### Sleep stage

뇌파의 경우 자동화 판독의 경우 N1은 3.2%, N2는 54.8%, N3는 19.0%였으며 검사자 판독의 경우 N1은 21.0%, N2는 48.1%, N3는 12.0%였고 non-REM수면의 모든 단계는 통계 적으로 유의미한 차이를 보였다. 한편 REM수면의 경우 자동 화 판독의 경우 17.6% 검사자 판독의 경우 17.2%로 통계학적 으로 유의미한 의미가 없었다(Table 2).

### Respiratory event

무호흡-저호흡 지수(AHI)의 경우 자동화 판독은 17.9회/ 시간, 검사자 판독은 13.6( $p<0.001$ ), 이중 무호흡 지수의 경 우 자동화 판독은 3.3회/시간 검사자 판독은 3.4회/시간( $p=0.164$ ), 저호흡 지수의 경우 자동화 판독은 12.9회/시간 검사 자 판독은 8.9회/시간( $p=0.001$ ) 소견을 보여 저호흡에서 차이 가 있었다. 호흡곤란 지수(respiratory disturbance index, RDI) 는 자동화 판독은 17.9회/시간, 검사자 판독에서는 14.0회/시간 으로 차이가 있었다( $p<0.001$ ).

자동화 판독의 중요한 문제 중 하나는 호흡이벤트의 분절 이었다(Fig. 1). 분절된 호흡이벤트의 분절만을 수정하는 것은 20분 이내의 시간만 소요되었으며, 해당 분절된 호흡이벤트 만 수정한 후 자동판독을 재시행한 결과, 수기판독과 차이가 10% 수준으로 줄어들었고, 이는 양군간 AHI 및 RDI 값이 차이가 없는 수준임을 확인하였다( $p=0.170$ ,  $p=0.140$ )(Table 3).

### Sleep stage in particular patient

두 명의 특정환자를 무작위로 선정해 각 판독단위(epoch)

**Table 1.** Baseline characteristics of the study population

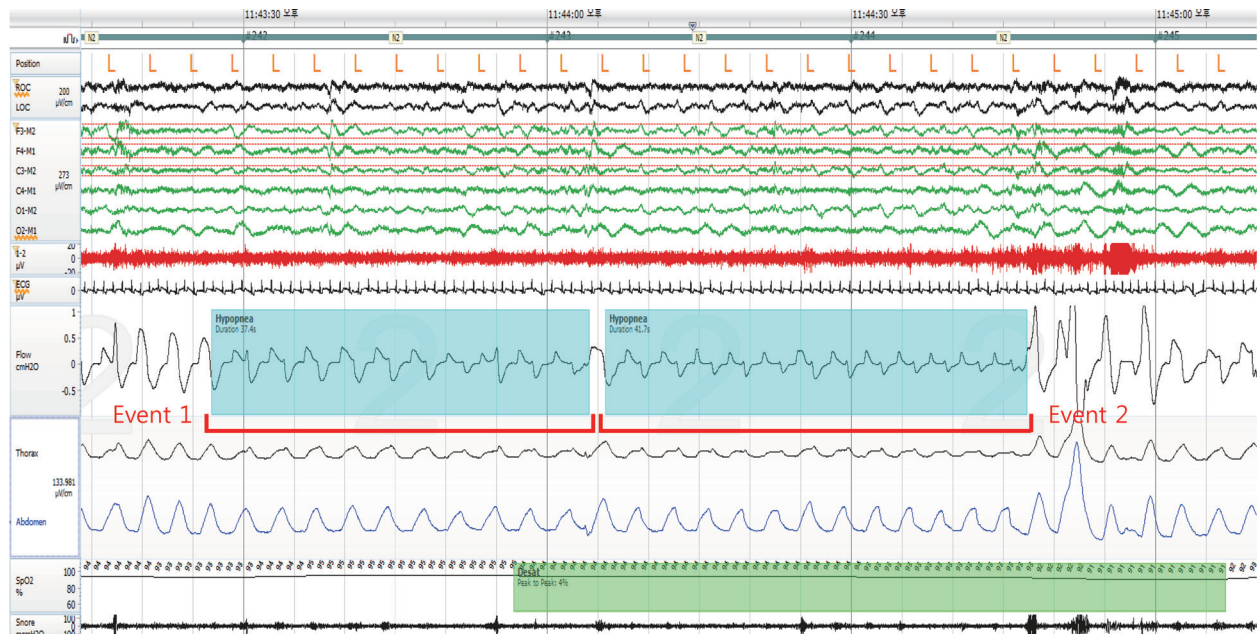
Category	Results
Number of patients	20
Gender (male:female)	15:5
Age (years)	49 (39–58)
Height (cm)	170.0 (158.3–171.8)
Weight (kg)	66.0 (61.0–84.8)
BMI (m <sup>2</sup> /kg)	24.9 (22.3–28.1)
Average SpO <sub>2</sub> (%)	94.1 (93.3–94.7)
Minimum SpO <sub>2</sub> (%)	84.5 (81.0–87.0)

BMI: body mass index, SpO<sub>2</sub>: denotes saturation of oxygen

**Table 2.** Comparison between automated scoring and manual scoring according to each category

Category	Auto scoring (n=20)	Manual scoring (n=20)	p-value*
TST (min)	340.8 (298.0–393.9)	299.3 (235.4–379.1)	0.001
Sleep latency (min)	17.9 (5.3–33.9)	18.0 (9.4–31.3)	0.619
REM latency (min)	75.8 (5.4–109.3)	139.0 (85.9–219.3)	0.001
WASO (min)	67.3 (34.5–104.4)	104.3 (35.8–151.2)	0.001
Sleep efficiency (%)	79.4 (67.5–85.4)	68.4 (54.1–84.4)	0.001
Sleep stage N1	3.2 (2.2–4.8)	21.0 (11.0–37.0)	<0.001
Sleep stage N2	54.8 (49.0–67.8)	48.1 (41.5–53.2)	0.015
Sleep stage N3	19.0 (12.9–27.8)	12.0 (6.1–21.3)	0.000
Sleep stage REM	17.6 (11.5–23.5)	17.2 (10.1–21.1)	0.249
AI (events/hours)	3.3 (1.2–12.6)	3.4 (1.1–8.1)	0.164
HI (events/hours)	12.9 (5.5–21.2)	8.9 (3.6–16.4)	0.001
AHI (events/hours)	17.9 (7.9–34.1)	13.6 (5.6–26.9)	<0.001
RDI (events/hours)	17.9 (7.9–34.1)	14.0 (5.6–26.9)	<0.001

\*Wilcoxon signed-rank test was used to compare values between autoscoring and manual scoring. TST: total sleep time, WASO: wake after sleep onset, AI: apnea index (events/hours), HI: hypopnea index (events/hours), AHI: apnea-hypopnea index (events/hours), RDI: respiratory disturbance index



**Fig. 1.** Examples of fragmented respiratory events. A discrepancy was originated from fragmented respiratory events. Note that the respiratory event that was counted as one respiratory event (event 1+event 2) in manual scoring was counted as two respiratory events (event 1 & event 2) in automated scoring.

**Table 3.** The results of the correction of the fragmented respiratory events and missed wake epochs

Category	Manual scoring (n=20)	Corrected scoring (n=20)	p-value*
TST (min)	299.3 (235.4–379.1)	282.7 (222.8–368.5)	0.024
Sleep latency (min)	18.0 (9.4–31.3)	20.3 (11.6–31.1)	0.807
REM latency (min)	139.0 (85.9–219.3)	92.7 (57.0–146.8)	0.009
WASO (min)	104.3 (35.8–151.2)	116.0 (45.4–166.2)	0.013
Sleep efficiency (%)	68.4 (54.1–84.4)	68.1 (50.4–84.7)	0.099
Sleep stage N1	21.0 (11.0–37.0)	2.3 (1.6–3.2)	<0.001
Sleep stage N2	48.1 (41.5–53.2)	56.4 (51.3–68.0)	0.005
Sleep stage N3	12.0 (6.1–21.3)	20.2 (14.9–28.9)	<0.001
Sleep stage REM	17.2 (10.1–21.1)	15.9 (12.5–22.3)	0.749
AI (events/hours)	3.4 (1.1–8.1)	3.2 (0.8–9.9)	1.000
HI (events/hours)	8.9 (3.6–16.4)	10.7 (4.4–19.3)	0.026
AHI (events/hours)	13.6 (5.6–26.9)	15.5 (5.8–29.0)	0.170
RDI (events/hours)	14.0 (5.6–26.9)	15.5 (5.8–29.0)	0.140

\*Wilcoxon signed-rank test was used to compare values between autoscoring and manual scoring. TST: total sleep time, WASO: wake after sleep onset, AI: apnea index (events/hours), HI: hypopnea index (events/hours), AHI: apnea-hypopnea index (events/hours)

의 자동화 판독 결과와 검사자 판독 결과를 일대일로 비교하여, 자동화 판독과 검사자 판독이 어디에서 차이를 보이는지 비교해 보았다(Table 4). 검사자 판독을 기준으로 하여 자동화 판독의 일치율 정도를 비교할 때, 첫 번째 환자는 전체적으로 83.6%, 두 번째 환자는 전체적으로 82.8%의 일치율을 보였다. 수면단계별로 비교해 보면, 각성시는 74.3%와 78.3%의 일치율을 보였고, N1 수면단계에서 13.7%와 5.6%로 낮은 일치율을 보였으며, N2 수면단계와 N3 수면단계는 각각 97.3%, 83.8% 그리고 90.2%와 99.0%의 일치율을 보였다.

REM수면단계는 95.2%와 81.4%의 일치율을 보였다.

## 고 찰

수면다원검사는 폐쇄성 수면무호흡증의 확진과 중증도 확인을 위해 필수적이며,<sup>11)</sup> 수면다원검사의 판독은 검사자에 의한 수기 판독이 표준화된 방법이다. 하지만 수기 판독은 시간이 많이 걸리고, 인력 비용이 발생하며, 해석에 주관성이 개입될 수 있다는 문제점도 있다.<sup>12)</sup> 이와 함께 최근 국내실정은



**Table 4.** Comparison between automated scoring and manual scoring according to each epoch of patients

Number of epochs	A <sub>auto</sub>	N1 <sub>auto</sub>	N2 <sub>auto</sub>	N3 <sub>auto</sub>	R <sub>auto</sub>	Number of matched	Number of epochs	Concordance rate (%)
Number of epoch (P1+P2)	204	32	713	405	370	1435	1724	83.32
P1								
A <sub>Manual</sub>	78	1	17	3	6	78	105	74.3
N1 <sub>Manual</sub>	10	13	31	2	39	13	95	13.7
N2 <sub>Manual</sub>	0	4	320	4	1	320	329	97.3
N3 <sub>Manual</sub>	0	0	18	166	0	166	184	90.2
R <sub>Manual</sub>	1	0	7	0	158	158	166	95.2
P2								
A <sub>Manual</sub>	101	2	4	0	22	101	129	78.3
N1 <sub>Manual</sub>	7	2	16	1	10	2	36	5.6
N2 <sub>Manual</sub>	6	3	279	29	16	279	333	83.8
N3 <sub>Manual</sub>	1	0	1	200	0	200	202	99.0
R <sub>Manual</sub>	0	7	20	0	118	118	145	81.4

P1: Patient #1, P2: Patient #2, A<sub>auto</sub>: awakening analyzed by auto scoring, N1<sub>auto</sub>: sleep stage N1 analyzed by auto scoring, N2<sub>auto</sub>: sleep stage N2 analyzed by auto scoring, N3<sub>auto</sub>: sleep stage N3 analyzed by auto scoring, R<sub>auto</sub>: sleep stage REM analyzed by auto scoring, A<sub>Manual</sub>: awakening analyzed by manual scoring, N1<sub>Manual</sub>: sleep stage N1 analyzed by manual scoring, N2<sub>Manual</sub>: sleep stage N2 analyzed by manual scoring, N3<sub>Manual</sub>: sleep stage N3 analyzed by manual scoring, R<sub>Manual</sub>: sleep stage REM analyzed by manual scoring, REM: rapid eye movement

수면다원검사의 급여화와 함께 수면다원검사 시행건수가 급격하게 증가하고 있으며, 이에 따른 검사자 및 검사실 부족의 문제가 심각한 현실이다. 이에 본 연구에서는 레벨 1 수면다원검사 장비중 레벨 II로 전용이 가능하면서 자동화 판독이 가능한 Nox A1장비를 선택하여, 검사실과 검사실인력을 줄일 수 있는지 확인해 보았다. 기존에도 유사한 연구가 있긴 하나, 실제 개발회사가 참여하지 않은 Nox A1(Nox Medical)과 Noxturnal 5.1.3 소프트웨어로 시행한 자동화 판독 결과의 검증은 본 연구가 처음인 것으로 판단된다.

논란은 있지만, 문헌에 따르면 자동화 판독과 검사자 판독의 일치율은 약 60~90% 정도로 알려져 있다.<sup>13-15)</sup> Oztürk 등<sup>16)</sup>은 일치율이 58%라고 보고하며 자동화 판독의 신뢰성에 의문을 제기하였다. Barreiro 등<sup>17)</sup>은 Somnostar α 4100를 이용하여 자동화 판독과 검사자 판독을 비교하였다. 그들 논문에서 Type II 수면검사장비를 통해 28명의 환자를 대상으로 자동화 판독과 검사자 판독을 비교한 결과, 호흡사건과 관련된 결과는 일치율이 높은 편이었고, 특히 무호흡과 무호흡-저호흡 지수가 높은 일치율을 보인다고 보고하였다. 본 저자들의 연구결과도 이와 유사하며, 무호흡-저호흡 지수는 비교적 일치하였고, 무호흡보다는 저호흡이 일치율이 높았다.

수면단계에 관련하여, 기존 논문에서는 non-REM수면 1, 2단계는 높은 일치율을 보이고, non-REM수면 3단계와 REM수면은 비교적 일치하지 않는 것으로 보고하였으나 본 저자들의 연구결과는 non-REM수면 1, 2, 3단계 모두 수기판독과 통계적으로 유의미한 차이를 보였고, REM수면은 차이를 보이지 않는 것으로 확인되었다. 이는 두 판독 프로그램의 판

독 알고리즘에 따른 차이인 것으로 사료된다. 또 다른 문헌에서는 자동화 판독은 점점 판독 정확도가 증가하고 있으나, 총 수면시간과 2단계 수면단계를 과소평가한다고 보고하며, 이는 자동화 판독이 sleep spindle과 k-complex를 잘 찾아내지 못한 것에서 그 원인을 찾고 있다.<sup>18)</sup> 이로 인해 1단계 수면단계는 과잉 평가되고, 3단계 수면단계와 REM수면은 큰 영향이 없다고 보고하였다. 그리고 또 다른 논문에서는 자동화 분석의 경우 일반적으로 무호흡-저호흡 지수를 과소 평가하는 경향이 있고, 호흡상태와 수면단계에서 임계값 설정을 어떻게 하느냐에 따라 검사자 판독과 비교해서 민감도와 특이도가 낮다고 보고하였다.<sup>19)</sup> 본 연구에서는 자동화 판독이 TST, N1, N2, N3가 전반적으로 다 과잉평가하였는데, 이는 Noxturnal 소프트웨어가 수면중간에 발생한 각성을 못잡아내는 오류가 찾아 나타난 결과로 보인다.

본 연구에 사용한 NOX A1 수면다원검사키기와 Noxturnal 소프트웨어를 사용하여 진행된 기존 문헌을 찾을 수 없으나, 같은 회사에서 유사한 알고리즘을 사용할 것으로 판단되는 T3 장비를 이용한 결과가 2015년에 보고되었다. 2015년 Nox T3를 이용하여 120명의 중증 폐쇄성수면무호흡 환자들을 자동화 판독과 검사자 판독을 시행한 연구결과가 있었고, 전체 평균 일치율은 91.7%로 알려져있다. 무호흡-저호흡 지수의 경우 자동화 판독은 67.3±24.3, 검사자 판독은 62.7±25.1로 91.3%의 일치율을 보였고, 무호흡 지수의 경우 자동화 판독은 36.9±25.9, 검사자 판독은 39.0±24.5으로 95.4%의 일치율을 보였으며, 저호흡 지수의 경우 자동화 판독은 25.9±17.3, 검사자 판독은 19.4±12.8로 84%의 일치율을 보였는데 무호

흡-저호흡 지수에 비해 상대적으로 낮은 일치율을 보였다.<sup>20)</sup> 또한 Younes 등<sup>21)</sup>의 연구결과는 숙련된 판독자 간의 수면다원검사 판독 일치율이 78.9%로 보고하였으며, 이런 사항들을 고려하면, Noxturnal 소프트웨어의 자동화 판독의 수준은 신뢰성이 있는 것으로 생각된다. 또한 본 연구에서 확인된 Noxturnal 소프트웨어의 문제인 수면 중 각성을 못 잡아내는 문제와, 한 개의 호흡사건을 여러 개로 분절시키는 문제를 해결하면(Table 3, Fig. 1), 적은 노력으로 통계학적으로 차이가 없는 수준의 자동화 판독결과를 보이므로, 자동화 판독은 환자 진료와 검사자 부족문제 해결에 도움이 될 것으로 생각된다. 하지만, 수면단계에 따른 변화가 더욱 중요한 요인으로 작용하는 수면무호흡외 다른 질환의 경우 자동화 판독시 주의 기울이는 것이 필요해 보인다.

본 연구의 한계점은 연구에 등록된 인원이 20명으로 각군 20개씩의 판독을 바탕으로 연구를 진행하였다는 것이다. 따라서 향후 보다 대규모 스터디가 필요해 보인다. 그럼에도 불구하고, 본 연구에서는 Nox 자동화 판독의 주된 문제점인 수면 중 각성을 수면지속으로 판독하는 오류 및 단일 호흡사건을 여러 개의 사건으로 분절한다는 문제점을 파악하였으며, 적은 노력을 들여 이들 일부 문제의 보완을 시행함으로 판독의 신뢰성을 높일 수 있는 방법을 제시하였다는 데에 본 연구의 의의가 있겠다.

본 연구에서 수면다원검사 자동화 판독은 검사자가 시행한 판독과 수면단계 판독에서 80% 이상의 일치율을 보였고, 특히 일부 간단한 호흡이벤트의 수정을 통해 통계적으로 문제 없는 수준의 신뢰성을 확보할 수 있었다. 이를 통해 자동화 판독의 가능성을 확인할 수 있었다. 수면다원검사의 급여화로 검사시행 건수가 늘어남에 따라 재정적 부담, 수면검사실 및 인력의 부족이 나타나는 현재 상황에서 레벨 2로의 전용과 자동화 판독이 가능한 수면다원검사기기는 이들 문제를 해결할 수 있는 실마리가 될 수 있을 것이다.

## Acknowledgments

Special thanks to all people, institutions and companies for supporting these research progresses.

Especially, I would like to express my gratitude and appreciation to CNUH Open Laboratory personnel, Dr. Jiyoung Choi and Juha Yong, and cooperative companies, Samdo Environment Co., Ltd. and S.I. Solutions Co., Ltd.

## Author Contribution

Conceptualization: Sang Chul Lim. Data curation: Hyung Chae Yang. Formal analysis: Hyung Chae Yang. Funding acquisition: Hyung Chae Yang. Investigation: Jong Yuap Seong. Methodology: Jong Min Park. Project administration: Hyung Chae Yang. Resources: Su Jeong Jeong. Software: Jisu Kim. Supervision: Sang Chul Lim. Validation: Chung Man Sung. Visualization: Su Jeong Jeong. Writing—original draft: Jong Yuap Seong. Writing—review & ed-

iting: Su Jeong Jeong.

## ORCID

Sang-Chul Lim <https://orcid.org/0000-0001-5610-1554>

## REFERENCES

- 1) Lee YC, Eun YG, Shin SY, Kim SW. Prevalence of snoring and high risk of obstructive sleep apnea syndrome in young male soldiers in Korea. *J Korean Med Sci* 2013;28(9):1373-7.
- 2) Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. *N Engl J Med* 2000;342(19):1378-84.
- 3) Canessa N, Castronovo V, Cappa SF, Aloia MS, Marelli S, Falini A, et al. Obstructive sleep apnea: Brain structural changes and neurocognitive function before and after treatment. *Am J Respir Crit Care Med* 2011;183(10):1419-26.
- 4) Yang HC, Jung EK, Yoon SH, Cho HH. The efficacy of drug induced sleep endoscopy using multimodality monitoring system. *PLoS One* 2018;13(12):e0209775.
- 5) Sung CM, Kim HC, Yang HC. The clinical characteristics of patients with an isolate epiglottic collapse. *Auris Nasus Larynx* 2020;47(3):450-7.
- 6) Yoon DW, Hong IH, Baik I, Shin HW. Evaluation of the feasibility and preference of Nox-A1 type 2 ambulatory device for unattended home sleep test: A randomized crossover study. *Sleep Biol Rhythms* 2019;17(3):297-304.
- 7) Xu L, Han F, Keenan BT, Kneeland-Szanto E, Yan H, Dong X, et al. Validation of the Nox-T3 portable monitor for diagnosis of obstructive sleep apnea in Chinese adults. *J Clin Sleep Med* 2017;13(5):675-83.
- 8) Cairns A, Wickwire E, Schaefer E, Nyanjom D. A pilot validation study for the NOX T3(TM) portable monitor for the detection of OSA. *Sleep Breath* 2014;18(3):609-14.
- 9) Epstein LJ, Kristo D, Strollo PJ Jr, Friedman N, Malhotra A, Patil SP, et al. Clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. *J Clin Sleep Med* 2009;5(3):263-76.
- 10) Berry RB, Brooks R, Gamaldo C, Harding SM, Lloyd RM, Quan SF, et al. AASM scoring manual updates for 2017 (version 2.4). *J Clin Sleep Med* 2017;13(5):665-6.
- 11) Kushida CA, Littner MR, Morgenthaler T, Alessi CA, Bailey D, Coleman J Jr, et al. Practice parameters for the indications for polysomnography and related procedures: An update for 2005. *Sleep* 2005;28(4):499-521.
- 12) Collop NA. Scoring variability between polysomnography technologists in different sleep laboratories. *Sleep Med* 2002;3(1):43-7.
- 13) Schaltenbrand N, Lengelle R, Toussaint M, Luthringer R, Carelli G, Jacquin A, et al. Sleep stage scoring using the neural network model: Comparison between visual and automatic analysis in normal subjects and patients. *Sleep* 1996;19(1):26-35.
- 14) Malhotra A, Younes M, Kuna ST, Benca R, Kushida CA, Walsh J, et al. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep* 2013;36(4):573-82.
- 15) Redline S, Budhiraja R, Kapur V, Marcus CL, Mateika JH, Mehra R, et al. The scoring of respiratory events in sleep: Reliability and validity. *J Clin Sleep Med* 2007;3(2):169-200.
- 16) Oztürk O, Mutlu LC, Sağcan G, Deniz Y, Cuhadaroglu C. [The concordance of manual (visual) scoring and automatic analysis in sleep staging]. *Tuberk Toraks* 2009;57(3):306-13.
- 17) Barreiro B, Badosa G, Quintana S, Esteban L, Heredia JL. [Comparison between automatic and manual analysis in the diagnosis of obstructive sleep apnea-hypopnea syndrome]. *Arch Bronconeumol* 2003;39(12):544-8.
- 18) Sforza E, Vandi S. Automatic Oxford-Medilog 9200 sleep staging

- scoring: Comparison with visual analysis. *J Clin Neurophysiol* 1996;13(3):227-33.
- 19) Carrasco O, Montserrat JM, Lloberes P, Ascascio C, Ballester E, Fornas C, et al. Visual and different automatic scoring profiles of respiratory variables in the diagnosis of sleep apnoea-hypopnoea syndrome. *Eur Respir J* 1996;9(1):125-30.
  - 20) Cachada N, Thomas M, Wharton S. Comparison of manual and automatic scoring of limited channel sleep studies: Nocturnal software correlates well with manual scoring in severe OSA. *Eur Respir J* 2017;50:PA2301.
  - 21) Younes M, Raneri J, Hanly P. Staging sleep in polysomnograms: Analysis of inter-scorer variability. *J Clin Sleep Med* 2016;12(6): 885-94.