Big data analysis of cyclic alternating pattern during sleep using deep learning

by

Simon Hartmann

M.Sc. (Electrical and Electronic Engineering),
Karlsruhe Institute of Technology, Germany, 2017

B.Sc. (Electrical and Electronic Engineering),
University of Ulm, Germany, 2014

Thesis submitted for the degree of

Doctor of Philosophy

in

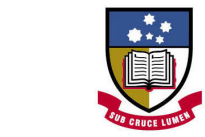
School of Electrical and Electronic Engineering
The University of Adelaide, Australia

December, 2021

Supervisors:

Associate Professor Mathias Baumert, School of Electrical and Electronic Engineering

Professor Derek Abbott, School of Electrical and Electronic Engineering



THE UNIVERSITY
of ADELAIDE

Abstract

Sleep scoring has been of great interest since the invention of the polysomnography method, which enabled the recording of physiological signals overnight. With the surge in wearable devices in recent years, the topic of what is high-quality sleep, how can it be determined and how can it be achieved attracted increasing interest. In the last two decades, cyclic alternating pattern (CAP) was introduced as a scoring alternative to traditional sleep staging. CAP is known as a synonym for sleep microstructure and describes sleep instability. Manual CAP scoring performed by sleep experts is a very exhausting and time-consuming task. Hence, an automatic method would facilitate the processing of sleep data and provide a valuable tool to enhance the understanding of the role of CAP.

This thesis aims to expand the knowledge about CAP by developing a high-performance automated CAP scoring system that can reliably detect and classify CAP events in sleep recordings. The automated system is equipped with state-of-the-art signal processing methods and exploits the dynamic, temporal information in brain activity using deep learning. The automated scoring system is validated using large community-based cohort studies and comparing the output to verified values in the literature. Our findings present novel clinical results on the relationship between CAP and age, gender, subjective sleep quality, and sleep disorders demonstrating that automated CAP analysis of large population based studies can lead to new findings on CAP and its subcomponents. Next, we study the relationship between CAP and behavioural, cognitive, and quality-of-life measures and the effect of adenotonsillectomy on CAP in children with obstructive sleep apnoea as the link between CAP and cognitive functioning in children is largely unknown. Finally, we investigate cortical-cardiovascular interactions during CAP to gain novel insights into the causal relationships between cortical and cardiovascular activity that are underpinning the microstructure of sleep.

In summary, the research outcomes in this thesis outline the importance of a fully automated end-to-end CAP scoring solution for future studies on sleep microstructure. Furthermore, we present novel critical information for a better understanding of CAP and obtain first evidence on physiological network dynamics between the central nervous system and the cardiovascular system during CAP.

Statement of Originality

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

The author acknowledges that copyright of published works contained within the thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed

09/11/2021

Date

Acknowledgments

I would like to express my deepest appreciation to my supervisor Associate Professor Mathias Baumert for his guidance, his patience, and his nurturing supervision. I am extremely grateful that Mathias relentlessly tried to create spare time in his busy schedule to listen to my ideas and that he was extremely supportive about exploring new paths. I cannot express how much I enjoyed working alongside Mathias and how much I appreciate that he shared his tremendous knowledge with me. I thank my co-supervisor Derek Abbott for his support and guidance in every stage of my candidature.

I would like to extend my sincere thanks to my collaborators Dr Raffaele Ferri (Oasi Research Institute - IRCCS, Troina, Italy), Associate Professor Oliviero Bruni (Sapienza University, Rome, Italy), and Professor Susan Redline (Harvard Medical School, Boston, USA) for sharing their immense knowledge in sleep research with me. I thank the Head of School Associate Professor Wen Soong and our postgraduate coordinator Dr Withawat Withayachumnankul for their support. Thanks also to my colleagues and the entire staff at the School of Electrical and Electronic Engineering. Particularly helpful to me during this time were my desk neighbours Simanto, Madhu, and Filip who provided much needed entertainment at every stage of my candidature. Thanks should also go to the entire GOstralia! team for their support prior to my candidature and to the University of Adelaide for their financial support during my candidature (GOstralia! University of Adelaide PhD Scholarship).

At this point I want to thank all the great people I was lucky to meet on my journey to this day and without whom I would not be where I am today. Several people extended a great amount of assistance during different stages of my life and provided very much appreciated distraction from the academic life. Moreover, various people significantly contributed to my personal growth by consistently supporting me, but also by repeatedly challenging my perspectives.

Finally, I would like to thank all my family members for their ongoing support. I am extremely grateful for my parents' relentless support and patience during my education, especially during the time when school was not my highest priority. Special thanks to my brother who reduced the pressure on me by carrying the burden of being the first-born and to my sister who opened up the academic path for me by being the first member in our family that attended a university and also by being the first one that received a doctor title. It seems

normal today but your decision to embark on this path had a significant impact on my life.

Simon Hartmann,
November 2021,
Adelaide, Australia.

Conventions

This thesis is typeset using the MiKTeX software. TeXstudio was used as text editor interfaced to MiKTeX. Ti*k*Z, RStudio, GNU Image Manipulation Program, and Matlab® were used to produce schematic diagrams and other drawings. Harvard style is used for referencing and citation in this thesis. Australian English spelling is adopted, as defined by the Macquarie Dictionary Eighth Edition, Macquarie Dictionary, 2020, 8th edn, Macquarie.

Publications

Journals

Hartmann, S., Ferri, R., Bruni, O. and Baumert, M. (2021), 'Causality of cortical and cardiovascular activity during cyclic alternating pattern in non-rapid eye movement sleep', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379**:20200248.

Shahrbabaki, S. S., Linz, D., Hartmann, S., Redline, S., and Baumert, M. (2021), 'Sleep arousal burden is associated with long-term all-cause and cardiovascular mortality in 8001 community-dwelling older men and women', *European Heart Journal* **42**, pp. 2088–2099.

Hartmann, S., Bruni, O., Ferri, R., Redline, S. and Baumert, M. (2020), 'Cyclic alternating pattern (CAP) in children with obstructive sleep apnea and its relationship with adenotonsillectomy, behavior, cognition, and quality-of-life', *Sleep* **44**(1), zsaa145.

DelRosso, L. M., Hartmann, S., Baumert, M., Bruni, O., Ruth, C. and Ferri, R. (2020), 'Non-REM sleep instability in children with restless sleep disorder', *Sleep Medicine* **75**, pp. 276–281.

Hartmann, S., Bruni, O., Ferri, R., Redline, S. and Baumert, M. (2020), 'Characterization of cyclic alternating pattern during sleep in older men and women using large population studies', *Sleep* **43**(7), zsaa016.

Hartmann, S. and Baumert, M. (2019), 'Automatic A-Phase detection of cyclic alternating patterns in sleep using dynamic temporal information', *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**(9), pp. 1695–1703.

Conference papers

Saha S, Hartmann S, Linz D, Sanders P, and Baumert M, (2019), 'A ventricular far-field artefact filtering technique for atrial electrograms', *2019 Computing in Cardiology (CinC)*, 10.23919/CinC49843.2019.9005813.

Hartmann, S. and Baumert, M. (2019), 'Improved A-phase detection of cyclic alternating pattern using deep learning', *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1842–1845.

Conference abstracts

Hartmann, S. and Baumert, M. (2020), '0392 The effect of benzodiazepine use on non-REM sleep instability in community-dwelling older men', *Sleep* **43**(Supplement_1), A150–A150.

Hartmann, S. and Baumert, M. (2020), '0393 The effect of trazadone use on non-REM sleep instability in community-dwelling older men', *Sleep* **43**(Supplement_1), A150–A151.

Hartmann, S. and Baumert, M. (2020), '0818 Cyclic alternating pattern as indicator for subjective sleep quality in community-dwelling older men', *Sleep* **43**(Supplement_1), A312–A312.

DelRosso, L. M., Hartmann, S., Baumert, M., Bruni, O. and Ferri, R. (2020), '0943 Increased non-REM sleep instability in children with restless sleep disorder', *Sleep* **43**(Supplement_1), A358–A358.

Patents

The University of Adelaide, (2019), 'A method and system for classifying sleep related brain activity', *AUS Patent Application Ser. No. 2019902074* (filed 2019-06-27). PCT stage.

Contents

Acronyms **215**

List of Figures

List of Tables

1 Introduction

ALTHOUGH sleep is commonly considered as the rest state of the human body, the brain is highly active during this restorative phase. Transient, intermittent perturbations representing sudden bursts of brain activity can be inspected in the electroencephalography (EEG) as part of the polysomnography method. The common approach to analyse an overnight screening is to visually score the recorded signals. However, a robust automatic software for sleep data classification that can be applied on large-scale, population-level databases of sleep studies would enhance the understanding of sleep. This chapter presents the objectives and motivations behind the research presented in this thesis on innovative sleep data processing. Furthermore, this chapter highlights the original contributions to existing knowledge obtained by the studies in this thesis.

1.1 Introduction

During treatment of his rare form of pancreatic cancer, Steve Jobs, the co-founder of Apple Inc., came across the fascinating intersection of biology and technology. Watching his son Reed pushing the boundaries of cancer research using cancer genome sequencing, Steve Jobs, who undoubtedly shaped the information age like no one else, predicted that the twenty-first century will be the era of biotechnology (Isaacson, 2011). His prediction also applies to technological advances in the medical field as the term biotechnology encompasses a wide field and not only technologies using biological systems and organisms. In this context, it is inevitable to mention the impact of artificial intelligence (AI) on the future of medical research and diagnostics. The application of AI has been a key focus in various medical areas and will present one of the greatest challenges in the upcoming decades.

One very popular sector for the usage of AI in medical research is sleep stage scoring due to its unambiguous rules and the large volume of datasets. The rules for sleep scoring are based on the scoring guide by Rechtschaffen and Kales (1968) and were later redefined in the current consensus published by the American Academy of Sleep Medicine (AASM) (Iber *et al.*, 2007). According to the consensus, sleep can be divided into 30-second epochs of wakefulness, states of high neuronal activity (rapid eye movement (REM)), and states of quiescence (non-rapid eye-movement (NREM)). As short-lasting events and transient power alterations in frequency sub-bands are neglected in the rules of sleep scoring, additional sleep parameters were defined to characterise sleep events such as arousals, K-complexes, or cyclic alternating pattern (CAP). This thesis focuses on enhancing the understanding of the role of CAP which can be described as a marker of sleep instability. It comprises oscillating brain waves defined as short EEG amplitude increases (<60 s) during NREM stages that are in tune with the rest of the body (Parrino, 2021). It is known as a synonym for the microstructure of sleep. Also, CAP has been linked to a large number sleep disorders in various studies but the role of CAP is not entirely understood yet. Its relationship to cognitive functioning and brain connectivity remains largely unknown and needs to be explored further. Moreover, CAP needs to be investigated in studies with larger samples as the tedious and time-consuming manual scoring impedes the analysis of large sleep cohorts.

Hence, automated scoring systems using AI may provide a solution to overcome the limitations of manual scoring. The greater goal of automated systems is to reduce the workload for sleep clinicians as manual scoring is often tedious and time-consuming and to

provide repeatable diagnostic outcomes. However, most of the automated scoring systems share common drawbacks such as the arduous search for the best architecture and parameters, database variability (Alvarez-Estevez and Fernández-Varela, 2020), data inefficiency (Phan *et al.*, 2020), the computational limits of deep learning (Thompson *et al.*, 2020), and the lack of practical viability (He *et al.*, 2019). In detail, the lack of transparency with regard to model interpretability, accountability in terms of patient safety, and data management issues increase the doubt in many clinicians to integrate AI-based technology in their environment (He *et al.*, 2019). In this thesis, an automated CAP detection system is proposed that was developed closely with leading sleep experts in this field to overcome aforementioned issues and to provide the opportunity to analyse CAP in large sleep cohorts. It is demonstrated that a big data approach can provide new insights into the role of CAP.

The following sections provide an overview over the research questions that are addressed in this thesis and present the contribution of the work completed in this thesis to existing knowledge in the field of sleep and sleep data processing.

1.2 Contextual statement

Sleep was regarded for a long time as a biological phenomenon with greatly reduced neuronal activity (Hobson *et al.*, 1978). The discovery of REM and NREM sleep phases in the 1950s and 60s reversed this assumption. In 1969, Rechtschaffen and Kales (1968) published the first guideline to score sleep in human subjects in order to enhance the understanding of sleep. Later, the rules by Rechtschaffen and Kales were replaced by the current consensus detailed in the AASM scoring manual (Iber *et al.*, 2007). However, short-lasting events such as K-complexes and transient power alterations in frequency bands are not included in the current AASM framework. Hence, further guidelines were released to score sleep events such as CAP. Commonly known as a synonym for sleep microstructure (Parrino, 2021), CAP describes periodically recurring short-lasting brain activations that are in tune with the rest of the body. However, the manual inspection of sleep signals in terms of CAP scoring causes some key issues. A human scorer is limited to visually prominent indicators in the signals. Hence, finding hidden relationships and correlations between the signals can be very time-consuming or even not feasible. Consequently, studies on CAP were limited to a small number of samples that can be scored manually. This opened the

research field to algorithms that can automatically evaluate sleep recordings. In this thesis, a fully automated CAP scoring system, including comprehensive pre-processing and artefact removal stages, is proposed.

Published work on automatically classifying CAP events in EEG is limited. First systems using polygraphic features (Rosa *et al.*, 1999) were introduced shortly before the turn of the millennium. Later, thresholding algorithms in combination with distinct feature selection (Navona *et al.*, 2002; Ferri *et al.*, 2012), or competitive machine learning algorithms (Mariani *et al.*, 2013; Mendonça *et al.*, 2018a) were applied to accurately detect CAP sequences. However, the majority of the systems fail to represent end-to-end solutions including pre-processing, artefact removal, multi-class classification, and post-processing and fail to deal with the bias introduced by imbalanced datasets. Here, a comprehensive end-to-end solution using long-short term memory (LSTM) network as representative for the recurrent neural network (RNN) class is developed. The proposed model demonstrates a significant improvement in accuracy and sensitivity when using information from the past through RNN as compared to previously proposed systems. Also, the feedback loop in the training process was optimised to be able to deal with imbalanced data as training set.

CAP has been characterised in small population samples in numerous previous studies. Large studies on CAP with more than 100 subjects are rare exceptions due to the immense workload associated with manual scoring. Although semi-automated scoring systems have been available lately to facilitate the scoring process by providing an initial scoring output which is then reviewed by a manual scorer, big data analysis has not been applied on large population-based cohort studies to expand the knowledge about the role of CAP. Hence, the studies presented in this thesis use the fully automated end-to-end CAP scoring system to analyse on large sleep cohorts. Firstly, the automated CAP scoring system is validated by characterising CAP in a large population based cohort and comparing the output to values from the literature. Moreover, we demonstrate that automated CAP analysis of large population based studies can lead to new findings on CAP and its subcomponents. Next, CAP is analysed in a large paediatric sleep studies to investigate the relationship between CAP and cognitive functioning in children as well as the effect of early adenotonsillectomy (eAT) on CAP in children with obstructive sleep apnea (OSA). The relation between CAP and cognitive functioning is a research field that has not been widely investigated yet.

Besides the characterisation of CAP in the general population and in pathologies, the interplay between cortical and cardiovascular activity during CAP has not yet been elucidated. Several studies have previously reported the activating effect of CAP on the autonomous

nervous system (ANS) such as heart rate and blood pressure (Kondo *et al.*, 2014; Ferini-Strambi *et al.*, 2000; Dorantes-Méndez *et al.*, 2018) as compared to the sustained stability between cortical activity and ANS during non-CAP periods (Parrino *et al.*, 2016). The discovered correlations in those studies display the significance of analysing the sympathetic activity alterations underpinning the microstructure of sleep. Thus, the final chapter of this thesis focuses on causal relationships between cortical events defined by CAP and autonomic cardiovascular control using Wiener-Granger causality (GC). Differences in cortical-cardiovascular interactions during CAP sequences with predominantly A1, A2, or A3-phases as compared to non-CAP sequences reveal novel cues about the underlying cortical-cardiovascular dynamics.

In summary, the rationale behind the studies conducted in this thesis was to expand the understanding of CAP by developing a high-performance end-to-end CAP scoring system and applying it on large sleep cohorts. This way, novel insights into the role of CAP as marker of sleep instability may be identified.

1.3 Overview of thesis

This thesis encompasses seven main chapters including the introduction, the literature review, four main parts containing original contribution, and the conclusion. The outline of each chapter is described as follows:

Chapter 1 comprises the current introductory chapter which serves as an introduction into the foci and key questions of this thesis.

Chapter 2 represents the literature review of the research field in this thesis. The literature review provides physiological background about sleep and introduces well-known guidelines to analyse sleep recordings. Moreover, it explains the fundamental concept of supervised classification and discusses previously reported models for automatic sleep stage classification and CAP detection.

Chapter 3 presents the high-performance A-phase detection algorithm that was developed within this thesis. The chapter describes the methodological concepts in detail and it provides an in-depth evaluation of the performance of the developed system.

Chapter 4 and Chapter 5 summarize the outcomes obtained by applying the previously developed CAP scoring algorithm on two large population studies. Chapter 4 assesses CAP in relation to age, gender, self-reported sleep quality, and the degree of sleep disruption in

large community-based cohort studies of older people. Chapter 5 explores in children with OSA the effect of adenotonsillectomy (AT) on CAP and the relationship between CAP and behavioural, cognitive, and quality-of-life measures.

Chapter 6 investigates the causal relationships between cortical and cardiovascular activity during CAP in NREM sleep using GC.

Chapter 7 summarizes the findings and contributions of this thesis and provides a short overview over future research directions and potentially succeeding work that builds on the achieved results in this thesis.

Appendix section contains additional information about the developed scoring algorithm (Appendix A) and the outcomes of automated CAP analysis on a sample of children with restless sleep disorder (Appendix B). Furthermore, the Appendix section includes a study (Appendix C) to quantify the arousal burden across large cohort studies and determine its association with long-term cardiovascular and overall mortality in men and women.

1.4 Statement of original contribution

Four peer-reviewed first-author journal articles which are presented in Chapters 3–6 resulted from the studies conducted towards this thesis. Additionally, Appendices A–C contain one conference paper and two peer-reviewed co-author journal articles. The methodology and the engineering framework in this thesis were developed solely by the author using MATLAB® and Python. The framework for statistical analysis in each study was generated solely by the author using the statistical computing software R. The original contributions of the author in each study comprised formulating the hypothesis, developing appropriate research methodology and testing the hypothesis using statistical analysis.

1.5 Data

A publicly available dataset on PhysioNet (CAP Sleep Database) and various sleep studies available online at the National Sleep Research Resource (NSRR) were used in this thesis. Furthermore, samples from a sleep study conducted at the Seattle Children's Hospital, Seattle, WA, USA, were part of this thesis. Following paragraphs explain in detail the datasets used in this thesis.

The publicly available CAP Sleep Database (Terzano *et al.*, 2001) on PhysioNet (Goldberger *et al.*, 2000), an open-source repository for physiological signal recordings, was used in Chapter 3 to train the developed detection system and in Chapter 6 to investigate cortical-cardiovascular interactions during CAP. The dataset comprises 108 polysomnographic recordings conducted at the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy. The recordings include 16 healthy subjects and 92 patients including 40 subjects with nocturnal frontal lobe epilepsy (NFLE), 22 with REM behaviour disorder, 10 with periodic leg movement, 9 with insomnia, 5 with narcolepsy, 4 with sleep-disordered breathing, and 2 with bruxism. Each recording contains sleep stage scoring according to the rules of Rechtschaffen & Kales (Rechtschaffen and Kales, 1968) and CAP scoring according to the atlas of CAP (Terzano *et al.*, 2001). The sampling rate for EEG signals in this dataset ranges from 100 Hz to 512 Hz. The REMlogic™ software Embla® was used to visualize and score the signals. No further information is available on technical specifications.

In Chapter 4, two multi-center sleep cohorts (Osteoporotic Fractures of Men (MrOS) and the Study of Osteoporotic Fractures (SOF)) provided by the NSRR were used to characterise CAP in large population-base cohorts (available online at the National Sleep Research Resource; sleepdata.org) (Zhang *et al.*, 2018). The long-standing cohort study MrOS aimed to investigate fracture risk in relation to bone mass, bone geometry, lifestyle, anthropometric and neuromuscular measures, and fall propensity from 2000 to 2005 (Orwoll *et al.*, 2005). 3,135 of the 5,994 participants were recruited for an ancillary sleep study (MrOS Sleep Study) to expand the understanding on the relationship between sleep architecture, sleep-disordered breathing, and cognition in men aged 65 or older (Blackwell *et al.*, 2011). Subjects who used mechanical devices or oxygen therapy during sleep were generally excluded from the study. Of the 3,135 men, 2,909 completed in-home overnight polysomnography (PSG) using unattended, portable devices (Blackwell *et al.*, 2011). On the other hand, 461 of the 4,727 women that participated in the latest visit cycle of the SOF study between 2002 and 2004 completed an unattended overnight 12-channel in-home PSG to investigate the association between sleep disturbances and cognitive impairment in community-dwelling older women (Spira *et al.*, 2008). Originally, SOF was designed to find new knowledge on the osteoporosis and ageing in women aged 65 or older (Cummings *et al.*, 1990). In MrOS, the Compumedics® Safiro Sleep Monitoring System was used to record in-home PSG whereas the Compumedics® Siesta Unit was used in SOF for in-home PSG. In both studies, set up of in-home PSG was performed by trained experts that followed standardized procedure manuals (available online at the National Sleep Research Resource; sleepdata.org). In MrOS, EEG was recorded using a sampling rate of 256 Hz and a high pass

hardware filter at 0.15 Hz. The sampling rate in SOF was set at 128 Hz and a high pass hardware filter with a cut-off frequency at 0.15 Hz was applied. In both studies, gold cup electrodes were deployed as sensors to record left and right central EEG (C3 and C4), and left and right mastoid EEG (A1 and A2).

The Childhood Adenotonsillectomy Trial (CHAT) (Marcus *et al.*, 2013) was part of the CAP analysis in Chapter 5 (available online at the National Sleep Research Resource; sleepdata.org) (Zhang *et al.*, 2018). The multi-center, single-blind, randomized, controlled trial was designed to determine whether eAT in children with mild to moderate obstructive sleep apnoea demonstrates greater improvement in cognitive, behavioral, quality-of-life, and sleep measures as compared to watchful waiting with supportive care (WWSC) (Marcus *et al.*, 2013). The recruited cohort of children between 5.0 and 9.99 years of age were randomly assigned to one of treatment strategies and underwent standardized full PSG with scoring at baseline and after a 7-month observation period. Certified sleep technicians trained in pediatric PSG performed PSG during overnight visits at one of the eligible clinical sites. Each clinical site used a standardized approach that was established at training. Each PSG unit was certified prior to the study commencement based on signal quality, sensor interface, de-identification and the ability to follow the CHAT PSG Standardizations. Hence, a various number of different equipments was deployed across testing sites ranging from Natus® units such as the Embla® N7000 and the Xltek® to Compumedics® units such as the E-series. Data was standardised by following regulated rules for the same montage, comparable sensors and sampling rates and filters across sites that were outlined in the PSG Manual of Procedures (available online at the National Sleep Research Resource; sleepdata.org). During each PSG, the left and right central EEG (C3 and C4) as well as the left and right mastoid EEG (M1 and M2) were recorded using a sampling rate of 200 Hz.

For the study on the association of sleep arousal burden with long-term cardiovascular and overall mortality in Appendix C, 4,795 subjects in the Sleep Heart Health Study (SHHS) were analysed in addition to 2,782 male participants of the MrOS sleep study and 424 female participants of the SOF study. The SHHS is a multi-center study to analyse the cardiovascular consequences of sleep-disordered breathing (Redline *et al.*, 1998). Altogether, 6,441 men and women aged 40 years and older were recruited for the first visit SHHS Visit 1 which took place between November 1, 1995 and January 31, 1998. In-home data collection was standardized using clear protocols and rigorous training resulting in good-quality multichannel PSG data (Redline *et al.*, 1998). In the SHHS, the portable sleep unit

Compumedics® SleepWatch PS was used for unattended in-home PSG. Two EEG channels (EEG and EEGsec) were recorded using a sampling rate of 125 Hz and a high pass hardware filter at 0.15 Hz.

Additionally, thirty-eight children who fulfilled restless sleep disorder (RSD) diagnostic criteria (DelRosso *et al.*, 2019) (23 boys and 15 girls, age range 5–17 years), 23 children with restless limb syndrome (RLS) (18 boys and 5 girls, age range 4–17 years) and 19 controls (10 boys and 9 girls, age range 5–18 years) from a local sleep study at the Seattle Children's Hospital, Seattle, WA, USA were analysed in regard to the sleep microstructure during NREM sleep in Appendix B. The study was approved by the local institutional review board. Children who were younger than 4 years, or use medications that alter sleep parameters, or shown signs of co-morbid sleep disorder, or have medical or psychiatric conditions known to affect sleep, or use caffeine were generally excluded. Left and right central EEG referenced to left and right mastoid EEG (C4M1 and C3M2) were recorded using a sampling rate of 128 Hz. No information on applied PSG equipment is available.

Literature review

2

SLEEP is nowadays considered as a common habitus among living organisms including humans. The physiological role of sleep for humans can be briefly summarized as the reorganization of neuronal activity. Furthermore, sleep deprivation or fragmentation increase the risk of cardiovascular diseases highlighting its overall recuperative role. This chapter provides a brief overview on sleep and the gold standard for sleep monitoring—the PSG method. Furthermore, the conceptual frameworks of classification algorithms are briefly summarized. With the aid of classification algorithms, recurring patterns in time signals evoked by sleep specific events can be detected and categorised.

2.1 Physiological background

Sleep is a recurring, biological pattern that is essential for human beings. Scientists have conducted numerous studies in the past with the aim to enhance our understanding of sleep. This thesis concentrates on extracting crucial information from sleep recordings. Hence, the following sections provide background information on sleep and how it can be measured. Moreover, different sleep scoring methods are presented including CAP, which is the main focus of this thesis.

2.1.1 Sleep

For decades, sleep has been considered as a period of greatly reduced cortical activity (Hobson *et al.*, 1978). It was regarded as a state of inactivity—the opposite state of wakefulness. However, sleep is not only a state of reduced activity but rather a process that consists of a specific architecture and initializes dramatic changes in brain electrophysiology, neurochemistry, and functional anatomy (Brand and Kirov, 2011). Moreover, its reversibility distinguishes sleep from other inactive states like hibernation, coma, or death (Deboer, 2015). Studies on sleep in animals have shown that typical signs of sleep, i.e. quiescence and an increased arousal threshold, are common among mammals and birds (Cirelli and Tononi, 2008). It can be identified that all animal species display sleep-like behaviour, i.e. a transition into a resting state (Deboer, 2015). However, more research needs to be conducted to be able to state that sleep is universal.

The primary function of human sleep still remains an unresolved question but extensive research in the last decades has shed more light on our understanding of sleep. It is known that the physiological role of sleep includes memory consolidation and re-consolidation facilitating the recapturing in waking state (Stickgold and Walker, 2005). Moreover, a recent study discovered that waste products of neural metabolism accumulated during the day are removed while sleeping (Xie *et al.*, 2013). It is understood that sleep is regulated by circadian and homoeostatic processes. The circadian oscillation generated by the suprachiasmatic nucleus of the hypothalamus is responsible for the distribution of sleep over the 24-h day ensuring a proper adjustment of internal rhythms to the daily light-dark cycle (Franken and Dijk, 2009). On the other hand, the result of sleep loss in an increased propensity to sleep and a compensation of the lost hours with a longer sleep period imply the existence of a homoeostatic control system (Benington, 2000).

The quality of sleep is primarily defined by three pillars: duration, continuity, and depth. Previous studies have shown that different forms of sleep deprivation have severe, negative impacts on the neurobehavioural functions of humans during wakefulness (Van Dongen *et al.*, 2003). Especially partial sleep deprivation defined as a diurnal sleep period less than five hours can result in a significantly diminished cognitive and motor performance (Pilcher and Huffcutt, 1996). Sleep disorders such as sleep apnoea or periodic limb movements have a negative effect on sleep continuity causing fragmented sleep. Increased objective and subjective sleepiness as well as decreased psychomotor performance can be found in subjects with a severely disturbed sleep (Bonnet and Arand, 2003). Furthermore, sleep continuity is a major contributor to daytime well being in older people (Carskadon *et al.*, 1982). Finally, slow wave sleep (SWS) as a marker for deep sleep appears to play an important role in cerebral recovery (Horne, 1992). During SWS, total brain connectivity decreases but functional clustering increases resulting in a higher neuronal synchrony (Boly *et al.*, 2012). Furthermore, SWS is an instrumental factor in the maintenance and consolidation of sleep (Dijk, 2009).

Sleep can be assessed using a variety of methods. The gold standard for sleep monitoring is the PSG method, which is explained in detail in the following section. Alternative monitoring methods include actigraphy, video-PSG, or sleep surveys. Actigraphy and sleep surveys are popular in settings where standard PSG is limited such as intensive care units or long-term assessment. A wireless solution is the actigraphy method which describes the body activity during sleep using accelerometers. Sleep questionnaires or diaries represent a non-invasive alternative as a simple quantitative measure of the subjective sleep quality. Finally, video-PSG is already an integral part of PSG studies in sleep centres or hospitals and is particularly useful in paediatric sleep studies to visually investigate the sleep behaviour post recording.

2.1.2 Polysomnography (PSG)

The origin of physiological recordings can be dated back to 1875, when the Scottish physiologist Richard Caton recorded the first time electrical activity in brains of rabbits and monkeys (Deak and Epstein, 2009). In 1924, Hans Berger recorded the first time electrical activity of the human brain using a string galvanometer (Hirshkowitz, 2015). Berger named his new method of monitoring electrical activity of the human brain "electroencephalogram (EEG)". Additionally, Berger was able to demonstrate that a distinct waveform in the range of 8–13 Hz, so called alpha waves, diminish at sleep onset (Hirshkowitz, 2002). These

first findings on physiological recordings would later evolve into the PSG method with the recording of multichannel electro-physiological signals of EEG, electrooculogram (EOG), and electromyography (EMG). Nowadays, PSG involves a wide range of physiological signals. The following paragraphs introduce commonly measured channels during PSG in a brief manner.

Electroencephalography (EEG)

EEG is a non-intracranial method to capture the cerebral activity. Shortly, EEG measures the electric field that is caused by transmembrane ion movements in neurons (Proekt, 2018). To determine the strength of the electrical field extracellularly, two electrodes are required. The voltage between the electrode at the point of interest and the reference electrode displays the potential difference. Importantly, EEG is only capable of capturing the combined contributions from myriad neuronal and glial sources but not the voltage state of a single neuron (Proekt, 2018). Hence, EEG provides no information about the activity of individual neurons but the activity of cortical pyramid cells that share the same spatial orientation. Therefore, high voltage is a sign for synchronous activity of a vast number of such cells.

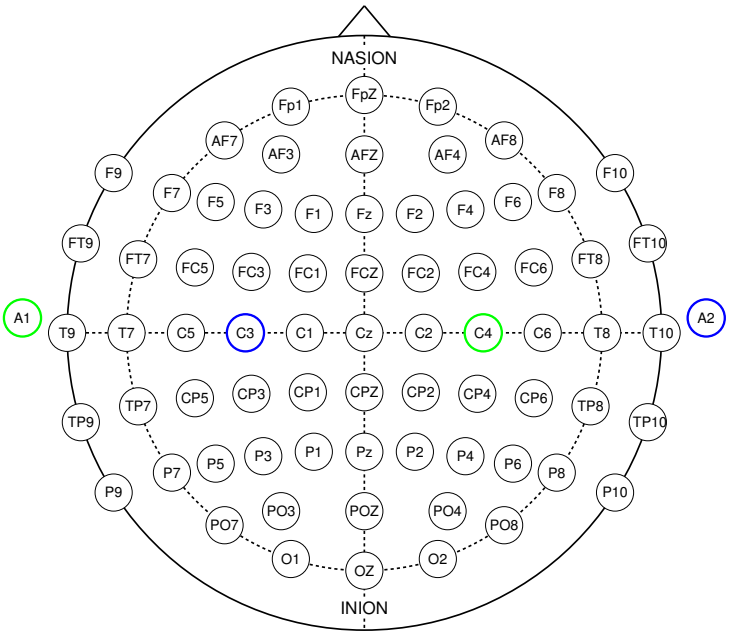


Figure 2.1. Illustration of the international 10–20 system. Display of the EEG electrode positions in the international 10–20 system. The channels for the central cortex area C4-A1 and C3-A2 are highlighted in green and blue, respectively.

Most commonly the reference electrode is placed on the subject to avoid long wires acting as antennas. Preferred choices are the mastoid or the ear. However, points on the body

Table 2.1. List of EEG frequency bands.

Band	Frequency (Hz)	Location	Characteristics
Delta	0.5–4	Frontal in adults, occipital in children	Slow-wave sleep (Stage 3 & 4), deep sleep
Theta	4–8	Fronto-central	Can be found in stage 1 and 2, drowsiness, associated with dreaming and meditative states
Alpha	8–12	Occipital	Typical for stage 1, posterior basic rhythm, associated with relaxed and calm states, eyes closed
Sigma	12–16	Fronto-central	Typical for stage 2, sleep spindle activity
Beta	16–32	Frontal and central, attenuated posteriorly	Associated with alert stage, eyes open

Detailed list of the five most common EEG frequency bands, their spectral range, their location, and their characteristics.

have the disadvantage that they rarely have a true zero potential. Other options include ground or a second electrode on the scalp as reference. The latter can only provide information about the difference in potential between two locations on the scalp. Moreover, the spatial placement of the electrodes plays an important role as the electrodes should cover the maximal possible area but also as many important underlying areas of the brain as possible. Thus, an international standard called the 10–20 system for electrode placement during EEG was developed. Figure 2.1 illustrates the electrode locations of the 10–20 system for EEG.

The traditional way of analysing EEG waveforms is the decomposition into spectral bands. Table 2.1 lists five of the most common frequency bands that are used to classify EEG waveforms. In general, EEG activity can be found in the bandwidth ranging from 0.5 Hz to 70 Hz. Gamma activity in the range of >32 Hz is only from interest in the relation to epilepsy but portrays low information in the context of sleep EEG (Worrell and Gotman, 2011).

Electrooculography (EOG) and electromyography (EMG)

EOG measures eye movement during sleep. An electrode is placed below and above the outer canthus of the right and left eye, respectively, and both electrodes are connected to a single reference electrode. This arrangement ensures the monitoring of horizontal and vertical eye movements. At a person's perceived sleep onset, EOG shows a change to slow eye movements (Carskadon and Dement, 2005). Moreover, EOG is an important source to classify REM phases. On the other hand, EMG records the electrical activity of skeletal muscles. In the context of PSG, EMG usually refers to chin muscle activity and

limb movement. To capture the activity of chin muscles, one electrode is placed above the jawline and one below. Two electrodes are placed over the anterior tibialis on the legs to monitor leg activity. Overal, EMG is important to determine the onset of REM sleep, to identify tooth clenching, and to detect sleep disorders such as periodic limb movement.

Cardio-respiratory signals

A typical PSG often measures various additional signals such as airflow, snoring, or thoracic and abdominal movement using respiratory inductance plethysmography (RIP). Here, we focus on the signals recording cardio-respiratory activity such as electrocardiography (ECG), photoplethysmogram (PPG) and RIP. The ECG represents the cardiac electrical activity. It projects the rhythmic contraction of the heart onto a time-voltage space. Before cardiac contraction, synchronized electrical current is spread through the heart muscle which is concurrently captured by electrodes placed on the body (Goldberger *et al.*, 2017). Pacemaker cells, specialized conduction tissue within the heart, and the heart muscle itself generate these electrical currents (Goldberger *et al.*, 2017). In detail, the cardiac cell membranes are negatively charged in their resting state but depolarize during cardiac stimulation (Becker, 2006). The return to the polarized resting membrane potential is called repolarization. Depolarization, the spread of a stimulus through the heart muscle, and repolarization result in five basic ECG waveforms (P, QRS, ST, T, and U) which combined represent the typical heartbeat seen in ECG. The P wave represents the depolarization of the atrial muscle, the QRS complex displays the ventricular depolarization, and ST, T, and U wave describe the ventricular repolarization (Goldberger *et al.*, 2017).

On the other hand, PPG is a non-invasive method to trace the blood circulation. Devices measuring PPG are equipped with a light source and a light detector that is placed either along the light source (reflectance mode) or at the contrary side of the light radiation separated by the tissue (transmission mode) (Castaneda *et al.*, 2018). The detected signal consists of a continuous proportion comprising emitted light that is partly absorbed by the tissue, partly reflected by the skin, and partly back-scattered by the underlying blood vessels as well as of a pulsating fraction as a result of changes in the blood flow volume (Roberts, 1982). The alternations in blood volume during the systolic and diastolic phase of the cardiac cycle produce alternations in detected light intensity enabling to monitor the pulse rate non-invasive (Tamura *et al.*, 2014). Normal PPG devices capture the blood flow using one single light source whereas pulse oximeters are able to additionally determine the blood oxygen level fusing the information from a red and a infra-red light source.

ECG and PPG are commonly used to determine variables to describe the autonomous nervous system. Common measures extracted from ECG are the heart rate (HR) and heart rate variability (HRV) which describe the time interval between R peaks and its variations. A direct quantity that can be derived from PPG is the pulse wave amplitude (PWA) describing vasoconstriction at the measurement site. Information about the wave propagation can be obtained from simultaneously collected ECG and PPG signals. The indirect measure pulse transit time (PTT), pulse arrival time (PAT), and pulse wave velocity (PWV) provide important information about the cardiovascular function status such as blood pressure and arterial stiffness (Liang *et al.*, 2019). The two measures describing the pulse wave propagation, PAT and PTT, differ in the end point of the wave propagation period. In detail, PAT describes the period from the R peak to a distal arterial site such as the fingertip whereas PTT is defined as the period from one arterial site such as the upper arm to another arterial site such as the fingertip (Mukkamala *et al.*, 2015).

RIP is a method to assess lung volume changes non-invasively (Wolf and Arnold, 2005). It can provide thoracic and abdomen cross sectional area changes using two elastic bands around the chest and the abdomen (Eberhard *et al.*, 2001) and is utilized to measure breathing during sleep as invasive methods may perturb the sleep process. In comparison to traditional breathing estimation techniques using thermal sensors or nasal pressure transducers, RIP demonstrates a reasonable alternative to estimate ventilatory parameters (Eberhard *et al.*, 2001). Thermistors exploit the change in temperature between exhaled and ambient air to estimate airflow whereas pressure transducers convert the nasal airflow pressure into an electrical signal using a nasal cannula (Redline *et al.*, 2007). Thermistors, which were the most common sensor to quantify airflow limitation in patients until recently, show a high accuracy in detecting complete cessations of airflow (apnoeas) but inaccuracies in estimating reduction of airflow (hypopnoeas) whereas nasal pressure transducers and RIP appear to be superior for hypopnoea detection (Redline *et al.*, 2007). Hence, a combination of multiple sensors provides the best information for apnoea and hypopnoea scoring in PSG in order to diagnose sleep related breathing disorders such as OSA.

2.1.3 Sleep scoring

In 1953, first observations of fast, jerky, and binocularly symmetrical eye movements during sleep were propagated (Aserinsky and Kleitman, 1953). Later, this work paved the ground for the first classification guideline of 30-second sleep epochs into phases with REM and

NREM published by Rechtschaffen and Kales (1968). The latter, in turn, was initially subdivided into four stages but was later modified to three distinct stages in the current consensus released by the AASM (Iber *et al.*, 2007).

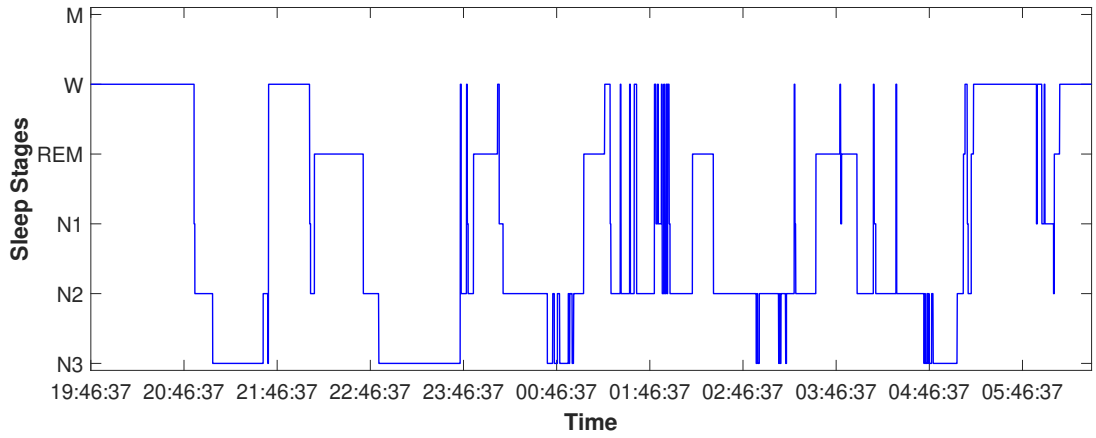


Figure 2.2. Sleep hypnogram as graphical representation of the sleep cycle. Example of a sleep hypnogram identifying the different stages of sleep: body movement (M), wake stage (W), rapid eye movement (REM) stage, stage N1 non-rapid eye movement (NREM) sleep, stage N2 NREM sleep, and stage N3 NREM sleep.

Stage N1 NREM sleep represents the transition into sleep following a wake stage (W) highlighted by a substantial decrease in alpha activity. It is commonly referred to as light sleep. Stage N2 NREM is characterized by sleep spindles and high activity in the theta EEG frequency band (Malhotra and Avidan, 2013). It can be regarded as an intermediate stage of sleep. Stage N3 NREM, also termed SWS, is described by high-amplitude slow waves. It symbolizes the deepest and most restorative sleep stage. In Rechtschaffen and Kales (1968), SWS was subdivided into stages N3 and N4 NREM sleep which were merged into one stage in the current consensus. The graphical representation of the sleep cycle as a function of time is called an hypnogram. Figure 2.2 displays an example of an hypnogram showing the nightly shifts between sleep stages.

One drawback of the traditional sleep scoring is the neglect of critically important short-lasting events and the dismissal of short stage shifts as each epoch is assigned one single label. Other scoring methods such as the arousal definition (ASDA, 1992) implement a more dynamic approach capturing short bursts in cortical activity. In the consensus-based rules published by ASDA (1992), the minimum duration criterion for arousals was established at three seconds. However, shorter EEG arousals (<3 seconds) commonly observed in children were not considered in the guideline (Grigg-Damberger *et al.*, 2007). Moreover, the proposed arousal definition fails to take account for slow-wave oscillations such as K-complexes. A more comprehensive approach to score sleep is defined in the atlas of

CAP (Terzano *et al.*, 2001). With its ability to capture sleep instability, sleep disturbance, or both, CAP provides crucial information about a subject's sleep in addition to the traditional sleep scoring.

2.1.4 Cyclic alternating pattern (CAP)

In 2001, a new guideline for recurrent observations in the sleep brain signal, the so called cyclic alternating pattern (CAP), was introduced (Terzano *et al.*, 2001). A CAP sequence is specified by three or more consecutive activation phases that are represented by transient, prominent events in the EEG data and only separated by a background phase (Terzano *et al.*, 2001). A-phases describe transient, phasic events that are distinct from the background, whereas B-phases portray intervals with lesser arousal level than A-phases but noticeably greater than the continuous background noise (Terzano and Parrino, 2000). Stereotypical A-phase patterns are delta bursts, vertex sharp transients, K-complex sequences, K-alpha, polyphasic bursts, intermittent alpha, and arousals (Terzano *et al.*, 2001). The activation periods can be solely found in the NREM sleep stage and are characterised by slower higher-voltage rhythms (A1), faster lower voltage rhythms (A3) or by a combination of both (A2) (Terzano and Parrino, 2005). Subtype A1 is associated with periods dominated by high EEG synchrony whereas A2 and A3 subtypes are connected to desynchronized patterns that are commonly found before and after REM sleep. Examples for each subtype are depict in Figure 2.3.

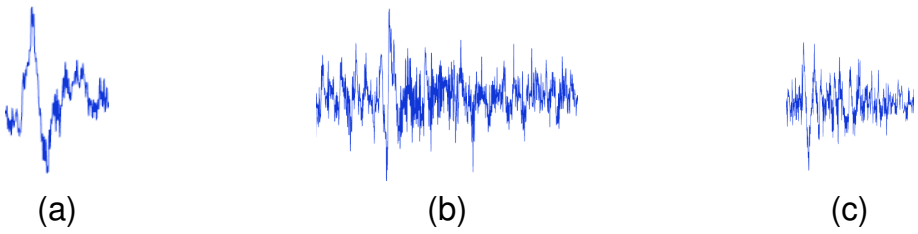


Figure 2.3. Typical examples of activation phase subtypes during cyclic alternating pattern events. (a) Subtype A1 is defined by slow high-voltage waves, (b) subtype A3 displays fast, low-voltage waves, and (c) subtype A2 is specified as a combination of both.

The incidence of subtypes depends on the time of the sleep cycle and the time within the sleep cycle. A sleep cycle, hereby defined as the time between sleep onset and the end of the first REM period and as the time between the ends of successive REM periods, is composed of a descending branch with progressive EEG synchronisation, a period

with high EEG synchrony, and an ascending branch with progressive EEG desynchronisation (Hobson *et al.*, 1978). A study by Terzano *et al.* (2000) demonstrated that the number of both subtypes with rapid EEG rhythms (A2 and A3) increased significantly the later the sleep cycle occurred whereas the amount of A1 subtypes did not alter across sleep cycles. Furthermore, Terzano *et al.* (2000) reported a higher incidence of A1 subtypes (90% of all A-phases) during sleep cycle states of progressive EEG synchronisation and high EEG synchrony whereas A2 and A3 subtypes displayed a higher incidence (64% of all A-phases) during the ascending branch with progressive EEG desynchronisation. In summary, A1 phases representing EEG synchrony are involved in the build-up and maintenance of deep sleep, while desynchronised EEG rhythms such as subtypes A2 and A3 are associated with the breakdown of SWS and the preparation of REM activity (Bruni *et al.*, 2010b).

The main indices to describe the prevalence of CAP in sleep data are the CAP rate and the subtype rates. The CAP rate defined as the percentage of NREM sleep occupied with CAP sequences follows a bimodal distribution with two peaks during adolescence and senescence, respectively (Parrino *et al.*, 2012) (see figure 2.4).

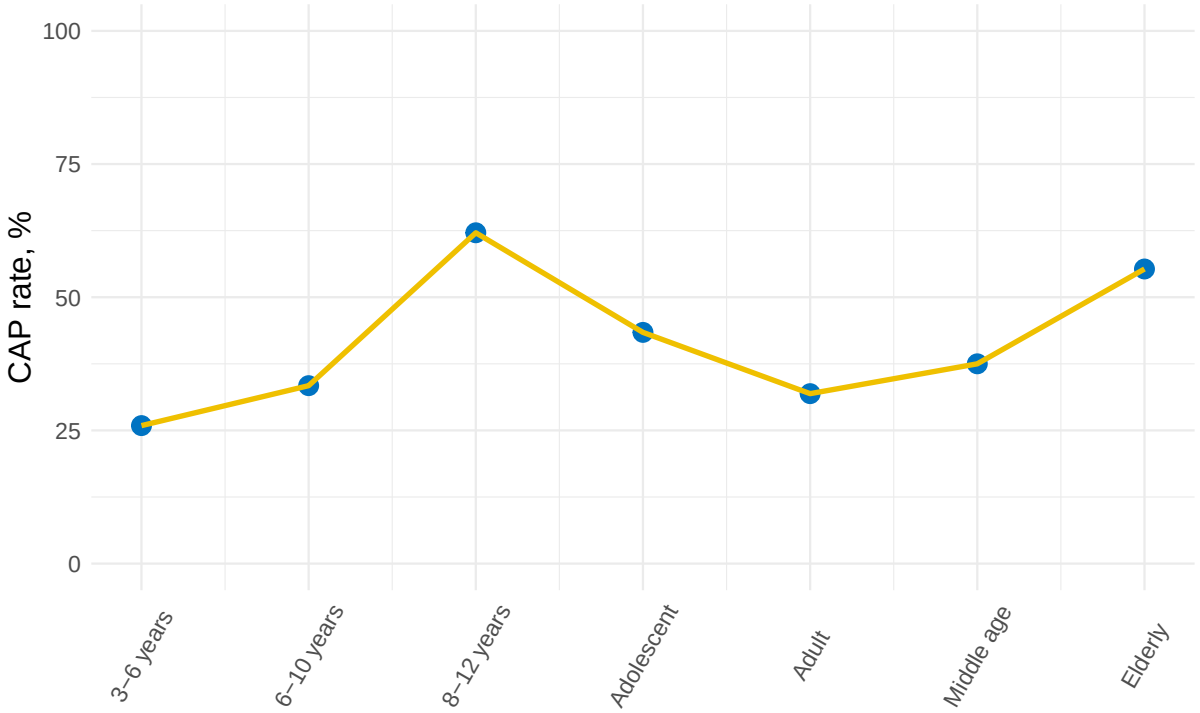


Figure 2.4. Age-related distribution of CAP rate. Age-related distribution of CAP rate from preschool age to senescence indicating the bimodal distribution with two peaks during adolescence and senescence, respectively.

CAP rate in children starts at 25.9% in pre-school age aged 3–6 years (Bruni *et al.*, 2005) and increases gradually until a peak at the peripubertal age (33.4% in school children aged 6–10 years (Bruni *et al.*, 2002); 62.1% in peripubertal children aged 8–12 years (Lopes *et al.*, 2005)). After the first peak during adolescence, CAP rate declines until young adulthood (43.4% in teenagers; 31.9% in young adults) (Parrino *et al.*, 1998). With increasing age during adulthood, CAP rate rises again until senescence (37.5% in middle aged and 55.3% in elderly) (Parrino *et al.*, 1998). A1 index, the number of A1-phases per hour of NREM sleep, follows a bell shape along the normal life span with a peak around school and peripubertal age whereas A2 and A3 indices emulate an inverse bell shape (Parrino *et al.*, 2012). The high ratio between A1 and A2/A3 subtypes in school-age children supports the idea that humans experience the highest sleep quality in this age range (6–10 years) as compared to other life stages (Parrino *et al.*, 2012).

CAP is regarded as a marker of sleep instability (Parrino *et al.*, 2012) and highly correlates with numerous sleep disorders in adults. Previous studies have shown that CAP rate increases as response to sleep disrupting disorders such as insomnia (Terzano *et al.*, 2003), upper airway resistance syndrome (Guilleminault *et al.*, 2007), sleep apnoea syndrome (Terzano *et al.*, 1996), periodic limb movement (Parrino *et al.*, 1996), and nocturnal frontal lobe epilepsy (Zucconi *et al.*, 2000; Parrino *et al.*, 2006). Also, pathologies such as depression (Farina *et al.*, 2003), Prader-Willi syndrome in adults (Priano *et al.*, 2006), and eating disorders (Della Marca *et al.*, 2004) elevate the amount of CAP during sleep. Sleep-promoting conditions such as narcolepsy (Terzano *et al.*, 2006; Ferri *et al.*, 2005c), CPAP treatment in obstructive sleep apnoea (Parrino *et al.*, 2005), and sleep recovery after prolonged wakefulness (Parrino *et al.*, 1993) have a lowering effect on the CAP rate. Further research has shown that CAP is not only influenced by sleep disorders but also operates as a modulator for sleep-related events (Parrino *et al.*, 1996, 2012).

Although CAP has been the main topic in various studies, further research is still needed to fill in the gaps of knowledge in certain areas. In terms of CAP in adults, the relationship between CAP and daytime sleepiness requires further investigations (Parrino *et al.*, 2012). Moreover, the relation between CAP and brain connectivity is a largely unexplored field that can provide novel knowledge on the understanding of underlying processes during sleep. In paediatric research, studies with large sample numbers as well as longitudinal studies that address CAP changes during development need to analysed in terms of CAP (Bruni *et al.*, 2010b). Furthermore, CAP needs to be explored in different sleep disorders of childhood such as insomnia, first-night effect, or hormonal dysfunction (Bruni *et al.*, 2010b). More

importantly, the relationship between CAP and cognitive functioning needs to be investigated in further studies. It has been the topic in initial studies but more studies need to be conducted to explore the role of CAP in learning processes and memory consolidation in children (Parrino *et al.*, 2012).

2.1.5 Physiological interplay during sleep

During sleep, the central nervous system (CNS) and the ANS are closely interrelated as it can be exemplary seen during sleep state transitions. The sympathetic predominance reaches relatively normal awake levels during REM sleep whereas it declines during NREM sleep and plunges during SWS when the autonomic balance is shifted toward parasympathetic activity (Baharav *et al.*, 1995; Mancia, 1993). Moreover, several studies have demonstrated that high-frequency interferences in the cortical system described as phasic events—synchronized or desynchronized—are correlated to activations in the ANS (de Zambotti *et al.*, 2018). Also, there is strong evidence that the cortex is stimulated in response to heartbeats, so called heartbeat-evoked potentials, concluding that the cortex is activated to process heart-related sensory inputs (Kern *et al.*, 2013; Park *et al.*, 2018). However, it is still questionable if the detected correlations are an actual response of the cortical system to changes in the ANS or just projections of the afferent signals from the brainstem regulating the cardio-respiratory system.

In response to a desynchronized phasic event, which is usually regarded as an arousal, the cardiac and respiratory activity increases (Trinder *et al.*, 2001b). The response of the body is mainly considered as a reflex-like reaction to reach quickly the full performance of the physiological systems in order to establish the opportunity to defend a threatening cause for the arousal (de Zambotti *et al.*, 2018). Furthermore, an arousal is from transient nature implying that the cardiovascular response is not dependent on the transition to wakefulness (Trinder *et al.*, 2003). Hence, the cardiovascular system in subjects, that suffer from a high occurrence of arousals during sleep, is frequently activated during sleep, resulting in a potentially higher cardiac mortality risk (Shahrbabaki *et al.*, 2021). Synchronized phasic events consist of spindles, delta bursts or K-complexes and are regarded as an arousal phenomenon, a sleep-promoting response on a stimulus or just a marker (Colrain, 2005). Recently published results show that K-complexes are associated with clear biphasic patterns in the heart rate reflecting a relationship between the brain and the cardiac system (de Zambotti *et al.*, 2016).

CAP as a marker of NREM-sleep instability has drawn increasing attention in recent years to enhance the understanding of the relationship and the coupling between the CNS and the ANS during sleep. Sequences of CAP cycles form an oscillating network with blood pressure, muscle tone, and heart rate whereas non-CAP periods display sustained physiological stability between CNS and ANS as no external or internal events challenge the sleep process (Parrino *et al.*, 2016). Kondo *et al.* (2014) have shown in their study that all A-phase subtypes demonstrate an increase in HR and blood pressure (BP) after onset. A study by Ferini-Strambi *et al.* (2000) of ten healthy subjects showed a significantly increased low-frequency component and a significantly decreased high-frequency component of HRV during CAP as compared to non-CAP. Ferri *et al.* (2000) supported those findings in their study of six normal children and adolescents, respectively. Dorantes-Méndez *et al.* (2018) investigated HRV during CAP in healthy subjects and NFLE patients. They detected significant shifts towards the low-frequency components of HRV in all subtypes but with a more pronounced shift in A3-phases as compared in A1 and A2. Bosi *et al.* (2018) reported in their study on the relation between A-phases and PWA changes after airway obstruction in patients with OSA a significant correlation between respiratory events combined with A-phases and respiratory events combined with PWA drops. In summary, the presented literature shows the importance of including CAP when investigating the relationship between CNS and ANS during sleep.

Studies on the dynamic interrelationship between CNS and ANS primarily applied bivariate non-causal methods. Novel theoretical frameworks and the availability of recorded multivariate time signals enable the investigation of causal relationships between cortical, cardiovascular and respiratory variables. Recent studies deploying linear, non-linear, or information theoretic approaches reveal the beta EEG frequency band as the centre of the information flow between EEG rhythms and as the central node of information transfer between brain and cardiovascular system (Faes *et al.*, 2014b, 2015, 2014a). Schiecke *et al.* (2019) reported a significant difference in directed interaction from beta and alpha frequency band to HRV as well as from HRV to beta frequency band using non-linear convergent cross mapping. To integrate the complex behaviour and interactions of particular organs into one system-wide network, the novel emergent field of network physiology was established (Bartsch *et al.*, 2015). The integrated network seeks to define an individual's psychophysiological state by applying a theoretical framework focused on the coordination and interactions among the different physiological systems (Ivanov *et al.*, 2016). There have been numerous studies on the link between CAP and sleep disorders but studies on the causal relationships between cortical, cardiovascular and respiratory variables during

CAP are limited. These novel methods to describe the dynamic interrelationship between CNS and ANS may be useful for analysing the level of decoupling during CAP, and consequently may highlight the importance of sleep microstructure when evaluating autonomic activity alteration during sleep.

2.2 Machine learning and deep learning in EEG analysis

Classification can be described as a process that assigns an observation to its related class from a set of subgroups using predictor features. The classification decision is called supervised in case a priori knowledge from labelled training data is available. On the contrary, unsupervised classification generates clusters without labelled data. Here, the focus is on supervised learning methods, in particular machine learning methods. With the recent surge in computing power and available data, machine learning algorithms soared in popularity for classification tasks based on images or time signals. The following sections provide a short introduction into the classification process and a brief summary of the most common supervised classification algorithms. Moreover, a comprehensive review of previously proposed EEG analysis methods in sleep staging, seizure detection, and CAP detection is conducted.

2.2.1 Classification techniques

The ability to cluster data into subgroups by finding similarities in the observations is essential to data analysis, data mining, and pattern recognition. Figure 2.5 displays the basic principle of a classification task. Firstly, an input batch of observations is passed into a classifier. Subsequently, the classifier categorizes the input batch into output classes based on a measure that was previously trained on labelled data. During training, the classification measure is optimised minimizing the discrepancy between classified samples and ground truth, the labelled data. The optimisation process is commonly modelled with a loss function.

A common problem in pattern recognition in sleep EEG is the imbalance in observations for each class. Imbalanced datasets (see Figure 2.6) pose a major problem for supervised learning as they introduce a bias into the decision finding process. A classifier trained on an imbalanced dataset favours the class of the most prominent label to achieve a high

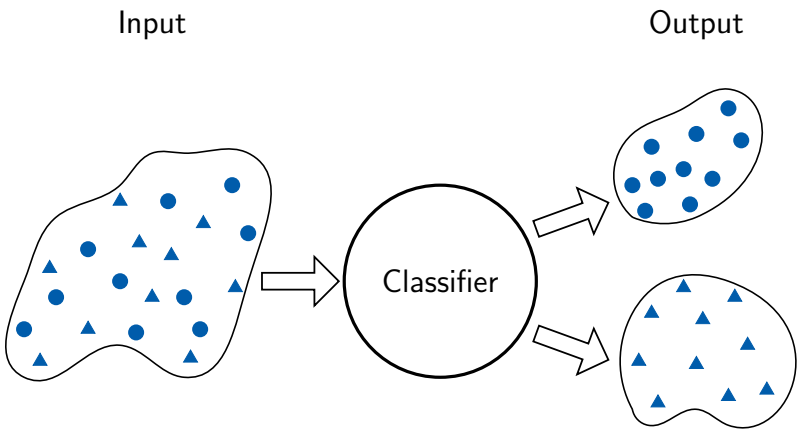


Figure 2.5. Schematic overview of a binary classification process. Simplified schematic of the information flow in a binary classification process and supervised learning.

accuracy. Standard methods to overcome the problem of imbalanced data are downsampling such as randomized removal of samples, upsampling such as data augmentation, and adaptation of the training algorithm.

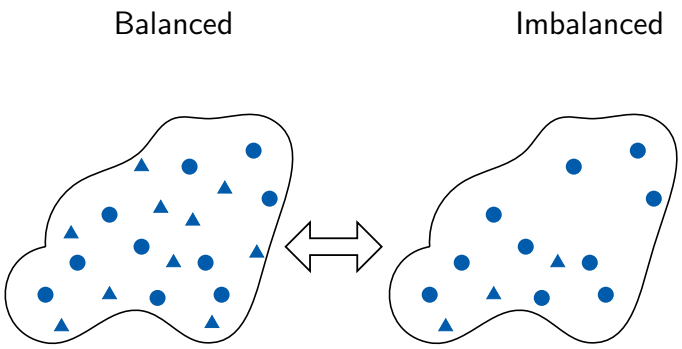


Figure 2.6. Balanced training data vs. imbalanced training data. Comparison of data composition for balanced and imbalanced training sets.

The proposed algorithms for classification tasks in sleep EEG ranges from simple dicriminant analysis to complex deep learning methods. The following paragraphs include short summaries of available supervised classification algorithms.

Linear Discrimant Analysis (LDA)

The linear discriminant analysis (LDA) algorithm is a statistical classification method that exploits the probability functions of all classes to separate unseen data. The probability density functions are estimated as multivariate normal distribution (Garrett *et al.*, 2003). The LDA method assumes that the covariance of all classes is the same, only the means vary between the classes. In the case of the quadratic discriminant analysis method, the covariances and the means differ. The observations are divided into different classes by a (D-1)-dimensional hyperplane with D representing the dimension of the input (Bishop, 2006).

k-Nearest Neighbours (kNN)

K-nearest neighbour (kNN) is a simple classification algorithm that classifies new observations according to the majority class of the surrounding *k*-nearest neighbours (Keller *et al.*, 1985). Euclidean distance is often selected as distance measure to obtain the nearest neighbours for each data point (Lorena *et al.*, 2011). The number of nearest neighbours *k* can be considered as a parameter for the degree of smoothing as a greater number of neighbours will result in fewer larger regions (Bishop, 2006). In case *k* is an even number, the distances of the sample to each neighbour are considered as tie-breaker (Keller *et al.*, 1985).

Support Vector Machine (SVM)

In machine learning, support vector machine (SVM) represents a complex but accurate algorithm in case the data fails to be linearly separable in base feature space (Singh *et al.*, 2016). To solve this problem, SVMs deploy kernel functions to transform the data to a space of higher dimension where the classes become linearly separable (Lorena *et al.*, 2011). After the transformation, SVMs aim to maximize the margin by minimizing the distance between the separating hyperplane and the nearest samples (Burges, 1998). The main limitations of SVMs is that they were originally formulated for two classes (Bishop, 2006). Several models to extend SVMs to multi-class problems have been proposed but additional parameters and constraints impose new challenges and can result in a large optimization problem or overfitting and underfitting, respectively (Aly, 2005).

Neural Network (NN)

Artificial neural network (ANN) are mapping architectures that assign an input batch of the dimension *n* to an output of the dimension *m* using sequential layers of neurons (Hecht-Nielsen, 1989). Neurons as single processing units are non-linear functions of a linear combination of the inputs using adaptive parameters (Bishop, 2006). The parameters for each neuron, the weights and biases, are updated through a backpropagation algorithm that aims to minimize the error between the predicted output and the target output. Multi-layer Perceptron ANNs are ANNs with multiple layers where neurons from neighbour layers are connected to each other.

Recurrent Neural Network (RNN)

RNNs are a class of ANN designed to learn sequential or time-varying patterns (Medsker and Jain, 2001). They contain stacked neural networks with a closed feedback loop to exploit current information as well as information from the past. Consequently, the state of the current time step referred to as hidden state depends on all previous hidden states (Graves and Schmidhuber, 2005). In 1997, Hochreiter and Schmidhuber (1997) introduced LSTM networks to resolve long-term dependencies that occur as a result of multiple stacked neural networks. With the use of memory cells, LSTMs are capable of deciding which information is relevant to control the information flow between the stacked neural networks.

Convolutional Neural Network (CNN)

Convolutional neural network (CNN) is a subclass of ANN designed to process multidimensional input data such as colour images (LeCun *et al.*, 2015). The architecture of CNNs is rooted in three main ideas: local receptive fields, shared weights, and spatial or temporal subsampling (Lecun *et al.*, 1998). A convolutional layer is regarded as a feature map containing units that are connected to a small subregion of the image or to local patches in the feature maps of the previous layer (LeCun *et al.*, 2015). All units in a feature map share the same set of weights called filter bank (LeCun *et al.*, 2015). Following a convolution layer is a pooling layer that performs subsampling on a small receptive field in the preceding convolution layer (Bishop, 2006). The advantage of CNNs as compared to ANNs is their ability of extracting local features of subregions to exploit the key property of images that nearby pixels share more information with each other than distant pixels (Bishop, 2006).

2.2.2 Automated EEG analysis applications

CAP scoring is mainly performed on sleep EEG signals as CAP describes periodically recurring brain activity rhythms during sleep. Other classification methods that require the same signal information are sleep staging and seizure detection systems. Hence, the majority of the systems in these fields share similar signal processing and classification methods. Here, a short review of automated sleep staging, seizure detection, and CAP detection methods is presented.

Sleep staging methods

With the introduction of the sleep scoring rules of Rechtschaffen and Kales, first automated digital sleep scoring methods were developed to reduce the time to score sleep and to potentially exploit new quantitative measures that are not covered in the Rechtschaffen and Kales guideline (Penzel and Conradt, 2000). Since then various systems were proposed differing in dataset, recording setup such as sampling rate, deployed PSG channels, preprocessing methods such as artefact removal, feature selection, and classification methods. Here, the focus is on systems utilizing PSG channels (EEG, EOG, and EMG) that are traditionally used for visual sleep scoring.

First semi-automated classification systems were initially developed on data obtained from animals such as rats and cats (Lim and Winters, 1980; Witting *et al.*, 1996; Itowi *et al.*, 1990). Simultaneously, hybrid sleep staging programs tested on human sleep data were proposed (Gaillard and Tissot, 1973). Later, statistical pattern recognition techniques such as interval histogram (Kuwahara *et al.*, 1988) or stochastic methods (Stanus *et al.*, 1987) and expert systems with a set of decision rules defined by sleep experts (Ray *et al.*, 1986) were deployed for automatic sleep stage scoring. With the surge in machine learning algorithms, first scoring systems based on ANNs were introduced (Schaltenbrand *et al.*, 1993; Grözinger *et al.*, 1995; Principe and Tome, 1989). Other applications investigated the ability to score sleep stages using machine learning algorithms such as SVM (Čić *et al.*, 2013), random forest (Memar and Faradji, 2018), and K-means clustering (Diykh *et al.*, 2016).

State-of-the-art sleep scoring methods exploit the classification power of deep learning methods such as CNNs and RNNs. Supratak *et al.* (2017) proposed a highly accurate model (accuracy of 86%) called DeepSleepNet combining CNNs with bidirectional LSTM that automatically learns significant features from different raw single-channel EEG without the aid of handcrafted features. Other proposed methods based on single-channel EEG use

CNNs in combination with deep ANNs (Tsinalis *et al.*, 2016) or multitaper spectral images as input (Vilamala *et al.*, 2017) achieving an accuracy of 71–76% and 84–88%, respectively. Chambon *et al.* (2018b) implemented a sleep stage classification system exploiting information from multiple EEG, EOG, and EMG channels by extracting spatio-temporal distributions using a time distributed CNN. In their analysis on automatic sleep stage scoring based on multivariate PSG data, Stephansen *et al.* (2018) applied data from approximately 3,000 recordings on LSTMs that were fed with features computed by CNNs from each EEG, EOG, and EMG. To target the the issues of low data variability and low data efficiency, deep transfer learning approaches such as finetuning (Phan *et al.*, 2020), domain adaptation (Chambon *et al.*, 2018a), or ensemble-based approaches (Alvarez-Estevez and Fernández-Varela, 2020) were proposed. Novel developments in automatic sleep stage scoring involve sequence-to-sequence classification that support the labelling of multiple epochs at once by sharing information between the epochs (Phan *et al.*, 2019; Seo *et al.*, 2020).

In summary, the development of automated sleep staging systems has progressed drastically in recent years. Considering the average sleep stage agreement of human scorers can range up to 85% (Rosenberg and Van Hout, 2021; Stephansen *et al.*, 2018), some advancements in automated sleep staging are able to outperform human scorers (Stephansen *et al.*, 2018). Deep learning models such as the models proposed by Supratak *et al.* (2017), Vilamala *et al.* (2017), Stephansen *et al.* (2018), and Phan *et al.* (2019) can achieve an agreement significantly higher than the inter-scorer reliability of human scorers. Also, the maturity of sleep staging systems is reflected in the large number of commercial products provided by PSG hardware and software producers. However, automated scoring systems are not yet widely used in clinical settings because of the aversion to technology, the user-unfriendliness, technical limitations, security and privacy issues, and the diverging interpretation of scoring rules (Fiorillo *et al.*, 2019).

Seizure detection

Automatic seizure detection is a valuable tool in treatment optimisation of epilepsy patients. Commonly, seizures are detected and tracked by the patient or the patient's family resulting in a lower accuracy due to inaccuracies in identifying or recalling an event (Ulate-Campos *et al.*, 2016). Seizure detection devices can provide a more accurate seizure quantification whereas seizure prediction devices can improve the quality-of-life by alerting

the patient of an upcoming seizure (Ulate-Campos *et al.*, 2016). One type of seizure detection devices uses information from EEG signals as input, similar to sleep staging and CAP detection systems. Other sensor types exploit information from surface EMG, electrodermal activity, ECG, or accelerometry (Ulate-Campos *et al.*, 2016).

An immense part of research on seizure detection in EEG is dedicated to the spatial set-up of electrodes and the development of wireless EEG measurement tools to enhance a patient's comfort and to decrease artefacts based on poor cable management. On the other hand, a part of research on seizure detection explores the application of machine learning methods and seeks to optimise feature selection. Recent studies examined the application of statistical features (Teixeira *et al.*, 2014), spectral features such as sub-band spectral power (Bandarabadi *et al.*, 2015), spatio-temporal phase correlation features (Parvez and Paul, 2016), 2-dimensional representation of short-time Fourier transform results (Cao *et al.*, 2017), or 2-dimensional images of raw EEG (Wei *et al.*, 2018). Moreover, various deep learning models have been investigated for seizure detection, such as fully connected neural network (FCNN) (Jang and Cho, 2019), CNN (Wei *et al.*, 2018), and RNN (Hussein *et al.*, 2019).

CAP detection algorithms

In terms of CAP scoring, the development of automated systems to automatically evaluate sleep microstructure started early after the release of the guidelines. Scientists at the University of Lisbon and at the University of Parma initiated the pursuit of an automated CAP scoring methods. First methods focused on finding spectral features to describe the sleep microstructure (Rosa *et al.*, 1999; Barcaro *et al.*, 1998; Ferri *et al.*, 2005b). Subsequently, the computed descriptors were passed to threshold detectors that were able to identify potential A-phases (Rosa *et al.*, 1999; Navona *et al.*, 2002; Barcaro *et al.*, 2004). Other decision making methods included fuzzy logic (Rosa and Allen, 1996) and maximum likelihood estimation (Lima and Rosa, 1997). A different approach was proposed by Largo *et al.* (2005) using wavelet transform for time-frequency analysis and an adaptive reservoir genetic algorithm for decision finding. An optimized version of the wavelet based scoring system was later implemented in the Somnologica Science 3.3 program as the product of combined efforts from the bioengineering department at the University of Lisbon (Portugal) and the Flaga-Embla research and development team (Reykjavik, Iceland) (Guilleminault *et al.*, 2006a).

The majority of automated CAP detection systems rely on handcrafted features selected by experts or validated in previous studies. Among the variety of features are energy features such as Teager energy operator (Machado *et al.*, 2015) or band power descriptors (Mariani *et al.*, 2011b), statistical features such as skewness or kurtosis (Mendez *et al.*, 2016), amplitude-based features such as Hjorth activity (Mariani *et al.*, 2011b) or similarity index (Niknazar *et al.*, 2015), spectral features such as spectral entropy (Karimzadeh *et al.*, 2015) or discrete Fourier transform (Machado *et al.*, 2018), and information theoretical features such as Lempel-Ziv complexity or sample entropy (Mendez *et al.*, 2016). Simultaneous to the trend in automatic sleep stage scoring, first automated systems to score CAP using ANNs were implemented shortly after the turn of the millennium (Mariani *et al.*, 2010). In their study on efficient automatic A-phase classifiers, Mariani *et al.* (2012) concluded that LDA and ANN are among the better-performing algorithms for CAP classification. Later, they showed that the computation of descriptors on variable window length resulting in signal excerpts with uniform spectral characteristics is a useful tool to increase CAP classification accuracy (Mariani *et al.*, 2013). Other machine learning algorithms such as kNN (Mendez *et al.*, 2016), SVM (Mariani *et al.*, 2011a), self-organizing map (Mendonça *et al.*, 2018a), and classification tree (Mendonça *et al.*, 2018a) were also successfully applied to identify A-phases.

Additionally, studies were conducted to assess the inter-scorer reliability in scoring CAP parameters as benchmark for automatic CAP scoring. Ferri *et al.* (2005b) reported in their study on four human scorers a Kendall W coefficient of concordance of 0.83 for CAP rate and 0.68 for the number of CAP sequences. Largo *et al.* (2019) showed a mutual agreement of approximately 70% for A-phase scoring. State-of-the-art methods show that they are able to score CAP events with a similar agreement as visual scorers. Mendonça *et al.* (2020a) demonstrated that their probabilistic model using Gaussian Mixture Model and symbolic dynamics reached a CAP cycle detection accuracy of 76% in fifteen normal subjects. Another method proposed by Mendonça *et al.* (2020b) using Matrix of Lags scored a CAP cycle detection accuracy of 77% in a population of nine healthy subjects and 4 subjects with sleep-disordered breathing. Using three layers of LSTM, Mendonça *et al.* (2020c) demonstrated in their next study on 15 normal subjects an overall average accuracy of 82% using features and 81% using pre-processed EEG signal for A-phase subtype classification during NREM, REM, and Wake phases. In their study on A-phase classification, Arce-Santana *et al.* (2020) applied a CNN-based classifier on log-spectrograms of single-channel EEG data achieving an accuracy of 88.1% for binary classification and 77.3% for subtype classification on nine healthy subjects. Dhok *et al.* (2020) reported in their study an average accuracy of 87.5%

on 2-seconds EEG segments from six normal subjects using Wigner-Ville based entropy features.

Another approach to classify CAP is the cardiopulmonary coupling analysis proposed by (Thomas *et al.*, 2005). In their first study on CAP estimation using ECG and cardiopulmonary coupling, Mendonça *et al.* (2018b) reported an accuracy of 77% using an ANN and a deep stacked autoencoder. In follow-up studies, the same research group demonstrated an accuracy of 91% (Mendonça *et al.*, 2021) and 94% (Mendonça *et al.*, 2020d) in relation to age-related CAP rate percentages using a CNN-based estimator. An additional feature of the algorithm was it implementation in a home monitoring device composed of a processing unit, a sensing module, and a display unit. This approach is similar to a previously suggested portable solutions for CAP analysis from the same authors (Mendonça *et al.*, 2019).

Limitations in previous work on automated CAP detection

The presented literature on automated CAP detection algorithms demonstrates several limitations. Firstly, all supervised classifiers were designed to be trained with a balanced dataset, i.e. a balanced number of representations for each class. However, the amount of CAP sequences or A-phases in a normal sleep recording is substantially smaller than the number of periods without CAP. Hence, a large number of background periods or periods without CAP need to be removed from the training dataset to achieve a balanced representation. Consequently, important information is lost when using balanced datasets. Moreover, performance metrics need to be adjusted to account for the bias that is introduced by imbalanced data. In summary, the issue of imbalanced data in CAP scoring has not yet been satisfactorily addressed.

Although a few deep learning models have been proposed in recent years, the full range of possibilities using deep learning has not yet been explored. Considering EEG channels are time signals that represent the brain activity at a certain location at a specific point in time, classification models that exploit the given information from the past and the present have access to more relevant information than other classifiers. In theory, this will result in a more accurate and precise classification decision. One approach to exploit past information in the classification are RNNs. They have rarely been part of previous CAP detection studies. Hence, the full capability of deep learning methods for automated CAP detection has not been exploited yet.

Various studies on CAP detection decided to score CAP sequences instead of single A-phases and to include REM epochs into their analysis. Both approaches may limit their utility in a clinical setting. Although the scoring of CAP sequences features a more inclusive approach and is preferable in real-time applications, the detection of single A-phases is more clinically relevant as it offers more information about the characteristics of A-phases, B-phases, and CAP cycles. The majority of systems developed by or with the support of leading clinical experts in this field are based on the detection of single A-phases to adhere to the rules of CAP sequences. This underlines the aspect that single A-phase detection is preferable in a clinical context. Regarding the inclusion of REM epochs into CAP detection, the CAP atlas (Terzano *et al.*, 2001) categorically excludes REM epochs from CAP scoring. Moreover, the lack of CAP scored REM epochs prevents the training of CAP detection classifier on REM data. Consequently, the scoring of CAP sequences instead of single A-phases and the inclusion of REM epochs prevent the comparison of performance measures between studies. Hence, this results in increased complexity when comparing methods for CAP detection and a more generalised approach is preferable.

2.3 Key Questions

Figure 2.7 summarizes the key questions addressed in this thesis. The agenda of this thesis was divided into four main work packages. Firstly, key issues regarding the automated scoring of CAP sequences were addressed. In this study, we propose a novel, automated system that addresses two main limitations of previous works - the bias of imbalanced data and the non-generalised approach. Our end-to-end solution uses a RNN to extract crucial information in the temporal behaviour of CAP. Also, state-of-the-art signal processing methods were implemented to reduce the influence of cardiac field artefact (CFA) and the loss function of the training process was optimised to deal with the biasing issue of imbalanced datasets.

Following, the proposed automated CAP scoring was validated using a long-standing cohort study by comparing the scoring output to values from the literature on the relation between CAP and markers of sleep disturbances. To the best of our knowledge, this is the first time that such a comparison was performed. It is also the first time that a study of this dimension (more than 2,000 participants) was analysed in terms of CAP. The study also demonstrates that automated CAP analysis of large population based cohorts can lead to new findings on CAP and its subcomponents. Afterwards, a study is presented that addresses the gaps

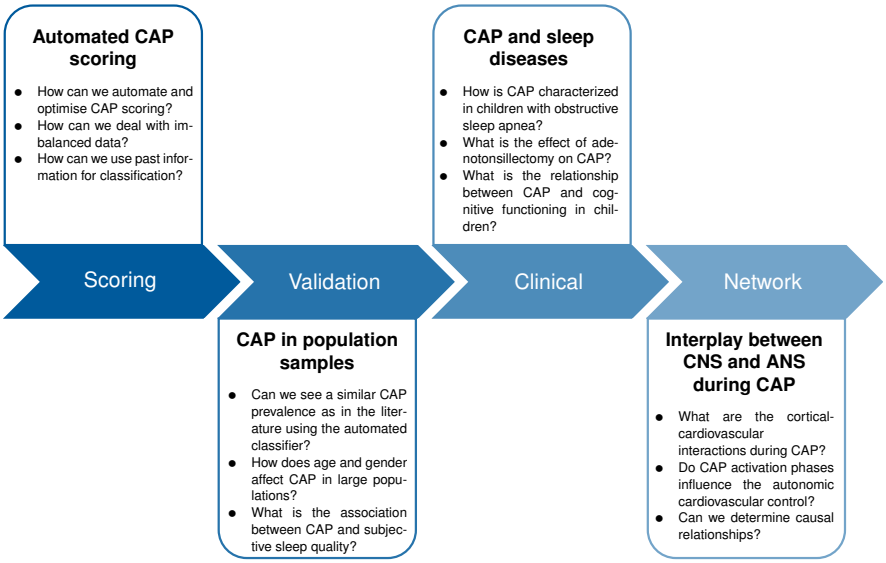


Figure 2.7. Summary of key questions addressed in this thesis. An overview of the key questions addressed in this thesis depicted as a flow chart.

of knowledge in the relation between CAP and cognitive functioning in children. Using the validated CAP classification system, the relationship between CAP and behavioural, cognitive, and quality-of-life measures as well as the effect of eAT on CAP are determined in children with OSA.

Finally, novel investigations into the interplay between CNS and ANS during CAP are conducted. As mentioned in Section 2.1.4, the relation between CAP and brain connectivity and the ANS, respectively, is not fully understood yet. In this study, we analyse a tool called GC based on its utility for probing the level of interplay between CNS and ANS during CAP. Our analysis provides first evidence on the causal interplay between cortical and cardiovascular activities during CAP.

Paper I

3

Automatic A-phase Detection of Cyclic Alternating Patterns in Sleep Using Dynamic Temporal Information

The content of this chapter is a modified version of the publication:

Hartmann, S. and Baumert, M. (2019), 'Automatic A-Phase detection of cyclic alternating patterns in sleep using dynamic temporal information', *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **27**(9), pp. 1695–1703.

Abstract

The identification of recurrent, transient perturbations in brain activity during sleep, so called cyclic alternating patterns (CAP), is of significant interest as they have been linked to neurological pathologies. CAP sequences comprise multiple, consecutive cycles of phasic activation (A-phases). Here, we propose a novel, automated system exploiting the dynamical, temporal information in electroencephalography (EEG) recordings for the classification of A-phases and their subtypes. Using recurrent neural networks (RNN), crucial information in the temporal behaviour of the EEG is extracted. The automatic classification system is equipped to deal with the biasing issue of imbalanced data sets and uses state-of-the-art signal processing methods to reduce inter-subject variation. To evaluate our system, we applied recordings of 16 normal, healthy subjects (aged 32.2 $\pm$ 5.4 years) and 30 subjects with nocturnal frontal lobe epilepsy (aged 31.0 $\pm$ 11.4 years) from the publicly available CAP Sleep Database on PhysioNet. Our results show that the RNN improved the detection accuracy by 3–5% and the F_1-score by approximately 7% on two data sets compared to a normal feed-forward neural network. Our system achieves a sensitivity of approximately 76–78% and F_1-score between 63–68%, significantly outperforming existing technologies. Moreover, its sensitivity for subtype classification of 60–63% (A1), 42–45% (A2), and 71–74% (A3) indicates superior multi-class classification performance for CAP detection. In conclusion, we have developed a fully automated high performance CAP scoring system that includes A-phase subtype classification. RNN classifiers yield a significant improvement in accuracy and sensitivity compared to previously proposed systems.

3.1 Introduction

Sleep is as an essential part of life for many species including humans. It comprises recurring alternating patterns of quiescence followed by high activity (rapid eye movement (REM) sleep) that appear to serve several vital functions including cellular restoration, memory consolidation, and brain clearance from metabolites, but the process in its entirety is incompletely understood.

To discern its macrostructure, sleep is classified in four different stages based on the American Academy of Sleep Medicine consensus guidelines (REM plus three non-REM (NREM) stages) (Iber *et al.*, 2007), which originated in the rules proposed by Rechtschaffen & Kales (Rechtschaffen and Kales, 1968). The classification is typically performed on thirty-seconds windows of the electroencephalography recording. Every window is scored by a trained expert into one of the four stages based on the prominent EEG pattern.

In 2001, the concept of cyclic alternating pattern (CAP) was introduced as an alternative way to characterise NREM sleep (Terzano *et al.*, 2001). A CAP sequence is defined as three or more consecutive activation phases (A-phases) which represent transient, prominent events separated by a B-phase, the so called background phase (see Figure 3.1). The A-phase is commonly restricted to sleep stages without rapid eye movement and characterised by slower high-voltage rhythms, faster lower-voltage rhythms or by both (Terzano and Parrino, 2005). Based on the respective proportion of the aforementioned waveforms, A-phases can be divided into three subtypes. Subtype A1 is dominated by EEG synchrony, i.e. high-voltage slow waves, whereas faster low-amplitude rhythms are more prevalent in subtype A3 (Terzano *et al.*, 2001). Subtype A2 contains a mixture of both waveforms. In general, CAP serves as indicator for sleep instability due to its correlation with sleep pathologies such as sleep disordered breathing or insomnia (Parrino *et al.*, 2012). A CAP reflects a stimulating character on physiological functions during the transition from wakefulness to sleep or during light sleep (Terzano and Parrino, 2000).

Currently, the scoring of CAP events is performed semi-manually by a trained expert using a software as visualisation tool to streamline A-phase detection. As this task can be exhausting and time-consuming, a robust, fully automated system for scoring A-phases could tremendously accelerate research in the field and pave the way towards clinical application.

In the past, a few studies were performed to implement automatic CAP detection, where the distinctive characteristics between the activation phase and the background phase were exploited to score CAP sequences (Mendez *et al.*, 2016). In early works, the alteration in signal amplitude averages between short and long time periods were conducted by various models mixed with further analysis (Navona *et al.*, 2002). Later, statistical or spectral features were extracted from EEG followed by either thresholding classification algorithms (Niknazar *et al.*, 2015) or competitive machine learning algorithms (Mendez *et al.*, 2016; Mariani *et al.*, 2012; Largo *et al.*, 2005; Mariani *et al.*, 2013; Mendonça *et al.*, 2018a). In terms of subtype recognition, research initially focused on the unique attributes of each subtype (Mendez *et al.*, 2016; Navona *et al.*, 2002). Only a few systems classify subtypes as

a whole in sleep recordings (Machado *et al.*, 2018). Importantly, all aforementioned classifiers rely solely on the information provided at a single time step, neglecting supplementary knowledge that may be contained in consecutive time steps.

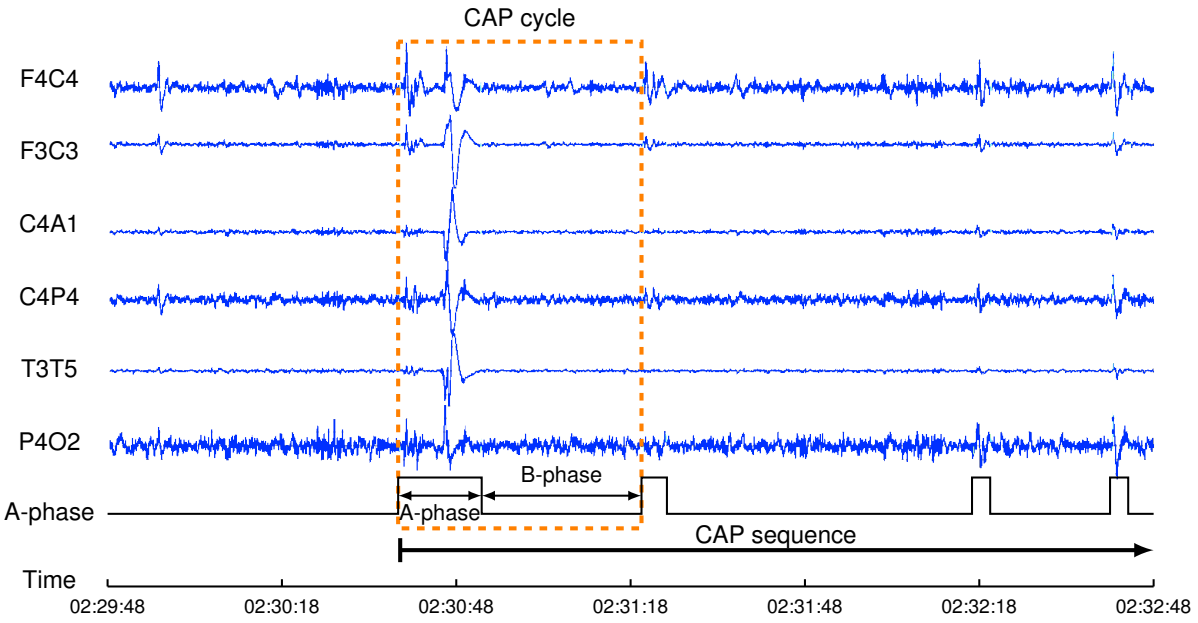


Figure 3.1. Short excerpt of a typical cyclic alternating pattern sequence. Multiple electroencephalography channels illustrating a cyclic alternating pattern (CAP) sequence which consists per definition of more than two CAP cycles (one CAP cycle is highlighted in the red square). A CAP cycle is composed of an activation phase (A-phase) and a background phase (B-phase). The A-phase in the highlighted time period displays a slow high-voltage rhythm compared to the B-phase which is typical for a certain type of CAP events.

Here, we propose a fully automated CAP scoring system, including comprehensive preprocessing and artefact removal stages, as well as feature selection that uses a recurrent neural network (RNN) classifier exploiting the dynamic temporal information in EEG. We selected long-short term memory network (LSTM) as representative for the RNN class. Quantifying the temporal EEG behaviour can yield more precise scoring of conventional sleep stages, see Dong *et al.* (2018) and Supratak *et al.* (2017). Additionally, the scoring process was expanded to a multi-class classification system determining the subtype of the detected A-phase. Section 3.2 explains the proposed method, the test environment and the entire automated system. The results obtained with different classifiers plus the comparison of performance measures for general A-phase detection and subtype classification are presented in Section 3.3. Finally, Section 3.4 discusses the results and limitations of the system followed by concluding remarks in Section 3.5.

Table 3.1. Statistics of sleep macrostructure and cyclic alternating pattern occurrence in data sets (total duration in seconds).

Subject	Wake & REM	NREM	A1	A2	A3	Total sleep time
16 normal subjects	153,890	328,040	24,027	10,633	16,062	481,930
30 NFLE subjects	270,840	627,631	50,791	27,754	51,397	898,471

REM, rapid eye movement; NREM, non-rapid eye movement; NFLE, nocturnal frontal lobe epilepsy

3.2 Materials and methods

This section explains the data flow from a raw EEG recording as input to the classification output. The entire system is divided into four major parts: pre-processing, feature extraction, classification and post-processing. In the pre-processing step, the raw data are prepared for the feature extraction stage by removing cardiac field artefacts and eye movement artefacts plus filtering. Based on the processed signal, multiple features in the time and frequency domain are calculated. The extracted feature set serves as input for the classifier in the classification stage. Finally, the output of the classifier is modified according to the CAP scoring rules.

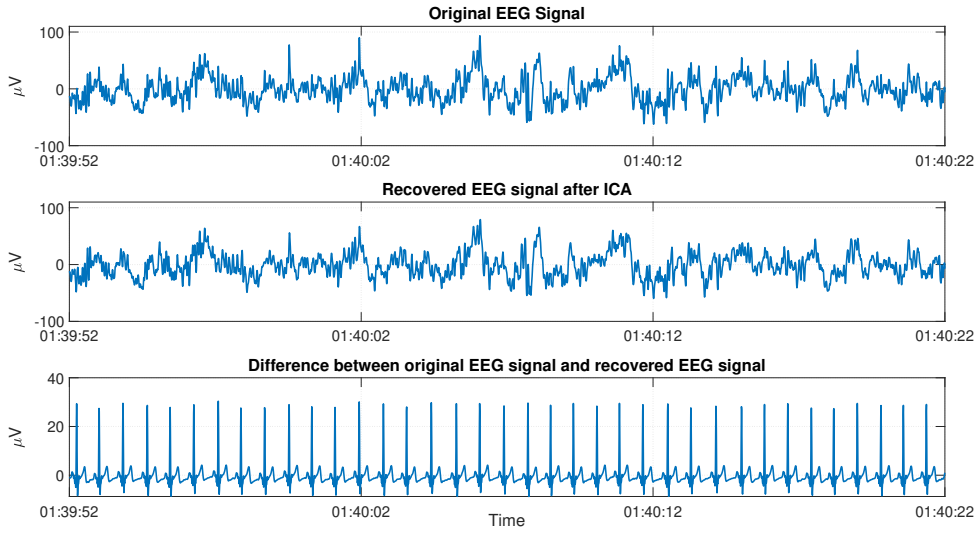


Figure 3.2. Illustration of the input and outcome of the cardiac field artefact removal method. Depiction of cardiac field artefact removal method in short signal window of subject n10: a) Contaminated original signal, b) recovered electroencephalography (EEG) signal from independent sources after independent component analysis, and c) difference between original and recovered EEG signal (superimposed electrocardiography signal on EEG channel).

3.2.1 CAP Sleep Database

We used the publicly available CAP Sleep Database on PhysioNet, an open-source repository for physiological signal recordings targeting various biomedical research fields (Goldberger *et al.*, 2000). The polysomnographic measurements were conducted by the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy. In particular, we selected 16 normal, healthy subjects (n1–n16) and 30 subjects with nocturnal frontal lobe epilepsy (NFLE) to benchmark our results against Mariani *et al.* (2013) and Machado *et al.* (2018). Subject n16 was included although the electrocardiography (ECG) recording is missing, which is necessary for the cardiac field artefact (CFA) removal explained in section 3.2.2. Usually the polysomnographic recordings contain at least one EEG channel (C3 or C4), multiple bipolar EEG channels and other parameter such as ECG or eye movement signals. For each recording an annotation file is provided containing manual scoring performed by expert neurologists. The scoring comprises sleep stages and CAP events according to the Rechtschaffen & Kales rules (Rechtschaffen and Kales, 1968) and the atlas of CAP scoring (Terzano *et al.*, 2001), respectively. The manual scoring serves as ground truth for the supervised learning of the event classifiers.

For algorithm evaluation, we selected the leave-one-out (LOO) method equivalent to Mariani *et al.* (2013) and Machado *et al.* (2018) as cross-validation approach. The LOO method is a k-fold cross-validation algorithm in which for each fold one subject is determined as test set and all remaining subjects are merged into the training set. The final performance measures are computed by summing up each individual validation value and dividing the sum by the number of instances. Summary statistics of the sleep macrostructure and CAP occurrence for each data set are listed in Table 3.1. Since CAP events only occur in NREM stages (Terzano *et al.*, 2001), the wake and REM periods were removed from the data. Thus, the numbers for NREM seconds in Table 3.1 represent the quantity of used samples for each data set.

3.2.2 EEG pre-processing

To guarantee a robust and precise classification output, the recordings of each subject have to be processed in advance to diminish any subject related variations. As signal channel, we selected a single EEG lead in each subject's set of recordings, which was either the C4-A1 or the C3-A2 channel. The sampling frequency varied across recordings between

100 Hz and 512 Hz, thus the data were resampled at a frequency of 128 Hz. Subsequently, the recordings were examined for CFA or eye movement artefact.

Cardiac field artefact removal

Typically, the cardiac electric field is superimposed on the cortical electric field resulting in more or less clearly visible peaks in EEG. To disentangle both sources in EEG, blind source separation (BSS) methods such as independent component analysis (ICA) can be applied. Here, we used a modified Wavelet-ICA method presented previously for the source separation of single-channel recordings (Mijovic *et al.*, 2010). The Wavelet-ICA combines the discrete wavelet transform, which generates an array of frequency sub-bands out of a 1-D signal, where the ICA method is used to find statistically independent sources in the wavelet scales. As mother wavelet, the coiflet wavelet was chosen due to its resemblance of the pulse waveform in ECG. Because of the mutual sampling frequency of 128 Hz, we selected a 6-level decomposition. To achieve good performance, during ICA, the recorded ECG is included as an input, adding to the existent cardiac field artefacts. After ICA, each independent component was recovered with the mixing matrix and correlated with the original ECG to identify the independent cardiac field source in the output signals. The CFA removal procedure was completed by determining the Pearson correlation coefficient between the ECG and the difference between EEG and the recovered signal of the non-ECG independent sources. If the coefficient value exceeded 0.75, the EEG was replaced by the recovered signal. Otherwise the algorithm proceeded without any cardiac field removal. An example of a contaminated signal and a denoised signal are shown in Figure 3.2.

Eye movement artefacts

The movement of the eyeballs can create additional noise in sleep EEG. Due to proximity, the electrooculogram (EOG), representing the eye movement, can be often found as an independent source in the EEG recordings. Thus, the aforementioned method for CFA removal is applied similarly to deal with eye movement artefacts. The EOG signal is added to the decomposed EEG signal before the ICA step. As a result, a set of independent sources is obtained containing signals free from any eye movement and a single eye movement signal.

Filtering

Before extracting signal features, the EEG was bandpass filtered with a finite impulse response (FIR) filter (0.5–30 Hz) and subsequently divided into five frequency bands. Using a least-squares linear phase FIR filter bank to separate the following EEG rhythms: *Delta* (0.5–4 Hz), *Theta* (4–8 Hz), *Alpha* (8–12 Hz), *Sigma* (12–16 Hz) and *Beta* (16–30 Hz).

3.2.3 Feature extraction

The following features were selected based on their performance reported in previous studies (Machado *et al.*, 2015; Karimzadeh *et al.*, 2015; Mariani *et al.*, 2011b). Every feature is calculated on a defined window length with partial overlapping. All features are centred on the current second resulting in a sample rate of 1 Hz.

Hjorth activity

The Hjorth activity is defined as the variance of the signal amplitude based on the discrete signal values s_i (Hjorth, 1970),

$$m_0 = \frac{1}{M}\sum_{i=1}^{M}(s_i - \mu)^2 \tag{3.1}$$

where M is the window length and $\mu = \frac{1}{M}\sum_{i=1}^{M} s_i$. In this study, m_0 was only determined in the delta band by using 3-s overlapped windows.

Shannon entropy

Given the probability p_i with i representing all amplitude values, the Shannon entropy is defined in this research as (Karimzadeh *et al.*, 2015):

$$H = -\sum_{i} p_i \log p_i. \tag{3.2}$$

The Shannon entropy was determined on the entire spectrum in 2-s signal windows with one second overlap.

Teager Energy Operator

The Teager Energy Operator was introduced for the instantaneous estimation of frequency and amplitude components (Cho *et al.*, 2014). The discrete version of the TEO, $\Psi[x[n]]$, is defined as follows, where $s[n]$ corresponds to a discrete sample of a time series:

$$\Psi[s[n]] = (s[n])^2 - s[n-1]s[n+1]. \tag{3.3}$$

The TEO feature was calculated for every spectral band in 2-s overlapping windows.

Band power descriptor

To evaluate the variation of power in different frequency bands, band descriptors were introduced in previous studies highlighting the transient spectral variations in a temporal range of 2–60 seconds (Mariani *et al.*, 2011b). At first, the signal in each band was squared and normalized with respect to the maximum power of the band. Afterwards, the mean power on windows of 2 seconds and 64 seconds were determined. The descriptor formula is shown in equation 3.4 (Mariani *et al.*, 2011b),

$$d_b(t) = \frac{e_s(t) - e_l(t)}{e_l(t)} \tag{3.4}$$

where $e_s(t)$ is the mean power of the short 2-s window, $e_l(t)$ the mean power of the long 64-s window and $d_b(t)$ the power descriptor for a specific frequency sub-band.

Differential EEG variance

The variance difference of the EEG was determined on 1-s non-overlapping windows calculating the difference between consecutive 1-s windows (Mariani *et al.*, 2011b).

Feature synchronisation

Prior to feature extraction, the raw EEG and the scoring data were synchronised to ensure perfectly aligned data. The starting time of the manual scoring was used as a reference for both inputs. In case the EEG recording started earlier than manual scoring, the EEG data prior to the scoring start was removed. In case the EEG recording started later than manual scoring, the entire recording was dismissed. After feature extraction, Wake and REM stages were removed. Hence, the first set of features describes the first second of the first manually scored NREM stage.

3.2.4 Classification algorithms

In this study, a RNN was designed to exploit the dynamical temporal information for the automatic A-phase detection of CAP cycles. We compared its performance to a conventional feed-forward neural network (NN). Before all classifiers were trained, the data were normalized to zero mean and unit variance. In case of the test sets, the mean and standard deviation of the training set was employed due to the lack of knowledge of the statistics of the sample. For A-phase detection, a binary output was used whereas four different output classes were predicted in the subtype classification case.

Artificial neural network - ANN

A shallow feed-forward NN with one hidden layer was implemented as standard artificial neural network. A feed-forward NN is based on consecutive functional transformations, which can exploit non-linearities by applying non-linear activation functions to the weighted inputs of each neuron (Bishop, 2006). The parameters of each neuron (weights and biases) are updated through backpropagation after comparing the outcome of the network to the labelled data with a defined loss function. For this study, the number of neurons of the single-layer feed-forward NN was fixed to 448 and a rectified linear unit (ReLU) was used as activation function. Furthermore, the total loss was calculated with the F_β-score (see Section 3.2.4).

Long-short term memory network - LSTM

A RNN is a subclass of NNs and differs from a feed-forward NN by passing the information from previous states to the next cell. Therefore, the state of the current time step referred to as hidden state depends on all previous hidden states (Graves *et al.*, 2013). This characteristic makes RNNs very useful for time series predictions such as speech recognition (Graves and Schmidhuber, 2005) due to the exploitation of the dynamic temporal behaviour However, a key problem of initial RNNs is the learning of long-term dependencies, which can lead to an increasingly difficult problem for the gradient based learning algorithm (Bengio *et al.*, 1994). To resolve long-term dependencies, LSTM RNN were proposed (Hochreiter and Schmidhuber, 1997). Instead of having a chain of feed-forward networks connected together, the cells of a LSTM provide the ability to decide which information is relevant. So called *memory cells* control and store information, see Figure 3.3. The memory cell is composed of the *input gate*, *forget gate*, *output gate*, and the *cell state*.

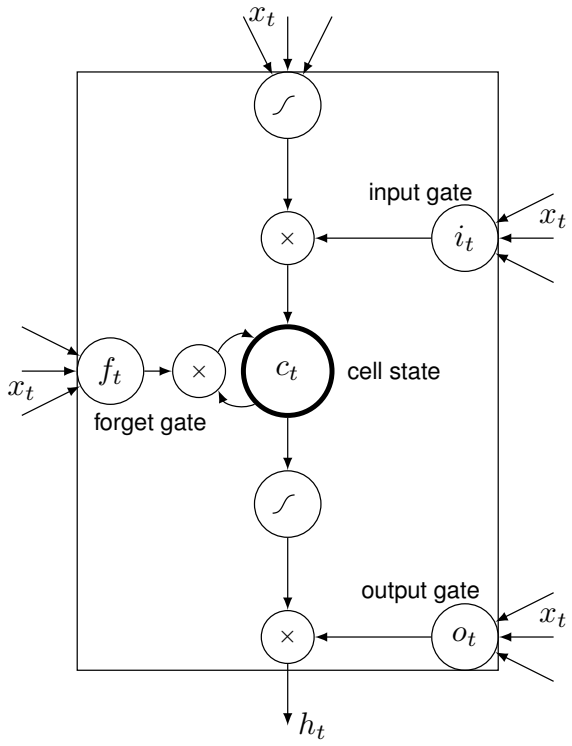


Figure 3.3. Schematic of a standard long-short term memory cell. Schematic of a standard long-short term memory cell highlighting the input gate, the forget gate, the cell state, and the output gate.

Each gate is controlling the information flow, hence all gates manipulate the current hidden state. The gates can be calculated by the following numerical calculations,

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{3.5a}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{3.5b}$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{3.5c}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{3.5d}$$

$$h_t = o_t \tanh(c_t) \tag{3.5e}$$

where σ is the logistic sigmoid function, $tanh$ the hyperbolic tangent, x_t the cell input, c_t the cell state, h_t the hidden state, W_{yz} the weights of gate z corresponding to the gate input y and b_z the bias of the gate z (Gers *et al.*, 2002). In this study, we create deep LSTM networks by stacking LSTM layers similar to those of feed-forward NNs. This results in connected layers with multiple cells per time step. The number of time steps, i.e. the sequence length, defines how much of the previous information is entered into the calculation of the

current state. The output of the last time step in the sequence is passed to a neural network layer with a ReLU as activation function. Finally, the class probabilities are calculated by a softmax layer following the fully connected layer. Consequently, the LSTM network and feed forward NN used in this study differ only in the LSTM layers, allowing us to evaluate the effect of the information gain of the LSTM memory.

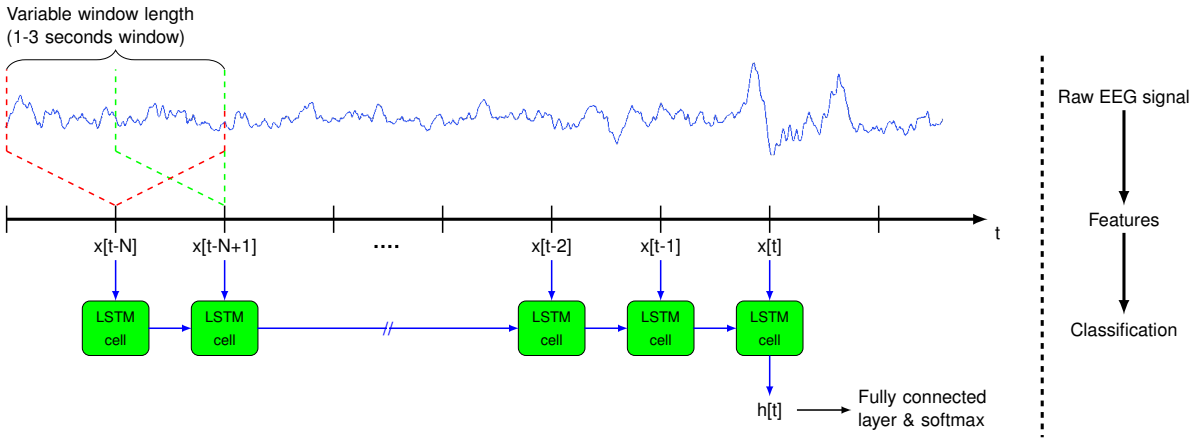


Figure 3.4. Schematic of the information flow in the proposed long-short term memory network architecture. Scheme of the information flow from a single electroencephalography trace to the classification outcome using a long-short term memory (LSTM) recurrent network: At first, features are calculated based on the raw data in short signal windows creating an input array with a feature vector for every second ($x[t]$ with $t \in \{1, 2, .., L\}$ where L is the length of the recording). For every time step, the LSTM classifier takes a sequence of N feature vectors ($x[t-N+1], ..., x[t]$) centred on the current time step where $N-1$ is the length of the desired past information. At the end, the hidden state ($h[t]$) of the current time step is linked to a fully connected layer followed by a softmax layer. Thus, for each time step the information from $N-1$ values of the past plus the current features are deployed for the classification.

F_β-score as cost function

Considering the statistics of the training set (see Table 3.1), the number of A-phases in each sleep recording is significantly lower than the number of background phases. As training of machine learning classifiers with common error functions such as the cross-entropy or means squared error relies on a balanced data set to receive an unbiased output, an imbalanced data set would result in a trained classifier that most likely chooses the class of the most prominent label to achieve a high accuracy. To avoid a biased classifier, the data set can be balanced either by removing samples of the more prominent class or by upsampling the under-represented class based on the available data. Alternatively, the training algorithm and the error function in particular could be adapted.

To train each neural network, we modified the objective error function by implementing the F_β-score as done previously for normal feed-forward NNs (Pastor-Pellicer *et al.*, 2013) and convolutional neural networks (Scherzinger *et al.*, 2017). The F_β-measure relies solely on the precision and sensitivity of the prediction. Both values are commonly used as objective quality measures for binary classification problems along with accuracy and specificity (see Section 3.2.6) (Sokolova and Lapalme, 2009). The positive β-value defines the prioritisation of precision or sensitivity. A β-value equal to 1 is commonly known as the F_1-score or F-score, determining the harmony of precision and sensitivity. The main requirement for an objective error function is the ability to calculate its gradient. Hence, equation 3.6 displays the derivation of the F_β-score, where o_i represents the output of the classifier, t_i describes the target label, and L the size of the data set (Pastor-Pellicer *et al.*, 2013).

$$\frac{\partial FM_\beta}{\partial o_i} = \frac{(1+\beta^2)\cdot t_i}{\sum_{j=1}^{L}(o_j+\beta^2\cdot t_j)} - \frac{(1+\beta^2)\cdot \sum_{j=1}^{L}(o_j\cdot t_j)}{\left[\sum_{j=1}^{L}(o_j+\beta^2\cdot t_j)\right]^2} \tag{3.6}$$

3.2.5 Classification post-processing

The outcomes of the classifiers were post-processed in accordance with the atlas for CAP scoring (Terzano *et al.*, 2001). Since A or B-phases can only occur in time periods longer than a second, isolated one-seconds prediction were replaced by their neighbouring values. Moreover, in terms of binary classification, the prediction of an A-phase longer than 60 seconds was reclassified again due to the high probability of containing multiple A-phases. For reclassification, a self-organising map with 500 epochs and unsupervised learning was used to cluster the particular time window again. These two steps were performed three times in a row to check for possible changes after the preceding iteration. This method proved effective in a previous study (Mariani *et al.*, 2013). Considering subtype classification, periods of one subtype longer than 60 seconds were split into shorter intervals. Additionally, activation phases with inconsistent labels of subtypes were set to the dominant subtype in this time period.

3.2.6 Classification performance measures

To compare the performance of the two A-phase classifiers, a set of performance measures for binary classification problems was calculated. The efficacy was evaluated based on the number of correctly identified events (true positives, t_p), the number of correctly recognized background phases (true negatives, t_n), and the number of seconds which were misidentified either as A-phase (false positive, f_p) or as background phase (false negative, f_n). Based on these, we quantified accuracy (ACC), sensitivity or true positive rate (TPR), precision or positive predictive value (PPV), true negative rate or specificity (SPC) and the F_1-score as follows (Sokolova and Lapalme, 2009):

$$\text{ACC} = \frac{t_p + t_n}{t_p + f_n + f_p + t_n} \tag{3.7a}$$

$$\text{TPR} = \frac{t_p}{t_p + f_n} \tag{3.7b}$$

$$\text{PPV} = \frac{t_p}{t_p + f_p} \tag{3.7c}$$

$$\text{F}_1 = \frac{2 \cdot TPR * PPV}{TPR + PPV} \tag{3.7d}$$

$$\text{SPC} = \frac{t_n}{f_p + t_n}. \tag{3.7e}$$

In the case of multi-class classification, the equivalent measures were computed using following formula where i represents the current class and l the total number of classes (Sokolova and Lapalme, 2009):

$$\text{ACC} = \frac{\sum_{i=1}^{l} \frac{t_p^i + t_n^i}{t_p^i + f_n^i + f_p^i + t_n^i}}{l} \tag{3.8a}$$

$$\text{TPR}^{\text{i}} = \frac{t_p^i}{t_p^i + f_n^i} \tag{3.8b}$$

$$\text{PPV}^{\text{i}} = \frac{t_p^i}{t_p^i + f_p^i} \tag{3.8c}$$

$$\text{F}_1{}^{\text{i}} = \frac{2 \cdot TPR^i * PPV^i}{TPR^i + PPV^i}. \tag{3.8d}$$

Table 3.2. List of long-short term memory network algorithm parameters.

Parameter	Value	Parameter	Value
Cell input (number of features)	13	Output classes	2
Epochs	100	Batch size	256
Learning rate	0.001	Learning rate drop period	20
Learning rate drop factor	0.1	β value	3
Dropout	No	L2 Regularization	No
Sequence length	14	Layer shape	[128, 64, 32]

3.2.7 Setup of test environment

Prior to the supervised training of the classifiers, the parameters for the LSTM algorithm were set. Table 3.2 displays the various parameters used in this study. For each NREM second a 14-seconds vector with 13 features each was generated resulting in a pool of two-dimensional arrays (see Figure 3.4). Subsequently, the pool of vectors for training was shuffled erasing any subject information. During training and testing, the cell state and hidden state of the LSTM cells were reset for each sequence to ensure that no previous information is carried on. Prior to evaluation, the vectors were not shuffled but fed sequentially to the trained system enabling to reconstruct the temporal structure of the classification output afterwards.

After training, each classifying technique was separately evaluated on both data sets and compared to the particular reference. In case of the subtype classification, only the LSTM method was validated. We also ratified the performance gain of using an imbalanced data set in combination with F_β-function instead of a balanced data set with a common error function. Moreover, the effects of removing CFA and eye movement artefact respectively were investigated.

The entire algorithm was implemented and executed in MATLAB® using the deep learning toolbox and the built-in NVIDIA GPU support. Training was performed on a NVIDIA GPU at the Phoenix High Performance Computing (HPC) service of the University of Adelaide. The NN and LSTM training took approximately 11 minutes and 4 hours, respectively. Subsequently, the classification of one entire overnight recording was performed in 5–10 seconds on a Intel Core i7 processor with 3.60 GHz.

Table 3.3. Comparison of performance measures for A-phase detection evaluated on CAP Sleep Database.

	TPR	SPC	ACC	F_1-score
16 normal subjects				
Mariani *et al.* (2013)	67.03 ± 4.56	89.63 ± 3.55	86.09 ± 3.14	-
Feed-forward ANN	73.04 ± 11.23	84.61 ± 7.28	82.66 ± 5.50	56.56 ± 6.82
LSTM	76.10 ± 14.47	88.49 ± 6.09	86.43 ± 4.62	63.46 ± 8.22
30 NFLE subjects				
Machado *et al.* (2018) - LDA	79.00 ± -	74.00 ± -	73.00 ± -	-
Machado *et al.* (2018) - k-NN	82.00 ± -	69.00 ± -	75.00 ± -	-
Machado *et al.* (2018) - SVM	79.00 ± -	76.00 ± -	76.00 ± -	-
Feed-forward ANN	73.42 ± 7.84	82.41 ± 7.08	80.40 ± 4.69	60.57 ± 7.13
LSTM	78.48 ± 8.66	86.50 ± 6.23	85.09 ± 4.54	67.66 ± 7.03

TPR, true positive rate; SPC, specificity; ACC, accuracy; ANN, neural network; LSTM, long-short term memory network; NFLE, nocturnal frontal lobe epilepsy; LDA, linear discriminant analysis; k-NN, k-nearest neighbor; SVM, support vector machine

3.3 Results

3.3.1 A-phase detection

Table 3.3 lists the performance measures of the two classifiers and the results of Mariani *et al.* (2013) as well as Machado *et al.* (2018) using the LOO method. The values represent the means plus the standard deviation of the five subjects. In case of Mariani *et al.* (2013) and Machado *et al.* (2018), no F_1 measure was specified. The results show that the sensitivity and precision of the RNN classifier is greater than the values of the feed-forward NN (TPR: 3–5%, F_1: 7%). Moreover, the overall accuracy of the RNN is increased compared to the feed-forward NN (ACC: 4%). Compared to the system presented in Mariani *et al.* (2013), the LSTM classifier a better for every objective measure (TPR: 9%, ACC: 0.5%) except the sensitivity (SPC: -1%). Considering the data set with NFLE subjects, RNN achieves a lower sensitivity (TPR: -0.5%) but higher specificity and accuracy (SPC: 10.5%, ACC: 9%) as compared to the SVM method in Machado *et al.* (2018). The values for the feed-forward NN are in turn decreased for every objective measure (TPR: -5%, SPC: -4%, ACC: -4.5%, F_1: -7%). Table 3.4 compares the results of the imbalanced data set to the results of a balanced data set. For training with a balanced data set, the crossentropy function was applied as cost function. The LSTM algorithm achieved higher sensitivity and precision when an imbalanced data set was applied (TPR: 1%, F_1: 6%). In case of the NN algorithm, the sensitivity is decreased using an imbalanced data set but the F_1-score indicates a significantly higher precision (TPR: -5%, F_1: 2%). Table 3.5 compares

Table 3.4. Comparison of performance measures for 16 normal subjects using imbalanced and balanced data sets.

	TPR	SPC	ACC	F_1-score
Imbalanced				
Feed-forward ANN	73.04 ± 11.23	84.61 ± 7.28	82.66 ± 5.50	56.56 ± 6.82
LSTM	76.10 ± 14.47	88.49 ± 6.09	86.43 ± 4.62	63.46 ± 8.22
Balanced				
Feed-forward ANN	78.32 ± 10.73	80.21 ± 8.61	79.77 ± 6.59	54.67 ± 7.23
LSTM	75.28 ± 12.00	83.90 ± 8.95	82.42 ± 7.09	57.41 ± 9.64

TPR, true positive rate; SPC, specificity; ACC, accuracy; ANN, neural network; LSTM, long-short term memory network

Table 3.5. Detailed list for performance measures of 10 normal subjects with and without artefact removal.

Subject	CFA removal			EOG removal			No artefact removal		
	TPR (%)	SPC (%)	ACC (%)	TPR (%)	SPC (%)	ACC (%)	TPR (%)	SPC (%)	ACC (%)
n1	70.17	94.51	90.69	55.98	95.62	89.39	63.04	95.25	90.19
n2	76.16	80.80	80.12	61.04	89.17	85.04	68.53	86.69	84.02
n4	88.71	86.50	86.78	87.37	87.20	87.22	86.61	87.47	87.36
n5	70.94	93.33	89.46	67.56	93.26	84.32	68.67	93.65	89.33
n7	95.56	79.08	81.26	90.97	84.32	85.20	94.17	80.63	82.42
n8	87.16	88.96	88.67	90.70	84.03	85.11	86.97	88.00	87.83
n10	60.92	79.88	76.95	55.32	85.56	80.89	61.29	81.78	78.61
n11	56.24	90.42	84.69	57.85	90.03	84.63	57.75	91.55	85.88
n13	78.37	87.42	85.54	82.60	84.13	83.81	86.30	82.49	83.28
n14	64.58	94.98	89.52	68.40	93.11	88.67	71.02	93.16	89.18
Mean ± std	74.88 ± 12.78	87.56 ± 6.00	85.37 ± 4.61	71.78 ± 14.70	88.64 ± 4.27	85.88 ± 2.65	74.44 ± 12.89	88.07 ± 5.26	85.81 ± 3.69

CFA, cardiac field artefact; EOG, electrooculography; TPR, true positive rate; SPC, specificity; ACC, accuracy

the results of ten normal subjects obtained with and without artefact removal. In all three cases, the LOO method was applied to the data set. Although statistics slightly varied for each subject there was no significant overall effect.

3.3.2 Subtype classification

Table 3.6 displays the multi-class performance measures of subtype classification with the LSTM method and the results of Machado *et al.* (2018). In Machado *et al.* (2018), no F_1 measure and no standard deviation were specified. Compared to the SVM method used in Machado *et al.* (2018), the LSTM classifier achieves a higher sensitivity for each class (background: 6%, A1: 1.5%, A2: 1.5%, A3: 50%). Considering the average accuracy, the

Table 3.6. Results for subtype classification performed on different data sets.

	Background		A1		A2		A3		Accuracy (%)
	TPR (%)	F_1-score (%)	TPR (%)	F_1-score (%)	TPR (%)	F_1-score (%)	TPR (%)	F_1-score (%)	
16 normal subjects									
LSTM	85.51 ± 7.68	90.23 ± 4.61	63.14 ± 12.46	46.61 ± 9.01	42.31 ± 15.88	32.96 ± 10.46	70.62 ± 17.95	60.32 ± 7.83	81.89 ± 6.84
30 NFLE subjects									
Machado *et al.* (2018) - LDA	73.00 ± -	-	66.00 ± -	-	37.00 ± -	-	18.00 ± -	-	68.00 ± -
Machado *et al.* (2018) - k-NN	71.00 ± -	-	59.00 ± -	-	31.00 ± -	-	16.00 ± -	-	70.00 ± -
Machado *et al.* (2018) - SVM	76.00 ± -	-	58.00 ± -	-	44.00 ± -	-	24.00 ± -	-	71.00 ± -
LSTM	82.30 ± 6.77	88.09 ± 3.62	59.54 ± 15.80	45.42 ± 10.05	45.40 ± 12.99	33.74 ± 7.82	74.25 ± 12.24	62.25 ± 9.50	78.27 ± 4.95

TPR, true positive rate; LSTM, long-short term memory network; NFLE, nocturnal frontal lobe epilepsy; LDA, linear discriminant analysis; k-NN, k-nearest neighbor; SVM, support vector machine

LSTM system outperforms the method in Machado *et al.* (2018) by 6%. The confusion matrices of both data sets for subtype classification are listed in Table 3.7.

3.4 Discussion

In this paper we describe an automated CAP classification system that exploits the dynamical temporal characteristics of the EEG signal using deep LSTM RNN. The classifiers were equipped with the F_β-score as objective error function to address the dilemma of using only a small percentage of the recordings due to the rare occurrence of A-phases. Consequently, a greater data set and the focus on the sensitivity and precision of the classification increased the number of correctly detected A-phases. Our comprehensive system consists of state-of-the-art signal processing methods to minimize the inter-subject variation. Furthermore, the detection system can be extended to a high-performance multi-class classification system labelling the detected A-phases with the predicted subtype.

As the results in Table 3.3 indicate, a RNN improves the sensitivity and precision of the classification compared to a conventional feed-forward NN. The NN and the LSTM only differ in the sequence of LSTM cells before the neural network layer. Thus, a LSTM can determine crucial information in a time sequence, resulting in a significantly improved detection of A-phases. Our overall comparison demonstrates that the LSTM classifier outperforms the system presented in Mariani *et al.* (2013). The numbers for the second data set indicate a more accurate and precise scoring than those achieved previously in Machado *et al.* (2018) although the sensitivity is marginally decreased. Note that no other previous studies were considered in our comparison, due to the difference in data sets. In summary, our scoring system using a LSTM classifier improves the performance of A-phase classification.

Table 3.7. Confusion matrix of subtype classification for a) 16 normal subjects and b) 30 NFLE subjects.

a)

		Target				Total
		B	A1	A2	A3	(PPV)
Output	B	**237,550**	**5,456**	**2,243**	**2,398**	**247,647**
		72.5%	1.7%	0.7%	0.7%	(95.9%)
	A1	**22,561**	**15,086**	**2,104**	**294**	**40,045**
		6.9%	4.6%	0.6%	0.1%	(37.7%)
	A2	**6,234**	**2,992**	**4,764**	**1,707**	**15,697**
		1.9%	0.9%	1.5%	0.5%	(30.3%)
	A3	**10,770**	**493**	**1,522**	**11,658**	**24,443**
		3.3%	0.2%	0.5%	3.6%	(47.7%)
Total (TPR)		**277,115** (85.7%)	**24,027** (62.8%)	**10,633** (44.8%)	**16,057** (72.6%)	

b)

		Target				Total
		B	A1	A2	A3	(PPV)
Output	B	**408,938**	**9,334**	**4,992**	**6,535**	**429,799**
		65.2%	1.5%	0.8%	1.0%	(95.1%)
	A1	**42,684**	**31,082**	**6,289**	**965**	**81,020**
		6.8%	5.0%	1.0%	0.2%	(38.4%)
	A2	**16,755**	**8,974**	**12,362**	**5,546**	**43,637**
		2.7%	1.4%	2.0%	0.9%	(28.3%)
	A3	**28,941**	**1,401**	**4,111**	**38,332**	**72,785**
		4.6%	0.2%	0.7%	6.1%	(52.7%)
Total (TPR)		**497,318** (82.2%)	**50,791** (61.2%)	**27,754** (44.5%)	**51,378** (74.6%)	

TPR, true positive rate; PPV, positive predictive value; NFLE, nocturnal frontal lobe epilepsy

The performance measures for subtype classification show promising results for LSTM to distinguish different activation phases of CAP. The multi-class system achieves a low sensitivity for subtype A2, presumably because of its nature of two mixed waveforms and extremely low prevalence in the training set. The sensitivity and F_1-score values for subtype A1 and A3 suggest a highly precise scoring for EEG rhythms with predominantly EEG synchrony or EEG desynchrony in comparison to Machado *et al.* (2018). However, Machado *et al.* (2018) applied full sleep recordings without removing REM or wake periods. As the CAP Sleep Database provides only CAP scoring annotations for NREM stages, it was presumed that during wake and REM periods no A-phases occur. Hence, wake and REM seconds were labelled as non-CAP phases, whereas according to the CAP atlas (Terzano *et al.*, 2001) A-phase features of desynchronisation can be seen during REM but are not part of a CAP sequence by definition. Consequently, this approach may have lead to a lower sensitivity for A3 phases, which resemble wake and REM phases. In this study, wake and REM periods were neglected due to the missing annotation information.

Evaluating the effect of the F_β-score as cost function, Table 3.4 shows a pivotal performance increase for the RNN algorithm regarding the sensitivity and precision. In case of the NN algorithm, the results display lower sensitivity but greater precision. In general, the numbers demonstrate that the usage of the F_β-score improves the accuracy of the A-phase classification by enabling the application of a greater data set and focusing on the sensitivity and precision of the scoring. The main issue when optimizing directly for the F_β-score itself is the shape of the cost function. Due to its non-convex shape, optimization algorithms can not consistently locate the global minimum of the error function. Approximation methods can help to erase this problem (Nan *et al.*, 2012).

Finally, the comparison between a data set with particular artefact removal and without artefact removal displays no significant overall difference. However, each subject shows different results for all three methods, implying that the effect of the classification process is small. CFA removal shows an increase in sensitivity for the majority (60%) of the subjects whereas the removal of eye movements indicates a decrease in correctly detected A-phases. However, artefact removal methods still appear to be favourable as they decrease the noise in the EEG signal.

One major limitation of the system is the reliance on recordings from the same data set. Our classifiers are trained on recordings performed by the same laboratory with the same setup. A dissimilar measurement setup may affect the classification results. Furthermore, the algorithm depends on a previously performed sleep stage scoring performed either manually by a sleep expert or automated by a classification algorithm. In case of the latter, a combined system could be developed to decrease the computing costs. Finally, the presented algorithm can not be deployed as a real-time application in its current implementation since the entire recording is required during the artefact removal step and future information is used for feature extraction. Both steps can be simply modified to make the system real-time applicable.

3.5 Conclusion

We developed a stand-alone, fully automated sleep scoring system to detect the A-phases of CAP events by exploiting the dynamical temporal behaviour of the EEG. The system is equipped with a deep LSTM network network, respectively. The usage of a time sequence and its dynamics improves the classification of A-phases. The system can also be applied to classify the different subtypes of activation phases.

3.6 Acknowledgement

This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide.

Paper II

Characterization of cyclic alternating pattern during sleep in older men and women using large population studies

4

The content of this chapter is a modified version of the publication:

Hartmann, S., Bruni, O., Ferri, R., Redline, S. and Baumert, M. (2020), 'Characterization of cyclic alternating pattern during sleep in older men and women using large population studies', *Sleep* **43**(7), zsaa016.

Abstract

Study Objectives: To assess the microstructural architecture of non-rapid eye movement (NREM) sleep known as cyclic alternating pattern (CAP) in relation to the age, gender, self-reported sleep quality, and the degree of sleep disruption in large community-based cohort studies of older people.

Methods: We applied a high-performance automated CAP detection system to characterize CAP in 2,811 men from the Osteoporotic Fractures in Men Sleep Study (MrOS) and 426 women from the Study of Osteoporotic Fractures (SOF). CAP was assessed with respect to age and gender and correlated to obstructive apnoea-hypopnoea index, arousal index (AI-NREM), and periodic limb movements in sleep index. Further, we evaluated CAP across levels of self-reported sleep quality measures using analysis of covariance.

Results: Age was significantly associated with the number of CAP sequences during NREM sleep (MrOS: $p = 0.013$, SOF: $p = 0.051$). CAP correlated significantly with AI-NREM (MrOS: $\rho = 0.30$, SOF: $\rho = 0.29$). CAP rate, especially the A2+A3 index, was inversely related to self-reported quality of sleep, independent of age and sleep disturbance measures. Women experienced significantly fewer A1-phases compared to men, in particular, in slow-wave sleep (N3).

Conclusions: We demonstrate that automated CAP analysis of large-scale databases can lead to new findings on CAP and its subcomponents. We show that sleep disturbance indices are associated with the CAP rate. Further, the CAP rate is significantly linked to subjectively reported sleep quality, independent from traditionally scored markers of sleep fragmentation. Finally, men and women show differences in the microarchitecture of sleep as identified by CAP, despite similar macro-architecture.

4.1 Statement of Significance

We report the prevalence of periodically occurring cortical activation phases in large population samples of older men and women. To the best of our knowledge, this effort represents the first time that cyclic alternating pattern (CAP) was scored and analysed with a high-performance automated detection system in large community-based cohort studies.

We ascertain the relationship with gender, age, self-reported sleep quality measures, and traditional polysomnographic indices of disordered sleep. Individuals experiencing a higher CAP rate, in particular, A2+A3-phases, report a lower sleep quality independent of apnoea-hypopnoea index, arousal index, and periodic limb movement index. In older populations, age is a significant predictor for non-rapid eye movement sleep fragmentation. We also reveal gender differences in the microarchitecture of sleep despite similar macro-architecture.

4.2 Introduction

Since Rechtschaffen and Kales (1968) published their scoring guide in 1968, sleep has been traditionally divided into states of high neuronal activity and quiescence, typically known as rapid eye movement (REM) and non-REM (NREM). The latter, in turn, is partitioned into three distinct stages according to the current consensus detailed in the American Academy of Sleep Medicine scoring manual (Iber *et al.*, 2007). One major drawback of these scoring rules is the neglect of short-lasting events such as K-complexes and transient power alterations in frequency bands (Terzano and Parrino, 2000). In the AASM framework, short periods of changes in cortical activation are only captured by the arousal definition (Iber *et al.*, 2007). Phasic events like K-complexes and delta bursts show arousal-like characteristics but they are not regarded as arousals when not related to short-term frequency increases in an electroencephalogram (EEG) (Terzano and Parrino, 2000). Hence, an additional sleep scoring atlas was devised including such recurring phasic events in brain activity under the name of cyclic alternating pattern (CAP) (Terzano *et al.*, 2001).

CAP analysis seeks to capture the microstructure of sleep. It focuses on short EEG amplitude increases (<60 s) that reappear periodically in NREM stages, separated by equally long time spans of lower-amplitude background activity (Terzano and Parrino, 2000). Such short events are called activation phases because of their high neural excitability and autonomic correlates. It is believed that recurring periods of activation during sleep represent time windows that facilitate sensory input for the brain and synchronize with physiological and pathological events (Terzano and Parrino, 2005). Hence, an increased CAP rate may occur in sleep disorders such as periodic limb movement disorder, sleep apnoea syndrome, or insomnia (Terzano and Parrino, 1993). In recent years, the role of CAP has been receiving enlarged clinical interest, but current evidence is limited to small studies focusing on particular disorders. The role and prevalence of CAP during sleep on the broader population remain still largely unknown.

In this study, we characterize for the first time CAP across large population samples. We describe the prevalence of CAP in older populations in relation to age and gender. Moreover, we explore the relationship between CAP and common disorders that have been associated with sleep fragmentation as well as self-reported sleep quality measures.

4.3 Methods

4.3.1 Definition of CAP

We defined CAP in agreement with Terzano *et al.* (2001) as sequences of at least two consecutive cycles that consist of an activation phase (A-phase) followed by the period between two repetitive A-phases, called B-phase (background). A-phases represent transient, phasic events that stand out from the background, whereas B-phases are thought to embody rebound deactivation reflecting active inhibition rather than passive recovery of the stationary baseline during NREM sleep (Terzano and Parrino, 2000). We defined A-phases or B-phases to last 2–60 s but did not limit the number of cycles per CAP sequence. In accordance with the CAP atlas, the time period between two CAP sequences was considered as non-CAP. The last A-phase prior to a non-CAP period was also defined as non-CAP as it does not form a cycle.

Typical patterns for A-phases include delta bursts, vertex sharp transients, K-complex sequences, K-alpha, polyphasic bursts, intermittent alpha, and arousals (Terzano *et al.*, 2001). Thus, A-phases consist of either high-voltage, slow waves or low-voltage, fast waves or a combination of both. High-voltage slow waves portray synchronized EEG patterns and low-amplitude fast rhythms represent desynchrony (Terzano and Parrino, 2000). Based on the content of these two frequency components, we subdivided A-phases into three subtypes. Subtype A1 is associated with periods where high EEG synchrony is prevalent, i.e. slow rhythms with high amplitudes. Desynchronized patterns are classified as A2 and A3 subtypes mostly occurring in time periods before and after REM sleep. They represent high-frequency rhythms with low amplitudes. As REM sleep includes mainly desynchronized A-phases located further apart than 60 s, we restricted CAP to NREM sleep in agreement with Terzano *et al.* (2001).

4.3.2 Automated A-phase detection and CAP quantification

We deployed our previously developed, highly precise automated system for CAP analysis, which is described in detail in the work of Hartmann and Baumert (2019). At the core of the system is a deep learning recurrent neural network (RNN) that was trained specifically to recognize A-phases in EEG recordings. The entire system is divided into four major parts: preprocessing, feature extraction, classification, and post-processing. In the preprocessing step, the raw signal of one central EEG channel is prepared to reduce intersubject variation by removing the cardiac field and eye movement artefacts. Based on the processed signal, multiple features in the time and frequency domain are calculated such as Hjorth activity, Shannon entropy, Teager Energy Operator, band power descriptor, and differential EEG variance. The extracted feature set serves as input for the RNN classifier in the classification stage. The classifier was trained with the F_β-score as loss function to increase the quantity of correctly detected A-phases and reduce the number of incorrectly classified periods. Finally, we post-processed the output of the A-phase detection system applying the aforementioned rules for CAP sequences. Isolated A-phases that did not form a sequence were removed from the scoring outcome. The training comprised 15 healthy participants and 24 participants with sleep disorders from a publicly available database (Goldberger *et al.*, 2000). The polysomnographic measurements in the training set including visual CAP scoring were conducted by the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy. The second-by-second A-phase inter-rater reliability between visual scoring and our system, quantified by the Cohen's kappa coefficient, was 0.53 on a set of 16 healthy participants and 0.56 on a set of 30 participants with nocturnal frontal lobe epilepsy. The event-based inter-rater reliability between human scorers ranges between 0.42 and 0.75 (Ferri *et al.*, 2005b).

Measurements recorded with a low bit rate or a low physical range were excluded because they often contain severe clipping leading to false classification results. In this study, we computed the CAP rate and subtype rate based on the following equations:

$$\text{CAP rate} = \frac{\text{total CAP time (in seconds)}}{\text{total NREM sleep (in seconds)}} \times 100 \tag{4.1a}$$

$$\text{A1 index} = \frac{\text{number of A1-phases}}{\text{total NREM sleep (in seconds)}} \times 3{,}600 \tag{4.1b}$$

$$\text{A2+A3 index} = \frac{\text{number of A2+A3-phases}}{\text{total NREM sleep (in seconds)}} \times 3{,}600 \tag{4.1c}$$

Subtypes A2 and A3 were merged into a single parameter due to their congruent nature.

We defined four consecutive sleep periods of 90-min duration each to investigate the relationship between CAP and sleep intervals during 6 h of sleep. Sleep stage scoring manually performed by trained sleep technicians during the implementation of the studies was used to identify NREM sleep. As both studies were conducted before the release of the AASM scoring manual, sleep stage scoring was performed in accordance to the criteria in the work of Rechtschaffen and Kales (1968). Sleep stages 3 and 4 were merged to a single stage (called here SWS) and sleep stage 1 was excluded due to the low occurrence in the majority of the participants.

4.3.3 Study samples: MrOS and SOF

For our analysis, we utilized data from two multi-center sleep cohorts: Osteoporotic Fractures in Men (MrOS) Study and Study of Osteoporotic Fractures (SOF). Both data sets were provided by the National Sleep Research Resource (available online at the National Sleep Research Resource; sleepdata.org) (Dean II *et al.*, 2016).

MrOS is a long-term cohort study designed to determine fracture risk in relation to multiple factors such as bone characteristics, lifestyle, anthropometric and neuromuscular measures, and fall propensity. In total, 5,995 men aged 65 or older were examined during a 25-month period from 2000 to 2002 followed by a second visit in 2005 (Orwoll *et al.*, 2005). The study was conducted at six clinical sites with the requirement that all participants needed to be able to walk without assistance and must not have had a bilateral hip replacement. As part of the MrOS cohort, 3,115 men were recruited for an ancillary sleep study (MrOS Sleep Study) including comprehensive overnight polysomnography (PSG), designed to identify the cardiovascular and health consequences of sleep disturbances (Blackwell *et al.*, 2011). Men who used mechanical devices or oxygen during sleep were excluded from the study. The baseline sleep exam (Visit 1) was conducted between 2003 and 2005 and a follow-up exam (Visit 2) was conducted between 2009 and 2012. We removed recordings with technically inadequate PSG or fewer than 3 h of good EEG quality resulting in 2,811 participants for Visit 1 and 933 participants for Visit 2.

SOF was designed to investigate the risk factor for hip fractures among older women (Cummings *et al.*, 1990). Women who were community-dwelling, 65 years or older, able to walk unassisted, and had no previous bilateral hip replacement were recruited during September 1986 and October 1988 in four metropolitan areas (Claman *et al.*, 2006). Within the latest visit cycle between 2002 and 2004, a subset of 461 women underwent an unattended

overnight 12-channel in-home PSG to evaluate the relationship of sleep disturbances to a number of health outcomes (Spira *et al.*, 2008). After discarding recordings with inadequate EEG quality by applying the same approach as for MrOS, 426 recordings from SOF were available for analysis.

4.3.4 Statistical methods

Statistical analysis was conducted using non-parametric tests based on the assumption that the CAP rate and subtype indices do not follow a normal distribution. For each statistical test, the significance level was adjusted to the number of variables under consideration using Bonferroni correction. All values are presented as median $\pm$ interquartile range (IQR).

We subdivided CAP rate data across both cohorts into quartiles to evaluate the association of CAP with the obstructive apnoea-hypopnoea index at 4% oxygen desaturation (OAHI), the arousal index (AI-NREM), and the periodic limb movement in sleep index (PLMSI) as clinical indicators of sleep fragmentation. The Jonckheere–Terpstra test was applied to identify a statistically significant trend between quartiles of CAP parameters and indices of disordered sleep. Spearman correlation coefficients were determined to examine the relationship between the aforementioned indices and CAP parameters.

Participants in MrOS were asked to score the quality of their sleep following PSG on a Likert scale of five items from light to deep, from short to long, and from restless to restful. In SOF, self-reported sleep quality after PSG was not measured. To investigate the effect of these measures on CAP, we applied the analysis of covariance (ANCOVA) with CAP rate, A1 index, and A2+A3 index as dependent variables, the three self-reported sleep quality measures as independent variables, and age, OAHI, AI-NREM, and PLMSI as covariates.

To explore the effect of age on CAP, multivariable regression was conducted with age, OAHI, AI-NREM, and PLMSI as independent predictors for CAP rate, A1 index, and A2+A3 index. Each multivariable regression was carried out separately in MrOS and SOF. To analyse the effect of gender on CAP, two normalized, age-matched subsets were sampled from the MrOS and SOF cohorts, respectively, comprising 220 men and women. Both subsets were restricted to participants with AHI less than 15, AI-NREM less than 25, and PLMSI less than 15. We selected the Mann–Whitney–Wilcoxon test for comparing the independent gender groups and sleep stages within both genders. To determine differences between sleep intervals within both genders, we used the Kruskal–Wallis test by ranks. Finally,

the reproducibility of the applied system was tested comparing matching participants from MrOS Sleep Visit 1 and Visit 2.

4.4 Results

The median age of participants was 76 years in MrOS (baseline exam) and 82 years in SOF. Participants in MrOS had a median BMI of 26.7 kg/m^2; women in SOF demonstrated a median BMI of 27.1 kg/m^2. MrOS participants experienced a median OAHI of 8.1/h as well as a median AI-NREM of 22.3/h. In SOF, the median OAHI was 6.7/h, and the median AI-NREM was 19.8/h. Both cohorts show a median duration of total NREM sleep of about 5 h (MrOS: 288 min [IQR: ±69.0], SOF: 286 min [IQR: ±78.0]). The total scored sleeping time in MrOS was 357.5 min (±83.8) and in SOF was 353 min (±95.0), which results in an NREM sleep percentage of 80.5% (±8.9) in MrOS and 81.7% (±9.8) in SOF.

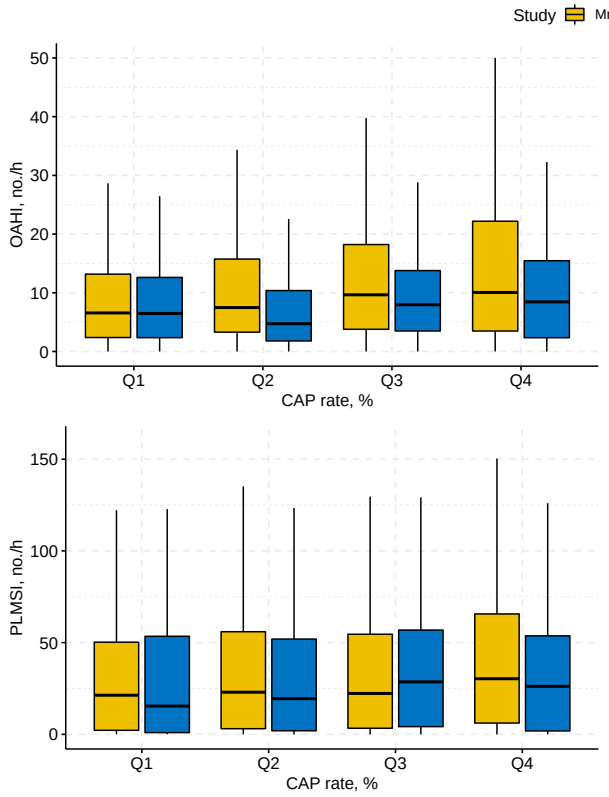
4.4.1 CAP and sleep fragmentation

In terms of CAP, the male cohort (MrOS) displayed an overall large amount of 57.0% (±21.5) NREM sleep occupied by periodically occurring phasic events. The even older female cohort (SOF) demonstrated similar values of CAP accounting for 54.1% (±26.1) of NREM sleep. Indices of disordered sleep (OAHI, PLMSI, and AI-NREM) increased significantly with increasing CAP in MrOS (Figure 4.1). SOF participants showed a similar relationship, except for PLMSI, which was slightly reduced in the last quartile compared to the previous quartile but followed a similar trend overall.

On average, 15.2 (±20.7) A1-phases occurred per hour of NREM sleep in MrOS and 13.1 (±18.5) in SOF. The A2+A3 index was substantially higher in both cohorts (46.6 [±31.1] in MrOS and 44.5 [±33.2] in SOF). The statistics on the relationship between indices of disordered sleep and A-phase subtypes are summarized in Figures 4.2 and 4.3, respectively. In MrOS, increasing A1 was associated with decreasing OAHI, AI-NREM, and PLMSI. In SOF, the same effect could be observed for AI-NREM. Conversely, in both cohorts, increasing A2+A3 was associated with higher OAHI, AI-NREM, and PLMSI.

4.4.2 CAP and self-reported sleep quality measures

Supplementary Figure 4.5 illustrates the results of the ANCOVA for all three self-reported sleep quality measures reported in MrOS with CAP rate, A1 index, and A2+A3 index as



MrOS

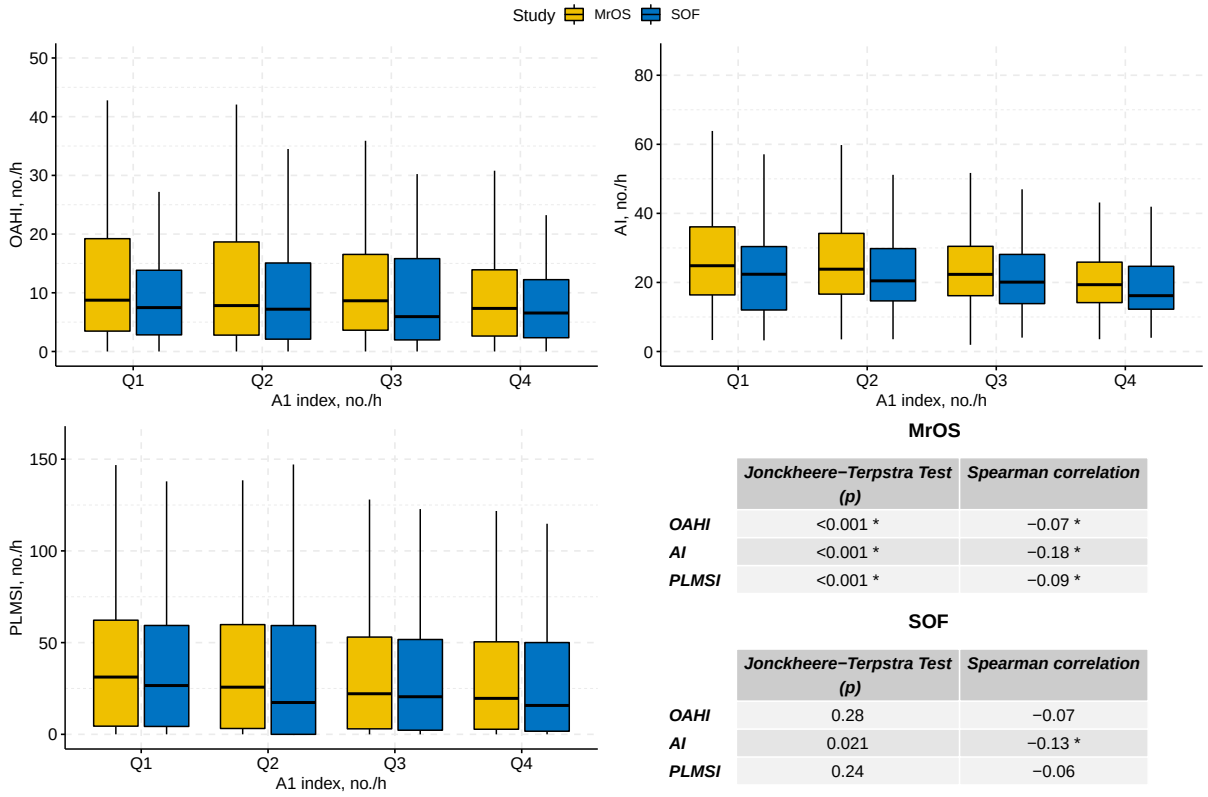
	Jonckheere–Terpstra Test (p)	*Spearman correlation*
OAHI	<0.001 *	0.13 *
AI	<0.001 *	0.30 *
PLMSI	<0.001 *	0.09 *

SOF

	Jonckheere–Terpstra Test (p)	*Spearman correlation*
OAHI	0.052	0.10 *
AI	<0.001 *	0.29 *
PLMSI	0.19	0.09

Figure 4.1. Indices of disordered sleep for CAP rate quartiles in MrOS and SOF. The relationship between indices of disordered sleep (OAHI, AI-NREM, and PLMSI) and CAP quartiles in Osteoporotic Fractures in Men (MrOS) Study and Study of Osteoporotic Fractures (SOF). Significance level (*p < 0.017) was adjusted according to the number of variables under consideration. OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index; PLMSI, periodic limb movement in sleep index.

dependent variables and age, AI-NREM, OAHI, and PLMSI as covariates. CAP rate decreased significantly with increasing quality of sleep for all three self-reported measures (light vs. deep: 58.8 ± 22.3% vs. 54.6 ± 20.5%, $p < 0.001$; short vs. long: 58.4 ± 21.4% vs. 55.1 ± 20.5%, $p < 0.001$; restless vs. restful: 59.4 ± 20.8% vs. 55.6 ± 21.0%, $p = 0.002$). The A1 index did not vary significantly across all three sleep quality parameters (light vs. deep: 12.9 ± 20.2 no./h vs. 17.7 ± 21.8 no./h, $p = 0.19$; short vs. long: 15.8 ± 20.4 no./h vs. 16.4 ± 17.6 no./h, $p = 0.76$; restless vs. restful: 15.1 ± 20.1 no./h vs. 15.8 ± 21.1 no./h, $p = 0.94$). Similar to the CAP rate, the A2+A3 index decreased with increasing values for each self-reported measure (light vs. deep: 49.0 ± 32.0 no./h vs. 41.3 ± 29.0 no./h, $p < 0.001$; short vs. long: 47.5 ± 28.2 no./h vs. 44.9 ± 30.3 no./h, $p < 0.001$; restless vs. restful: 48.8 ± 31.4 no./h vs. 42.9 ± 30.0 no./h, $p < 0.001$). Detailed ANCOVA results including all three self-reported sleep quality measures with AI-NREM as dependent variable are listed in Supplementary Tables 4.3, 4.4, 4.5, and 4.6.



MrOS

	Jonckheere–Terpstra Test (p)	*Spearman correlation*
OAHI	<0.001 *	−0.07 *
AI	<0.001 *	−0.18 *
PLMSI	<0.001 *	−0.09 *

SOF

	Jonckheere–Terpstra Test (p)	*Spearman correlation*
OAHI	0.28	−0.07
AI	0.021	−0.13 *
PLMSI	0.24	−0.06

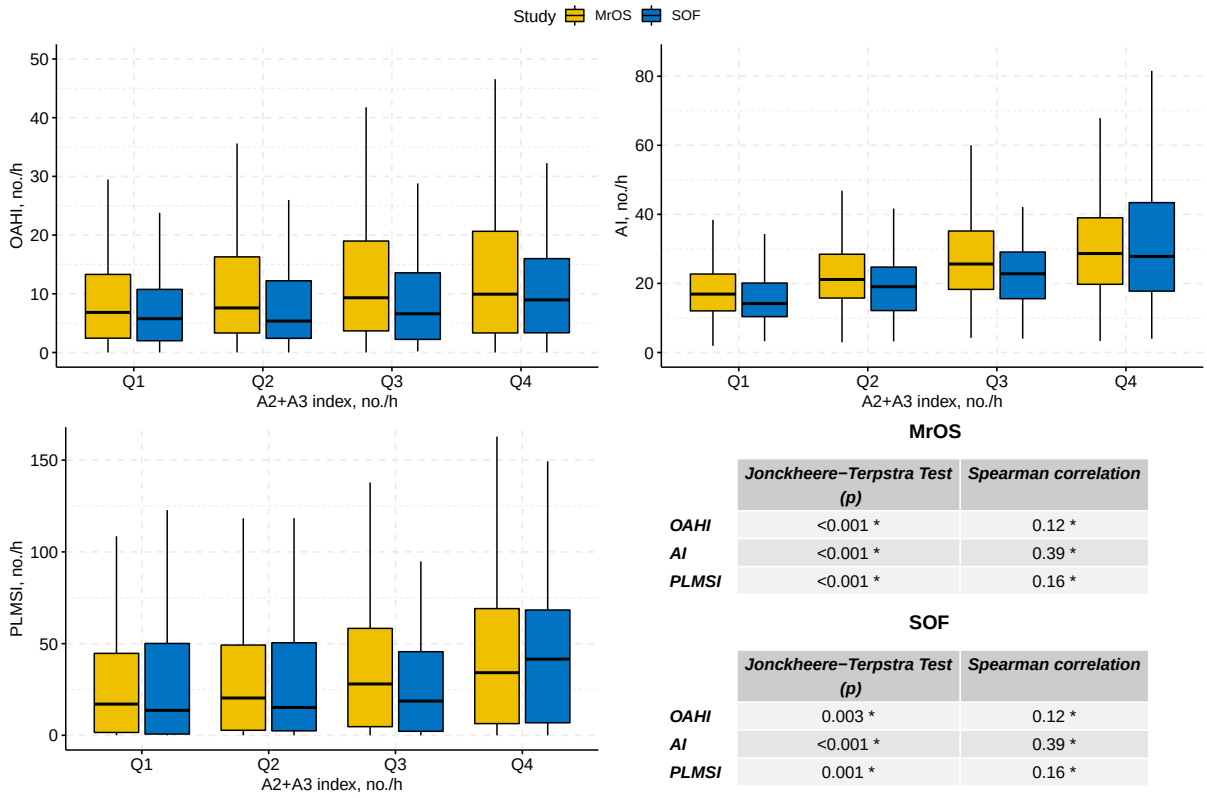
Figure 4.2. Indices of disordered sleep for A1 index quartiles in MrOS and SOF. The relationship between indices of disordered sleep (OAHI, AI-NREM, and PLMSI) and A1 index quartiles in Osteoporotic Fractures in Men (MrOS) Study and Study of Osteoporotic Fractures (SOF). Significance level (* = p < 0.017) was adjusted according to the number of variables under consideration. OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index; PLMSI, periodic limb movement in sleep index.

4.4.3 CAP and age

Several multivariable regression models were evaluated to investigate the effect of age on CAP rate, A1 index, and A2+A3 index in MrOS and SOF, respectively (Table 4.1).

Age and AI-NREM were significantly associated with CAP rate in MrOS (age: B = 0.13, p = 0.013; AI-NREM: B = 0.36, $p < 0.001$), whereas only AI-NREM was significantly associated with CAP rate in SOF (age: B = -0.50, p = 0.051; AI-NREM: B = 0.40, $p < 0.001$). Neither OAHI or PLMSI was associated with CAP in either cohort. The overall model fit for CAP rate was R2= 0.10 for both cohorts.

Neither age nor OAHI was significantly associated with the A1 index in MrOS or SOF. In MrOS, AI-NREM and PLMSI were significantly negatively associated with the A1 index (PLMSI: B = -0.03, $p < 0.001$; AI-NREM: B = -0.23, $p < 0.001$), whereas in SOF the only



MrOS

	Jonckheere−Terpstra Test (p)	*Spearman correlation*
OAHI	<0.001 *	0.12 *
AI	<0.001 *	0.39 *
PLMSI	<0.001 *	0.16 *

SOF

	Jonckheere−Terpstra Test (p)	*Spearman correlation*
OAHI	0.003 *	0.12 *
AI	<0.001 *	0.39 *
PLMSI	0.001 *	0.16 *

Figure 4.3. Indices of disordered sleep for A2+A3 index quartiles in MrOS and SOF. The relationship between indices of disordered sleep (OAHI, AI-NREM, and PLMSI) and A2+A3 index quartiles in Osteoporotic Fractures in Men (MrOS) Study and Study of Osteoporotic Fractures (SOF). Significance level (*p < 0.017) was adjusted according to the number of variables under consideration. OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index; PLMSI, periodic limb movement in sleep index.

significant association was for AI-NREM (AI-NREM: B = -0.20, $p < 0.01$). The overall model fit for the A1 index prediction was R2 = 0.05 in MrOS and R2 = 0.02 in SOF.

Finally, AI-NREM and PLMSI were significantly negatively associated with the frequency of A2+A3-phases in MrOS (PLMSI: B = -0.07, $p < 0.001$; AI-NREM: B = -0.70, $p < 0.001$) and SOF (PLMSI: B = -0.07, $p = 0.014$; AI-NREM: B = -0.72, $p < 0.001$). Age and OAHI were not associated with A2+A3 in either cohort. The overall model fit for A2+A3 index prediction was R2 = 0.16 in MrOS and R2 = 0.17 in SOF.

4.4.4 CAP and gender

A subset of 110 participants with identical age distribution and no severe sleep disorders (AHI <15, AI-NREM <25, and PLMSI <15) was sampled from the MrOS cohort and the SOF cohort to compare both sexes while eliminating the age influence. The statistics for

Table 4.1. Multivariable Regression Results Predicting CAP Rate, A1 Index, and A2+A3 Index in MrOS and SOF.

	MrOS										
	CAP rate, %				A1 index, no./h				A2+A3 index, no./h		
Variables	B	SE	t	*p*	B	SE	t	*p*	B	SE	t
Age, years	0.13	0.05	2.48	0.013*	-0.05	0.05	-0.92	0.36	0.18	0.08	2.36
AI-NREM, no./h	0.36	0.03	14.05	<0.001*	-0.23	0.03	-8.81	<0.001*	0.70	0.04	18.77
OAHI, no./h	0.02	0.02	0.69	0.49	-0.02	0.03	-0.81	0.42	-0.07	0.04	-1.86
PLMSI, no./h	0.01	0.01	1.47	0.14	-0.03	0.01	-3.90	<0.001*	0.07	0.01	6.01
Constant	37.02	3.90	9.49	<0.001*	29.79	4.03	7.39	<0.001*	17.57	5.70	3.08
R2	0.10				0.05				0.16		
Adj. R2	0.10				0.05				0.16		
	F(4,2806) = 75.60		*p* <0.001		F(4,2806) = 36.17		*p* <0.001		F(4,2806) = 131.90		*p* <0.001
	SOF										
	CAP rate, %				A1 index, no./h				A2+A3 index, no./h		
Variables	B	SE	t	*p*	B	SE	t	*p*	B	SE	t
Age, years	-0.52	0.26	-1.96	0.051	0.12	0.25	0.49	0.62	-0.77	0.35	-2.21
AI-NREM, no./h	0.40	0.07	5.88	<0.001*	-0.20	0.06	-3.06	<0.01*	0.72	0.09	7.91
OAHI, no./h	0.07	0.08	0.95	0.34	-0.04	0.07	-0.60	0.62	0.06	0.10	0.59
PLMSI, no./h	0.031	0.02	1.19	0.23	-0.02	0.02	-0.81	0.42	0.07	0.03	2.46
Constant	85.35	21.62	3.95	<0.001*	12.74	20.56	0.62	0.54	94.33	28.73	3.28
R2	0.11				0.03				0.18		
Adj. R2	0.10				0.02				0.17		
	F(4,421) = 12.98		*p* <0.001		F(4,421) = 3.62		*p* <0.01		F(4,421) = 23.33		*p* <0.001

Results of each multivariable regression predicting cyclic alternating pattern (CAP) rate, A1 index, and A2+A3 index in Osteoporotic Fractures in Men (MrOS) Study and women in Study of Osteoporotic Fractures (SOF) with age as independent variable and the obstructive apnoea–hypopnoea index (OAHI) at 4% oxygen desaturation, the arousal index (AI-NREM) in NREM sleep, and the periodic limb movement in sleep index (PLMSI) as additional independent variables. B, an estimate of beta coefficient; SE, standard error of beta coefficient. Significance level: $p < 0.017$ (adjusted to the number of variables under consideration).

each subset on sleep parameters and indices of sleep disturbance are listed in Table 4.2. Mann–Wilcoxon U-test shows a significantly lower A1 index in women compared to men (A1 index: $p = 0.036$). Men had a significantly higher percentage of stage 2 sleep compared to women, but a lower percentage of SWS (S2: $p = 0.002$, SWS: $p < 0.001$).

Figure 4.4 displays A1 and A2+A3 indices for gender groups across sleep intervals and NREM stages. Both men and women display a decrease in A1-phases throughout the night (sleep interval: $p < 0.001$), whereas men experienced more A1-phases compared to women (gender: $p < 0.001$). The A2+A3 index did not show any variations between men and women and remained constant throughout the night, (gender: $p = 0.95$, sleep interval:p = 0.11). Regarding sleep stages, the A1 index was significantly higher in men in both stages. The number of A1-phases increased from S2 to SWS (gender: $p < 0.001$, sleep stage: $p < 0.001$). On the contrary, both genders experienced less A2+A3-phases in SWS compared to S2 (gender: $p = 0.42$, sleep stage: $p < 0.001$).

Table 4.2. MrOS vs. SOF with identical age distribution.

	Age, years	AI-NREM, no./h	OAHI, no./h	PLMSI, no./h	CAP rate, %	A1 index, no./h	A2+A3 index, no./h	Scored sleep time, min	Percentage of sleep in stage 2, %	Percentage of sleep in stage 3/4, %
	Median (IQR)									
Men in MrOS	82.0 (±4.0)	15.4 (±5.9)	4.5 (±6.9)	1.2 (±4.9)	51.3 (±18.9)	17.3 (±21.6)	37.8 (±21.5)	369.0 (±83.8)	58.4 (±14.0)	13.9 (±14.3)
Women in SOF	82.0 (±4.0)	14.2 (±9.3)	4.6 (±6.3)	1.1 (±7.4)	48.9 (±27.1)	13.7* (±18.8)	38.6 (±25.0)	356.4* (±90.0)	52.5* (±18.6)	20.8* (±19.0)

Comparison between men in Osteoporotic Fractures in Men (MrOS) Study and women in the Study of Osteoporotic Fractures (SOF) using a subsample with identical age distribution. OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index; PLMSI, periodic limb movement in sleep index.
Significance level: $p < 0.05$.

4.4.5 Reproducibility test

The histograms of the CAP rate for MrOS Visit 1 and Visit 2 (mean difference: 6 years) illustrate the reproducibility of the automated system for CAP detection (Supplementary Figure 4.6). Both histograms demonstrate an identical distribution with a minor shift for Visit 2 (Mann–Wilcoxon U-test: $p = 0.091$; Spearman correlation: $\rho = 0.38$, $p < 0.001$). Detailed values for CAP parameters and traditional polysomnographic indices of sleep disturbance for matched participants in the baseline and follow-up study are tabulated in Supplementary Table 4.7.

4.5 Discussion

Our analysis showed that age is independently associated with the CAP rate in older populations. Multivariable regression analysis, adjusting for sleep disturbance indices such as AI-NREM, OAHI, and PLMSI, showed that CAP rate in MrOS and SOF indicated a significant or close to significant association with age. Although all participants were within a narrow age range (68–90 years), our findings are consistent with previous studies that have shown that across the entire age spectrum CAP rate follows generally a U-shaped function of age (Parrino *et al.*, 1998) and thus continuously increases with age in older populations. On the contrary, the frequency of A1-phases decreases linearly with age. The negative beta coefficients in the multivariable regression analysis for A1-phases in this study confirm this behaviour. Interestingly, multivariable regression in SOF demonstrated also a negative association between CAP rate and age, while this association was positive in MrOS. This could be caused by the higher age in SOF, considering that the CAP rate diminishes at very high ages (Parrino *et al.*, 2012), or reflect gender differences in CAP rates with advanced ageing.

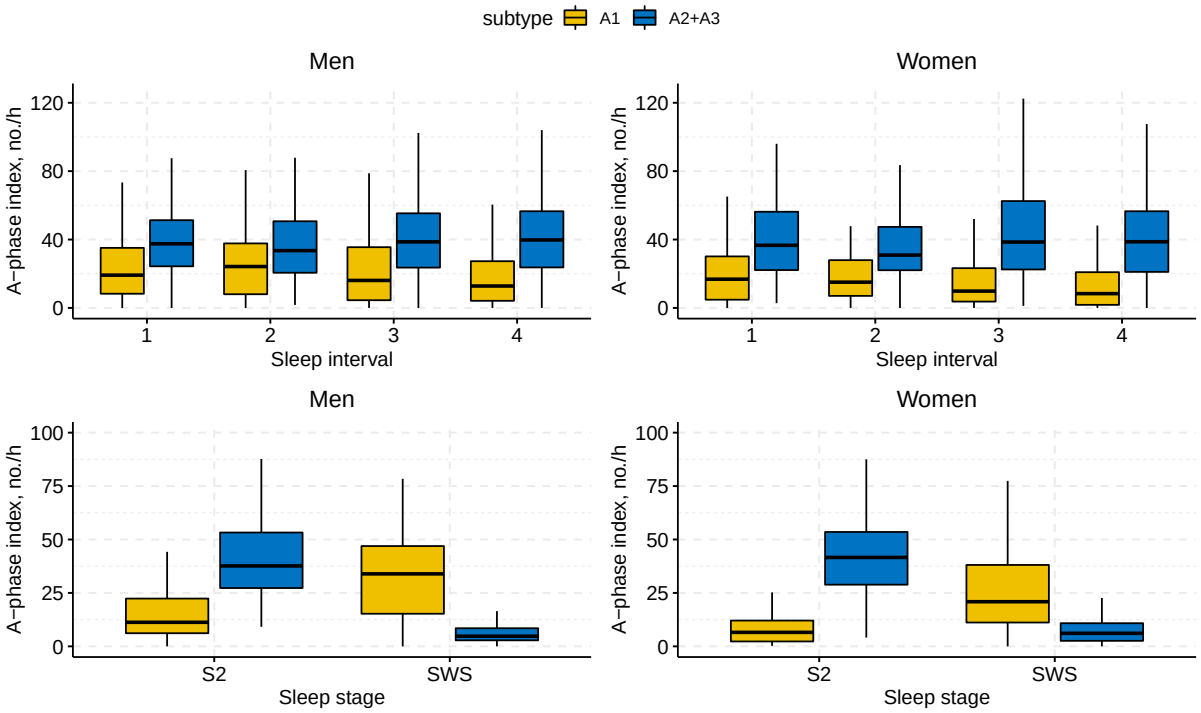


Figure 4.4. Analysis of A-phase subtypes during sleep intervals and non-rapid eye movement stages based on gender. Average number of A-phase subtypes per hour of none-rapid eye movement (NREM) sleep across consecutive 90-min intervals (top), as well as NREM sleep stages (bottom) for men (left) and women (right) with identical age distribution in Osteoporotic Fractures in Men (MrOS), Study and Study of Osteoporotic Fractures (SOF), respectively.

Further, women appeared to experience fewer A1-phases per hour throughout the night. CAP rate and A2+A3 index were comparable throughout the night for both men and women, resulting in a higher A1-to-A2+A3 ratio for men. Women did not show any significant variations in indices for sleep disturbance (AI-NREM, OAHI, and PLMSI) with CAP compared to men. In S2 and SWS, men exhibited a higher number of A1-phases compared to women, whereas the A2+A3 remained approximately identical to women. Although women demonstrated a higher percentage of SWS (which is characterized by a higher A1 index), men still experienced a higher number of A1-phases throughout the night. One can speculate that older women show less periodic activity in lower EEG frequency bands than older men due to more isolated or monomorphic events. This gender difference may provide clinical insight into the contradictory observations that compared to men, women have more SWS but report more frequent concerns over sleep quality (Redline *et al.*, 2004; Bixler *et al.*,

2009). Our data suggest that the microarchitecture of SWS in men and women may differ. Future research is needed to examine whether differences in the A1 index may explain gender differences in sleep quality.

Considering the number of A-phase subtypes per hour reported in this study, the A1 index tends to be lower than reported in the works of Parrino *et al.* (2012) and Maestri *et al.* (2015), whereas the A2+A3 indices are comparable. One reason for the lower frequency of A1-phases could be the higher age of the participants in MrOS and SOF compared to a previous study (Maestri *et al.*, 2015) as the frequency of A1-phases steadily declines with age. Another reason could be the low representation of older people in the data set used to train our automated CAP detection system.

Our data confirm the link between CAP and markers of sleep disturbances, suggesting that respiratory and leg movement events and increased arousals fragment the sleep microstructure. Across the large population samples, we observed a significant correlation between CAP and AI-NREM. Arousals have by definition a broad overlap with the characters of subtypes A2 and A3 and thereby CAP. Previous studies have shown that the majority of arousals (87%) appear within a CAP sequence (Terzano *et al.*, 2002). Moreover, 95% of subtypes A3 and 62% of subtypes A2 meet the scoring requirements for arousals (Parrino *et al.*, 2001). We observed only a moderate relationship between the A2+A3 index and AI-NREM, possibly because arousals were scored manually while CAP events were detected automatically using our system. Due to possibly low representations of arousals in general and overlaps with A-phases in the training set of our detection system, high variations in correlations with the consistent automated scoring of CAP events can be expected. Furthermore, our study confirms the connection between CAP and sleep pathologies such as sleep-disordered breathing or PLMS disorder, respectively. Our analysis depicts an inverse linear association between OAHI and the A1 subtype and, conversely, a positive association between OAHI for A2+A3-phases, analogous to results found in children with OSAS (Kheirandish-Gozal *et al.*, 2007). This shift in the subtype ratio raises sleep instability and has a severe negative impact on the NREM sleep microstructure (Parrino *et al.*, 2012). Data from middle-aged persons with OSAS also support our findings (Terzano *et al.*, 1996). According to Terzano *et al.* (1996), B-phases appear to be connected to episodes of breathing cessations, whereas A-phases seem to be linked to the recovery of effective breathing. Regarding PLMS, the majority of limb movements was reported to occur with the onset of A-phases and follow the periodicity of CAP (Parrino *et al.*, 1996). Our results demonstrate

a significant increase of PLMS in people with higher CAP rates. Thus, individuals with a high CAP occurrence are more likely to experience disruptive sleep.

An additional finding of our study is the inverse relationship between CAP rate, in particular, the frequency of fast low-amplitude EEG rhythms (A2+A3), and self-reported sleep quality that is independent of clinical markers of sleep disturbances. Our results show that the CAP rate, mainly the frequency of A2+A3-phases, is reduced in older men who report good sleep quality. ANCOVA of the A1 index did not show any significant relationship with self-reported sleep quality measures, possibly due to the low occurrence of A1-phases. Our results are in line with previously reported outcomes on correlations between CAP and self-reported sleep quality, mostly quantified by means of visual analogue scales (VAS). Terzano *et al.* (1990) suggested the first time a possible relation between CAP and self-reported sleep quality in their study on the influence of increasing levels of acoustic perturbation on sleep structure. Such a negative correlation between CAP rate and self-reported sleep quality has been confirmed by larger studies in subsequent years (Svetnik *et al.*, 2010; Ozone *et al.*, 2008; Terzano *et al.*, 2003). We show in our analysis that a negative correlation between CAP and self-reported sleep quality is independent of age and sleep disturbance reflected in OAHI, AI-NREM, and PLMSI. Our results also show a strong relationship between AI-NREM and self-reported sleep measures although objective sleep quality measures derived from PSG have shown not to be suitable predictors for individually reported quality of sleep especially in older adults (Kaplan *et al.*, 2017; Buysse *et al.*, 1991). Our data are in agreement with the correlation between AI and VAS reported by Terzano *et al.* (2003).

To the best of our knowledge, this effort represents the first time that CAP was scored and analysed in large community-based cohort studies. Commonly, the scoring of CAP is performed manually, which is a tedious and time-consuming task for the scorer, considering that one recording consists of multiple hours of sleep. The low number of sleep technicians trained in CAP scoring and the immense volume of work required have likely been barriers that prevented CAP studies with large numbers of participants in the past. Previous studies on CAP were limited to 10–50 recordings with rare exceptions such as the work of Terzano and Parrino (1993) that included 400–500 persons. Here, we evaluated in total 3,237 participants (MrOS: 2,811, SOF: 426) using a high-performance automated detection algorithm that enabled in an unprecedented examination of CAP in elderly male and female populations.

A limitation of this study is the older age of the participants, precluding assessment across the full age range. Nonetheless, sleep disorders and quality are of particular relevance in

older populations. Previous studies have shown that sleep fragmentation is highly prevalent among older people (Carskadon *et al.*, 1982; Bonnet and Arand, 2007). Another limitation pertains to the accuracy of our developed automated detection system. Although the system has demonstrated outstanding performance in comparison to manual scoring as the gold standard (Hartmann and Baumert, 2019), the results depend on the training data set, i.e. it may be biased to the human expert who scored the training data. Visual scoring may allow a human scorer to consider subject-specific variations, whereas our system will strictly score events based on the representations of events in the training set. On the other hand, the inter-rater concordance for manual CAP scoring is approximately 83% (Ferri *et al.*, 2005b), reflecting some inconsistency even between human scorers. Hence, imperfections in our automated detection system can also been observed in scoring results from multiple human experts. The high reproducibility of our system is evident when comparing repeated measurements between MrOS Sleep Visit 1 and Visit 2. The statistics demonstrate an identical CAP rate distribution for Visit 1 and Visit 2 with a non-significant shift in Visit 2 due to the time gap of 6 years. Also, the automated scoring results are easily reproducible as the automated analysis of one recording takes only a few seconds, unlike manual scoring results. Furthermore, the method was implemented in MATLAB, a numerical computing environment, mostly using a built-in function that makes it easy to reproduce.

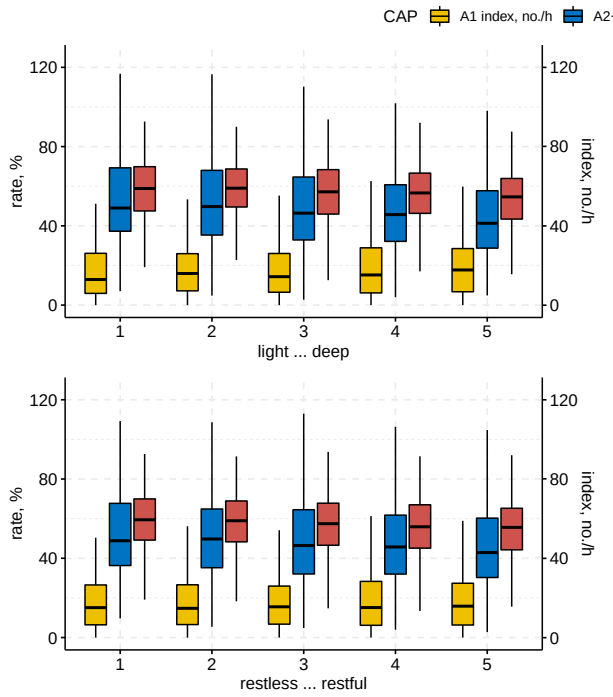
In sum, these findings confirm in large community-based cohort studies that CAP scoring serves as an important indicator for sleep quality and sleep fragmentation in older populations. Moreover, the results provide fundamental data on the variation of CAP in older adults, providing the bases for future studies of the relationship of CAP with health outcomes. Finally, automated scoring systems such as the algorithm employed in this study can assist in analysing CAP in large populations.

4.6 Funding

This study was partly supported by a grant from the Australian Research Council (DP0663 345). The NSRR is supported by grant number HL114473 from the National Heart, Lung, and Blood Institute (NHLBI), NIH. The NHLBI provides funding for the MrOS Sleep ancillary study “Outcomes of Sleep Disorders in Older Men” under the following grant numbers: R01 HL071194, R01 HL070848, R01 HL070847, R01 HL070842, R01 HL070841, R01 HL070837, R01 HL070838, and R01 HL070839. The Study of Osteoporotic Fractures (SOF) is supported by the National Institutes of Health funding. The National Institute on

Aging provides support under the following grant numbers: R01 AG005407, R01 AR35582, R01 AR35583, R01 AR35584, R01 AG005394, R01 AG027574, R01 AG027576, and R01 AG026720.

4.7 Supplemental material



ANCOVA
(significance level: p < 0.017
adjusted according to the number of variables
under consideration)

	light ... deep	*short ... long*	*restless ... restful*
CAP rate, %	<0.001 *	<0.001 *	0.002 *
A1 index, no./h	0.19	0.76	0.94
A2+A3 index, no./h	<0.001 *	<0.001 *	<0.001 *

Figure 4.5. CAP parameters in relation to subjective sleep quality measures in MrOS. The relationship between cyclic alternating pattern (CAP) rate, A1 index, and A2+A3 index and subjective sleep quality measures in Osteoporotic Fractures in Men Sleep Study (MrOS). Significance level (* = $p < 0.017$) was adjusted according to the number of variables under consideration.

Table 4.3. Analysis of covariance (ANCOVA) for effect of different levels from light to deep sleep on CAP rate, A1 index, and A2+A3 index in MrOS.

	CAP rate, %			A1 index, no./h			A2+A3 index, no./h		
	Df	F	*p*	Df	F	*p*	Df	F	*p*
Independent									
light . . . deep	4	5.9	<0.001*	4	1.5	0.19	4	10.2	<0.001*
Covariate									
Age, years	1	22.2	<0.001*	1	7.1	<0.01*	1	33.8	<0.001*
AI-NREM, no./h	1	264.7	<0.001*	1	116.8	<0.001*	1	429.3	<0.001*
OAHI, no./h	1	0.3	0.60	1	0.2	0.70	1	6.2	0.01*
PLMSI, no./h	1	2.3	0.13	1	15.1	<0.001*	1	35.6	<0.001*

Results of analysis of covariance (ANCOVA) in Osteoporotic Fractures in Men (MrOS) Study with cyclic alternating pattern (CAP) rate, A1 index, and A2+A3 index as dependent variables, the subjective sleep quality measure light to deep sleep as independent variable and age, the obstructive apnoea–hypopnoea index at 4% oxygen desaturation (OAHI), the arousal index (AI-NREM) in NREM sleep and the periodic limb movement in sleep index (PLMSI) as covariates. MrOS, Osteoporotic Fractures in Men; B, estimate of beta coefficient; SE, standard error of beta coefficient; OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index; PLMSI, periodic limb movement in sleep index.
significance level: $p < 0.017$ (adjusted to the number of variables under consideration)

Table 4.4. Analysis of covariance (ANCOVA) for effect of different levels from short to long sleep on CAP rate, A1 index, and A2+A3 index in MrOS.

	CAP rate, %			A1 index, no./h			A2+A3 index, no./h		
	Df	F	*p*	Df	F	*p*	Df	F	*p*
Independent									
short . . . long	4	4.9	<0.001*	4	0.2	0.94	4	6.8	<0.001*
Covariate									
Age, years	1	22.5	<0.001*	1	7.3	<0.01*	1	33.0	<0.001*
AI-NREM, no./h	1	264.4	<0.001*	1	122.7	<0.001*	1	434.7	<0.001*
OAHI, no./h	1	0.2	0.62	1	0.2	0.65	1	6.4	0.01*
PLMSI, no./h	1	2.0	0.16	1	15.2	<0.001*	1	35.2	<0.001*

Results of analysis of covariance (ANCOVA) in Osteoporotic Fractures in Men (MrOS) Study with cyclic alternating pattern (CAP) rate, A1 index, and A2+A3 index as dependent variables, the subjective sleep quality measure short to long sleep as independent variable and age, the obstructive apnoea–hypopnoea index at 4% oxygen desaturation (OAHI), the arousal index (AI-NREM) in NREM sleep and the periodic limb movement in sleep index (PLMSI) as covariates. MrOS, Osteoporotic Fractures in Men; B, estimate of beta coefficient; SE, standard error of beta coefficient; OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index; PLMSI, periodic limb movement in sleep index.
significance level: $p < 0.017$ (adjusted to the number of variables under consideration)

Table 4.5. Analysis of covariance (ANCOVA) for effect of different levels from restless to restful sleep on CAP rate, A1 index, and A2+A3 index in MrOS.

	CAP rate, %			A1 index, no./h			A2+A3 index, no./h		
	Df	F	*p*	Df	F	*p*	Df	F	*p*
Independent									
restless ... restful	4	4.4	0.002*	4	0.5	0.94	4	5.9	<0.001*
Covariate									
Age, years	1	23.2	<0.001*	1	7.0	<0.01*	1	33.8	<0.001*
AI-NREM, no./h	1	261.6	<0.001*	1	124.5	<0.001*	1	429.3	<0.001*
OAHI, no./h	1	0.3	0.59	1	0.2	0.70	1	6.2	0.01*
PLMSI, no./h	1	2.1	0.15	1	15.8	<0.001*	1	35.6	<0.001*

Results of analysis of covariance (ANCOVA) in Osteoporotic Fractures in Men (MrOS) Study with cyclic alternating pattern (CAP) rate, A1 index, and A2+A3 index as dependent variables, the subjective sleep quality measure restless to restful sleep as independent variable and age, the obstructive apnoea–hypopnoea index at 4% oxygen desaturation (OAHI), the arousal index (AI-NREM) in NREM sleep and the periodic limb movement in sleep index (PLMSI) as covariates. MrOS, Osteoporotic Fractures in Men; B, estimate of beta coefficient; SE, standard error of beta coefficient; OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index; PLMSI, periodic limb movement in sleep index.
significance level: $p < 0.017$ (adjusted to the number of variables under consideration)

Table 4.6. Analysis of covariance (ANCOVA) for effect of subjective sleep quality measures on AI-NREM in MrOS.

	AI-NREM, no./h								
	Df	F	*p*		Df	F	*p*		Df	F	*p*
Dependent											
light ... deep	4	37.8	<0.001*	short ... long	4	32.3	<0.001*	restless ... restful	4	37.7	<0.001*
Covariate											
Age, years	1	50.7	<0.001*	Age, years	1	54.6	<0.001*	Age, years	1	56.2	<0.001*
CAP rate, %	1	339.4	<0.001*	CAP rate, %	1	338.4	<0.001*	CAP rate, %	1	335.9	<0.001*
OAHI, no./h	1	724.9	<0.001*	OAHI, no./h	1	718.9	<0.001*	OAHI, no./h	1	721.1	<0.001*
PLMSI, no./h	1	55.2	<0.001*	PLMSI, no./h	1	55.3	<0.001*	PLMSI, no./h	1	54.0	<0.001*

Results of analysis of covariance (ANCOVA) in Osteoporotic Fractures in Men (MrOS) Study with the arousal index (AI-NREM) in NREM sleep as dependent variable, the three subjective sleep quality measures as independent variables and age, cyclic alternating pattern (CAP) rate, the obstructive apnoea–hypopnoea index at 4% oxygen desaturation (OAHI), and the periodic limb movement in sleep index (PLMSI) as covariates. MrOS, Osteoporotic Fractures in Men; B, estimate of beta coefficient; SE, standard error of beta coefficient; OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index; PLMSI, periodic limb movement in sleep index.
significance level: $p < 0.017$ (adjusted to the number of variables under consideration)

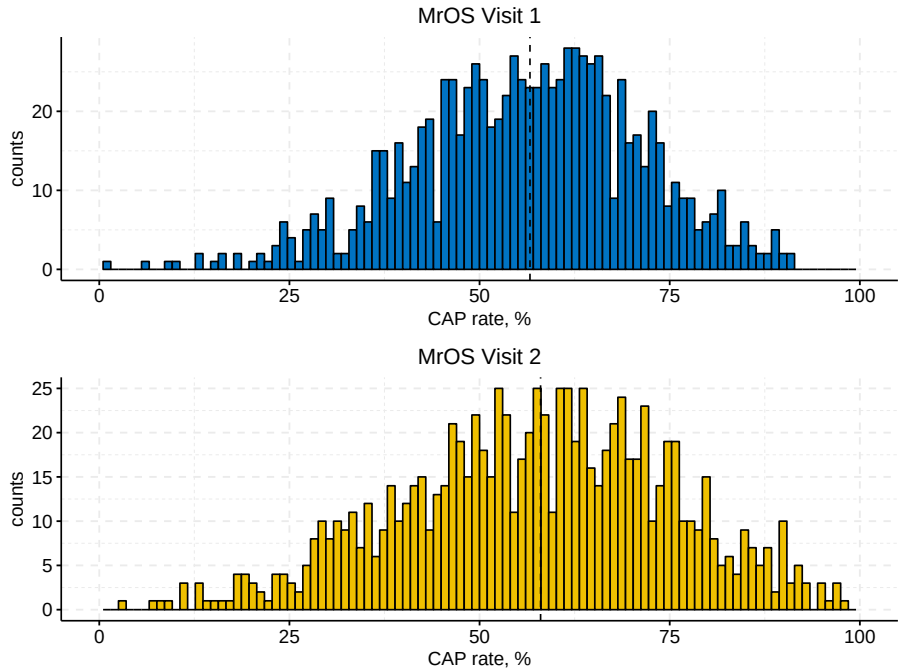


Figure 4.6. Histograms of CAP rate for MrOS Visit 1 and Visit 2. Histograms of cyclic alternating pattern (CAP) rate for identical subjects in Osteoporotic Fractures in Men (MrOS) Study Visit 1 and Visit 2. The median value for each distribution is illustrated by the dotted lines.

Table 4.7. MrOS Visit 1 and Visit 2 with identical subjects.

	Age, years	AI-NREM, no./h	OAHI, no./h	CAP rate, %	A1 index, no./h	A2+A3 index, no./h
	Median (IQR)					
MrOS Visit 1	74.0 (±7.0)	21.9 (±16.0)	8.0 (±13.1)	56.6 (±19.6)	16.0 (±20.3)	45.9 (±29.5)
MrOS Visit 2	80.0* (±7.0)	22.7 (±18.5)	7.9 (±15.4)	58.0 (±24.2)	16.9 (±22.9)	44.6 (±38.3)

Comparison of clinical markers of sleep disturbance between identical subjects in Osteoporotic Fractures in Men (MrOS) Study Visit 1 and Visit 2. MrOS, Osteoporotic Fractures in Men; OAHI, obstructive apnoea–hypopnoea index; AI-NREM, arousal index.
* significance level: $p < 0.05$

Paper III

Cyclic alternating pattern in children with obstructive sleep apnoea and its relationship with adenotonsillectomy, behaviour, cognition, and quality of life

The content of this chapter is a modified version of the publication:

Hartmann, S., Bruni, O., Ferri, R., Redline, S. and Baumert, M. (2020), 'Cyclic alternating pattern (CAP) in children with obstructive sleep apnea and its relationship with adenotonsillectomy, behavior, cognition, and quality-of-life', *Sleep* **44**(1), zsaa145.

Abstract

Study Objectives: To determine in children with obstructive sleep apnoea (OSA) the effect of adenotonsillectomy (AT) on the cyclic alternating pattern (CAP) and the relationship between CAP and behavioural, cognitive, and quality-of-life measures.

Methods: CAP parameters were analysed in 365 overnight polysomnographic recordings of children with mild-to-moderate OSA enrolled in the Childhood Adenotonsillectomy Trial (CHAT), randomized to either early AT (eAT) or watchful waiting with supportive care (WWSC). We also analysed CAP in a subgroup of 72 children with moderate OSA (apnoea-hypopnoea index > 10) that were part of the CHAT sample. Causal mediation analysis was performed to determine the independent effect of changes in CAP on selected outcome measures.

Results: At baseline, a higher number of A1 phases per hour of sleep was significantly associated with worse behavioural functioning (caregiver Behaviour Rating Inventory of Executive Function (BRIEF) Global Executive Composite (GEC): ρ = 0.24, p = 0.042; caregiver Conners' Rating Scale Global Index: ρ = 0.25, p = 0.036) and lower quality of life (OSA-18: ρ = 0.27, p = 0.022; PedsQL: ρ = -0.29, p = 0.015) in the subgroup of children with moderate OSA, but not across the entire sample. At 7-months follow-up, changes in CAP parameters were comparable between the eAT and WWSC arms. CAP changes did not account for significant proportions of variations in behavioural, cognitive, and quality-of-life performance measures at follow-up.

Conclusions: We show a significant association between the frequency of slow, high-amplitude waves with behavioural functioning, as well as the quality of life in children with moderate OSA. Early AT in children with mild-to-moderate OSA does not alter the microstructure of non-rapid eye movement sleep compared with watchful waiting after an approximately 7-month period of follow-up.

5.1 Statement of Significance

Obstructive sleep apnoea (OSA) in children severely affects their behaviour and most likely increases the risk of developing cardiovascular disease. The adverse effect of OSA on

the sleep macrostructure of children has been extensively investigated, but its impact on non-rapid eye movement (NREM) sleep microstructure remains unclear. To ascertain the relationship between cyclic alternating pattern (CAP), characterizing sleep microstructure, and adenotonsillectomy (AT), behaviour, cognition, and quality of life, we investigated 365 overnight polysomnographic recordings of children with mild-to-moderate OSA enrolled in the Childhood Adenotonsillectomy Trial (CHAT). Children with moderate OSA who experience a high frequency of slow, high-amplitude waves (A1 phases) display worse behavioural functioning and score lower in caregiver-rated quality of life. At 7-months follow-up, children who underwent AT showed no differences in CAP measures compared with children in the watchful waiting group, indicating that early AT does not yield an additional benefit in terms of NREM sleep instability; in both treatment arms, CAP rate, especially the A1 index, was increased at follow-up.

5.2 Introduction

Obstructive sleep apnoea (OSA) is the most severe form of upper airway disease during sleep and found among 1% to 4% of children (Lumeng and Chervin, 2008). Compared with unaffected children, those with OSA are more likely to present with behavioural problems (Ali *et al.*, 1993) and are at increased risk for developing cardiovascular and metabolic risk factors such as systemic hypertension (Marcus *et al.*, 1998) and higher levels of C-reactive protein (CRP) (Larkin *et al.*, 2005). First-line treatment for childhood OSA is commonly adenotonsillectomy (AT) as enlarged tonsils and adenoids often result in narrowing the upper airway structure or pharyngeal collapse, leading to snoring and periods of apnoea and hypopnoea Typically, apnoeas and hypopnoeas affect the quality and quantity of restorative sleep due to subsequent sleep stage transitions to lighter sleep, arousals, or periods of wakefulness.

By convention, sleep is scored by the rules of Rechtschaffen and Kales (1968) adopted and modified by the American Academy of Sleep Medicine (AASM) scoring manual (Iber *et al.*, 2007). In the AASM scoring, electroencephalography (EEG) arousals are defined as abrupt changes in EEG of a minimum duration of 3 seconds and do not consider 1 to 2-second activities observed in children (Grigg-Damberger *et al.*, 2007), which consequently may limit the clinical utility of the conventional assessment of macrostructure in children. Cyclic alternating pattern (CAP) analysis is a method to assess the microstructure of sleep. It captures dynamic changes in EEG amplitude and frequency that recur periodically in non-rapid eye movement (NREM) stages (Terzano *et al.*, 2001). Such sequences of recurring

activation phases represent periods of high neural excitability with an intermittent background period in between (Terzano and Parrino, 2000). As they coincide with physiological and pathological events, CAP provides insights into the fragmentation of NREM sleep not possible with traditional sleep staging. Previous studies on the relationship between CAP and pathologies such as sleep disordered breathing (SDB), narcolepsy, and neuropsychological disabilities in paediatric populations concluded that children with these pathologies almost always show less synchronized slow-wave activity than healthy children (Bruni *et al.*, 2010b).

The aim of this study was to investigate the relationship between NREM sleep microarchitecture and behavioural, cognitive, and quality-of-life measures in children with mild-to-moderate OSA and the effect of AT on these measures. We report data from the randomized controlled Childhood Adenotonsillectomy Trial (CHAT; ClinicalTrials.gov identifier: NCT00560859). Previous analysis on the CHAT study has shown a larger decrease in arousals, a larger decrease in the percentage of N1 sleep, and a greater increase in the percentage of N2 sleep but no change in N3 or rapid eye movement (REM) sleep following surgery compared with watchful waiting and supportive care (Marcus *et al.*, 2013). Further studies reported a similar reduction in the number of arousals post-surgery (Suen *et al.*, 1995; Frank *et al.*, 1983; Brouillette *et al.*, 1982). However, it is currently unclear whether AT improves the NREM microstructure represented by CAP in children with OSA. To comprehensively probe NREM microstructure, we determined CAP in this sample. We also analysed a subgroup of children enrolled in CHAT who presented with moderate OSA (apnoea-hypopnoea index [AHI] > 10) to test whether treatment results in a greater improvement in polysomnographic (PSG) findings and CAP parameters.

5.3 Methods

5.3.1 Definition and detection of CAP

In agreement with the atlas and rules for scoring CAP published by Terzano *et al.* (2001), we defined CAP as sequences of at least two consecutive CAP cycles. A CAP cycle consists of one activation phase (A phase) that represents transient, phasic events and an intervening background phase (B phase) that separates two successive A phases. Stereotypical A-phase patterns are delta bursts, vertex sharp transients, K-complex sequences, K-alpha, polyphasic bursts, intermittent alpha, and arousals (Terzano *et al.*, 2001). We subdivided

A phases into periods of slow high-amplitude waves (A1), fast low-amplitude EEG rhythms (A3), or a mixture of both (A2). A1 phases portray synchronized EEG patterns, while A3 phases represent desynchrony (Terzano and Parrino, 2000). We defined A phases or B phases to last 2–60 seconds and restricted their occurrence to only NREM periods. Thus, REM periods and the periods between two CAP sequences were considered non-CAP. In this study, each CAP sequence was terminated by an A phase that was assigned to the following non-CAP period.

We analysed CAP in overnight EEG recordings utilizing our previously developed, highly precise automated system for CAP analysis (Hartmann and Baumert, 2019). Our CAP detection system is a deep learning recurrent neural network that was trained with manually scored recordings from children. In the first step, the EEG channel is filtered and processed to remove powerline interference, noise, and cardiac field artefacts. Time and spectral features (Hjorth activity, Shannon entropy, Teager energy operator, band power descriptor, and differential EEG variance) were extracted from the processed signal and passed into the classifier as input values. The extracted features were sampled at 1 Hz, yielding a classification output that indicates whether the current second is part of an A phase or not and if so, what kind of A phase. We selected the F_β-score as loss function for training to deal with the imbalanced dataset and increase the sensitivity and precision of the classification. Per the previously defined rules for CAP sequences, we post-processed the output of the A-phases detection system, i.e. isolated A phases and B phases, less than two CAP cycles, and the terminating A phases were removed.

We used 19 recordings of healthy children, 15 recordings of healthy adults, and 24 recordings of adults with sleep disorders as training set to cope with the inhomogeneous EEG characteristics of children and adults. Our system has previously demonstrated a second-by-second A-phase inter-rater reliability, quantified by the Cohen's kappa coefficient, of 0.53 on a set of 16 healthy participants and 0.56 on a set of 30 participants with nocturnal frontal lobe epilepsy (Hartmann *et al.*, 2020). On the contrary, the event-based inter-rater reliability between human scorers ranges between 0.42 and 0.75 (Ferri *et al.*, 2005b). The specifics of the training data and additional information on the performance of the deployed automated scoring system are provided in supplementary section 5.7.1.

Here, we used the left and the right central EEG channels, re-referenced to the mastoid channels, in each PSG recording only counting overlapping A phases to increase the sensitivity of the classification. We defined the CAP rate as the percentage of NREM sleep that is covered by CAP sequences. Subtype indices represent the number of A1 and A2 + A3,

respectively, per hour of NREM sleep. Subtypes A2 and A3 were merged into a single parameter due to their congruent nature. Additionally, the number, duration, and percentage of A1 and A2 + A3 phases were measured as well as the duration of B phases, CAP cycles, and CAP sequences.

5.3.2 Childhood Adenotonsillectomy Trial

We used overnight PSG from the CHAT, a multi-center, single-blind, randomized, controlled trial designed to analyse the efficacy of early AT (eAT) on children with mild-to-moderate OSA. The study tested whether children with mild-to-moderate OSA, randomized to eAT, demonstrate greater improvement in cognitive, behavioural, quality-of-life, and sleep measures at 7-months follow-up than children who were randomly assigned to watchful waiting with supportive care (WWSC) (Marcus *et al.*, 2013). Children between 5.0 and 9.99 years of age diagnosed with OSA (OAI, number of obstructive apnoeas per hour of sleep, $\geq$1, or AHI $\geq$ 2), tonsillar hypertrophy $\geq$ 1, and cleared for surgery by an otolaryngologist were recruited in six clinical sites in the United States (Redline *et al.*, 2011).

We removed recordings with fewer than 3 hours of good EEG quality and those from one specific clinical site (clusterid 13) due to equipment-related issues (n = 23). In consequence, 365 overnight recordings were evaluated (179 eAT and 186 WWSC). Also, we evaluated a subgroup of 72 children (38 eAT and 34 WWSC) with moderate SDB (AHI > 10) out of the 365 recordings as those children are likely to show a greater improvement in PSG findings (Marcus *et al.*, 2013) and CAP parameters. The dataset is available at the National Sleep Research Resource (NSRR) (available online at the National Sleep Research Resource; sleepdata.org) (Zhang *et al.*, 2018).

5.3.3 Outcome measures

The primary outcome measure of CHAT was the Attention/ Executive (A/E) Functioning Domain Index from the Developmental Neuropsychological Assessment (NEPSY). Secondary outcome measures include indices of behaviour, sleep symptoms, generic and disease-specific quality of life, PSG measures of sleep apnoea, anthropometric measures, and blood pressure.

As the relationship between CAP and the child's behaviour and cognitive performance is of great interest (Bruni *et al.*, 2012; Bruni and Ferri, 2009), we analysed the causal mediation

of changes in the A/E Functioning Domain Index and secondary cognitive and behavioural outcomes (Behaviour Rating Inventory of Executive Function [BRIEF] Global Executive Composite [GEC] T score, and Conners' Rating Scale Global Index T score) by variations in CAP parameters (Redline *et al.*, 2011). We also included the key quality-of-life measure in CHAT, the caregiver-rated total score from the Paediatric Quality of Life Inventory (PedsQL), as well as the disease-specific quality-of-life total score on the 18-item OSA-18 assessment tool. To assess symptoms of the OSA syndrome, we added the Paediatric Sleep Questionnaire sleep-related breathing disorder scale (PSQ-SRBD) to the list of outcome measures. Higher scores on the BRIEF GEC T score, Conners' Rating Scale Global Index T score, OSA-18 score, and PSQ-SRBD scale indicate worse functioning, worse quality of life, or greater severity, respectively. On the contrary, higher scores on NEPSY A/E Functioning score and PedsQL caregiver-rated score represent better functioning and better quality of life, respectively. Finally, we included in the causal mediation analysis the change in AHI defined as the number of all obstructive and mixed apnoeas, plus hypopnoeas associated with either a $\geq$ 3% desaturation or electroencephalographic arousal, per hour of sleep, and the periodic limb movement sleep index (PLMSI) defined as the number of periodic limb movement (PLM) per hour of NREM.

5.3.4 Statistical methods

Statistical analysis was conducted using non-parametric tests assuming that CAP rate and subtype indices do not follow a normal distribution. The relationship between baseline measurements of CAP parameters and age, arousal index during NREM (AI-NREM), AHI, PLMSI, behavioural, cognitive, as well as quality-of-life performance measures was examined using the Spearman correlation coefficient. For each statistical test, the significance level was $p < 0.05$.

For the evaluation of treatment-specific changes in CAP parameters, including CAP rate, subtype indices, total number of subtypes, subtype percentages, mean duration of subtype phases, the average duration of B phases, the average duration of CAP cycles, and the average duration of CAP sequences, we applied two-factor repeated-measures analysis of variance (ANOVA) with time and treatment as factors and adjusting for the stratification factors of age, race, weight status, and study site.

We used the causal mediation analysis described by Imai *et al.* (2010), to identify the independent effect of CAP changes on outcome measures (Figure 5.1). The model describes

the total effect of treatment as a sum of the mediated effect using an independent mediator and the direct effect. Here, we analysed three mediation models with CAP rate, A1 index, and A2 + A3 index, respectively, as mediators. We used a linear regression model where the pretreatment covariates were identical to those used in the repeated-measures ANOVA. We log-transformed the change of AHI due to its non-normal distribution. We applied the mediation package for R with 1,000 non-parametric bootstrap resamples (Tingley *et al.*, 2014). Results include the estimates of average causal mediation effects, the average direct effects, the total effects, and the proportion of the mediated effects with 95% confidence intervals and the associated *p* values.

5.4 Results

Table 5.1 details the baseline characteristics for each outcome measure and their respective Spearman correlation value with CAP rate, A1 index, and A2 + A3 index for children with mild-to-moderate OSA (Table 5.1a) and the subgroup of children with moderate OSA (Table 5.1b).

In children with mild-to-moderate OSA at baseline, AI-NREM significantly correlated with the A2 + A3 index (ρ = 0.33, p = <0.001). PLMSI showed significant negative correlations with the A1 index and positive correlation with the A2 + A3 index (A1 index: ρ = -0.14, p = 0.009; A2 + A3 index: ρ = 0.15, p = 0.003). Other outcome measures did not show any significant correlations with CAP rate and subtype indices, respectively.

In the baseline subgroup of children with AHI > 10, age was significantly correlated with the CAP rate and the A1 index (CAP rate: ρ = 0.24, p = 0.047; A1 index: ρ = 0.28, p = 0.018) and significantly inversely correlated with the A2 + A3 index (ρ = -0.29, p = 0.014). AI-NREM showed significant correlations with the A2 + A3 index (ρ = 0.35, p = <0.001). On the contrary, A1 index demonstrated significant correlations with AHI (ρ = 0.24, p = 0.040) and significant inverse correlations with PLMSI (ρ = -0.27, p = 0.021). Regarding cognitive, behavioural, and quality-of-life measures, the A1 index demonstrated significant correlations with the caregiver BRIEF GEC T score (ρ = 0.24, p = 0.042), the caregiver Conners' Rating Scale Global Index T score (ρ = 0.25, p = 0.036), and the OSA-18 score (ρ = 0.27, p = 0.022). Finally, the PedsQL caregiver-rated score was significantly inversely correlated with the CAP rate and the A1 index (CAP rate: ρ = -0.25, p = 0.036; A1 index: ρ = -0.29, p = 0.015).

Table 5.1. Distributions and Spearman's correlation between CAP parameters and age, PSG sleep disturbance indices, and behavioural, cognitive, and quality-of-life measures at baseline in (a) children with mild-to-moderate OSA and (b) children with moderate OSA (AHI > 10).

a) Children with mild-to-moderate OSA (n = 365, 190 boys and 175 girls)							
		Spearman correlation					
	Baseline	CAP rate		A1 index		A2+A3 index	
	Median ± IQR	ρ	p	ρ	p	ρ	P
Age, years	6.0 ± 3.0	0.09	0.08	0.10	0.059	-0.02	0.69
BMI, kg/m^2	17.2 ± 6.5	0.06	0.26	0.10	0.06	-0.14	<0.01*
AI-NREM, no./h	7.8 ± 4.4	0.00	0.99	-0.07	0.20	0.33	<0.001*
AHI, no./h	4.6 ± 6.2	0.03	0.54	0.01	0.79	0.04	0.42
PLMSI, no./h	0.6 ± 2.4	-0.09	0.082	-0.14	0.009*	0.15	0.003*
NEPSY Attention/Executive Functioning Scaled Score	102.0 ± 3.0	-0.06	0.27	-0.09	0.099	0.05	0.33
BRIEF Global Executive Composite Total T Score							
Caregiver rating	49.0 ± 2.5	-0.04	0.39	-0.02	0.69	-0.05	0.31
Teacher rating	56.0 ± 17.0	-0.03	0.60	-0.02	0.70	-0.01	0.82
Conners' Rating Scale Global Index Total T Score							
Caregiver rating	50.0 ± 14.0	-0.04	0.49	-0.03	0.59	-0.03	0.61
Teacher rating	51.0 ± 17.0	-0.07	0.25	-0.04	0.53	-0.11	0.077
PedsQL caregiver-rated total score	81.7 ± 23.8	0.01	0.92	-0.01	0.86	0.06	0.22
PSQ-SRBD score	0.5 ± 0.26	-0.05	0.35	-0.03	0.54	-0.05	0.37
OSA-18 total score	51.0 ± 24.5	-0.01	0.82	-0.01	0.90	-0.01	0.78
b) Children with moderate OSA (AHI >10) (n = 72, 40 girls and 32 boys)							
		Spearman correlation					
	Baseline	CAP rate		A1 index		A2+A3 index	
	Median ± IQR	ρ	p	ρ	p	ρ	P
Age, years	6.0 ± 2.0	0.24	0.047*	0.28	0.018*	-0.29	0.014*
BMI, kg/m^2	17.6 ± 6.9	0.22	0.06	0.26	0.03*	-0.18	0.14
AI-NREM, no./h	9.6 ± 5.6	0.04	0.74	0.00	0.97	0.35	<0.01*
AHI, no./h	15.0 ± 7.8	0.22	0.065	0.24	0.040*	-0.18	0.13
PLMSI, no./h	0.8 ± 2.9	-0.22	0.064	-0.27	0.021*	0.14	0.25
NEPSY Attention/Executive Functioning Scaled Score	98.0 ± 22.8	-0.17	0.14	-0.20	0.093	0.10	0.39
BRIEF Global Executive Composite Total T Score							
Caregiver rating	47.0 ± 12.5	0.21	0.071	0.24	0.042*	-0.06	0.59
Teacher rating	57.5 ± 22.8	0.00	0.99	0.00	0.99	-0.10	0.47
Conners' Rating Scale Global Index Total T Score							
Caregiver rating	49.0 ± 16.0	0.23	0.051	0.25	0.036*	0.00	0.99
Teacher rating	52.0 ± 21.0	0.09	0.53	0.12	0.39	-0.15	0.28
PedsQL caregiver-rated total score	82.4 ± 29.4	-0.25	0.036*	-0.29	0.015*	0.11	0.34
PSQ-SRBD score	0.5 ± 0.3	0.07	0.56	0.12	0.32	-0.11	0.35
OSA-18 total score	56.0 ± 28.5	0.22	0.06	0.27	0.022*	-0.10	0.38

AI-NREM, Non-rapid eye movement sleep (NREM) Arousal Index; AHI, Obstructive Apnoea-Hypopnoea Index (AHI) >= 3% - number of [obstructive apnoeas] and [hypopnoeas with >=3% oxygen desaturation or arousal] per hour of sleep; PLMSI, Number of Periodic limb movement (PLM) per hour of non-rapid eye movement sleep (NREM); NEPSY, Developmental Neuropsychological Assessment (NEPSY); BRIEF, Behaviour Rating Inventory of Executive Function; PedsQL, Paediatric Quality of Life Inventory; PSQ-SRBD, Paediatric Sleep Questionnaire sleep-related breathing disorder scale, OSA-18, Obstructive Sleep Apnoea-18 assessment tool.

*significance level: $p < 0.05$

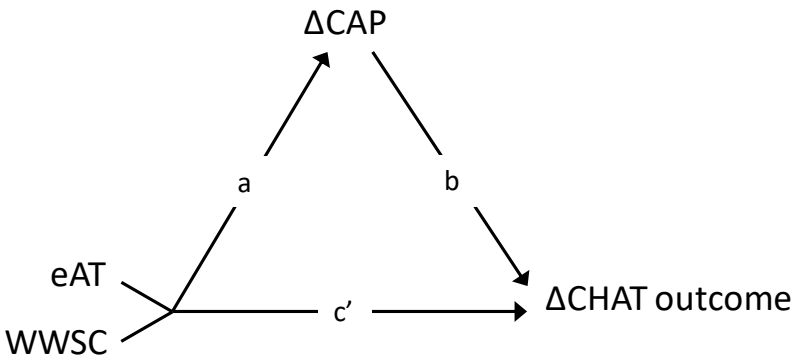


Figure 5.1. Causal mediation model to identify the independent effect of CAP changes on CHAT outcome measures.. Causal mediation diagram illustrating the direct effect of treatment on the CHAT outcome (c′), the effect of treatment on ΔCAP as mediator (a), and the effect of the ΔCAP as mediator on the CHAT outcome (b). The product of the paths (a) and (b) equals the indirect effect on the CHAT outcome.

5.4.1 The effect of AT on CAP

Table 5.2 details the change of CAP parameters from baseline to follow-up for children in the eAT and WWSC arms for the entire CHAT sample. The median CAP rate increased in both treatment groups by 1%–4%. The eAT group demonstrated a marginally higher increase but the interaction between treatment and time was not significant (p = 0.37). Similar outcomes were observed for both subtype indices. The A1 index increased by 3.8% from baseline to follow-up PSG for children undergoing eAT, whereas children in the WWSC arm displayed an increase of 1.1%. Both groups showed a similar trend for the A2 + A3 index with an overall lower increase (eAT: 1.3%, WWSC: 0.1%). However, repeated-measures ANOVA indicated no significant interaction between treatment and time for either subtype index (A1 index: p = 0.58, A2 + A3 index: p = 0.24). In line with the results for CAP rate and subtype indices, other CAP parameters did not show any significant interactions between treatment and time either.

Table 5.3 lists the change of CAP parameters from baseline to follow-up PSG for children in the eAT and WWSC arms in the subgroup of children with AHI > 10. CAP rate was increased in both treatment groups by 1%–2% with no significant interaction between treatment and time (p = 0.84). The A1 index showed only for the WWSC group an increase of 1%. On the contrary, the A2 + A3 index was only in the eAT group decreased by 1%. Repeated measures ANOVA of both subtype indices demonstrated no significant interaction between treatment and time (A1 index: p = 0.94, A2 + A3 index: p = 0.31). In line

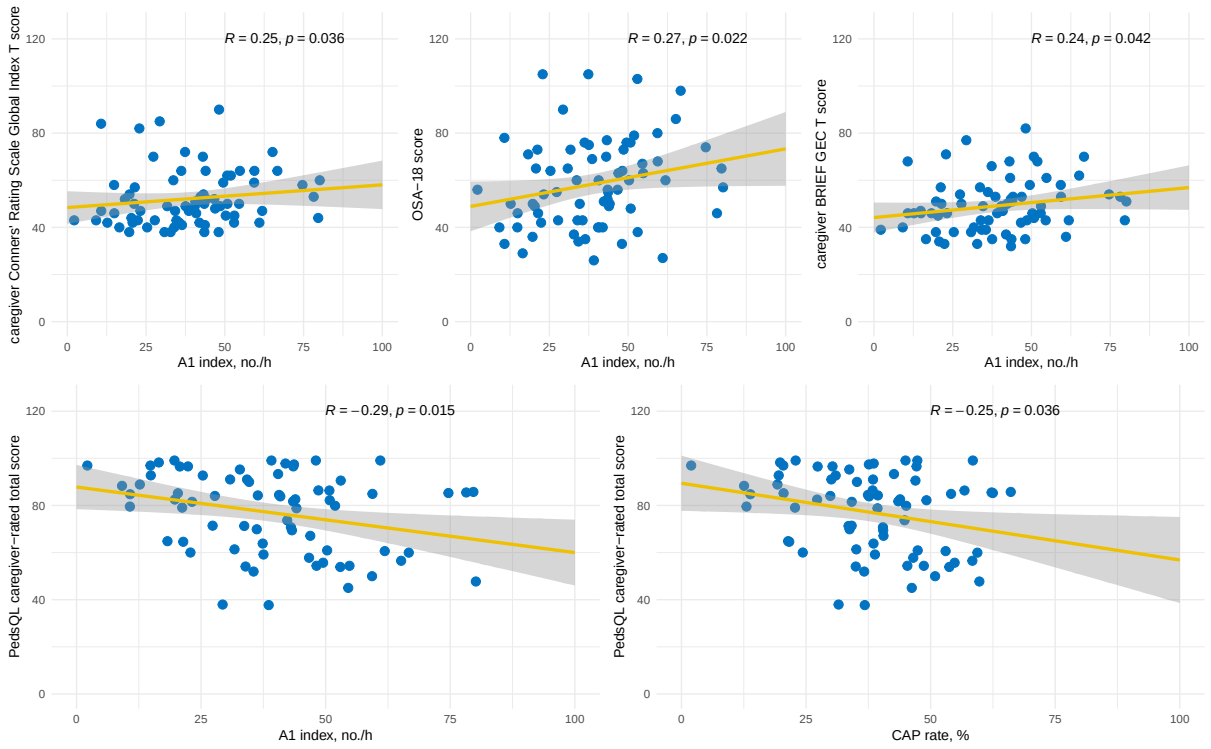


Figure 5.2. Significant Spearman's correlation in children with moderate OSA (AHI > 10). Scatter plots with linear regression line to highlight significant Spearman's correlation between the A1 index and the caregiver BRIEF GEC T score, the caregiver Conners' Rating Scale Global Index T score, the OSA-18 score, and the PedsQL caregiver-rated score as well as between the CAP rate and the the PedsQL caregiver-rated score at baseline in children with moderate OSA (AHI > 10).

with the results for CAP rate and subtype indices, other CAP parameters did not show any significant interaction between treatment and time either.

5.4.2 The effect of CAP changes on behaviour, cognitive performance, and quality of life

Supplementary Table 5.6 displays the results of the mediation analysis with CAP rate as the mediator in the entire study sample. The total effects of treatment were significant for Conners' Rating Scale Global Index scores, the caregiver BRIEF GEC score, the PedsQL caregiver-rated total score, the PSQ-SRBD scale, the OSA-18 score, and the AHI. No significant average mediation effects were identified for CAP rate. Similarly, neither A1 index nor A2 + A3 index showed a significant average mediation effect (Supplementary Tables 5.7 and 5.8). We obtained identical outcomes in the subgroup of children with moderate OSA.

Table 5.2. Comparison of CAP parameters between both randomized arms (eAT and WWS) at follow-up for children with mild-to-moderate OSA.

CAP measures	Early Adenotonsillectomy (n = 179)		Watchful Waiting (n = 186)		$p^{\dagger}$		
	Baseline	Change from Baseline to 7 months	Baseline	Change from Baseline to 7 months	Treatment‡	Time§	Treatment *Time‖
CAP rate, %	38.8 (±18.1)	3.5 (±19.7)	37.8 (±22.3)	1.2 (±16.9)	0.12	<0.01*	0.37
A1, no.	243.0 (±143.5)	21.0 (±148.5)	237.0 (±197.5)	11.0 (±130.8)	0.31	<0.01*	0.53
A2+A3, no.	41.0 (±36.0)	7.0 (±35.5)	33.0 (±36.0)	0.5 (±30.5)	0.045*	0.068	0.28
A1, %	84.5 (±11.1)	-0.7 (±9.2)	86.2 (±12.4)	0.7 (±11.8)	0.22	0.69	0.31
A2+A3, %	15.5 (±11.1)	0.0 (±0.1)	13.8 (±12.4)	0.0 (±0.1)	0.22	0.69	0.31
A1 mean duration, s	4.0 (±0.4)	0.0 (±0.5)	4.1 (±0.5)	0.0 (±0.5)	0.84	0.39	0.60
A2 + A3 mean duration, s	12.5 (±2.5)	0.2 (±3.1)	12.4 (±2.7)	0.1 (±3.3)	0.78	0.018*	0.52
A1 index, no./h	38.7 (±22.7)	3.8 (±23.0)	37.7 (±30.3)	1.1 (±19.8)	0.35	<0.01*	0.58
A2 + A3 index, no./h	6.4 (±5.6)	1.3 (±5.6)	5.3 (±6.2)	0.1 (±5.4)	0.036*	0.093	0.24
B duration, s	27.4 (±2.5)	-0.5 (±3.4)	27.3 (±3.4)	-0.5 (±3.4)	0.41	<0.01*	0.85
CAP cycle duration, s	30.8 (±2.4)	-0.4 (±3.3)	30.8 (±3.2)	-0.5 (±3.0)	0.40	<0.01*	0.79
CAP sequence duration, s	184.6 (±59.5)	13.9 (±73.5)	181.9 (±79.2)	3.8 (±66.8)	0.25	0.044*	0.19

†Repeated measures analysis of variance (ANOVA) adjusting for the stratification factors of age (5–7 vs. 8–9 years old), race (African American vs. other), weight status (overweight/obese vs. non-overweight), and study site.
‡Effect of treatment (eAT vs. watchful waiting) on CAP measures after controlling for the effect of time (baseline vs. follow-up).
§Effect of time (baseline vs. follow-up) on CAP measures after controlling for the effect of treatment (eAT vs. watchful waiting).
‖Effect of the treatment * time interaction on CAP measures.
*Significance level: $p < 0.05$.

The results of the mediation analysis with CAP rate and subtype indices as the mediator for children with AHI > 10 are presented in Supplementary Tables 5.9, 5.10, and 5.11.

5.5 Discussion

We show that children with moderate OSA (AHI > 10) demonstrate a significant association between a higher frequency of slow high-amplitude rhythms, the so-called A1 phases, and worse behavioural functioning and lower quality of life rated by their caregivers at baseline. However, CAP rate and both subtype indices did not account for significant proportions of changes in behavioural, cognitive, and quality-of-life performance measures plus changes in AHI and PLMSI at follow-up. This outcome may be explained by the negligible changes in CAP parameters at follow-up within children with moderate OSA. Considering the entire CHAT sample, we demonstrate that children with mild-to-moderate OSA display elevated CAP rates, specifically more frequent A1 phases, at 7-months follow-up. However, this

Table 5.3. Comparison of CAP parameters between both randomized arms (eAT and WWS) at follow-up for children with moderate OSA (AHI > 10).

CAP measures	Early Adenotonsillectomy (n = 38)		Watchful Waiting (n = 34)		$p^{\dagger}$		
	Baseline	Change from Baseline to 7 months	Baseline	Change from Baseline to 7 months	Treatment‡	Time§	Treatment *Time‖
CAP rate, %	38.7 (±17.0)	1.3 (±17.8)	37.9 (±15.3)	1.8 (±22.5)	0.77	0.30	0.84
A1, no.	262.0 (±155.0)	-2.0 (±180.8)	240.5 (±152.0)	5.0 (±179.3)	0.79	0.25	0.95
A2+A3, no.	44.5 (±41.3)	-4.0 (±43.0)	37.0 (±23.8)	-0.5 (±25.5)	0.96	0.99	0.29
A1, %	83.9 (±13.0)	0.2 (±10.9)	85.2 (±15.7)	1.8 (±12.4)	0.86	0.27	0.94
A2+A3, %	16.1 (±13.0)	-0.2 (±10.9)	14.8 (±15.7)	-1.8 (±12.4)	0.86	0.27	0.94
A1 mean duration, s	4.1 (±0.3)	0.1 (±0.4)	4.1 (±0.6)	0.0 (±0.4)	0.73	0.48	0.67
A2 + A3 mean duration, s	12.5 (±3.0)	0.6 (±3.2)	12.4 (±2.4)	0.5 (±2.4)	0.083	0.068	0.48
A1 index, no./h	40.7 (±25.5)	-0.1 (±26.1)	36.8 (±23.4)	0.8 (±23.9)	0.83	0.23	0.94
A2 + A3 index, no./h	7.0 (±6.3)	-0.7 (±6.7)	5.9 (±4.1)	-0.2 (±4.5)	0.95	0.87	0.31
B duration, s	26.7 (±2.2)	-0.1 (±2.7)	27.7 (±3.1)	-0.5 (±3.9)	0.88	0.51	0.37
CAP cycle duration, s	30.3 (±2.0)	-0.1 (±2.9)	30.8 (±3.0)	-0.5 (±3.5)	0.81	0.49	0.33
CAP sequence duration, s	195.4 (±64.4)	0.3 (±70.7)	180.4 (±69.4)	8.3 (±57.5)	0.62	0.50	0.72

†Repeated measures analysis of variance (ANOVA) adjusting for the stratification factors of age (5–7 vs. 8–9 years old), race (African American vs. other), weight status (overweight/obese vs. non-overweight), and study site.
‡Effect of treatment (eAT vs. watchful waiting) on CAP measures after controlling for the effect of time (baseline vs. follow-up).
§Effect of time (baseline vs. follow-up) on CAP measures after controlling for the effect of treatment (eAT vs. watchful waiting).
‖Effect of the treatment * time interaction on CAP measures.
*Significance level: $p < 0.05$.

increase is independent of the treatment as no significant interaction between treatment and time was found.

At baseline, children with moderate OSA demonstrated a positive correlation between A1 index and AHI. Previous studies in children with SDB have shown opposing findings, i.e. a decrease of A1 subtypes, mainly during N3 sleep compared with children without SDB (Kheirandish-Gozal *et al.*, 2007; Lopes and Guilleminault, 2006). However, both studies investigate the changes in CAP between children with OSA and controls or children with mild SDB (AHI of $\geq$ 1 but < 5/hr) whereas our analysis was performed within children with AHI > 10. Additionally, our results may reflect the fact that children often do not respond to apnoeas with a noticeable EEG arousal (McNamara *et al.*, 1996) but instead with a short 1 to 3-second burst primarily in the theta band (Scholle and Zwacka, 2001). This is in contrast to adults who terminate obstructive apnoeas regularly with an arousal stimulus (Berry and Gleeson, 1997). In a previous study, children with SDB demonstrated a positive correlation between the intelligence quotient and the percentage of A2, and the A2 index, respectively, suggesting that respiratory events in children with SDB elicit arousal-like protective

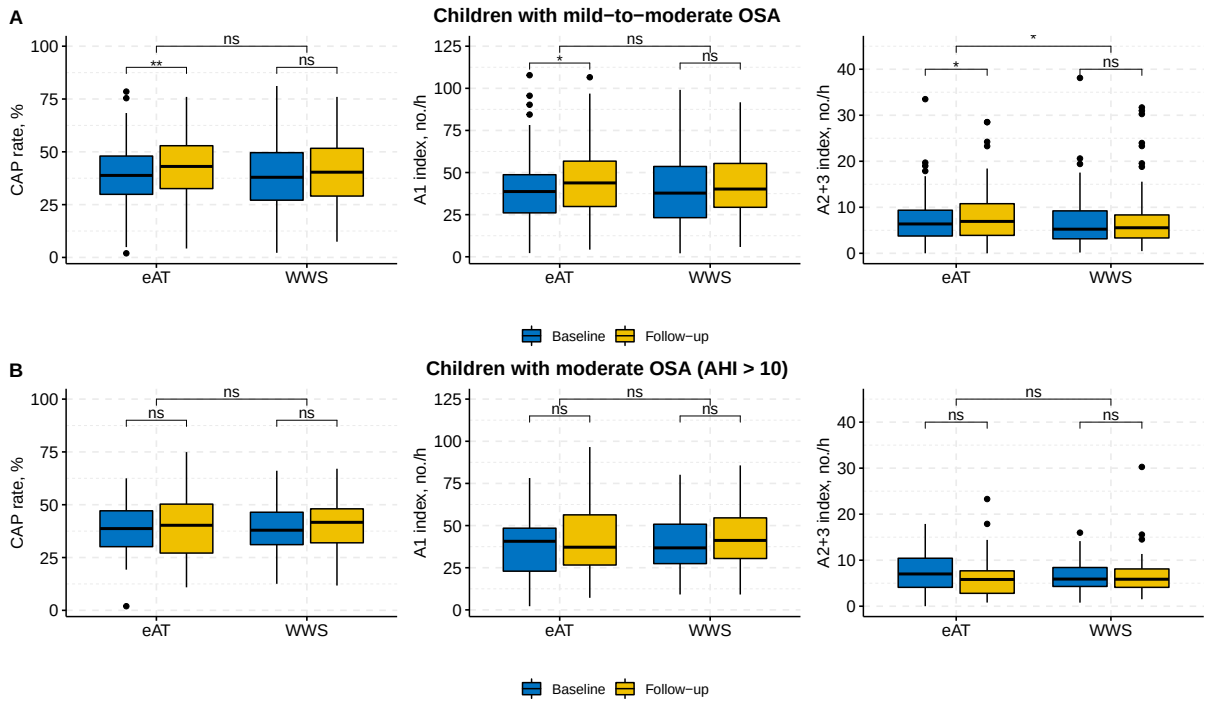


Figure 5.3. Comparison of major CAP parameters between both randomized arms (eAT and WWS) at follow-up. CAP parameters (CAP rate, A1 index, and A2+3 index) in (A) children with mild-to-moderate OSA and (B) children with moderate OSA (AHI > 10) who underwent early adenotonsillectomy (eAT) versus watchful waiting (WWS) at baseline and follow-up polysomnography grouped by treatment arm. Data are presented as mean ± sd.

responses to preserve neurodevelopment to the detriment of NREM sleep instability (Miano *et al.*, 2011). The higher age of children in that study (9.1 ± 2.3 years) may explain the shift from an increase in A1 phases to more arousal-like A2 phases. The authors also reported a positive correlation between the A1 index during N3 and mean overnight oxygen saturation, indicating an increased occurrence of A1 phases in response to respiratory events. In summary, children tend to react to internal or external disturbing stimuli with a protective mechanism that minimizes the negative effects on neurodevelopment (Lopes and Marcus, 2007), which may potentially explain the increase in slow high-amplitude rhythms defined as A1 phases in our study.

Children with moderate OSA also demonstrated a significant association between the A1 index and the caregiver BRIEF GEC T score, the caregiver Conners' Rating Scale Global Index T score, the OSA-18 score, and the caregiver-rated PedsQL. The correlation between the number of A1 phases per hour of sleep and the disease-specific quality-of-life OSA-18 score is most likely a result of the significant association between A1 index and AHI.

Bruni *et al.* (2007) have previously reported that a high frequency of A1 phases is significantly associated with poor behaviour in children with Asperger syndrome. Our results corroborate these findings by revealing a significant association between poorer behavioural functioning, i.e. higher scores in the caregiver BRIEF GEC T score and the caregiver Conners' Rating Scale Global Index T score and more frequent A1 phases. Furthermore, we demonstrate a significant association between lower children's quality of life rated by their caregivers and a higher A1 index. Potentially, these results may reflect an indirect effect of the significant correlations between the A1 index and AHI and PLMSI, respectively, in children with moderate OSA. One can only speculate what the likely causality is in this context as CAP describes external and internal factors that interfere with the sleep process but does not reveal the source of perturbation (Parrino *et al.*, 2012). A very likely scenario is that a higher degree of sleep disturbance indicated by the A1 index and CAP rate result in poor behaviour during the day because of drowsiness and attention deficits.

The CAP rate results obtained with our automated system are in line with previously reported values scored manually in children with SDB. Both treatment arms demonstrated a median CAP rate of 38%–39% during the baseline visit, congruent to the values of children with disordered sleep breathing of the same age range reported by Kheirandish-Gozal *et al.* (2007). The increase in CAP rate and A1 phases observed in the CHAT sample at 7-months follow-up agrees with the previously described pattern of A-phase occurrence during childhood development. The CAP rate increases from preschool age (3–6 years) to school age (6–10 years), peaking during peripubertal age at a previously reported value of around 60% and subsequently decreases during young adulthood (Bruni *et al.*, 2005, 2002; Lopes *et al.*, 2005; Parrino *et al.*, 1998). The subgroup of children with moderate OSA did not show such an increase in CAP rate, especially the number of A1 phases did not change. One can speculate that an increase in CAP sequences due to maturation compensated the expected decrease in the number of CAP sequences due to AHI normalization.

The comparison between treatment arms did not show any significant difference in CAP rate between children undergoing surgery and watchful waiting. A previous study on the treatment of OSA in children with rapid maxillary expansion indicated an increase in CAP rate during slow-wave sleep that was associated with a rise in A1 phases per hour of sleep (Miano *et al.*, 2009). However, follow-up PSG were only recorded for children in the treatment group, preventing the assessment of the time effect and developmental changes in CAP in children without intervention. Hence, those results may reflect the overall increase of CAP sequences and A1 phases with age rather than the effect of treatment, which is in line with

our findings. In CHAT, normalization in PSG findings (defined as AHI < 2 or an obstructive apnoea index score of <1 event per hour) was found not only in 79% of the children after AT but also in 46% of the children randomized to WWSC (Marcus *et al.*, 2013). Moreover, the CHAT study reported no significant improvement in cognitive functioning for children after AT compared with children after WWSC. This result is in agreement with other studies that reported no significant improvement in cognitive functioning after AT (Kohler *et al.*, 2009; Montgomery-Downs *et al.*, 2005) as neurocognition in children is strongly linked to other biomarkers such as the level of CRP (Gozal *et al.*, 2007) or urinary neurotransmitters (Kheirandish-Gozal *et al.*, 2013). The marginal difference in AHI normalization between treatment arms and the lack of improvement in cognitive functioning most likely explains the lack of improvement in CAP parameters after AT compared with watchful waiting.

We also investigated the contribution of changes in CAP to changes in neurocognitive, behavioural, and quality-of-life measures. Although we observed a significant association between the A1 index and behavioural and quality-of-life measures in children with moderate OSA, mediation analysis does not suggest that the changes in cognitive, behavioural, and quality-of-life performance scores due to treatment are attributable to changes in CAP. This observation is supported by the negligible changes in CAP parameters between baseline and follow-up in children with moderate OSA. When considering the entire CHAT sample, mediation analysis yielded an identical outcome, probably due to the predominance of mild cases.

Our study is limited by the original CHAT study design that constraints enrolment to children with an OAI $\geq$ 1 or AHI $\geq$ 2. A control group of normal children would enable comparative analyses and ranking of CAP parameters. Furthermore, the time span of seven months between baseline and follow-up is relatively short potentially limiting the effect of surgery on CAP parameters. Sequelae related to hypoxemia may require more time than seven months to resolve (Marcus *et al.*, 2013). Because of the lack of longitudinal data, the ideal follow-up period for maximal recovery has not been scientifically determined yet. Another limitation is the accuracy of our developed automated detection system. Although it has already shown strong performance (Hartmann and Baumert, 2019) and high reproducibility (Hartmann *et al.*, 2020), the performance for subtype detection is limited by the low number of manually scored paediatric PSG available for machine learning.

In conclusion, we show a significant association between the frequency of slow, high-amplitude waves and the behavioural functioning, as well as the quality of life in children

with moderate OSA. Early AT in children with mild-to-moderate OSA does not alter the microstructure of NREM sleep compared with watchful waiting.

5.6 Funding

The CHAT study was supported by grants (HL083075, HL083129, UL1-RR-024134, and UL1 RR024989) from the National Institutes of Health. S.R. was partially supported by NIH R35 HL135818. Data access was supported by the National Sleep Research Resource (R24 HL114473) and contract NHLBI (75N92019C00011).

5.7 Supplemental material

5.7.1 Supplemental information on automated scoring system

As mentioned in section 5.3.1, we used 19 recordings of healthy children, 15 recordings of healthy adults, and 24 recordings of adults with sleep disorders as training set to cope with the inhomogeneous EEG characteristics of children and adults.

The 19 recordings of healthy children contained 10 normal healthy preschool-aged children (6 girls and 4 boys, mean age 4.6 years, range 3–6 years) (Bruni *et al.*, 2005) and nine out of 10 healthy school-aged children (6 boys and 4 girls, mean age 8.3 $\pm$ 1.5 years, range 6–10 years) (Bruni *et al.*, 2002). One school-aged subject was deleted from the data set as the recording contains a large number of artefacts and clipping. Subjects in both children subsets had normal sleep habits confirmed by parental and child interviews, sleep habits and sleep disorders questionnaires, sleep diaries obtained for 15 days before the experiment, and a complete physical examination. Also, no children had any serious physical or neurological or psychiatric disorder nor history of major sleep problems and none was taking medication at the time of the recording. As recording equipment, a polysomnography digital system (Embla N7000, Medcare, Iceland) and its related software (Somnologica, Medcare, Iceland) was used in both studies. More detailed information on the electrode placement and clinical setup are listed in Bruni *et al.* (2005) and Bruni *et al.* (2002). CAP scoring was performed manually according to the criteria published in Terzano *et al.* (2001) using the sleep software Somnologica (Medcare) as aid for visual detection of A-phases.

Table 5.4. Results for subtype classification performed on paediatric dataset.

	Background		A1		A2		A3		Accuracy (%)
	TPR (%)	F_1-score (%)	TPR (%)	F_1-score (%)	TPR (%)	F_1-score (%)	TPR (%)	F_1-score (%)	
Trained with 29 adults	63.25 ± 10.83	75.45 ± 8.94	76.36 ± 10.62	30.14 ± 8.92	35.37 ± 16.47	10.14 ± 6.15	67.38 ± 15.36	33.68 ± 18.25	63.83 ± 10.36
Trained with 29 adults and 19 kids	85.32 ± 10.99	90.14 ± 7.82	56.00 ± 24.10	38.84 ± 12.98	27.10 ± 14.13	17.90 ± 8.92	69.54 ± 13.55	41.73 ± 13.55	82.62 ± 10.40

TPR, true positive rate; LSTM, long-short term memory network

The 15 recordings of healthy adults and the 24 recordings of adults with sleep disorders are part of the publicly available CAP Sleep Database on PhysioNet (Terzano *et al.*, 2001; Goldberger *et al.*, 2000). The CAP Sleep Database comprises 108 polysomnographic recordings conducted at the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy. The dataset consists of recordings of 16 healthy subjects and 92 patients, including 40 subjects with NFLE (nfle), 22 with REM behaviour disorder (rbd), 10 with periodic leg movement (plm), 9 with insomnia (ins), 5 with narcolepsy (narco), 4 with sleep-disordered breathing (sdb), and 2 with bruxism (brux). Here, 15 healthy subjects (n1–n15), 4 subjects each with NFLE (nfle1–nfle4), REM behaviour disorder (rbd1–rbd4), PLM (plm1–plm4), insomnia (ins1–ins4), narcolepsy (narco1–narco4), or SDB (sdb1–sdb4) were included in the training data set. The included subjects were 45.2 ± 19.2 years old at the time of the recording with 17 subjects being females and 22 subjects being males. CAP scoring was carried out according to the atlas of CAP scoring (Terzano *et al.*, 2001).

Prior to this study, we compared our system trained with aforementioned adult recordings to the entire presented children and adult training data set to analyse the scoring performance improvement when including paediatric data. We selected the leave-one-out (LOO) method as cross-validation approach. For each fold one child was determined as test set and all remaining chidren and all adults were merged into the training set. For training with the adult subset, each fold contained one child as test set and only all adult recordings as training data. Table 5.4 lists the multi-class performance measures of our system trained with the two training sets. Table 5.5 displays the confusion matrices of both training sets for subtype classification. The results show that the inclusion of paediatric data into the training set of the classifier improves the average scoring accuracy on paediatric data by 18.8%. Moreover, the average F_1-score for each subtype increases by 8–15% (background: 14.7%, A1: 8.7%, A2: 7.8%, A3: 8.1%).

Table 5.5. Confusion matrix of subtype classification on paediatric dataset using classifier trained with a) 29 adults and b) 29 adults plus 19 children.

a)

		Target				Total
		B	A1	A2	A3	(PPV)
Output	B	**257,993**	**2,618**	**824**	**813**	**262,248**
		56.9%	0.6%	0.2%	0.2%	(98.4%)
	A1	**91,452**	**22,346**	**2,249**	**624**	**116,671**
		20.2%	4.9%	0.5%	0.1%	(19.2%)
	A2	**37,350**	**2,985**	**2,546**	**857**	**43,738**
		8.2%	0.7%	0.6%	0.2%	(5.8%)
	A3	**23,696**	**1,104**	**1,281**	**4,760**	**30,841**
		5.2%	0.2%	0.3%	1.0%	(15.4%)
Total		**410,491**	**29,053**	**6,900**	**7,054**	
(TPR)		(62.8%)	(76.9%)	(36.9%)	(67.5%)	

b)

		Target				Total
		B	A1	A2	A3	(PPV)
Output	B	**351,177**	**8,933**	**2,090**	**1,612**	**363,812**
		77.4%	2.0%	0.5%	0.4%	(96.5%)
	A1	**32,979**	**17,099**	**1,524**	**211**	**51,813**
		7.3%	3.8%	0.3%	0.0%	(33.0%)
	A2	**8,509**	**1,666**	**1,923**	**439**	**12,537**
		1.9%	0.4%	0.4%	0.1%	(15.3%)
	A3	**17,826**	**1,355**	**1,363**	**4,792**	**25,336**
		3.9%	0.3%	0.3%	1.1%	(18.9%)
Total		**410,491**	**29,053**	**6,900**	**7,054**	
(TPR)		(85.6%)	(58.9%)	(27.9%)	(67.9%)	

TPR, true positive rate; PPV, positive predictive value

5.7.2 Supplemental tables and figures

Following pages contain the results of the causal mediation analysis with CAP rate, A1 index, and A2+3 index as mediator for primary and secondary outcomes in children with mild-to-moderate OSA and in children with moderate OSA (AHI >10).

Table 5.6. Causal mediation analysis with CAP rate as mediator for primary and secondary outcomes in children with mild-to-moderate OSA.

	ACME			ADE			Total Effect†			Proportion Mediated‡		
	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*
NEPSY Attention/Executive Functioning Scaled Score (n = 365)	0.05	(-0.15 to 0.37)	0.59	-1.74	(-4.40 to 0.85)	0.17	-1.69	(-4.31 to 0.96)	0.19	-0.03	(-0.56 to 0.38)	0.70
BRIEF Global Executive Composite Total T Score												
Caregiver rating (n = 359)	0.04	(-0.11 to 0.26)	0.61	3.50	(1.65 to 5.29)	<0.001	3.54	(1.77 to 5.29)	<0.001	0.01	(-0.03 to 0.09)	0.61
Teacher rating (n = 192)	0.05	(-0.34 to 0.56)	0.76	2.84	(-0.34 to 6.41)	0.076	2.89	(-0.31 to 6.43)	0.072	0.02	(-0.22 to 0.33)	0.74
Conners' Rating Scale Global Index Total T Score												
Caregiver rating (n = 360)	0.00	(-0.18 to 0.20)	0.92	2.65	(0.66 to 4.71)	0.008	2.66	(0.72 to 4.71)	0.008	0.00	(-0.09 to 0.09)	0.92
Teacher rating (n = 195)	0.16	(-0.20 to 0.75)	0.43	4.23	(0.94 to 7.80)	0.018	4.39	(1.00 to 7.94)	0.016	0.04	(-0.06 to 0.21)	0.44
PedsQL caregiver-rated total score (n = 364)	0.07	(-0.15 to 0.43)	0.53	-5.23	(-8.09 to -2.41)	<0.001	-5.16	(-7.93 to -2.31)	<0.001	-0.01	(-0.10 to 0.03)	0.53
PSQ-SRBD score (n = 362)	0.00	(-0.00 to 0.00)	0.72	0.24	(0.20 to 0.28)	<0.001	0.24	(0.20 to 0.28)	<0.001	0.00	(-0.01 to 0.02)	0.72
OSA-18 total score (n = 361)	0.01	(-0.28 to 0.34)	0.95	16.68	(13.00 to 20.47)	<0.001	16.69	(13.03 to 20.37)	<0.001	0.00	(-0.02 to 0.02)	0.95
AHI (n = 365)	0.00	(-0.01 to 0.01)	0.72	0.21	(0.13 to 0.28)	<0.001	0.21	(0.14 to 0.28)	<0.001	-0.01	(-0.05 to 0.03)	0.72
PLMSI (n = 365)	0.05	(-0.04 to 0.19)	0.30	0.54	(-0.37 to 1.45)	0.20	0.59	(-0.31 to 1.49)	0.17	0.09	(-0.47 to 0.77)	0.42

The casual mediator was CAP rate. ACME, average causal mediation effect; ADE, average direct effect; NEPSY, Developmental Neuropsychological Assessment (NEPSY); BRIEF, Behaviour Rating Inventory of Executive Function; PedsQL, Paediatric Quality of Life Inventory; PSQ-SRBD, Paediatric Sleep Questionnaire sleep-related breathing disorder scale, OSA-18, Obstructive Sleep Apnoea-18 assessment tool; AHI, Obstructive Apnoea-Hypopnoea Index (AHI) >= 3% - number of [obstructive apnoeas] and [hypopnoeas with >=3% oxygen desaturation or arousal] per hour of sleep; PLMSI, Number of Periodic limb movement (PLM) per hour of non-rapid eye movement sleep (NREM).

†The total effect is decomposed into the ACME and ADE.

‡The contribution of the mediated effect as a proportion of the total effect is shown.

Table 5.7. Causal mediation analysis with A1 index as mediator for primary and secondary outcomes in children with mild-to-moderate OSA.

	ACME			ADE			Total Effect†			Proportion Mediated‡		
	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*
NEPSY Attention/Executive Functioning Scaled Score (n = 365)	0.01	(-0.16 to 0.28)	0.82	-1.69	(-4.40 to 1.01)	0.20	-1.69	(-4.41 to 1.05)	0.21	0.00	(-0.41 to 0.29)	0.85
BRIEF Global Executive Composite Total T Score												
Caregiver rating(n = 359)	0.02	(-0.08 to 0.21)	0.73	3.52	(1.74 to 5.25)	<0.001	3.54	(1.74 to 5.25)	<0.001	0.01	(-0.03 to 0.07)	0.73
Teacher rating (n = 192)	0.10	(-0.21 to 0.68)	0.53	2.78	(-0.50 to 6.34)	0.096	2.89	(-0.49 to 6.59)	0.092	0.04	(-0.21 to 0.46)	0.58
Conners' Rating Scale Global Index Total T Score												
Caregiver rating (n = 360)	0.00	(-0.16 to 0.20)	0.95	2.66	(0.84 to 4.67)	0.008	2.66	(0.82 to 4.69)	0.008	0.00	(-0.07 to 0.08)	0.95
Teacher rating (n = 195)	0.15	(-0.18 to 0.75)	0.43	4.24	(0.84 to 7.53)	0.014	4.39	(0.94 to 7.81)	0.012	0.03	(-0.07 to 0.22)	0.43
PedsQL caregiver-rated total score (n = 364)	0.06	(-0.16 to 0.38)	0.59	-5.22	(-8.08 to -2.56)	<0.001	-5.16	(-7.99 to -2.54)	<0.001	-0.01	(-0.09 to 0.04)	0.59
PSQ-SRBD score (n = 362)	0.00	(-0.00 to 0.00)	0.79	0.24	(0.20 to 0.28)	<0.001	0.24	(0.20 to 0.28)	<0.001	0.00	(-0.01 to 0.02)	0.79
OSA-18 total score (n = 361)	0.01	(-0.23 to 0.31)	0.92	16.67	(12.93 to 20.23)	<0.001	16.69	(12.91 to 20.27)	<0.001	0.00	(-0.01 to 0.02)	0.92
AHI (n = 365)	0.00	(-0.01 to 0.01)	0.94	0.21	(0.13 to 0.29)	<0.001	0.21	(0.14 to 0.29)	<0.001	0.00	(-0.03 to 0.03)	0.94
PLMSI (n = 365)	0.07	(-0.07 to 0.20)	0.41	0.55	(-0.40 to 1.53)	0.26	0.59	(-0.36 to 1.56)	0.22	0.07	(-0.77 to 1.00)	0.49

The casual mediator was A1 index. ACME, average causal mediation effect; ADE, average direct effect; NEPSY, Developmental Neuropsychological Assessment (NEPSY); BRIEF, Behaviour Rating Inventory of Executive Function; PedsQL, Paediatric Quality of Life Inventory; PSQ-SRBD, Paediatric Sleep Questionnaire sleep-related breathing disorder scale, OSA-18, Obstructive Sleep Apnoea-18 assessment tool; AHI, Obstructive Apnoea-Hypopnoea Index (AHI) >= 3% - number of [obstructive apnoeas] and [hypopnoeas with >=3% oxygen desaturation or arousal] per hour of sleep; PLMSI, Number of Periodic limb movement (PLM) per hour of non-rapid eye movement sleep (NREM).

†The total effect is decomposed into the ACME and ADE.

‡The contribution of the mediated effect as a proportion of the total effect is shown.

Table 5.8. Causal mediation analysis with A2+A3 index as mediator for primary and secondary outcomes in children with mild-to-moderate OSA.

	ACME			ADE			Total Effect†			Proportion Mediated‡		
	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*
NEPSY Attention/Executive Functioning Scaled Score (n = 365)	0.11	(-0.12 to 0.50)	0.44	-1.79	(-4.34 to 0.87)	0.17	-1.69	(-4.21 to 0.95)	0.22	-0.06	(-0.97 to 0.97)	0.58
BRIEF Global Executive Composite Total T Score												
Caregiver rating (n = 359)	0.07	(-0.07 to 0.33)	0.36	3.47	(1.67 to 5.20)	<0.001	3.54	(1.78 to 5.26)	<0.001	0.02	(-0.02 to 0.11)	0.36
Teacher rating (n = 192)	-0.27	(-0.98 to 0.25)	0.31	3.16	(-0.30 to 6.54)	0.068	2.89	(-0.36 to 6.29)	0.106	-0.09	(-0.93 to 0.68)	0.40
Conners' Rating Scale Global Index Total T Score												
Caregiver rating (n = 360)	0.02	(-0.15 to 0.28)	0.81	2.64	(0.63 to 4.84)	0.016	2.66	(0.73 to 4.78)	0.008	0.01	(-0.06 to 0.16)	0.81
Teacher rating (n = 195)	-0.10	(-0.75 to 0.43)	0.73	4.49	(1.25 to 8.12)	0.010	4.39	(1.07 to 7.88)	0.01	-0.02	(-0.27 to 0.13)	0.73
PedsQL caregiver-rated total score (n = 364)	-0.18	(-0.55 to 0.09)	0.24	-4.99	(-7.50 to -2.04)	<0.001	-5.16	(-7.78 to -2.14)	<0.001	0.03	(-0.02 to 0.13)	0.24
PSQ-SRBD score (n = 362)	0.00	(-0.00 to 0.01)	0.47	0.24	(0.20 to 0.28)	<0.001	0.24	(0.21 to 0.28)	<0.001	0.01	(-0.01 to 0.03)	0.47
OSA-18 total score (n = 361)	0.01	(-0.28 to 0.58)	0.90	16.68	(12.60 to 20.41)	<0.001	16.69	(12.65 to 20.51)	<0.001	0.00	(-0.02 to 0.04)	0.90
AHI (n = 365)	0.00	(-0.01 to 0.00)	0.21	0.21	(0.14 to 0.29)	<0.001	0.21	(0.13 to 0.28)	<0.001	-0.02	(-0.07 to 0.01)	0.21
PLMSI (n = 365)	-0.07	(-0.21 to 0.03)	0.20	0.66	(-0.29 to 1.57)	0.19	0.59	(-0.37 to 1.54)	0.24	-0.12	(-1.92 to 1.11)	0.40

The casual mediator was A2+A3 index. ACME, average causal mediation effect; ADE, average direct effect; NEPSY, Developmental Neuropsychological Assessment (NEPSY); BRIEF, Behaviour Rating Inventory of Executive Function; PedsQL, Paediatric Quality of Life Inventory; PSQ-SRBD, Paediatric Sleep Questionnaire sleep-related breathing disorder scale, OSA-18, Obstructive Sleep Apnoea-18 assessment tool; AHI, Obstructive Apnoea-Hypopnoea Index (AHI) >= 3% - number of [obstructive apnoeas] and [hypopnoeas with >=3% oxygen desaturation or arousal] per hour of sleep; PLMSI, Number of Periodic limb movement (PLM) per hour of non-rapid eye movement sleep (NREM).
†The total effect is decomposed into the ACME and ADE.
‡The contribution of the mediated effect as a proportion of the total effect is shown.

Table 5.9. Causal mediation analysis with CAP rate as mediator for primary and secondary outcomes in children with moderate OSA (AHI >10).

	ACME			ADE			Total Effect†			Proportion Mediated‡		
	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*
NEPSY Attention/Executive Functioning Scaled Score (n = 72)	-0.09	(-1.31 to 1.21)	0.92	-3.70	(-10.53 to 3.24)	0.24	-3.79	(-10.39 to 3.24)	0.22	0.02	(-0.88 to 0.82)	0.97
BRIEF Global Executive Composite Total T Score												
Caregiver rating (n = 71)	0.14	(-0.48 to 1.48)	0.68	1.98	(-2.69 to 6.03)	0.42	2.12	(-2.28 to 6.15)	0.36	0.07	(-1.44 to 1.50)	0.78
Teacher rating (n = 37)	-0.21	(-2.60 to 2.04)	0.84	1.94	(-8.10 to 12.96)	0.68	1.73	(-7.31 to 11.80)	0.67	-0.12	(-2.03 to 1.78)	0.80
Conners' Rating Scale Global Index Total T Score												
Caregiver rating (n = 71)	0.03	(-0.43 to 0.97)	0.75	1.42	(-2.21 to 5.13)	0.42	1.43	(-2.23 to 5.42)	0.39	0.02	(-1.00 to 1.51)	0.76
Teacher rating (n = 37)	-0.41	(-3.46 to 1.71)	0.76	4.07	(-5.10 to 14.16)	0.40	3.66	(-4.69 to 12.69)	0.39	-0.11	(-2.08 to 1.45)	0.67
PedsQL caregiver-rated total score (n = 72)	-0.10	(-1.72 to 1.01)	0.86	-5.97	(-13.66 to 1.53)	0.11	-6.07	(-13.77 to 1.45)	0.10	0.02	(-0.49 to 0.70)	0.85
PSQ-SRBD score (n = 71)	0.00	(-0.01 to 0.02)	0.83	0.24	(0.15 to 0.32)	<0.001	0.25	(0.15 to 0.33)	<0.001	0.00	(-0.05 to 0.07)	0.83
OSA-18 total score (n = 71)	0.16	(-0.86 to 2.44)	0.72	19.32	(8.64 to 29.12)	<0.001	19.48	(9.22 to 29.52)	<0.001	0.01	(-0.06 to 0.15)	0.72
AHI (n = 72)	0.00	(-0.05 to 0.05)	0.96	0.53	(0.27 to 0.81)	<0.001	0.53	(0.27 to 0.81)	<0.001	0.01	(-0.10 to 0.10)	0.96
PLMSI (n = 72)	-0.01	(-0.22 to 0.30)	0.97	-0.18	(-2.07 to 1.55)	0.81	-0.19	(-2.02 to 1.50)	0.82	0.06	(-1.11 to 1.38)	0.89

The casual mediator was CAP rate. ACME, average causal mediation effect; ADE, average direct effect; NEPSY, Developmental Neuropsychological Assessment (NEPSY); BRIEF, Behaviour Rating Inventory of Executive Function; PedsQL, Paediatric Quality of Life Inventory; PSQ-SRBD, Paediatric Sleep Questionnaire sleep-related breathing disorder scale, OSA-18, Obstructive Sleep Apnoea-18 assessment tool; AHI, Obstructive Apnoea-Hypopnoea Index (AHI) >= 3% - number of [obstructive apnoeas] and [hypopnoeas with >=3% oxygen desaturation or arousal] per hour of sleep; PLMSI, Number of Periodic limb movement (PLM) per hour of non-rapid eye movement sleep (NREM).
†The total effect is decomposed into the ACME and ADE.
‡The contribution of the mediated effect as a proportion of the total effect is shown.

Table 5.10. Causal mediation analysis with A1 index as mediator for primary and secondary outcomes in children with moderate OSA (AHI >10).

	ACME			ADE			Total Effect†			Proportion Mediated‡		
	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*
NEPSY Attention/Executive Functioning Scaled Score (n = 72)	-0.04	(-1.27 to 0.93)	0.98	-3.75	(-11.12 to 3.05)	0.23	-3.79	(-11.17 to 2.51)	0.21	0.01	(-0.53 to 1.04)	0.97
BRIEF Global Executive Composite Total T Score												
Caregiver rating (n = 71)	0.09	(-0.61 to 1.17)	0.79	2.03	(-2.77 to 6.65)	0.45	2.12	(-2.37 to 6.63)	0.39	0.04	(-1.31 to 1.25)	0.87
Teacher rating (n = 37)	-0.54	(-3.59 to 1.52)	0.64	2.27	(-7.65 to 13.41)	0.63	1.73	(-7.56 to 12.12)	0.68	-0.31	(-2.97 to 1.69)	0.73
Conners' Rating Scale Global Index Total T Score												
Caregiver rating (n = 71)	-0.01	(-0.51 to 0.77)	0.83	1.45	(-2.47 to 5.13)	0.45	1.43	(-2.56 to 5.19)	0.44	-0.01	(-1.26 to 1.04)	0.86
Teacher rating (n = 37)	0.58	(-3.90 to 1.53)	0.65	4.24	(-4.26 to 14.15)	0.34	3.66	(-4.11 to 12.56)	0.35	-0.16	(-1.99 to 1.29)	0.62
PedsQL caregiver-rated total score (n = 72)	-0.05	(-1.54 to 1.12)	0.89	-6.02	(-13.99 to 1.40)	0.11	-6.07	(-13.98 to 1.31)	0.098	0.01	(-0.56 to 0.57)	0.88
PSQ-SRBD score (n = 71)	0.00	(-0.01 to 0.02)	0.96	0.25	(0.17 to 0.33)	<0.001	0.25	(0.16 to 0.33)	<0.001	0.00	(-0.06 to 0.07)	0.96
OSA-18 total score (n = 71)	0.11	(-1.09 to 2.65)	0.83	19.37	(9.40 to 29.73)	<0.001	19.48	(9.72 to 29.90)	<0.001	0.01	(-0.07 to 0.16)	0.83
AHI (n = 72)	0.00	(-0.05 to 0.04)	0.96	0.53	(0.27 to 0.80)	<0.001	0.53	(0.27 to 0.80)	<0.001	0.00	(-0.12 to 0.09)	0.96
PLMSI (n = 72)	-0.01	(-0.23 to 0.25)	0.97	-0.18	(-1.88 to 1.47)	0.82	-0.19	(-1.78 to 1.45)	0.82	0.04	(-1.35 to 1.04)	0.94

The casual mediator was A1 index. ACME, average causal mediation effect; ADE, average direct effect; NEPSY, Developmental Neuropsychological Assessment (NEPSY); BRIEF, Behaviour Rating Inventory of Executive Function; PedsQL, Paediatric Quality of Life Inventory; PSQ-SRBD, Paediatric Sleep Questionnaire sleep-related breathing disorder scale, OSA-18, Obstructive Sleep Apnoea-18 assessment tool; AHI, Obstructive Apnoea-Hypopnoea Index (AHI) >= 3% - number of [obstructive apnoeas] and [hypopnoeas with >=3% oxygen desaturation or arousal] per hour of sleep; PLMSI, Number of Periodic limb movement (PLM) per hour of non-rapid eye movement sleep (NREM).
†The total effect is decomposed into the ACME and ADE.
‡The contribution of the mediated effect as a proportion of the total effect is shown.

Table 5.11. Causal mediation analysis with A2+A3 index as mediator for primary and secondary outcomes in children with moderate OSA (AHI >10).

	ACME			ADE			Total Effect†			Proportion Mediated‡		
	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*	Mean	Estimate (95% CI)	*p*
NEPSY Attention/Executive Functioning Scaled Score (n = 72)	0.14	(-1.05 to 1.80)	0.78	-3.93	(-10.72 to 2.80)	0.25	-3.79	(-10.68 to 2.87)	0.28	-0.04	(-1.57 to 1.27)	0.85
BRIEF Global Executive Composite Total T Score												
Caregiver rating (n = 71)	0.14	(-0.66 to 1.41)	0.66	1.98	(-2.66 to 6.18)	0.46	2.12	(-2.24 to 6.59)	0.40	0.07	(-1.27 to 1.94)	0.73
Teacher rating (n = 37)	-0.01	(-1.60 to 1.96)	0.99	1.73	(-7.67 to 12.74)	0.67	1.73	(-7.20 to 12.03)	0.67	0.00	(-1.45 to 1.22)	0.95
Conners' Rating Scale Global Index Total T Score												
Caregiver rating (n = 71)	0.32	(-0.36 to 1.87)	0.47	1.11	(-2.98 to 4.91)	0.61	1.43	(-2.32 to 5.51)	0.49	0.22	(-2.18 to 2.60)	0.67
Teacher rating (n = 37)	0.04	(-2.19 to 2.22)	0.96	3.61	(-5.12 to 12.64)	0.38	3.66	(-4.46 to 12.60)	0.38	0.01	(-1.38 to 2.17)	1.00
PedsQL caregiver-rated total score (n = 72)	0.28	(-0.90 to 1.96)	0.64	-6.35	(-14.00 to 1.42)	0.11	-6.07	(-13.60 to 1.20)	0.12	-0.05	(-0.73 to 0.50)	0.66
PSQ-SRBD score (n = 71)	0.00	(-0.02 to 0.01)	0.99	0.25	(0.17 to 0.33)	<0.001	0.25	(0.17 to 0.33)	<0.001	0.00	(-0.06 to 0.07)	0.99
OSA-18 total score (n = 71)	-0.15	(-1.80 to 1.90)	0.94	19.63	(9.47 to 30.11)	<0.001	19.48	(9.81 to 29.78)	<0.001	-0.01	(-0.10 to 0.11)	0.94
AHI (n = 72)	0.00	(-0.05 to 0.08)	0.87	0.53	(0.27 to 0.81)	<0.001	0.53	(0.29 to 0.80)	<0.001	0.00	(-0.10 to 0.17)	0.87
PLMSI (n = 72)	-0.07	(-0.21 to 0.03)	0.20	0.66	(-0.29 to 1.57)	0.19	0.59	(-0.37 to 1.54)	0.24	-0.12	(-1.92 to 1.11)	0.40

The casual mediator was A2+A3 index. ACME, average causal mediation effect; ADE, average direct effect; NEPSY, Developmental Neuropsychological Assessment (NEPSY); BRIEF, Behaviour Rating Inventory of Executive Function; PedsQL, Paediatric Quality of Life Inventory; PSQ-SRBD, Paediatric Sleep Questionnaire sleep-related breathing disorder scale, OSA-18, Obstructive Sleep Apnoea-18 assessment tool; AHI, Obstructive Apnoea-Hypopnoea Index (AHI) >= 3% - number of [obstructive apnoeas] and [hypopnoeas with >=3% oxygen desaturation or arousal] per hour of sleep; PLMSI, Number of Periodic limb movement (PLM) per hour of non-rapid eye movement sleep (NREM).
†The total effect is decomposed into the ACME and ADE.
‡The contribution of the mediated effect as a proportion of the total effect is shown.

Paper IV

Causality of cortical and cardiovascular activity during cyclic alternating pattern in non-rapid eye movement sleep

The content of this chapter is a modified version of the publication:

Hartmann, S., Ferri, R., Bruni, O. and Baumert, M. (2021), 'Causality of cortical and cardiovascular activity during cyclic alternating pattern in non-rapid eye movement sleep', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **379**:20200248.

Abstract

The dynamic interplay between central and autonomic nervous system activities plays a pivotal role in orchestrating sleep. Macrostructural changes such as sleep stage transitions or phasic, brief cortical events elicit fluctuations in neural outflow to the cardiovascular system, but the causal relationships between cortical and cardiovascular activities, underpinning the microstructure of sleep are largely unknown. Here, we investigate cortical–cardiovascular interactions during the cyclic alternating pattern (CAP) of non-rapid eye movement sleep in a diverse set of overnight polysomnograms. We determine the Granger causality in both 507 CAP and 507 matched non-CAP sequences to assess the causal relationships between electroencephalography (EEG) frequency bands, and respiratory, and cardiovascular variables (heart period, respiratory period, pulse arrival time, and pulse wave amplitude) during CAP. We observe a significantly stronger influence of delta activity on vascular variables during CAP sequences where slow, low-amplitude EEG activation phases (A1) dominate than during non-CAP sequences. We also show that rapid, high-amplitude EEG activation phases (A3) provoke a more pronounced change in autonomic activity than A1 and A2-phases. Our analysis provides the first evidence on the causal interplay between cortical and cardiovascular activities during CAP. Granger causality analysis may also be useful for probing the level of decoupling in sleep disorders.

6.1 Background

Sleep is a state of reduced consciousness that serves several purposes, including energy conservation, metabolic brain waste clearance, modulation of immune responses, memory consolidation and re-consolidation and preservation of mental well-being (Zielinski *et al.*, 2016). During sleep, the body is kept in homeostasis (Benington, 2000); cortical and subcortical brain structures display rich dynamics orchestrating an array of physiological processes that affect the neural outflow to the cardiovascular system, among other things (de Zambotti *et al.*, 2018). External triggers may perturb the cardiovascular system and, along with interoceptive processes, information may be relayed back to the brain,

forming complex closed-loop control systems (Silvani *et al.*, 2015; Silvani and Dampney, 2013). Traditionally, sleep is primarily scored by evaluating electroencephalography (EEG) activity, while chin muscle tone and eye movement may be added to distinguish rapid eye movement (REM) from non-REM (NREM) sleep (Rechtschaffen and Kales, 1968). The former is characterised by random rapid movements of the eyes while the latter exhibits three distinct stages that cover the transitions from high neuronal activity during light sleep to quiescence during deep sleep (Iber *et al.*, 2007). During REM sleep, the sympathetic activity intensifies but declines again below the level of wakefulness after transitioning to NREM sleep (Somers *et al.*, 1993; Trinder *et al.*, 2001a), causing hypotension and bradycardia (Mancia, 1993). The close relationship between the central nervous system (CNS) activity and autonomic nervous system (ANS) activity during sleep is also apparent during phasic, transient events such as arousals or K-complexes (de Zambotti *et al.*, 2018). Both heart rate and blood pressure are easily accessible markers of ANS activity and rise rapidly after the onset of sleep arousal, which is defined as the rapid shift from slow EEG waves to high frequencies in the α or β band (Silvani *et al.*, 2015; Trinder *et al.*, 2003). K-complexes that are defined as synchronised phasic events principally found in the thalamocortical circuitry (Amzica and Steriade, 2002) demonstrate a biphasic pattern in the cardiac response highlighting the association between the CNS and ANS (de Zambotti *et al.*, 2016).

A more comprehensive approach describing transient, phasic perturbations during sleep is the cyclic alternating pattern (CAP) analysis. CAP consists of periodically recurring arousal-related phasic events called activation phases (A-phases) that interrupt the slow cortical activity of NREM sleep. CAP is defined as a sequence of at least two consecutive cycles of an A-phase and the following background period (B-phase) terminated by an A-phase that is considered to be non-CAP (Terzano *et al.*, 2001). By definition, A- and B-phases are restricted to intervals of 2-60 s. Typical patterns of A-phases include δ bursts, vertex sharp transients, K-complex sequences, K-alpha, polyphasic bursts, intermittent alpha, and arousals (Terzano *et al.*, 2001). A-phases can be categorised into three subtypes based on their dominant frequency component and EEG synchrony (Terzano and Parrino, 2000). Subtype A1 represents high-voltage, slow waves with high EEG synchrony. On the contrary, A3 subtypes display low-amplitude, fast rhythms with low EEG synchrony. Subtype A2 is defined as a mixture of both but is often merged with A3 into one subtype (A2+3) because of their congruent nature (Smerieri *et al.*, 2007).

CAP is an important indicator of NREM-sleep instability (Parrino *et al.*, 2012). CAP combines information on CNS and ANS activity. CAP sequences represent continuing arousal

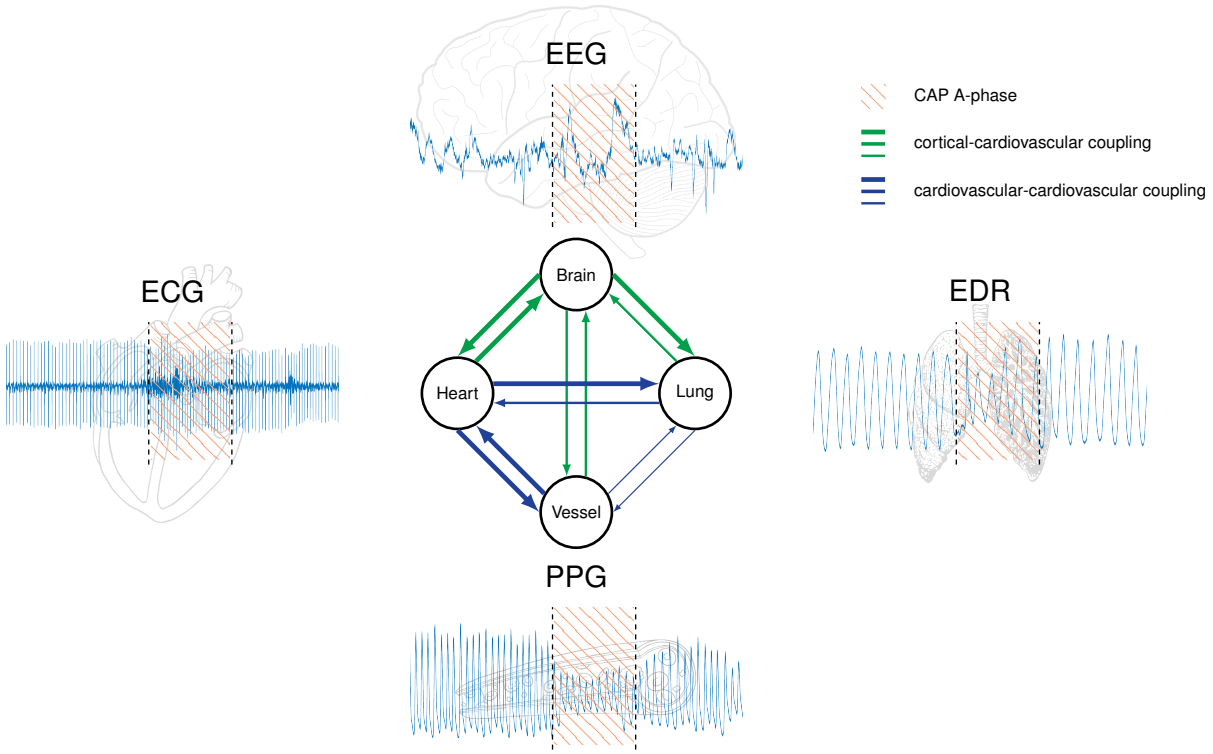


Figure 6.1. Schematic illustration of cortical and cardiovascular coupling investigated during cyclic alternating pattern (CAP) activation phases (A-phase).. Schematic illustration of cortical and cardiovascular coupling investigated during cyclic alternating pattern (CAP) activation phases (A-phase). EEG, electroencephalography; ECG, electrocardiography; PPG, photoplethysmography; EDR, electrocardiogram-derived respiration.

oscillations (Terzano *et al.*, 1988), including the activating effect on the ANS such as heart rate and blood pressure (Kondo *et al.*, 2014), whereas non-CAP periods display sustained stability between the CNS and ANS (Parrino *et al.*, 2016). Several studies have assessed the relationship between CAP and cardiovascular dynamics. Ferini-Strambi *et al.* (2000) reported a significantly increased low-frequency component and significantly decreased high-frequency component of heart rate variability (HRV) during CAP compared with non-CAP in 10 healthy subjects. A study by Ferri *et al.* (2000) of six normal children and adolescents supports the findings of altered sympathovagal balance. In their studies on HRV during CAP in healthy subjects and patients with nocturnal front lobe epilepsy (NFLE), Dorantes-Méndez *et al.* (2018) demonstrated a comparable significant shift towards the low-frequency components of HRV with a more pronounced shift in A3-phases than in A1- and A2-phases. Furthermore, a study on the relation between EEG changes defined by A-phases and the pulse wave amplitude (PWA) after airway obstruction in patients with obstructive sleep apnoea demonstrates a significant correlation between respiratory events combined with

A-phases and respiratory events combined with PWA drops (Bosi *et al.*, 2018). In summary, these findings indicate significant sympathetic activity alterations underpinning the microstructure of sleep.

Physiological signals recorded non-invasive during sleep can be easily exploited to capture the coupling between cortical and autonomic activations. The signal-processing methods used to quantify these relationships range from multivariate linear and non-linear techniques to entropy and information dynamics-based approaches (Pereda *et al.*, 2005). Most studies investigate either the coupling between variables of autonomic functions such as HRV and blood pressure (Silvani *et al.*, 2008) or the interactions between the cortical and autonomic variables such as EEG frequency bands and HRV (Jurysta *et al.*, 2003). Faes et al. used a wide spectrum of methods ranging from Wiener-Granger causality (GC) analysis (Faes *et al.*, 2014b) to linear and non-linear models (Faes *et al.*, 2015) to an information-theoretic approach (Faes *et al.*, 2014a) to reveal causal interactions between brain–brain and brain–heart nodes during sleep. Novel methods use non-linear convergent cross mapping (Schiecke *et al.*, 2019) or the maximum information coefficient in combination with synthetic data generation models (Catrambone *et al.*, 2019) to investigate the brain–heart interplay. Recently, the field of network physiology emerged with the objective of providing new insights into the dynamic interactions between multiple subsystems of the CNS and ANS (Ivanov *et al.*, 2016; Bashan *et al.*, 2012). This involves assembling hierarchical organisations for various physiological states based on the network interactions between brain waves and organs (Bartsch *et al.*, 2015).

In this work, we report on causal relationships between cortical events defined by CAP and autonomic cardiovascular control; see Figure 6.1. We determine brain activity using the energy in five EEG frequency bands and track autonomic activity changes based on electrocardiogram-derived respiration (EDR), heart period (HP), pulse arrival time (PAT), and PWA, the last two being closely related to arterial blood pressure changes. We use GC differences in cortical–cardiovascular interactions during CAP sequences with predominantly A1-, A2- or A3-phases in comparison with non-CAP sequences.

6.2 Materials and methods

6.2.1 Database

We utilised data from the publicly available CAP Sleep Database on PhysioNet (Terzano *et al.*, 2001; Goldberger *et al.*, 2000), which comprises 108 polysomnographic recordings conducted at the Sleep Disorders Center of the Ospedale Maggiore of Parma, Italy. Each recording contains at least three EEG channels, two electromyography (EMG) channels, respiration signals, and one electrocardiogram (ECG). Annotation files contain the scoring performed by expert neurologists, comprising sleep stages and CAP events according to the Rechtschaffen & Kales (Rechtschaffen and Kales, 1968) rules and the atlas of CAP scoring (Terzano *et al.*, 2001), respectively. The dataset consists of recordings of 16 healthy subjects and 92 patients, including 40 subjects with NFLE (nfle), 22 with REM behaviour disorder (rbd), 10 with periodic leg movement (plm), 9 with insomnia (ins), 5 with narcolepsy (narco), 4 with sleep-disordered breathing (sdb), and 2 with bruxism (brux).

For our analysis, we selected recordings containing one central EEG channel (C4-A1), one ECG channel, and one photoplethysmography (PPG) channel. The signal quality was reviewed, and recordings with bad signal quality based on clipping and signal-to-noise ratio were manually removed to achieve congruent results. As a result, 55 patients were selected for our analysis (3 healthy, 4 ins, 3 narco, 28 nfle, 7 plm, 10 rbd). The included subjects were 43.3 $\pm$ 19.5 years old at the time of the recording with 40% being female. From each recording, we extracted the first three A-phases of each CAP sequence plus the preceding 30 seconds and the successive 10 seconds. If the fourth A-phase was located within the subsequent 10-second period, the sequence was removed. Furthermore, if the ECG channel or PPG channel segments were corrupted because of motion artefacts, the sequence was manually removed. For each CAP sequence, a complementary non-CAP sequence of identical length and in the same sleep stage was extracted. In total, 507 sequences were extracted: 350 sequences with predominantly A1-phases, 78 sequences with predominantly A2-phases, and 79 sequences with predominantly A3-phases. On average, each subject contributed 9.2 $\pm$ 5.4 valid sequences to the analysis. An example of a CAP sequence is displayed on the left side in Figure 6.2.

6.2.2 Signal processing

For each CAP and non-CAP sequence, we computed the EDR, HP, PAT, and PWA using ECG and PPG. Prior to signal extraction, ECG and PPG were bandpass filtered using a finite impulse response (FIR) with a filter order of 20 and cut-off frequencies at 5–40 Hz and 0.5–20 Hz, respectively. Because of missing thoracic or abdominal signals in the majority of selected subjects, we derived the respiratory waveform from ECG using the peak-to-trough QRS amplitude (Babaeizadeh *et al.*, 2011). Subsequently, we removed outliers that were outside the range of mean value $\pm$ 2 SD in a sliding, non-overlapping 30-second window. The continuous EDR signal was computed using cubic spline interpolation with an interpolation interval of 0.25 s.

We defined HP as the R–R interval representing the time between two consecutive R peaks. We used the Pan-Tompkins algorithm for detection of R peaks in the ECG (Pan and Tompkins, 1985). The timing of each R peak and the time gap between R peaks were subsequently used to determine the continuous HP signal using cubic spline interpolation with an interpolation interval of 0.25 s.

PAT quantifies the time a pulse wave takes from the heart to reach a distal site, mostly the fingertip. Commonly, the R peak in the ECG is selected as an approximation of the start, and the systolic peak in PPG is chosen as the end of the interval (Elgendi *et al.*, 2019). Here, we used the R peak and the time the PPG reaches 2/3 of the systolic peak as references for the calculation of PAT because a fiducial mark closer to half of the systolic upstroke is clinically more relevant as it portrays the start point of the pulse wave propagation across the artery (Liang *et al.*, 2019). A continuous PAT signal was created by interpolating the discrete values of the time period between each R peak and its corresponding reference at the PPG waveform using cubic splines and an interpolation interval of 0.25 s.

We defined PWA as the amplitude difference between the systolic peak and its diastolic base in each pulse wave (Liu *et al.*, 2019). The upper envelope representing the systolic peaks was calculated using peak detection in a time span of 1/16 s and cubic spline interpolation with an interpolation interval of 0.25 s. The same approach was applied to determine the lower envelope. Consequently, we computed PWA as the difference between the upper and lower envelope and normalised it.

To capture cortical activity, we computed the energy in five EEG frequency bands: δ (0.5-4 Hz), θ (4-8 Hz), α (8-12 Hz), σ (12-16 Hz), and β (16-25 Hz). Two-way least-squares finite impulse response (FIR) filtering with a filter order of 3*(sample rate/low cut-off frequency)

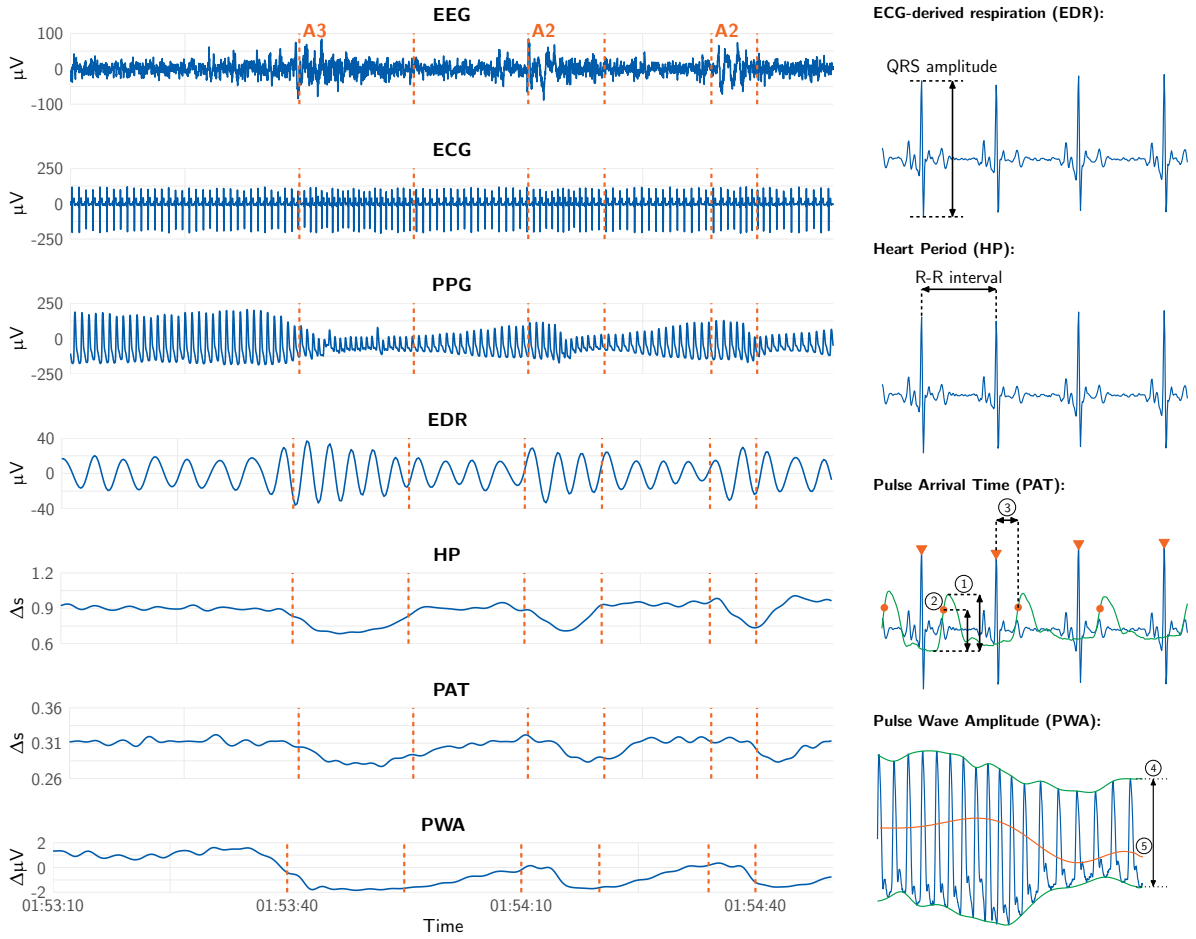


Figure 6.2. Example of a cyclic alternating pattern (CAP) sequence. Example of a cyclic alternating pattern (CAP) sequence in the electroencephalogram (EEG) of subject plm9 including two A2-phases and one A3-phase plus the preceding 30 seconds and the subsequent 10 seconds (left side). The algorithms to determine electrocardiogram-derived respiration (EDR), heart period (HP), pulse arrival time (PAT), and pulse wave amplitude (PWA) in electrocardiography (ECG) and photoplethysmography (PPG) are illustrated on the right-hand side. EDR is determined based on the QRS amplitude, whereas HP is derived from the R–R interval. PAT ③ is computed as the period from the R peak to two-thirds ② of the systolic upstroke ①. PWA is defined as the amplitude difference ④ between the upper and lower envelope of the PPG and the subsequently normalised value ⑤

was used to decompose the EEG into its bands. The band energy for each sequence was defined as the sum of the squared magnitudes using non-overlapping windows of 0.25 s. All autonomic function and EEG variables were computed using a sample rate of 4Hz. Furthermore, the autonomic variables were filtered using a bandpass filter with cut-off frequencies at 0.08 Hz and 0.4 Hz. Figure 6.2 illustrates the algorithms used to extract the aforementioned variables.

Moreover, we plotted the ten seconds after plus the five seconds before each onset of the first A-phase in each extracted sequence to analyse the dynamic changes in brain and cardiovascular activity. For each variable we used the average of a 5-second time window before the start of the plot as a reference. We selected only the first A-phase of each CAP sequence to prevent any long-lasting influence of preceding A-phases.

6.2.3 Granger causality

The GC was proposed by N. Wiener in 1956 and later formalised by C. W. J. Granger (Granger, 1969). It is a very powerful and popular tool to estimate causal relationships and the directions of information propagation between the time series of a multivariate set. Over the years GC has been extended to improve the accuracy of the causality prediction, including information and frequency domain methods (Porta and Faes, 2016). Time series X_i is called the Granger cause to time series X_j if the prediction of X_j significantly improves in case the history of X_i is included (Bressler and Seth, 2011).

Suppose $\boldsymbol{X}$ is a multivariate stochastic process with M scalar time series $X_1, ..., X_M$. Traditional GC analysis uses a vector autoregressive model (VAR) with a time lag of P

$$\boldsymbol{X}(n) = \sum_{k=1}^{P} \boldsymbol{A}(k)\boldsymbol{X}(n-k) + \boldsymbol{\epsilon}(n), \tag{6.1}$$

where $\boldsymbol{X}(n) = [X_1(n)...X_M(n)]^T$ contains the discrete samples of each time series at the time instant n, $\boldsymbol{A}(k)$ are $M \times M$ coefficient matrices, and $\boldsymbol{\epsilon}$ are the residuals. In the first step of GC calculation, the VAR model in Equation 6.1 is fitted to X_j including the history of all time series in $\boldsymbol{X}$. The representation of the fitted VAR model with a time lag of P is given by

$$X_j(n) = \sum_{k=1}^{P} \boldsymbol{A}_j(k)\boldsymbol{X}(n-k) + \epsilon_j(n), \tag{6.2}$$

where $\boldsymbol{A}_j(k)$ is the coefficient matrix for time series X_j. In the next step, the VAR model in Equation 6.1 is fitted to X_j including the history of all time series in $\boldsymbol{X}$ except X_i. The resulting model estimates the contribution of X_i to the prediction of X_j:

$$X_j(n) = \sum_{k=1}^{P} \boldsymbol{A}_j^*(k)\boldsymbol{X}_{/X_i}(n-k) + \epsilon_j^*(n), \tag{6.3}$$

where $\boldsymbol{X}_{/X_i}$ is the vector process without X_i, $\boldsymbol{A}_j^*$ contains the model coefficients, and ϵ_j^* are the model residuals. Both $\boldsymbol{A}_j^*$ and ϵ_j^* are in general different from $\boldsymbol{A}_j(k)$ and ϵ_j in Equation 6.2. Hence, the GC from X_i to X_j is defined as

$$\mathrm{GC}_{X_i \to X_j} = \ln \frac{\mathrm{var}(\epsilon_j^*(n))}{\mathrm{var}(\epsilon_j(n))}, \tag{6.4}$$

where $\epsilon_j^*(n)$ and $\epsilon_j(n)$ are the prediction errors and the $\mathrm{var}(\cdot)$ operator is the statistical variance. $\mathrm{GC}_{X_i \to X_j}$ is always positive and only describes the causal effects $X_i \to X_j$, which are independent from any other time series and from the reversed relationship $X_j \to X_i$.

In this study, $\boldsymbol{X}$ contained the signals described in section 6.2.2. For each sequence, the back and forth GC from every EEG frequency band to EDR, HP, PAT, and PWA was calculated. Also, the GC between EDR, HP, PAT, and PWA was computed.

For our analysis, we used the MATLAB® toolbox developed by Schiatti *et al.* (2015). We decided to use the traditional GC method provided in the toolbox instead of the extended GC method as instantaneous causal relations are not expected in our analysis. The model order was selected according to the Bayesian information criterion (BIC) within the range 1-20, minimizing the BIC figure of merit.

6.2.4 Statistical methods

We conducted Wilcoxon signed-rank tests to determine significant differences in GC between CAP and non-CAP segments (significance level: $p < 0.01$). Moreover, we assessed statistically significant non-zero GC values by comparing $\epsilon_j^*(n)$ and $\epsilon_j(n)$ using the Fisher F-test (significance level: $p < 0.01$) (Porta *et al.*, 2012). Prior to the F-test, we verified that the requirement of a normal distribution is given by conducting the Kolmogorov-Smirnov test. In case the distribution was non-normal, we transformed the non-normally distributed data using Box-Cox transformation.

6.3 Results

6.3.1 A-phase onset analysis

The changes in brain activity and autonomic variables before and after the first A-phase onset in each CAP sequence are summarized in Figures 6.3, 6.4, and 6.5, respectively. Figure 6.3 contains the plots for CAP sequences starting with an A1-phase (350 sequences) whereas Figure 6.4 and 6.5 illustrate the course of change for CAP sequences starting with an A2 or A3-phase, respectively (78 and 79 sequences, respectively). For each CAP sequence we extracted an equivalent non-CAP segment to compare the physiological behaviour between CAP and non-CAP periods.

After the onset of the first A1-phase in a CAP sequence, δ energy demonstrates a steep incline (highest median: 229.5% at 1.25 s after onset) whereas δ activity in non-CAP sequences remains stable across the time window. Differences in energy after the onset of an A1-phase can also be seen in the θ band (highest median: 96.5% at 0.50 s after onset). PWA indicates a change after the A1-phase onset as compared with non-CAP periods with a steady decline to -43.1% at 8.00 s. PAT displays a similar behaviour, reaching its valley at 7.25 s (-1.0%), whereas HP appears to start declining before the onset and reaches its minimum earlier than PWA and PAT (lowest median: -2.2% at 2.75 s before onset). No changes in autonomic cardiovascular activation can be seen in non-CAP periods. Respiration does not display notable changes.

Similar patterns in the EEG frequency bands can be seen after the onset of A2-phases (highest median for δ: 591.3% at 0.50 s after onset) but with more energy distributed in the θ and α frequency bands compared with A1-phases (highest median for θ: 177.0% at 0.25 s after onset, the highest median for α: 118.4% at 0.25 s after onset). All cardiovascular variables declined following the onset of an A2-phase, whereas they remain stable in non-CAP sequences. HP reaches its valley of -3.2% at 4.25 s after onset whereas PAT and PWA steadily decline to -1.4% and -112.7%, respectively, at 7.50 s and 8.50 s, respectively.

In line with A1- and A2-phases, δ energy demonstrates the highest increase among the EEG frequency bands after the onset of the first A3-phase (highest median for δ: 300.9% at 0.50 s after onset). Among high-frequency bands, the increase in energy is reasonably evenly distributed (71.9–123.5%), but β energy displays a delayed peak at 1.25 s. No substantial changes in any frequency band occur during non-CAP sequences across the time window (-13.9%–15.4%). The three cardiovascular variables have a more pronounced decline following the onset of an A3-phase as compared to the other two subtypes; in

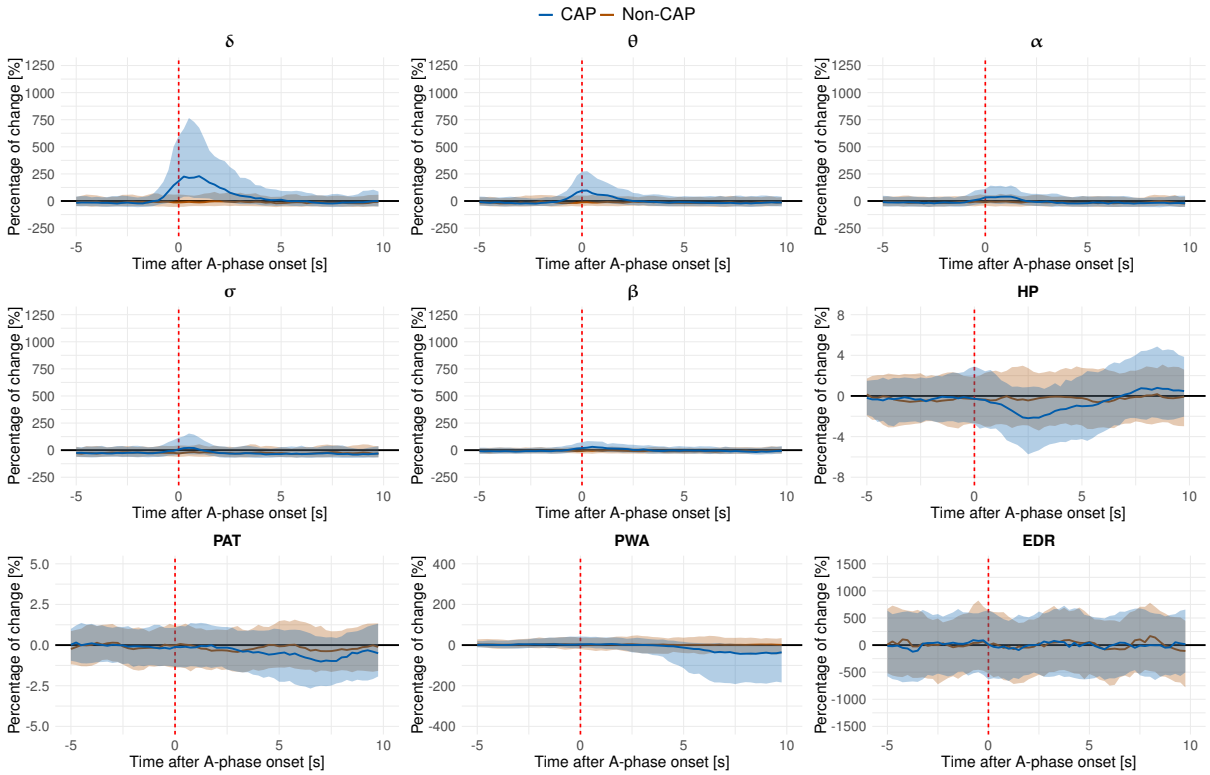


Figure 6.3. Summary of relative changes in electroencephalography (EEG) frequency bands and autonomic functions after the A1-phase onset. Summary of relative changes in electroencephalography (EEG) frequency bands and autonomic functions (heart period (HP), pulse arrival time (PAT), pulse wave amplitude (PWA), and electrocardiogram-derived respiratory (EDR)) after the onset of an A1-phase initiating a cyclic alternating pattern (CAP) sequence as compared with non-CAP periods. Each plot displays the median and the 25th and 75th percentiles of 350 A1-phases.

particular, PWA demonstrates a steep decline with a minimum of -190.4% at 8.25 s. Further, HP and PAT reach their valleys of -3.8% and -2.3%, respectively, at 4.25 s and 7.75 s, respectively.

6.3.2 Causality analysis

Figures 6.6, 6.7, and 6.8 display the results of the GC analysis for CAP sequences where A1-phases dominate (350 sequences with at least two A1-phases), sequences where A2-phases dominate (78 sequences with at least two A2-phases), and sequences where A3-phases dominate (79 sequences with at least two A3-phases).

During CAP sequences with predominant A1-phases, cortical activity appears to have a more pronounced causal impact on the autonomic activity than the reverse direction. In

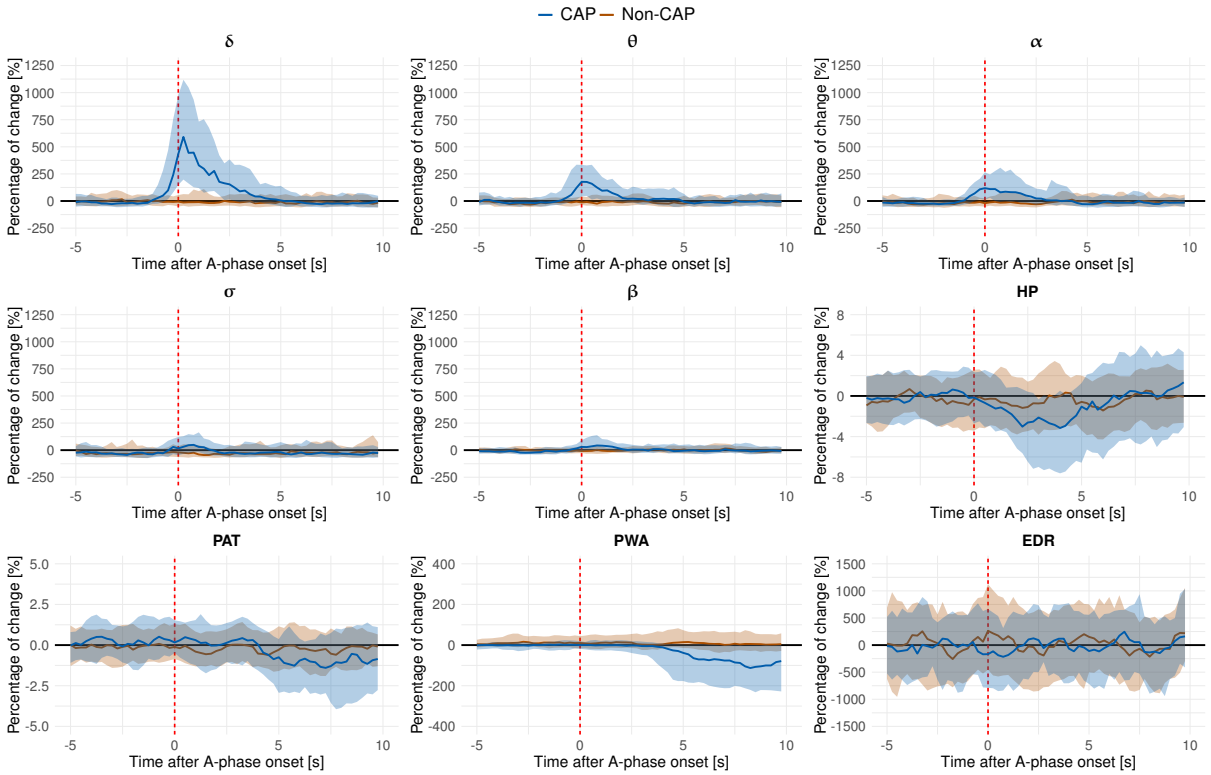


Figure 6.4. Summary of relative changes in electroencephalography (EEG) frequency bands and autonomic functions after the A2-phase onset. Summary of relative changes in electroencephalography (EEG) frequency bands and autonomic functions (heart period (HP), pulse arrival time (PAT), pulse wave amplitude (PWA), and electrocardiogram-derived respiratory (EDR)) after the onset of an A2-phase initiating a cyclic alternating pattern (CAP) sequence as compared with non-CAP periods. Each plot displays the median and the 25th and 75th percentiles of 78 A2-phases.

particular, δ activity plays a major role in CAP sequences with predominantly A1-phases as it demonstrates a high GC value throughout all nodes. Figure 6.6 shows a significantly higher $\delta \rightarrow \mathrm{PAT}$ and $\delta \rightarrow \mathrm{PWA}$ interaction during CAP periods than during non-CAP periods ($p<0.001$). A larger number of $\delta \rightarrow \mathrm{PAT}$ and $\delta \rightarrow \mathrm{PWA}$ GC connections were statistically significant in CAP sequences with predominantly A1-phases than in non-CAP segments ($\delta \rightarrow \mathrm{PAT}$: 27% in CAP vs. 21% in non-CAP, $\delta \rightarrow \mathrm{PWA}$: 41% in CAP vs. 29% in non-CAP). Additionally, PAT appears to be significantly more influenced by θ and σ during CAP segments than during non-CAP segments ($p<0.001$), with 29% and 29%, respectively, of the sequences indicating a significant GC connection (non-CAP: 19% and 23%, respectively).

Figure 6.7 displays the relation between cortical activity and autonomic activity for CAP periods with predominantly A2-phases. Similar to sequences with predominantly A1-phases,

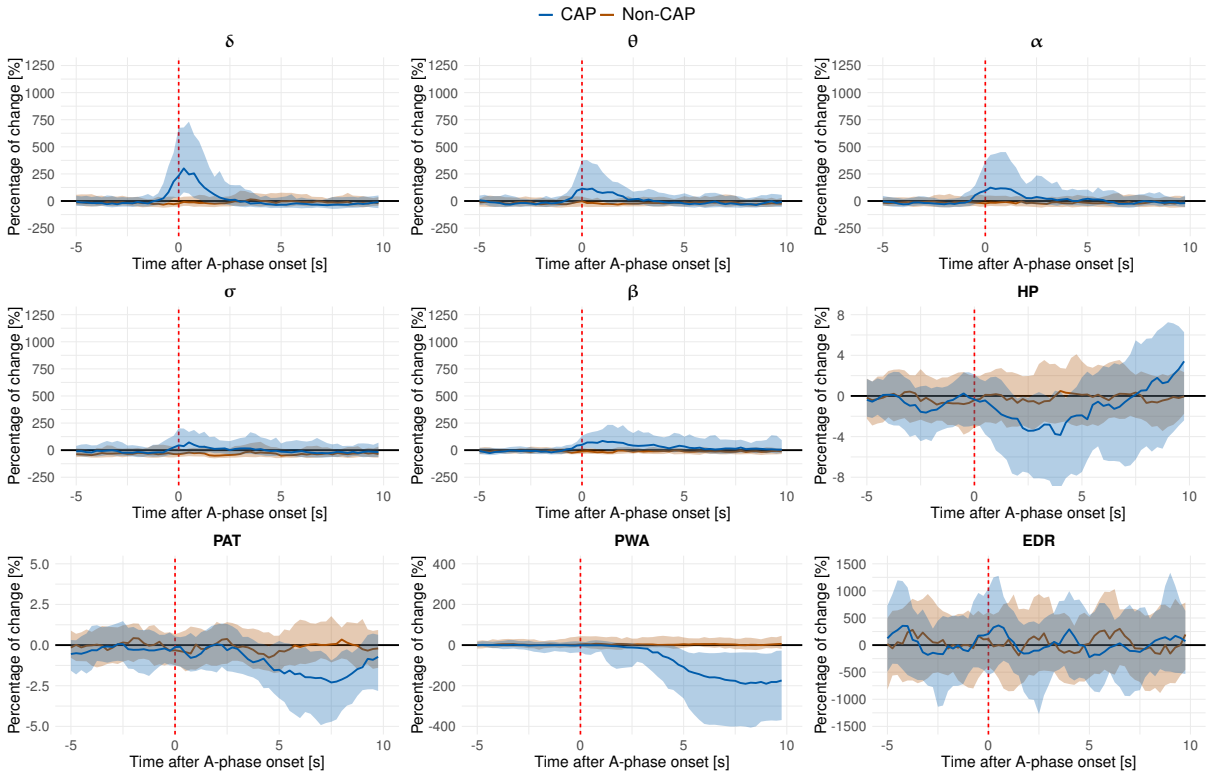


Figure 6.5. Summary of relative changes in electroencephalography (EEG) frequency bands and autonomic functions after the A3-phase onset. Summary of relative changes in electroencephalography (EEG) frequency bands and autonomic functions (heart period (HP), pulse arrival time (PAT), pulse wave amplitude (PWA), and electrocardiogram-derived respiratory (EDR)) after the onset of an A3-phase initiating a cyclic alternating pattern (CAP) sequence as compared with non-CAP periods. Each plot displays the median and the 25th and 75th percentiles of 79 A3-phases.

cortical activity appears to have a more substantial impact on autonomic activity than vice versa. Of note, Figure 6.7 shows a significant difference for the $\alpha \rightarrow \mathrm{PWA}$ interactions during CAP periods with predominantly A2-phases than during non-CAP periods (p=0.008), where approximately one-third of all $\alpha \rightarrow \mathrm{PWA}$ interactions in CAP sequences demonstrated a significant non-zero GC (non-CAP: 17%). No further significant differences between CAP periods with predominantly A2-phases and non-CAP were found.

On the contrary, autonomic activity appears to have a stronger causal impact on cortical activity during CAP sequences with predominantly A3-phases than vice versa. Figure 6.8 displays a significantly higher $\mathrm{PWA} \rightarrow \beta$ interaction during CAP periods than during non-CAP periods (p<0.001). Approximately one-quarter of all sequences demonstrated a significant non-zero GC value for $\mathrm{PWA} \rightarrow \beta$ connections during CAP sequences with predominantly A3-phases, as compared with 11% during non-CAP periods.

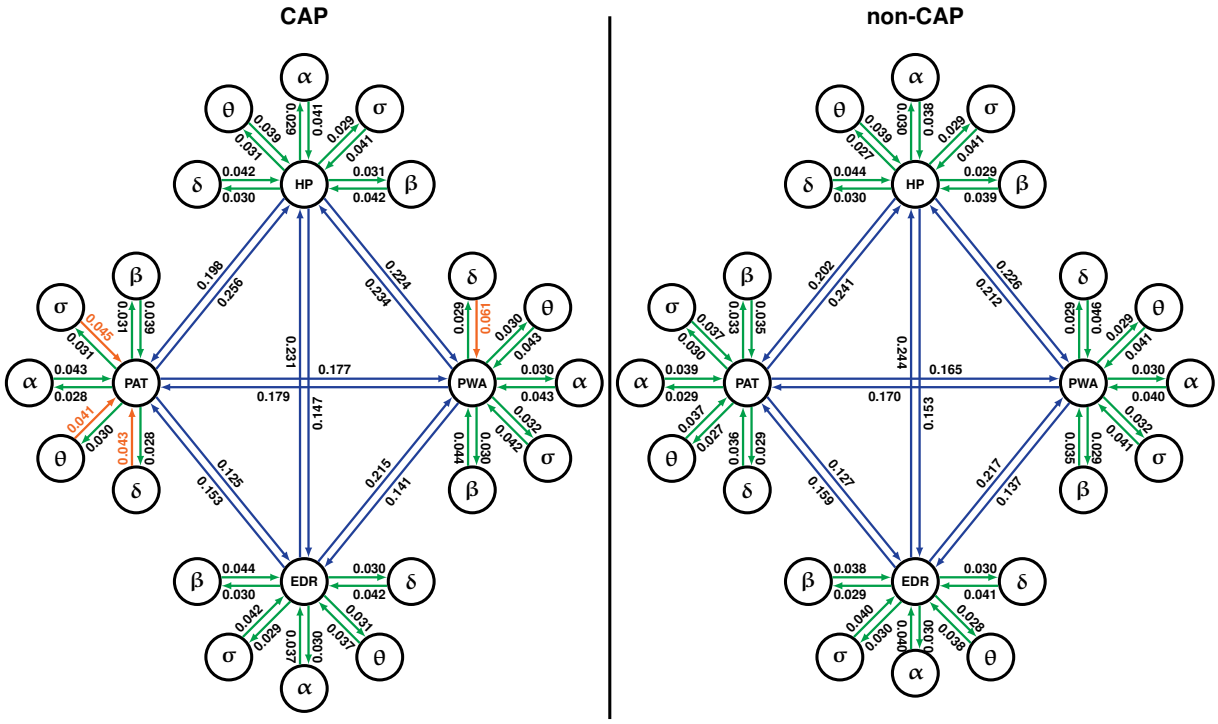


Figure 6.6. Graph of Wiener-Granger causality (GC) interactions between cortical–cardiovascular and cardiovascular–cardiovascular nodes during A1-phases. Graph of Wiener-Granger causality (GC) interactions between cortical–cardiovascular and cardiovascular–cardiovascular nodes. GC values for cyclic alternating pattern (CAP) sequences with predominantly A1-phases are shown on the left, GC values for equivalent non-CAP periods are shown on the right. Each connection displays the GC value averaged across all 350 sequences. Connections in green represent cortical-cardiovascular coupling and connections in blue describe cardiovascular-cardiovascular coupling. Connections highlighted in red demonstrate a significant difference in GC between CAP and non-CAP at a significance level of $p < 0.01$. HP, heart period; PAT, pulse arrival time; PWA, pulse wave amplitude; EDR, electrocardiogram-derived respiration

6.4 Discussion

In this study, we show that short bursts of cortical δ activity reflected by A1-phases cause an increase in vascular activity during sleep, which is likely to provoke a surge in arterial blood pressure. This suggests that A1-phases represent cortical events that, importantly, are not covered by the conventional American Sleep Disorders Association (ASDA) arousal definition (ASDA, 1992) but are responsible for arousal-like autonomic activations. Rapid, low-amplitude events defined as A3-phases appear to be preceded by increased autonomic activity. Moreover, a more pronounced impact on autonomic functions can be seen during A3-phases as compared to the other two subtypes.

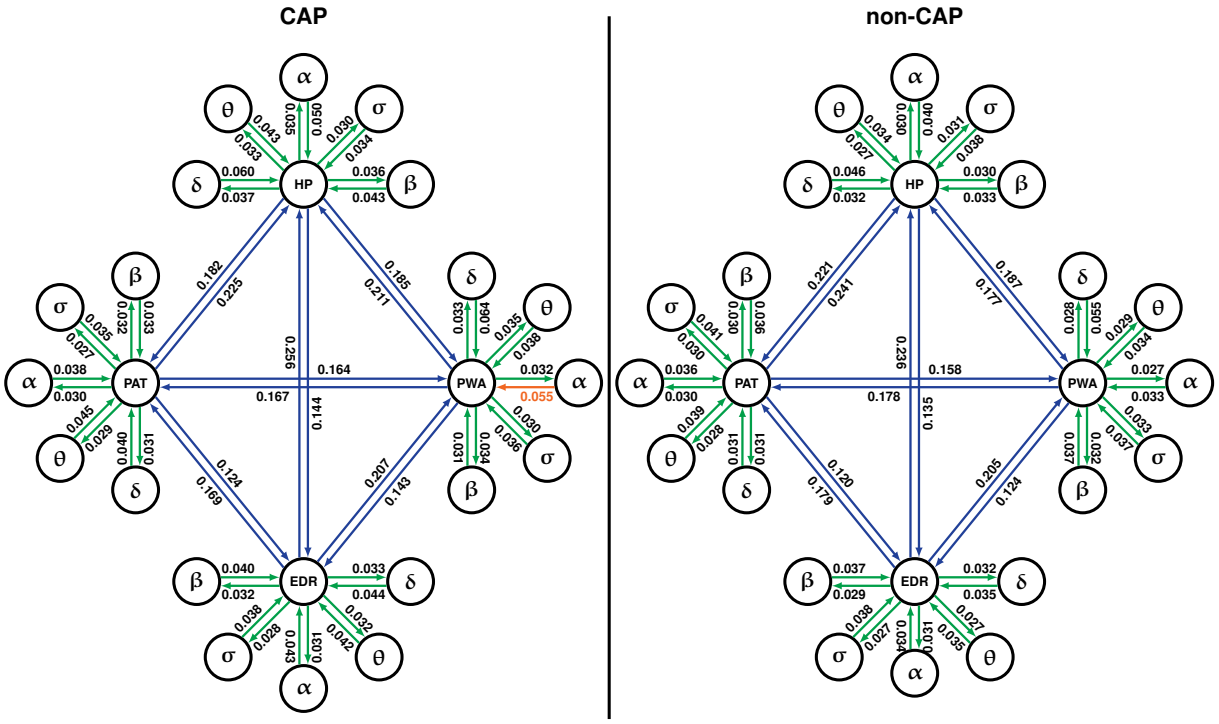


Figure 6.7. Graph of Wiener-Granger causality (GC) interactions between cortical–cardiovascular and cardiovascular–cardiovascular nodes during A2-phases. Graph of Wiener-Granger causality (GC) interactions between cortical–cardiovascular and cardiovascular–cardiovascular nodes. GC values for cyclic alternating pattern (CAP) sequences with predominantly A2-phases are shown on the left, GC values for equivalent non-CAP periods are shown on the right. Each connection displays the GC value averaged across all 78 sequences. Connections in green represent cortical-cardiovascular coupling and connections in blue describe cardiovascular-cardiovascular coupling. Connections highlighted in red demonstrate a significant difference in GC between CAP and non-CAP at a significance level of $p < 0.01$. HP, heart period; PAT, pulse arrival time; PWA, pulse wave amplitude; EDR, electrocardiogram-derived respiration

Ensemble-average analysis of the onset of the initial A-phase in CAP sequences reveals a sudden increase in δ and θ activity during A1-phases. At the onset of A3-phases, all frequency bands display elevated energy levels with δ and θ activity rising prior to the onset of the A3-phase, whereas β activity peaks later during the A3-phase. This is in agreement with the CAP atlas (Terzano *et al.*, 2001), which specifies A1-phases as rhythms with more than 80% slow-wave activity whereas A3-phases contain a higher percentage of fast rhythms but with 50% of slow-wave activity at the beginning of the A-phase. Also, our analysis demonstrates a more pronounced effect on autonomic variables after the onset of A3-phases as compared to A1- and A2-phases. The minimum of HP and PWA after the onset of an A3-phase is substantially lower than after an A1-phase (HP: -2.2% after A1 vs. -3.8% after A3, PWA: -43.1% after A1 vs. -190.4% after A3) and PAT displays a slightly

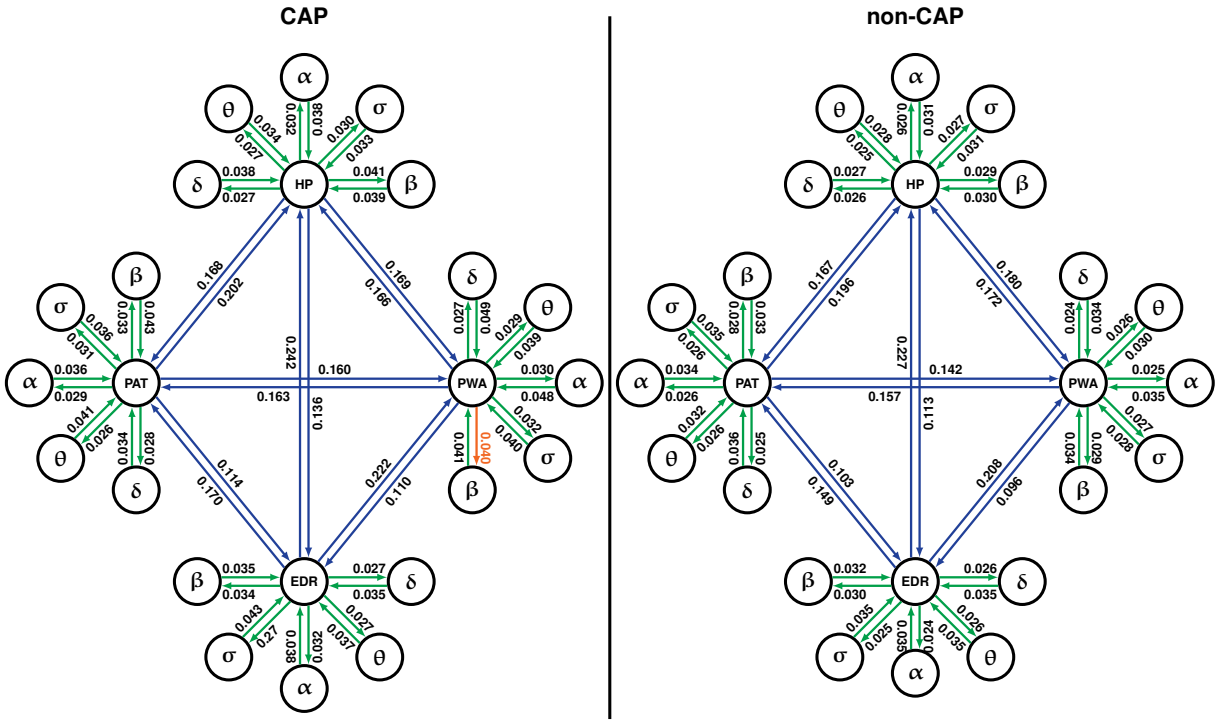


Figure 6.8. Graph of Wiener-Granger causality (GC) interactions between cortical–cardiovascular and cardiovascular–cardiovascular nodes during A3-phases. Graph of Wiener-Granger causality (GC) interactions between cortical–cardiovascular and cardiovascular–cardiovascular nodes. GC values for cyclic alternating pattern (CAP) sequences with predominantly A3-phases are shown on the left, GC values for equivalent non-CAP periods are shown on the right. Each connection displays the GC value averaged across all 79 sequences. Connections in green represent cortical-cardiovascular coupling and connections in blue describe cardiovascular-cardiovascular coupling. Connections highlighted in red demonstrate a significant difference in GC between CAP and non-CAP at a significance level of $p < 0.01$. HP, heart period; PAT, pulse arrival time; PWA, pulse wave amplitude; EDR, electrocardiogram-derived respiration

lower minimum for A3-phases (PAT: -1.0% after A1 vs. -2.3% after A3). This supports the findings of previous studies (Gonzalez-Salazar *et al.*, 2014; Dorantes-Méndez *et al.*, 2018) and is in line with the original definition of A-phases.

As A1-phases are defined in the CAP atlas as arousal equivalents and comprise K-complexes and δ bursts (Terzano *et al.*, 2001), they can be categorised as subcortical or autonomic arousals. Sforza et al. showed in their study that δ bursts and K-complexes induce similar arousal responses as microarousals or phases of transitory activation (Sforza *et al.*, 2000). However, the latter arousal types show an increase in heart rate prior to the arousal onset, whereas the increase in heart rate for δ bursts and K-complexes appeared with a short latency. Our results corroborate these findings and demonstrate a clear causal connection between the δ activity in A1-phases and the preceding sympathetic activation. In 27% and

41%, respectively, of the 350 CAP sequences with at least two A1-phases, the shift in δ energy most likely Granger-causes a surge in blood pressure, resulting in a shorter PAT and a drop in PWA, respectively. Consequently, our findings provide further evidence that A1-phases are closely linked to an increase in sympathetic activity. However, the impact on autonomic functioning is less pronounced than for A3-phases, which can be assumed to represent conventional arousals in this case (Hartmann *et al.*, 2020).

Additionally, our GC analysis reveals a significant difference in the $\mathrm{PWA} \rightarrow \beta$ interaction during CAP periods with predominantly A3-phases than during non-CAP periods. This could potentially be a result of the fact that spontaneous arousals show cardiovascular activation prior to the first signs of arousal in the EEG (Trinder *et al.*, 2012; Baumert *et al.*, 2010). Such an increase in cardiovascular activity can be caused by external perturbations as CAP can be triggered by external conditions (Parrino *et al.*, 2012). On the other hand, as a high percentage of arousals are preceded by slow waves such as K-complexes (Halász *et al.*, 2004), ASDA arousals do not take into account the slow element of the double EEG activation (slow and rapid) that is typical in NREM sleep (Terzano *et al.*, 2002). Thus, autonomous activations could be erroneously allocated before the onset of cortical activation (Bonnet and Arand, 1997) by not considering the pre-arousal slow-wave component. In our analysis, 22.8% of the CAP sequences with at least two A3-phases contained an A1-phase prior to an A3-phase, which is remarkably similar to the number of sequences with significant non-zero GC values for $\mathrm{PWA} \rightarrow \beta$ interactions (24%). Hence, a preceding burst in delta activity may be the cause of the activation of the autonomic system, which in turn foreshadows a shift in energy towards higher frequencies in brain activity.

In our study, we investigated subjects of both genders with different sleep pathologies and this might influence the results. However, we should consider that the neural mechanisms that constitute CAP are independent of gender or sleep pathologies since it has been demonstrated that these factors modulate the quantitative aspects of CAP but not its intrinsic generation mechanisms (Ferri *et al.*, 2005a). Studies have demonstrated that the physiological fluctuations of CAP are accompanied by subtle, but significant, changes in the balance between the sympathetic and vagal components of the autonomic system in different age groups (Ferri *et al.*, 2000) and in pathological patients who present similar changes in the dynamics of the heart rate reflected by the modification of A-phase dynamics (Leon-Lomeli *et al.*, 2014). Another study reported an increase in heart rate with higher values of low-frequency power in the A-phases with no differences between healthy subjects and patients with epilepsy (Dorantes *et al.*, 2015). All these studies demonstrate that

the neural mechanisms of CAP related to the heart–brain axis are independent from sleep pathologies.

One limitation of our study is the assessment of cortical–cardiovascular interactions using traditional GC analysis. The application of GC requires wide-sense stationary data (Bressler and Seth, 2011), which are commonly not given in physiological signals such as EEG and HRV. One suggested approach is to shorten the analysed time segments (Ding *et al.*, 2000). The analysed sequences in this study represent a short fraction of an overnight recording and thus can be assumed locally stationary. Additionally, traditional GC analysis as it is applied in this study is based on linear regression models, which do not capture non-linear effects. The assessment of non-linear effects can be achieved with non-linear GC methods such as kernel-based methods (Marinazzo *et al.*, 2008) or with transfer entropy as an information-theoretic causality measure (Faes *et al.*, 2015). Moreover, previous studies have shown that linear models work extremely well in neuroscience applications (McIntosh and Gonzalez-Lima, 1994) and can detect similar brain–heart networks to non-linear models in normal undisturbed sleep (Faes *et al.*, 2015). Additionally, the modelling of non-linear systems using linear VAR systems can be sufficiently achieved by increasing the order of the fitted model adequately (Winterhalder *et al.*, 2005). Finally, it is important to note the limitations of the GC measure. GC fails to measure an underlying causal mechanism, but it measures a causal effect, which is the reduction in prediction error (Barrett and Barnett, 2013). Hence, the significant cortical–cardiovascular interactions detected in this study do not represent underlying anatomical connections, but they reflect the directed influence from one system to another (Greco *et al.*, 2019). For a more detailed decomposition of the reduction in prediction error similar to the information-theoretic assessment, extended frameworks of the predictability framework allow accounting for the influence from the target process itself, from other processes in the network and from the interaction with other processes (Faes *et al.*, 2016).

One additional limitation of our study is the small number of recordings available in the CAP Sleep Database. A larger number of subjects resulting in a greater number of investigated sequences would enable a more comprehensive analysis of the interplay between CNS and ANS during CAP. A possible solution is presented by automated CAP scoring algorithms that are capable of scoring large cohort studies (Hartmann and Baumert, 2019; Hartmann *et al.*, 2020, 2021). Furthermore, the lack of healthy subjects in our analysis and the combination of various sleep pathologies is potentially limiting the significance of the

presented results. However, the limited number of recordings with the required polysomnographic channels of each pathology in the CAP Sleep Database prevented a more in-depth analysis of the interaction of CNS and ANS for each subcategory. Hence, our study can only provide a general insight into the dynamic interplay of CNS and ANS during CAP. Finally, another limitation is the age distribution of the investigated dataset with mostly middle-aged or older subjects. As the CAP rate follows a U-shaped curve with a high number of A1-phases in young people and a high number of A2+3-phases in older people (Parrino *et al.*, 1998), the interplay between CNS and ANS during A-phase subtypes may also change throughout different age groups. We demonstrated in this study that the delta activity in A1-phases is the cause for activation of autonomic functions in middle-aged to older subjects. However, the rate of A1 subtypes is low in these age groups compared with younger subjects. In children, A1-phases are regarded as 'anti-arousals' (Hirshkowitz, 2002) and support the homeostatic process to maintain the restorative function of sleep (Bruni *et al.*, 2010b). Hence, additional studies on the interplay between CNS and ANS during CAP in children could highlight differences in the autonomic response and the connection between delta activity and autonomic functions in children.

6.5 Conclusion

GC analysis between cortical and cardiovascular activation during CAP in NREM sleep demonstrates the important role of CAP in the assessment of physiological states during sleep. We show that A1-phases in CAP sequences, in particular the burst in delta activity, are closely linked to an increase in vascular activity during NREM sleep. We also show that A3-phases have a more pronounced impact on autonomic functions than A1- and A2-phases. These findings highlight the importance of sleep microstructure when evaluating autonomic activity alteration during sleep. Furthermore, we demonstrate the potential of GC analysis for gaining further information on CAP.

6.6 Data Accessibility

All data in this study are available for free download from public repositories. The CAP Sleep Database can be downloaded from physionet.org.

6.7 Authors' Contributions

SH conceived and designed the study, carried out the experiments, performed the data analysis, and drafted the manuscript. MB conceived and designed the study, and performed the data analysis. OB and RF contributed to the analysis and interpretation of the data. All authors read and approved the manuscript.

6.8 Funding

This study was partly supported by a grant from the Australian Research Council (DP0663345).

Conclusion

THIS chapter summarizes the conclusions of this dissertation and highlights the original contributions of each chapter. Additionally, potential future directions are discussed.

7

7.1 Conclusion and thesis summary

Previously proposed solutions to automatically detect CAP sequences in sleep recordings have demonstrated promising results in regard to accuracy and sensitivity. However, the proposed systems fail to consider critical information from the past to classify CAP events either, nor do they demonstrate clinical relevance as they were not applied on large clinical studies. This thesis has proposed a comprehensive end-to-end CAP detection method including pre-processing, artefact removal, multi-class classification using LSTM networks, and post-processing. Moreover, the studies presented in this thesis assess for the first time sleep microstructure in large population-based cohort studies, and investigate the causal relationships between cortical events defined by CAP and autonomic cardiovascular control. This chapter summarizes the key findings described in the thesis and discusses possible future directions for further research towards enhancing the understanding of sleep microstructure.

The automatic detection of sleep events such as CAP is of great interest as the manual inspection of sleep recordings is very tedious and time-consuming. In Chapter 3, a standalone, fully automated scoring system to detect the A-phases of CAP events is presented. The system is equipped with a deep LSTM network to exploit the dynamical temporal behaviour demonstrating the suitability of sequential information to improve the classification of A-phases (see Appendix A). Moreover, state-of-the-art signal processing methods were included to eliminate the influence of cardiac and ocular artefact in EEG. Artefact removal has shown to have a positive effect on scoring accuracy in individual cases although the overall impact was only marginal according to the results of the test set. Additionally, the classifier was adapted to the conditions of imbalanced distributions between CAP and non-CAP events in training data by optimizing the cost function. Thus, the proposed system presents a fully automated high performance solution to classify A-phases and their subtypes.

To add evidence towards the understanding that CAP scoring is an important indicator for sleep quality and sleep fragmentation, we sought to quantify the prevalence of CAP in large population samples of older men and women. The study presented in Chapter 4 analyses the first time CAP in large community-based cohort studies using the previously developed automated detection system. The results demonstrate that subjectively reported sleep quality correlates with the CAP rate, independent from traditionally scored markers of sleep fragmentation. Moreover, the results provide important insights into the behaviour of CAP in relation to age and gender. Hence, the article in Chapter 4 demonstrates the

importance of considering CAP in sleep analysis and the viability of our automated CAP scoring system to assist in analysing CAP in large populations.

High frequency of sleep disturbances during pre-adolescent age in children can have severe impact on their cognitive development and their behavioral functioning. In this thesis, the effect of sleep disorders such as OSA and RSD on the sleep microstructure of children was investigated. As reported in Chapter 5, children with moderate OSA display a significant association between the frequency of slow, high-amplitude waves and the behavioral functioning, as well as the quality of life. Our results indicate that eAT as first-line treatment for childhood OSA has no altering effect on sleep microstructure as compared to watchful waiting. Additionally, the findings reported in Appendix B suggest the presence of an increased sleep instability in children with RSD due to a higher CAP rate and significantly shorter CAP cycles in comparison to normal children. These findings underline the importance of further studies on the role of CAP during childhood.

To advance the understanding on the underpinning physiological mechanisms of sleep microstructure, it is important to study the cortical–cardiovascular interactions during CAP events. The study presented in Chapter 6 aimed at assessing the causal relationships between EEG frequency bands, respiratory, and cardiovascular variables during CAP using GC. Our findings show that the onset of A-phase subtypes show an increasing impact on autonomic functions the higher the percentage of fast, low-amplitude rhythms. Also, we report a close link between slow, high-amplitude waves and increased vascular activity during NREM sleep. This highlights in particular the importance of including CAP in analysing adverse effects of sleep fragmentation on cardiovascular health. As shown in Appendix C, nocturnal arousal burden, which is closely related to A3-phases, is associated with long-term cardiovascular and all-cause mortality in women and to a lesser extent in men. Finally, the outcomes demonstrate the value of GC analysis in regard to the relationship between CAP and ANS.

7.2 Future directions

A major limitation of the developed automated CAP scoring system is the high inter-subject variance of the scoring output. Hence, the main goal of future studies should be the reduction of the inter-subject variance with the aid of feature engineering or increased EEG normalization. The inclusion of a time feature describing the time period between the last wake or REM phase and the current time could potentially lead to an improved performance

as sleep experts regularly consider this relationship during scoring. Another idea would be the analysis of features that are critical during the decision finding process of the classifier using reverse engineering methods such as integrated gradients (Sundararajan *et al.*, 2017) and feature activation visualisation (Erhan *et al.*, 2009). This could result in a better understanding of what kind of features encompass relevant information for the classification process. In Chapter 3, the developed CAP detection system was trained on data from the CAP Sleep Database. Hence, the trained classifier relies on recordings performed with the same setup. To address the issue of database variability, a more heterogeneous training set could result in a more robust classifier in terms of recording set-up and cohort characteristics. As the results in Section 5.7.1 demonstrate, the addition of data from different age cohorts or sleep clinics increase the accuracy of the automated classifier. One approach to overcome the problem of database variability is based on an ensemble of local models which are trained on data with different equipment set-up or with a different age distribution (Alvarez-Estevez and Fernández-Varela, 2020). Subsequently, the output decision is based on the majority between models. A clinically more relevant approach could include output probabilities for each local model providing the possibility to remove models with a low certainty about the estimated output. Moreover, a collection of diverse datasets enables the addition of transfer learning techniques such as domain adaptation or ensemble training. In terms of deep learning techniques, the addition of attention-based recurrent layers (Bahdanau *et al.*, 2014) for improved feature selection or reinforcement learning (Mnih *et al.*, 2015) to overcome the need of labelled data may provide methods to improve the classification performance. Also, an extended pre-processing stage decreasing the variance between EEG recording from differing sleep centres could improve the robustness of the classifier.

Chapter 4 assessed CAP the first time in large community-based cohort-studies of older people. As a result, new insights into the relationship between age and CAP were obtained. Future studies with community-based cohorts from every age group would supplement existing knowledge about the influence of age on sleep microstructure. Additionally, it would establish standards as reference for healthy subjects in future studies as well as populations with neurological disorders. Future studies on CAP have been planned, including (i) a study on the influence of CAP on delayed verbal memory and overnight memory retention in collaboration with the University of Stanford, Stanford, USA, (ii) a study on transcutaneous vagal nerve stimulation effects on sleep microstructure in veterans in collaboration with the University of Florida, Gainesville, USA, and the Oasi Research Institute-IRCCS, Troina, Italy, and (iii) collaborative studies with the Flinders Medical Centre, Adelaide, Australia.

The interest in collaborations to analyse CAP in large studies demonstrates the demand for a consumer-friendly roll out of our detection software. A potential solution could be achieved in cooperation with commercial manufactures of sleep analysis programs. An integration into commonly used sleep screening diagnostic software would facilitate the access to our CAP scoring system. Another future solution could be the integration into a wearable device. Our system could provide additional information about sleep microstructure in combination with a home-screening device such as EEG headbands or in-ear EEG.

Finally, Chapter 6 investigates pre-scored data from the CAP Sleep Database to assess causal relations between cortical and cardiovascular activity. The application of automated CAP scoring algorithms that are capable of scoring large cohort studies such as the system presented in Chapter 3 would allow an analysis of a larger number of subjects resulting in a more detailed picture about the cortical-cardiovascular coupling during CAP sequences. Moreover, it would eliminate the possible limitations due to subject's age or health record.

7.3 Closing statement

This chapter summarized the major findings and conclusions of this thesis followed by recommendations for future work. This thesis has made a number of contributions towards enhancing the understanding of sleep microstructure by developing a high-performance end-to-end detection system and applying it on large sleep cohorts. The work herein is unique and original, laying groundwork for future analyses of sleep microstructure in healthy populations and cohorts with pathologies.

Appendix

A

Improved A-phase Detection of Cyclic Alternating Pattern Using Deep Learning

The content of this chapter is a modified version of the publication:

Hartmann, S. and Baumert, M. (2019), 'Improved A-phase detection of cyclic alternating pattern using deep learning', *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1842–1845.

Abstract

In recent years, machine learning algorithms have become increasingly popular for analysing biomedical signals. This includes the detection of cyclic alternating pattern (CAP) in electroencephalography recordings. Here, we investigate the performance gain of a recurrent neural network (RNN) for CAP scoring in comparison to standard classification methods. We analysed 15 recordings (n1–n15) from the publicly available CAP Sleep Database on PhysioNet to evaluate each machine learning method. A long short-term memory (LSTM) network increases the accuracy and F_1-score by 0.5–3.5% and 3.5–8%, respectively, compared to commonly used classification algorithms such as linear discriminant analysis, k-nearest neighbour or feed-forward neural network. Our results show that by using a LSTM classifier the quantity of correctly detected CAP events can be increased and the number of wrongly classified periods reduced. RNNs significantly improve the precision in CAP scoring by taking advantage of available information from the past for deciding current classification.

Introduction

Since the beginning of electronic computers, novel methods have been developed to train the new artificial workforce aiming to solve the fundamental problem of detecting patterns in large data (Kononenko, 2001; Bishop, 2006). The ability to learn how to automatically recognize regularities without being explicitly programmed for it was then subsumed under the term machine learning (Park *et al.*, 2018). Machine learning methods can be roughly divided into statistical methods such as *k*-nearest neighbour (k-NN), discriminant analysis like linear discriminant analysis (LDA), Bayesian classifiers such as decision trees and artificial neural networks including deep learning methods (Kononenko, 2001).

Besides the typical fields of application in computer vision like object recognition or face detection, machine learning algorithms were soon transferred to biomedicine due the existence of large data sets and various imaging systems. One major application field for machine learning in the area of biomedical research is the decoding of neuronal activity either to control prosthetic devices or to enhance the understanding of the human brain (Darvishi *et al.*, 2017; Yarkoni *et al.*, 2011). Considering sleep research, machine

learning algorithms are used to automatically classify sleep stages or detect neurophysiological patterns. A prominent example for such events are cyclic alternating pattern (CAP).

CAP was introduced in 2001 as alternative concept to characterize non-rapid eye movement (NREM) sleep (Terzano *et al.*, 2001). A CAP sequence is defined by two or more CAP cycles, which themselves, consist of an activation phase (A-phase) which represents transient, prominent events and a background phase (B-phase). CAP can by definition only occur in sleep stages without rapid eye movement (REM). The more relevant activation phase is characterized by slower high-voltage rhythms, faster lower-voltage rhythms or by both (Terzano and Parrino, 2005).

The first implemented automated CAP detection systems concentrated on the distinctive characteristics between the activation phase and the background phase (Mendez *et al.*, 2016). After the alteration in signal amplitude averages between short and long time periods was applied as indicator, statistical or spectral features were extracted from electroencephalography (EEG) in combination with either thresholding classification algorithms (Niknazar *et al.*, 2015) or competitive machine learning algorithms (Mendez *et al.*, 2016; Mariani *et al.*, 2012). However, previous methods investigated only shallow classifier such as LDA, k-NN, support vector machines (SVM) or feed-forward neural networks (NN) as machine learning methods.

Here, we show that a deep learning method such as long short-term memory network (LSTM) increases the scoring performance of automatic CAP detection methods. The more sophisticated deep learning algorithms effectively craft new features out of the input resulting in important information for detection. After the method is explained and the various classification algorithms under test are described in detail, the evaluation results on a publicly available data set are presented. Finally, the numbers are discussed including limitations of the proposed method.

Materials and Methods

Database

As data set, we used polysomnographic recordings of 15 normal, healthy subjects (n1–n15) from the CAP Sleep Database on PhysioNet, which is an open-source repository for recorded physiologic signals (Goldberger *et al.*, 2000). We selected normal, healthy subjects to exclude additional information from sleep pathologies. The list of physiological

Table A.1. Statistics of sleep macrostructure and CAP occurrence for subjects n1 - n15 in seconds.

Subject	Wake	REM	NREM	A1	A2	A3	Total sleep time
n1	1170	7170	25860	2217	747	1135	34410
n2	4290	4530	21150	1188	688	1239	30000
n3	4080	5640	20250	656	631	1043	30030
n4	6690	5940	16740	986	356	893	30330
n5	300	6960	22950	2863	328	784	30240
n6	1740	7920	21240	1871	976	1414	31200
n7	1950	7410	19860	1616	565	480	29550
n8	3780	5910	20070	949	465	1876	30030
n9	3540	6780	20280	1036	377	678	30840
n10	1980	6450	17130	1489	336	922	25800
n11	1680	11400	18450	1724	583	796	31590
n12	930	8910	19680	1064	153	573	29700
n13	5820	5430	17310	1628	1040	1041	29100
n14	4710	4950	19290	1037	1234	1209	29040
n15	1170	5940	22170	1449	1046	1244	29310
Total	43830	101340	302430	21773	9525	15327	451170

REM, rapid eye movement; NREM, non-rapid eye movement

signals comprises generally at least one EEG channel (C3 or C4), multiple bipolar EEG channels and other parameter such as electrocardiography (ECG) or eye movement signals. Additionally to each subject's recordings, an annotation file including sleep staging and CAP scoring is provided. An expert neurologists performed the manual scoring according to the Rechtschaffen & Kales rules (Rechtschaffen and Kales, 1968) and the atlas of CAP scoring (Terzano *et al.*, 2001). The manual CAP scoring is regarded as ground truth during the supervised training process of the various classification algorithms. The data contains a total of 7519.5 minutes of scored sleeping time. Since CAP events only occur in NREM stages (Terzano *et al.*, 2001), the data set comprises 5040.5 minutes of scoring-relevant data of which 15.4% are A-phases and 84.6% pertain to background periods. Summary statistics of the sleep macrostructure and CAP occurrence for each subject and in total are listed in table A.1.

Pre-processing

Each polysomnographic recording was processed before extracting features in the signal to ensure a homogeneous input for the classification step. The existent central EEG recording in each subject's polysomnogram (either C4-A1 or C3-A2) was used as channel for the scoring process. The data were resampled at a frequency of 128 Hz due to a sampling frequency range of 100–512 Hz in the data set. Afterwards, a bandpass FIR filter was applied to reject information outside the common EEG frequency range (0.5–30 Hz). Subsequently, the filtered data were divided into following EEG rhythms by applying a least-squares linear phase FIR filter bank: *Delta* (0.5–4 Hz), *Theta* (4–8 Hz), *Alpha* (8–12 Hz), *Sigma* (12–16 Hz) and *Beta* (16–30 Hz).

Feature extraction

The following features were selected based on their performance reported in previous studies (Machado *et al.*, 2015; Karimzadeh *et al.*, 2015; Mariani *et al.*, 2011b): Hjorth activity, Shannon entropy, Teager Energy Operator (TEO), band power descriptor, and differential EEG variance. All features are computed for each second resulting in a sample rate of 1 Hz. Firstly, the Hjorth activity was calculated for the delta band using 3s-overlapping windows. Then, Shannon entropy and Teager Energy Operator were derived from each frequency band in 2-s signal windows. To extract the temporal power shifts in the frequency bands, a band power descriptor was computed based on 2–60 second windows. Finally, the variance difference of the EEG was determined on 1-s non-overlapping windows.

Classifier

Linear Discriminant Analysis

The LDA algorithm separates data by determining the probability function of all classes. In general, the classifier estimates the probability density functions as a multivariate normal distribution (Garrett *et al.*, 2003). In the case of linear discriminant analysis, the covariance of all classes is the same, only the means vary between the classes. The resulting classifier divides the feature space into different classes by a $(D-1)$-dimensional hyperplane with D representing the dimension of the input (Bishop, 2006). We selected a linear discriminant analysis classifier with the dimension size equivalent to the number of extracted features.

k-Nearest Neighbours

The k-nearest neighbours classification rule is based on the simple idea of selecting the majority class of the surrounding *k*-nearest neighbours as classification output for an unclassified input vector (Keller *et al.*, 1985). The greater the number of neighbours are considered, the fewer larger regions are created forming k into a parameter for the degree of smoothing (Bishop, 2006). In the binary case, an odd number for k avoids a tie for classification, otherwise the distances of the sample to each neighbour are considered as tie-breaker (Keller *et al.*, 1985).

Neural Network

Artificial neural networks comprise a wide field of different variations. Here, we implemented a shallow feed-forward neural network with one hidden layer. In general, a neural network can be regarded as a mapping architecture which will assign an input of the dimension n to an output of the dimension m by using a sequential structure of rows with processing units (Hecht-Nielsen, 1989). Such a processing unit is called neuron and creates initially a linear combination of the input variables combined with weights and a bias (Bishop, 2006). A non-linear activation function is subsequently wrapped around the linear combination to exploit non-linearities in the data. In a backpropagation neural network, the parameters for each neuron are updated using the backpropagation algorithm i.e. the error between the predicted output and the target output is propagated backwards from the last layer to the first instance modifying the parameters with the objective to minimize the error. The error is calculated using a loss function like cross-entropy or means squared error. In this study, the implemented neural network contains a single fully connected layer with 640 neurons, a rectified linear unit (ReLU) as activation function, cross-entropy as loss function and a softmax layer as output.

Long Short-term Memory Network

Long Short-term Memory networks are a subclass of recurrent neural networks (RNN) which exploit information from the past as well as the current input to create a prediction or classification. A basic RNN contains time-sequentially stacked neural networks which pass the information from previous states to the current network. Hence, the state of the last time step relies not only on the current input but also on the information of previous hidden states (Graves *et al.*, 2013). Importantly, LSTM networks differ from basic RNNs by

resolving long-term dependencies which are inevitable in a large chain of feed-forward networks (Bengio *et al.*, 1994). A LSTM cell possesses the ability to decide which information is transferred to the next cell by simply erasing irrelevant data (Graves *et al.*, 2013). The sequence length of the RNN determines how much of the previous temporal information is exploited for the computation of the current state. In this study, we used a LSTM network with two layers of 128 and 64 cells respectively. Following the LSTM layers, a fully connected layer equivalent to the layer used in the neural network and a softmax output layer are added. The parameters of each layer are updated using the backpropagation algorithm and cross-entropy as cost function.

Post-processing

To predict activation phases in accordance with the atlas for CAP scoring, the outcomes of the classifiers were modified as it is described in Mariani *et al.* (2012). Isolated one-second classifications were assigned to their neighbour values and predicted A-phases longer than 60 seconds were reclassified with a self-organizing neural network.

Performance measures

The classification performance was quantified based on standard measures for binary classification: accuracy (ACC), true positive rate or sensitivity (TPR), specificity or true negative rate (SPC) and the F_1-score. Each measure depends on the number of correctly identified events (true positives, t_p), the number of correctly recognized background phases (true negatives, t_n), and the number of seconds which were misidentified either as A-phase (false positive, f_p) or as background phase (false negative, f_n). Hence, the objective measures are computed as follows (Sokolova and Lapalme, 2009):

$$\mathrm{ACC} = \frac{t_p + t_n}{t_p + f_n + f_p + t_n} \tag{A.1a}$$

$$\mathrm{TPR} = \frac{t_p}{t_p + f_n} \tag{A.1b}$$

$$\mathrm{F_1} = \frac{2 \cdot t_p}{2 \cdot t_p + f_n + f_p} \tag{A.1c}$$

$$\mathrm{SPC} = \frac{t_n}{f_p + t_n}. \tag{A.1d}$$

Table A.2. Average performance measures of each classification method on data set.

	TPR	SPC	ACC	F_1-score
LDA	59.01±10.34	86.31±7.16	82.14±5.09	49.99±4.55
k-NN	73.03±6.69	81.35±7.42	80.08±6.02	52.92±6.82
Feed-forward NN	79.28±7.53	79.70±8.78	79.58±7.23	54.65±8.02
LSTM	76.15±10.69	84.20±8.59	82.95±7.40	58.23±10.36

TPR, true positive rate; SPC, specificity; ACC, accuracy; NN, neural network; LSTM, long-short term memory network; LDA, linear discriminant analysis; k-NN, k-nearest neighbor

Test environment

For training and testing of the classification algorithms, the Leave-one-out (LOO) method was applied. The LOO method is a k-fold cross-validation algorithm in which for each fold one subject is determined as test set and all remaining subjects are merged into the training set. Thus, the classifier is trained k times on dataset D/D_t and tested on D_t, where D represents the entire dataset, k is the total number of subjects in the set (here: 12) and $t \in 1, 2, ..., k$ (Kohavi, 1995). After looping through every subject, the performance measures are determined by summing up each individual validation value and divide the sum by the number of instances. Prior to training, the quantity of training data for each class was levelled to obtain an unbiased classifier.

Results

In table A.2, the performance measures of all four classification methods are listed. The values represent the means plus the standard deviation of the 15 subjects. The results show that the LSTM method achieves the highest accuracy and F_1-score (ACC:0.5–3.5%, F_1-score:3.5–8%). Regarding the sensitivity values, the NN outperforms the LSTM method by 3% and the remaining two algorithms. Especially, the LDA classifier scores a low number of A-phases correctly resulting in a poor sensitivity of 59%. However, the LDA algorithm obtains the highest specificity value indicating that predominantly background phases were classified. Compared to the NN, the k-NN achieves a higher accuracy and specificity but the sensitivity and F_1-score are decreased by ∽6% and ∽2.5%, respectively.

Discussion

In this study, we show that a deep learning LSTM network increases the performance of automatic CAP classification. Consequently, previous EEG information can be important for

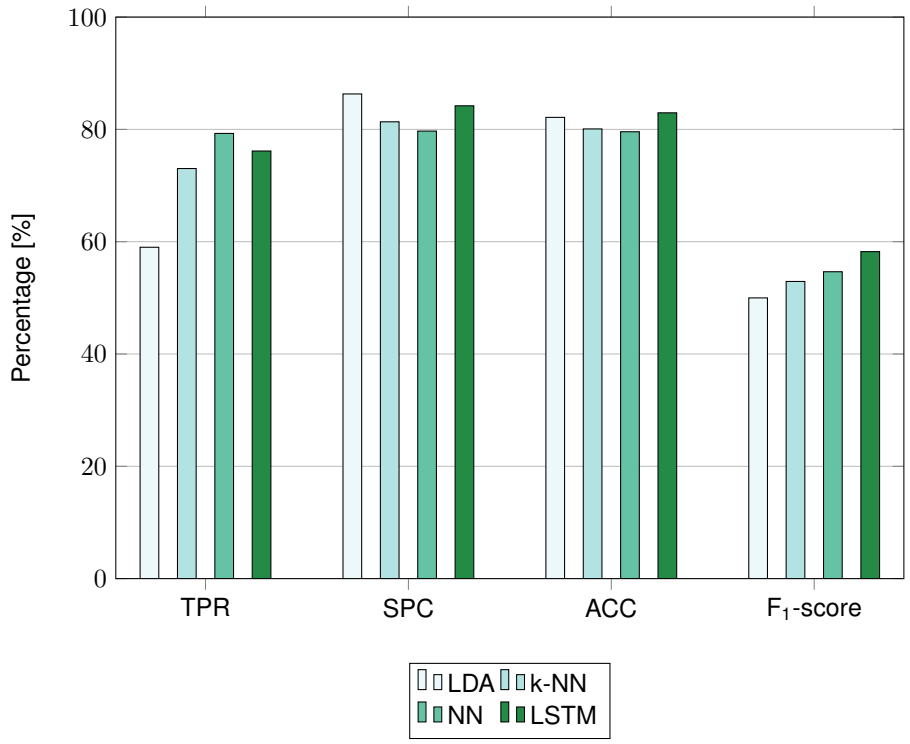


Figure A.1. Comparison of LDA, k-NN, NN and LSTM based on results from data set. Comparison of LDA, k-NN, NN and LSTM based on results from data set

the scoring of A-phases in sleep recordings. To evaluate the LSTM method, we compared it to common shallow classification methods used in this field like LDA, k-NN and feed-forward NN.

As results in table A.2 and A.3 indicate, LSTM networks increase the accuracy and F_1-score by 0.5–3.5% and 3.5–8%, respectively. Especially the high F_1-score demonstrates that the LSTM algorithm detects more precisely A-phases of CAP events than the remaining classifiers tested. Considering the statistics of the data set (table A.2), the proportion of A-phases during night is relatively low making a precise scoring substantial for further CAP analysis. In contrast, the LDA algorithm achieves a comparably good accuracy value but mostly due to a high specificity value indicating that the classifier predominantly classifies background phases, which is counteractive in case of CAP scoring.

Furthermore, the feed-forward NN scores a higher quantity of correct activation phases compared to LSTM but at the expense of precision, indicated by the lower F_1-score. Finally, the performance values for the k-NN algorithm show that it is an efficient alternative to artificial neural networks for CAP detection. The training process for both neural network algorithms exceeds largely the k-NN method in both time and computing resources. Especially, the LSTM network demands a large amount of computing resources during the

Table A.3. Detailed list of performance measures for each classification method.

Subject	LDA				k-NN				NN				LSTM			
	TPR (%)	SPC (%)	ACC (%)	F_1-score (%)	TPR (%)	SPC (%)	ACC (%)	F_1-score (%)	TPR (%)	SPC (%)	ACC (%)	F_1-score (%)	TPR (%)	SPC (%)	ACC (%)	F_1-score (%)
n1	46.63	92.78	85.53	50.31	67.07	86.86	83.75	56.46	76.98	85.28	83.97	60.15	72.73	89.03	86.47	62.82
n2	49.32	86.96	81.44	43.82	63.13	78.14	75.94	43.50	78.15	70.50	71.62	44.70	61.94	78.91	76.42	43.54
n3	64.73	84.69	82.40	45.72	75.40	80.78	80.16	46.53	84.65	79.96	80.50	49.84	82.88	80.32	80.62	49.48
n4	62.29	86.89	83.79	49.22	78.50	82.36	81.87	52.21	87.77	76.66	78.06	50.24	87.06	83.23	83.71	57.44
n5	54.91	89.69	83.67	53.77	75.42	83.43	82.04	59.22	79.39	83.27	82.60	61.21	79.27	85.40	84.34	63.65
n6	48.28	93.39	84.50	55.10	67.37	89.28	84.97	63.84	74.75	87.04	84.62	65.69	72.89	92.27	88.45	71.32
n7	83.72	72.78	74.22	46.10	90.82	68.31	71.27	45.44	94.66	65.02	68.92	44.51	93.19	74.28	76.78	51.39
n8	49.15	91.11	84.34	50.32	72.36	86.99	84.63	60.30	84.73	81.83	82.30	60.71	75.90	89.39	87.21	65.71
n9	65.20	89.86	87.35	51.20	75.35	88.57	87.23	54.57	82.34	87.12	86.63	55.64	83.29	90.46	89.73	62.29
n10	66.97	70.09	69.61	40.46	72.30	65.69	66.71	40.12	71.63	61.13	62.75	37.23	64.32	60.97	61.49	34.01
n11	58.17	86.64	81.87	51.82	70.06	76.78	75.65	49.10	68.68	80.64	78.63	51.87	55.91	88.22	82.80	52.16
n12	50.56	93.92	90.01	47.71	70.22	89.89	88.12	51.59	74.69	91.05	89.58	56.37	68.77	93.27	91.06	58.11
n13	72.16	78.58	77.25	56.85	78.39	74.31	75.16	56.72	83.01	73.59	75.55	58.50	88.43	78.56	80.61	65.46
n14	54.69	88.53	82.46	52.80	65.90	83.97	80.73	55.10	65.93	89.00	84.86	60.98	70.76	91.76	87.99	67.90
n15	58.41	88.79	83.68	54.65	73.20	84.87	82.91	59.06	81.81	83.43	83.16	62.07	84.94	86.98	86.63	68.17
Mean ± std	59.01±10.34	86.31±7.16	82.14±5.09	49.99±4.55	73.03±6.69	81.35±7.42	80.08±6.02	52.92±6.82	79.28±7.53	79.70±8.78	79.58±7.23	54.65±8.02	76.15±10.69	84.20±8.59	82.95±7.40	58.23±10.36

TPR, true positive rate; SPC, specificity; ACC, accuracy; NN, neural network; LSTM, long-short term memory network; LDA, linear discriminant analysis; k-NN, k-nearest neighbor

training process to calculate each parameter. Nevertheless, the classification process itself using a trained classifier needs negligibly more amount of time compared to the remaining classifiers.

Acknowledgement

This work was supported with supercomputing resources provided by the Phoenix HPC service at the University of Adelaide.

Appendix

B

Non-REM sleep instability in children with restless sleep disorder

The content of this chapter is a modified version of the publication:

DelRosso, L. M., Hartmann, S., Baumert, M., Bruni, O., Ruth, C. and Ferri, R. (2020), 'Non-REM sleep instability in children with restless sleep disorder', *Sleep Medicine* **75**, pp. 276–281.

Abstract

Study objectives: Restless sleep disorder (RSD) is a newly recognized condition characterized by motor movements involving large muscle groups with frequent repositioning or bed sheets disruption. We analysed cyclic alternating pattern (CAP) in these children, a marker of sleep instability that might be associated with the motor episodes of RSD and may play a role in their daytime symptoms.
Methods: Polysomnographic recordings from thirty-eight children who fulfilled RSD diagnostic criteria (23 boys and 15 girls), 23 children with restless legs syndrome (RLS, 18 boys and 5 girls) and 19 controls (10 boys and 9 girls) were included. For CAP analysis, a previously developed, highly precise automated system, based on a deep learning recurrent neural network, was used.
Results: Age and gender were not statistically different between groups. RSD patients showed a lower percentage of A3 CAP subtypes than controls (median 9.8 vs. 18.2, $p = 0.0089$), accompanied by shorter duration of the B phase of the CAP cycle (median 28.2 vs. 29.8 in controls, 30.2 in RLS, $p = 0.005$) and shorter CAP cycle duration than both controls and RLS subjects (median 33.8 vs. 35.0 in controls, 35.8 in RLS, $p = 0.002$). Finally, RSD children also showed a longer duration of CAP cycle sequences, when compared to controls (median 172.7 vs. 141.9, $p = 0.0063$).
Conclusions: In conclusion, our study indicates that NREM sleep EEG shows an increased instability in RSD; these findings add to the current knowledge on the mechanisms of this newly recognized sleep disorder and suggest that sleep instability might be a favouring mechanism for the emergence of the motor episodes characterizing RSD.

Introduction

Restless sleep has been mentioned in the literature in the past few decades in association with overall detrimental effect on wellbeing (Prendergast *et al.*, 2016), cognition (Qureshi *et al.*, 2014), ADHD (Greenhill *et al.*, 1983), psychiatric conditions (Simonds and Parraga, 1984), etc. Only recently it has been identified as a disorder in children (Del-Rosso *et al.*, 2018). Restless sleep disorder (RSD) is a condition characterized by motor

movements involving large muscle groups often described by parents as frequent repositioning or disruption of the bed sheets (DelRosso *et al.*, 2018, 2019). The movements are sleep related, occurring throughout the night and associated with daytime impairment (DelRosso *et al.*, 2018). Despite the increasing knowledge on the clinical and polysomnographic aspects of RSD in the last couple of years (DelRosso *et al.*, 2019; DelRosso and Ferri, 2019), its mechanisms remain elusive and need clarification. We strongly suspect that sleep instability associated with the frequent nocturnal movements in children with RSD may play a role in the daytime symptoms of sleepiness, fatigue or inattention. We decided to analyse cyclic alternating pattern (CAP) in these children and compare them with controls and children with restless legs syndrome (RLS) to further identify patterns of sleep instability.

CAP is an endogenous and physiological rhythm occurring in NREM sleep characterized by a quasi-periodic EEG activity with sequences of transient electro–cortical activations (phase A of the cycle) that clearly interrupt the background EEG activity (phase B of the cycle) (Terzano *et al.*, 1985). These sequences are repeated several times during the night and organized in a recurrent pattern interrupted by the presence of a stable sleep period, without oscillations, called non-CAP (NCAP), longer than 60 s. Sequences of CAP are orderly distributed in NREM sleep (Terzano *et al.*, 1988), and the percentage of CAP time to NREM sleep time (CAP rate) is considered to be a physiologic marker of NREM sleep instability (Parrino *et al.*, 2012). CAP A phases are subdivided into different subtypes: A1, A2 and A3, based on their frequency content (Terzano *et al.*, 2001; Parrino *et al.*, 2001), with the A1 subtype composed prevalently by slow-waves (EEG synchrony), A3 with prevalence of fast EEG activities (EEG desynchrony), and A2 presenting a combination of both (Terzano *et al.*, 2001).

The aim of this study was to analyse CAP in a group of children and adolescents with RSD and to compare it to that of age-matched normal controls and patients with RLS, in order to test our original hypothesis of the presence of a sleep instability in RSD and to observe eventual differences and similarities with RLS.

Methods

Subjects

Thirty-eight children who fulfilled RSD diagnostic criteria (DelRosso *et al.*, 2019) (23 boys and 15 girls, age range 5–17 years), 23 children with RLS (18 boys and 5 girls, age

range 4–17 years) and 19 controls (10 boys and 9 girls, age range 5–18 years) from our database were included. Exclusion criteria were: age younger than 4 years, use of medications that altered sleep parameters (i.e. antihistamines, antidepressants, antiepileptics, etc.), presence of co-morbid sleep disorder (i.e. obstructive sleep apnoea, central sleep apnoea, parasomnias, behavioural insomnia), medical or psychiatric conditions known to affect sleep (uncontrolled eczema, asthma, pain, neurodevelopmental disorders, genetic syndrome, neuromuscular disorders) or use of caffeine.

The sample size, even if not very large, was decided on the basis of practical considerations on the real possibilities of recruitment of patients and on the availability of PSG recordings of good general quality and with both signals from the central leads (left and right) without important interruptions during NREM sleep. All children underwent PSG. The study was approved by the local institutional review board.

Polysomnography

PSG was recorded following the AASM standards (Berry *et al.*, 2015) and included EEG (two frontal, two central, and two occipital channels, referred to the contralateral mastoid); electrooculogram, electromyogram (EMG) of the submentalis muscle, EMG of the right and left tibialis anterior muscles, respiratory signals, effort signals for thorax and abdomen, oximetry, capnography, a single lead ECG, video and audio recording. Calibrations were performed per routine standard by the sleep technician. Epochs and all sleep parameters were scored by a certified sleep technologist and board certified sleep physician, according to standard criteria (Berry *et al.*, 2015).

Automated CAP A subtype detection and CAP parameter computation

We deployed a previously developed, highly precise automated system for CAP analysis, which has been reported in detail elsewhere (Hartmann and Baumert, 2019). The detection system achieves on average a second-by-second A phase inter-rate reliability of 0.55, quantified by the Cohen's kappa coefficient, which is among the event-based inter-rater reliability between human scorers of 0.42–0.75 (Ferri *et al.*, 2005b). Briefly, the detection system consists of a deep learning recurrent neural network (RNN) that was trained with manually scored recordings from children to recognize CAP A subtypes in paediatric EEG recordings. Measurements with severe clipping, which can be often found in recordings with

a low bit rate or a low physical range, and unipolar signals do not meet the requirements of the automated A phase detection system. In the first step, the raw signal of one central EEG channel is filtered and processed to remove power line interference, high-frequency noise, and cardiac activity and eye movement artefacts. Multiple features in the time and frequency domain such as Hjorth activity, Shannon entropy, Teager Energy Operator (TEO), band power descriptor, and differential EEG variance are subsequently extracted from the processed signal and passed as input to the RNN classifier. The F_β-score was used as cost function to deal with the biasing issues of a low number of A phase representations in overnight EEG recordings and to concentrate on the sensitivity and precision of the scoring. Further, the precision of the detection system was increased by selecting only A phases that were detected in two central EEG channels. Finally, we post-processed the output of the A subtype detection system by applying the rules for CAP sequences (Parrino *et al.*, 2012; Terzano *et al.*, 2001) and computed the following parameters, for each recording:

a) CAP Rate: percentage of non REM sleep time occupied by CAP sequences;
b) A1 subtype: percentage of A1 subtypes among the total A subtypes;
c) A2 subtype: percentage of A2 subtypes among the total A subtypes;
d) A3 subtype: percentage of A3 subtypes among the total A subtypes;
e) A1 subtype mean duration;
f) A2 subtype mean duration;
g) A3 subtype mean duration;
h) B phase mean duration;
i) CAP cycle (A + B phase) mean duration;
j) CAP sequence mean duration.

Statistical analysis

The comparison of the gender composition of the groups of subjects was carried out by means of the chi-square test. Because of the non-normal distribution of several variables, non-parametric statistics were used (Siegel, 1956). For the comparison between the three groups of subjects, the Kruskal-Wallis ANOVA was computed, followed by post-hoc comparisons of mean ranks of all pairs of groups and p-values (Mann-Whitney test) associated with each comparison were obtained. Finally, correlations were analysed by computing the Spearman rank correlation coefficient

Table B.1. Comparison of sleep architecture variables obtained in the 3 groups of subjects.

	Controls (n=19)		RSD (n=38)		RLS (n=23)		Kruskal-Wallis ANOVA		Mann-Whitney post-hoc		
	Median	IQ range	Median	IQ range	Median	IQ range	$H_{(2,80)}$	*p*	1 vs. 2	2 vs. 3	1 vs.3
Time in bed, min	513.0	(484.5–537.0)	510.5	(474.0–555.0)	504.0	(482.5–544.0)	0.224	NS			
Sleep period time, min	480.0	(445.0–499.5)	470.0	(438.5–507.0)	478.5	(446.5–496.0)	0.070	NS			
Total sleep time, min	432.0	(386.0–470.5)	445.5	(413.5–494.5)	452.0	(379.0–484.0)	1.336	NS			
Sleep latency, min	23.8	(11.5–42.0)	24.0	(10.0–50.0)	22.0	(13.5–39.0)	0.042	NS			
First R latency, min	136.3	(103.0–177.5)	91.5	(67.0–153.5)	161.0	(116.5–232.0)	7.929	0.019	0.036	NS	0.012
Stage shifts/hour	11.5	(8.2–15.4)	10.2	(7.5–11.5)	12.0	(9.1–15.7)	3.120	NS			
Awakenings/hour	4.4	(3.3–6.8)	3.6	(1.7–5.2)	4.5	(2.9–6.0)	4.666	NS			
Sleep efficiency, %	86.7	(76.6–91.2)	91.3	(81.2–93.2)	89.6	(80.6–93.2)	3.849	NS			
Stage W, %	8.1	(4.3–13.0)	4.3	(2.4–7.0)	4.7	(2.7–9.3)	6.269	0.0435	0.012	NS	NS
Stage N1, %	5.5	(4.0–8.7)	5.0	(3.3–7.1)	5.7	(3.2–11.2)	0.836	NS			
Stage N 2, %	41.9	(37.1–52.8)	46.1	(39.5–49.2)	45.0	(38.4–47.8)	1.122	NS			
Stage N3, %	24.0	(20.3–26.2)	22.6	(18.7–26.0)	23.2	(20.4–30.7)	2.031	NS			
Stage R, %	15.3	(10.3–19.9)	21.2	(17.2–26.4)	14.7	(9.0–17.3)	15.736	0.0004	0.0008	NS	0.00026

IQ = interquartile range; NS = not significant

Results

The gender composition of the three groups was not statistically different (chi-square = 3.28, NS). The groups did not differ also for age (Kruskal-Wallis ANOVA $H_{(2,80)}$ = 2.814, NS).

Table B.1 reports the comparison between sleep architecture variables obtained in the 3 groups of subjects. RSD children showed a REM sleep latency significantly shorter and percentage of REM sleep significantly higher than those of the other two groups; moreover, they had an amount of wakefulness after sleep onset significantly lower than that of controls.

Regarding CAP variables, reported in Table B.2, RSD patients showed a lower percentage of A3 subtypes than controls (but not a lower number per hour), accompanied by shorter duration of the B phase of the CAP cycle and shorter CAP cycle duration than both controls and RLS subjects. Finally, RSD children also showed a longer duration of CAP cycle sequences, when compared to controls.

As age-related changes of CAP parameters can be expected in children (Parrino *et al.*, 2012; Bruni *et al.*, 2002, 2005), the correlation between CAP cycle duration and age was further analysed in the three groups of subjects by drawing scatterplots of these two variables. Figure B.1 shows such an analysis showing a general decrease with age of this parameter; in addition, the regression line pertaining to the RSD group is evidently below the lines obtained in the other two groups at all ages, with a tendency to converge towards the right side of the graph, because of a steeper decrease in the lines relative to the RLS and control groups; however, only in controls this correlation reached a statistical significance (Spearman rank correlation coefficient $\rho = 0.456$, $p < 0.05$).

Table B.2. Comparison of CAP variables obtained in the 3 groups of subjects.

	Controls (n=19)		RSD (n=38)		RLS (n=23)		Kruskal-Wallis ANOVA		Mann-Whitney post-hoc		
	Median	IQ range	Median	IQ range	Median	IQ range	$H_{(2,80)}$	*p*	1 vs. 2	2 vs. 3	1 vs.3
CAP Rate, %	24.5	11.3–29.2	33.4	20.2–45.9	25.9	13.3–36.4	5.373	NS			
A1 index, n/hour	23.9	14.0–28.5	32.3	19.6–46.6	22.0	10.8–36.9	5.822	NS			
A2 index, n/hour	1.6	0.9–4.5	1.2	0.5–2.8	1.2	0.5–1.8	2.164	NS			
A3 index, n/hour	3.7	1.9–7.6	3.0	1.8–4.3	2.5	1.6–7.3	0.859	NS			
A1 subtype, %	73.9	69.8–82.2	84.5	74.3–88.9	77.3	69.5–82.6	5.720	NS			
A2 subtype, %	7.2	3.5–10.5	4.1	1.7–7.7	5.6	2.4–10.4	3.754	NS			
A3 subtype, %	18.2	14.3–20.7	9.8	7.3–14.7	15.6	8.0–22.2	7.168	0.028	0.0089	NS	NS
A1 duration, s	4.6	4.3–4.8	4.7	4.5–5.0	4.6	4.4–5.0	1.904	NS			
A2 duration, s	6.0	5.7–6.3	5.9	5.3–6.8	6.1	5.5–6.7	0.033	NS			
A3 duration, s	9.3	8.4–10.3	8.6	7.8–9.7	8.6	8.0–9.7	2.952	NS			
B duration, s	29.8	28.6–30.7	28.2	26.2–29.7	30.2	27.6–31.3	10.560	0.005	0.0083	0.0074	NS
CAP cycle duration, s	35.0	34.5–36.2	33.8	31.8–34.9	35.8	33.1–37.3	12.491	0.002	0.0061	0.0026	NS
CAP sequence duration, s	141.9	127.3–151.3	172.7	143.1–211.7	151.2	126.7–178.9	7.844	0.0198	0.0063	NS	NS

IQ = interquartile range; NS = not significant

The top panel of Figure B.2 shows the distribution of onset-to-onset intervals between all consecutive A subtypes during light sleep (N1 and N2), in the three groups of subjects. The bottom panel of the same figure shows this distribution during sleep stage N3. In both panels it is possible to note the tendency of RSD to show higher values for the shorter intervals between approximately 15 and 35 s, although statistical significance was not reached for any of the graph points.

For RSD children only, we tested the correlation between the total number of movements/hour and some CAP variables by means of the Spearman's rank correlation coefficient and were unable to find any significant correlation with very low coefficient values (0.009 for CAP rate, -0.048 for A1 index, -0.082 for A2 index, and 0.091 for A3 index). Finally, we have analysed the occurrence of movements during CAP or NCAP periods of NREM sleep in the RSD group and have found that movements occurred within CAP sequences and associated with the A phases in 33.4% of cases and during NCAP periods in the remaining 66.6%. When occurring within CAP, movements were associated to A1 subtypes in 31.9% of cases, to A2 in 10.4% of cases, and to A3 subtypes in 57.8% of cases.

Discussion

Sleep has important restorative functions that include memory, energy saving, hormone regulation, homeostasis, sympathetic/parasympathetic balance among others. The term non-restorative sleep has for long time been associated with not feeling rested in the morning and feeling that sleep was restless or of poor quality (Ohayon, 2005). Children with

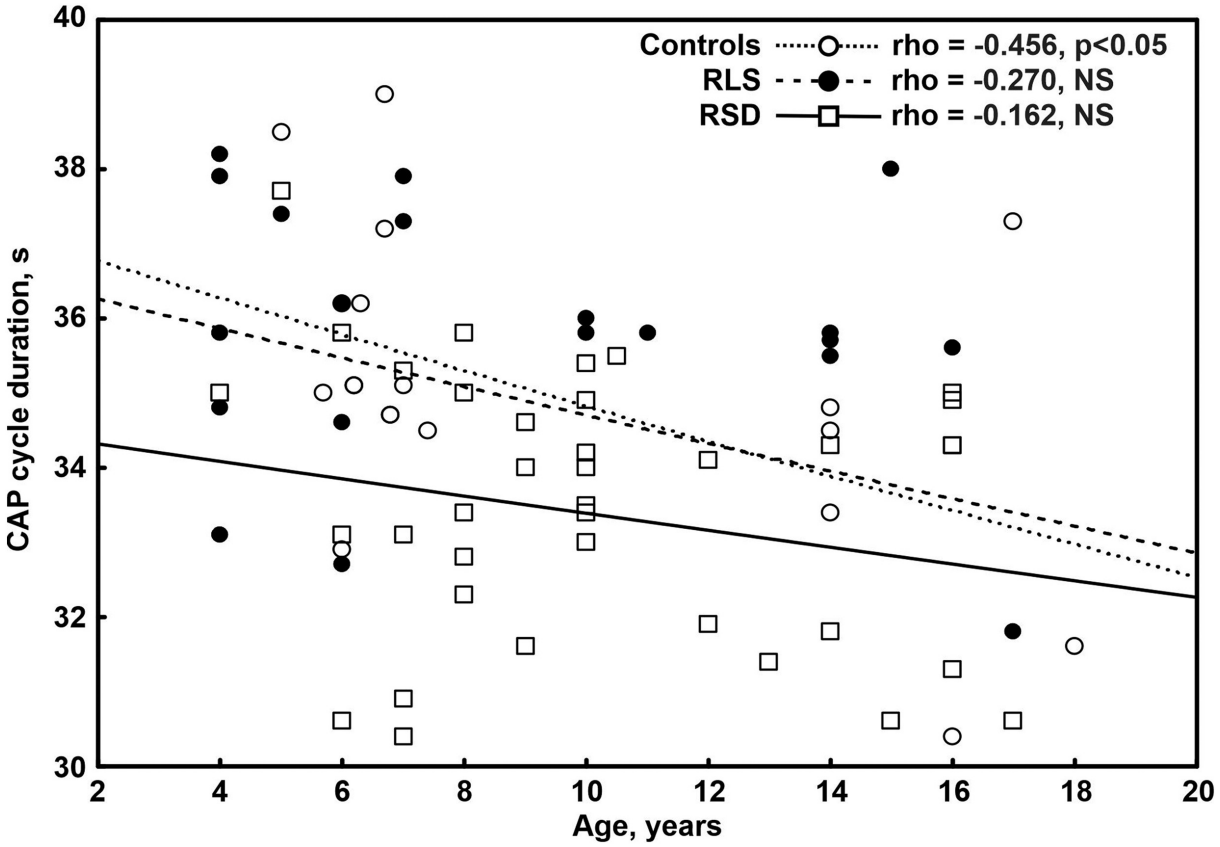


Figure B.1. Age-related changes of CAP cycle duration in the three groups of subjects.. Age-related changes of CAP cycle duration in the three groups of subjects. Also the linear regression line is shown for each group, along with the respective data points; however, the correlation was analysed by means of the Spearman rank correlation coefficient.

RSD, therefore not only show evidence of restless sleep but also evidence of daytime impairment (DelRosso *et al.*, 2018, 2019; DelRosso and Ferri, 2019).

In this study, we have found subtle but important changes in CAP that support our original hypothesis of the presence of an increased sleep instability in children with RSD. In fact, along with a tendency in showing increased CAP rate (which did not reach statistical significance), children with RSD showed clearly shorter CAP cycles, especially because of the shorter return to the lower-amplitude baseline (phase B), as well as longer sequences of CAP cycles. In addition, we found a decrease in the percentage of A3 subtypes (arousals) within these long sequences, as a consequence of the increase in the number/hour of A1 subtypes and not to a real decrease in the number of A3 (Table B.2). This also confirms that evaluating CAP goes beyond the simpler information that can be obtained by analysing only arousals; the classification of CAP subtypes encompasses the consideration of both sleep instability or maintenance (subtypes A1) and sleep fragmentation (subtypes A2 and

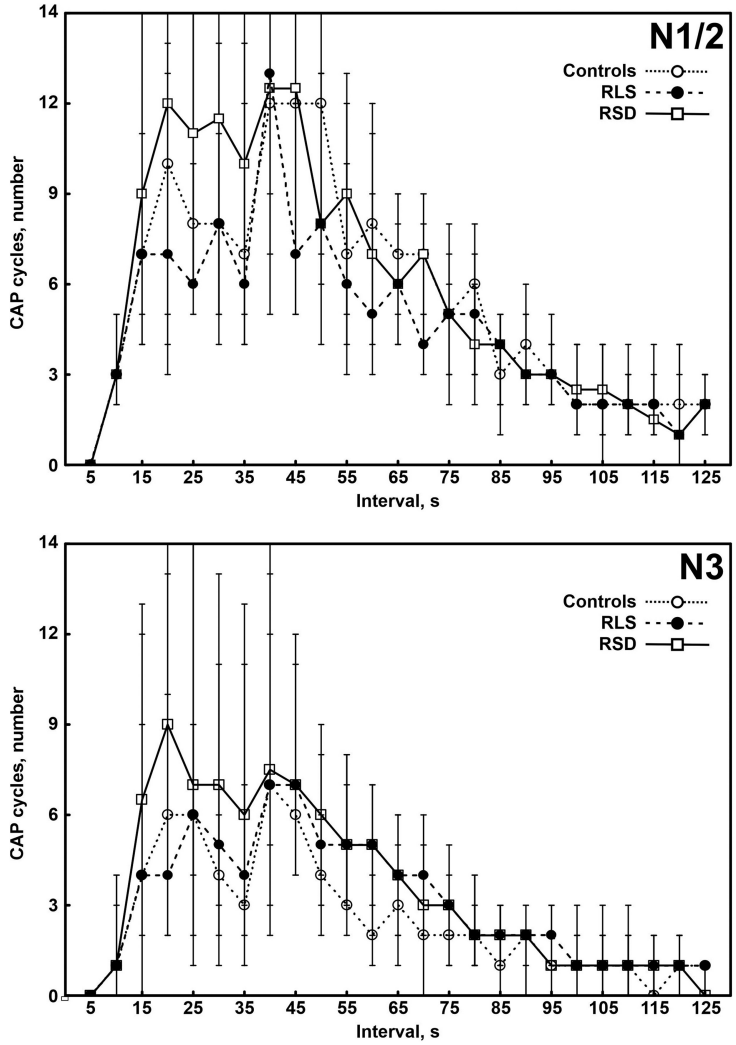


Figure B.2. Time structure of CAP cycles in the three groups of subjects. Data shown as median (circles and squares) and interquartile range (whiskers).. Time structure of CAP cycles in the three groups of subjects. Data shown as median (circles and squares) and interquartile range (whiskers).

A3) (Parrino *et al.*, 2001), with the added value of the consideration of the time structure of NREM sleep microarchitecture (Ferri *et al.*, 2006), in a comprehensive way.

All data reported above seem to indicate the presence of long periods of NREM sleep instability in these children, composed essentially by slow-wave containing transients (A1 subtypes represented 84.5% of all CAP A phases, see Table B.2). It is possible to also speculate that this NREM sleep instability perhaps does not allow them to maintain a long and stable slow-wave sleep, especially in the first sleep cycle, thus favouring the occurrence of REM sleep earlier than in the other children (Table B.1).

NREM sleep instability has been found to be associated with the emergence of movement or behavioural episodes in other sleep disorders in children. In particular, an increased

presence of instability during slow-wave sleep, characterized by an excessive number of short returns to the low-amplitude background EEG activity has been found to occur in children with sleep terrors (Bruni *et al.*, 2008) and sleep walking (Guilleminault *et al.*, 2006b). Similarly, an increase in NREM sleep instability has been indicated as a possible trigger for rhythmic movement disorder in children (Manni *et al.*, 2004).

Thus, the fast oscillatory pattern of EEG potential amplitude picked up by our CAP analysis seems to be similar to that reported in these studies on parasomnia and movement disorders; however, it cannot be used to further speculate on the ultimate nature of RSD, especially if it has to be considered a sleep movement disorder. In this respect, we should notice here that CAP in RSD was found to be different also from that of RLS children who, in contrast with the results of adult studies (Ferri *et al.*, 2010; Manconi *et al.*, 2012), did not seem to show and abnormal amount of CAP in their sleep. This is not surprising because, while in adult sleep disorders CAP tends to show a general increase (with the exception of narcolepsy), in children the most common global CAP change found has been its reduction (or no change) (Parrino *et al.*, 2012). Briefly, decreased CAP rate has been reported in narcolepsy (Ferri *et al.*, 2009), benign epilepsy with rolandic spikes (Bruni *et al.*, 2010a), drug-resistant epilepsy with severe mental retardation (Pereira *et al.*, 2012a), attention-deficit hyperactivity disorder (Miano *et al.*, 2006), Asperger syndrome (Bruni *et al.*, 2007), Prader-Willi syndrome (Verrillo *et al.*, 2009), and Fragile-X syndrome (Miano *et al.*, 2008). Conversely, increased CAP rate has been found in sleep enuresis (Soster *et al.*, 2017). No clear global CAP rate changes (but changes in its subtype distribution) have been reported in autism (Miano *et al.*, 2007), Down syndrome (Miano *et al.*, 2008), different types of lesional or non-lesional drug-resistant epilepsy (Pereira *et al.*, 2012b), and dyslexia (Bruni *et al.*, 2009). Unclear results have finally been reported in childhood sleep apnoea (Kheirandish-Gozal *et al.*, 2007; Miano *et al.*, 2009).

It is particularly interesting to note that movements were not found to be correlated with CAP in our analysis and tended to occur also during NCAP periods (the percentage of movements occurring during CAP was strikingly similar to that of CAP rate e the percentage of NREM sleep time occupied by CAP sequences). However, when occurring within CAP sequences they were most often associated with the CAP A2 and A3 subtypes that are more similar to arousals (if not overlapping with them) (Parrino *et al.*, 2001); this parallels our previous observation that movements in RSD can be followed by awakenings (Del-Rosso *et al.*, 2019) and confirms their sleep-disrupting properties which are probably at the basis of daytime consequences reported in these children. With this new study, we can

speculate that probably also the NREM instability might play a role in determining the above mentioned daytime consequences.

Probably, the main limitation of this study is the number of subjects included, especially in the control group, because practical limitations connected with the cost of PSG and the fact that this study was not funded did not allow us to recruit more participants. However, a preliminary power and sample size analysis would have been very difficult to perform because of the novelty of the condition under analysis (RSD) and, also in children with RLS, CAP has not been analysed in the past. Notwithstanding this important limitation, we were able to find statistically significant differences between the groups pointing at the parameters that were hypothesized to be abnormal in these patients and thus confirming our original hypothesis.

It should also be mentioned here, briefly, that the results of this study are a further confirmation of the validity of the automatic approach used to quantify CAP, previously reported in detail elsewhere (Hartmann and Baumert, 2019).

In conclusion, our study indicates that NREM sleep EEG shows an increased instability in RSD, similar to that already reported for other parasomnia or movement disorders of childhood and characterized by long runs of fast oscillations between transient activation phases (A subtypes) and shorter than normal returns to the background low-amplitude EEG activity (B phase of the CAP cycle). Even if with this study it is not possible to speculate further, these findings add to the current knowledge on the mechanisms of this newly recognized sleep disorder and suggest that sleep instability might be a favouring mechanism for the emergence of the motor episodes characterizing RSD.

Credit authorship contribution statement

Lourdes M. DelRosso: Contributed with writing and editing the manuscript. **Simon Hartmann**: Contributed to Methodology, Resources, Software. **Mathias Baumert**: Contributed to Methodology, Resources, Software. **Oliviero Bruni**: Has collaborated with Data curation, Manuscript auditing, Proof reading. **Chris Ruth**: Contributed with Data curation, Analysis. **Raffaele Ferri**: Contributed with the Conceptualization, Data curation, Formal analysis, Writing and editing.

Disclosure statement

This was not an industry-supported study and was partially supported by a fund from the Italian Ministry of Health “Ricerca Corrente” (RC n. 22751598) (R.F.) and by a grant from the Australian Research Council (DP0663345) (S.H., M.B.).

Appendix

C

Sleep arousal burden is associated with long-term all-cause and cardiovascular mortality in 8001 community-dwelling older men and women

The content of this chapter is a modified version of the publication:

Shahrbabaki, S. S., Linz, D., Hartmann, S., Redline, S., and Baumert, M. (2021), 'Sleep arousal burden is associated with long-term all-cause and cardiovascular mortality in 8001 community-dwelling older men and women', *European Heart Journal* **42**, pp. 2088–2099.

Abstract

Aims: To quantify the arousal burden (AB) across large cohort studies and determine its association with long-term cardiovascular (CV) and overall mortality in men and women.

Methods and results: We measured the AB on overnight polysomnograms of 2782 men in the Osteoporotic Fractures in Men Study (MrOS) Sleep study, 424 women in the Study of Osteoporotic Fractures (SOF) and 2221 men and 2574 women in the Sleep Heart Health Study (SHHS). During 11.2 $\pm$ 2.1 years of follow-up in MrOS, 665 men died, including 236 CV deaths. During 6.4 $\pm$ 1.6 years of follow-up in SOF, 105 women died, including 47 CV deaths. During 10.7 $\pm$ 3.1 years of follow-up in SHHS, 987 participants died, including 344 CV deaths. In women, multivariable Cox proportional hazard analysis adjusted for common confounders demonstrated that AB is associated with all-cause mortality [SOF: hazard ratio (HR) 1.58 (1.01–2.42), p = 0.038; SHHS-women: HR 1.21 (1.06–1.42), p = 0.012] and CV mortality [SOF: HR 2.17 (1.04–4.50), p = 0.037; SHHS-women: HR 1.60 (1.12–2.28), p = 0.009]. In men, the association between AB and all-cause mortality [MrOS: HR 1.11 (0.94–1.32), p = 0.261; SHHS-men: HR 1.31 (1.06–1.62), p = 0.011] and CV mortality [MrOS: HR 1.35 (1.02–1.79), p = 0.034; SHHS-men: HR 1.24 (0.86–1.79), p = 0.271] was less clear.

Conclusions: Nocturnal AB is associated with long-term CV and all-cause mortality in women and to a lesser extent in men.

Graphical Abstract

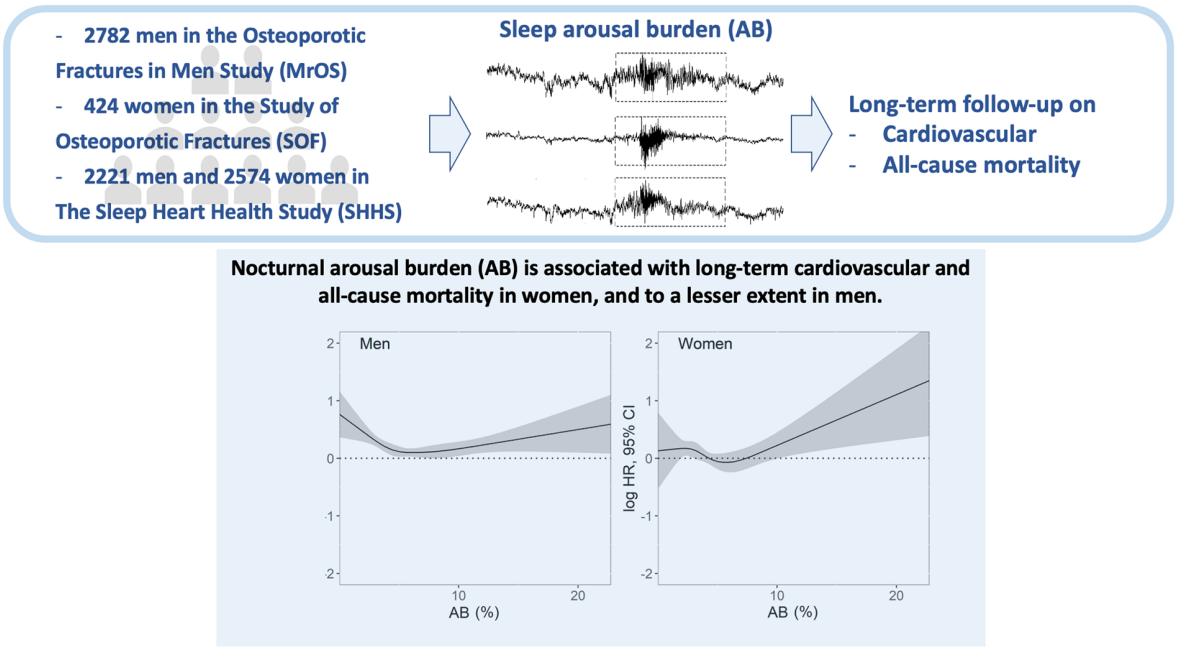


Figure C.1. Graphical abstract. The burden of cortical arousals on overnight polysomnograms was quantified across three large cohort studies and its association with long-term mortality was investigated. An increased cardiovascular and all-cause mortality was observed in women who experienced a high arousal burden. The association was weaker in men.

Introduction

Insufficient sleep is associated with cardiovascular (CV) disease and fatal CV outcomes (Fan *et al.*, 2020). A meta-analysis revealed a U-shaped association between self-reported sleep duration and all-cause and CV mortality; both short (≤6 h/day) and long (>8 h/day) self-reported sleep duration are associated with mortality risk (Wang *et al.*, 2019). In heart failure patients, objectively assessed sleep duration using overnight polysomnography (PSG) shows an inverse linear association with mortality and does not follow the U-shaped association observed with self-reported sleep (Reinhard *et al.*, 2013). Indeed, the length of sleep alone does not reflect critical neurophysiological aspects such as sleep quality, sleep continuity, and sleep depth (Javaheri *et al.*, 2018). For example, sleep irregularity (Huang *et al.*, 2020), difficulties initiating sleep, and non-restorative sleep are associated with increased risk of mortality, irrespective of sleep duration (Li *et al.*, 2014).

Brief intrusions of unconscious wakefulness, so-called cortical arousals, are a normal feature of sleep. They occur spontaneously or are elicited by sleep-disordered breathing (SDB) and periodic limb movements during sleep, trauma, pain, temperature, light, and traffic noise (Linz *et al.*, 2018). Arousals, irrespective of the underlying mechanism, impact heart rate, blood pressure, and cardiac haemodynamics acutely (Nalivaiko *et al.*, 2007), but, when frequent, may also disrupt the circadian rhythm of the CV system, which is associated with unfavourable metabolic profiles, such as higher blood pressure, dysregulated blood lipids, and insulin resistance (Wang *et al.*, 2019). Additionally, traffic noise, particularly when occurring during night, has been shown to be a risk factor of CV disease through increased levels of stress hormones and vascular oxidative stress (Münzel *et al.*, 2020; Kröller-Schön *et al.*, 2018).

Clinically, the cause and rate of arousal occurrence are assessed using overnight PSG (Bennett *et al.*, 1998). The arousal index (AI), i.e. the number of arousals per hour of sleep, is often used to quantify the level of sleep fragmentation (Smurra *et al.*, 2001). High AI values are associated with daytime sleepiness (Bennett *et al.*, 1998), poor sleep quality (Smurra *et al.*, 2001), and increased emotional and physical fatigue in patients with SDB. In addition to the rate of arousals, the duration of individual arousal events may further contribute to the extent of sleep fragmentation (Boselli *et al.*, 1998; Trinder *et al.*, 2003; Nigro and Rhodius, 2005). The clinical significance of detailed characterization of the arousal burden (AB) on CV and all-cause mortality remains unknown.

The objective of this study was to determine the nocturnal AB and its association with long-term CV and all-cause mortality in men and women in the broader population. Using a simple index that combines arousal frequency, duration, and total sleep time (TST) measured on home PSG, we ascertained the prognostic value of AB for CV and overall mortality in 2782 male participants of the Osteoporotic Fractures in Men Study (MrOS) Sleep study, 424 female participants in the Study of Osteoporotic Fractures (SOF), and 4795 individuals in the Sleep Heart Health Study (SHHS). We hypothesized that a high AB is associated with long-term mortality.

Methods

Study populations

The MrOS Sleep study recruited 3135 participants (Baumert *et al.*, 2019), of which 2892 (92.2%) had PSG data available for analysis. The SOF sleep study recruited 461 participants (Baumert *et al.*, 2019), 453 (98.3%) with available PSG. The SHHS included 6841 participants (Redline *et al.*, 1998), with 5791 (89.9%) with available PSG. Follow-up in MrOS was 11.2 $\pm$ 2.1 years, 6.4 $\pm$ 1.6 years in SOF, and 10.7 $\pm$ 3.1 years in SHHS. Cardiovascular and all-cause mortality were assessed in all studies. For recruitment and follow-up details, see the Supplementary material online and Supplementary material online, Figure C.6.

In-home overnight polysomnography and sleep scoring

Overnight PSG and sleep scoring were performed using standard methodologies. For details, see the Supplementary material online.

Characterization of arousal burden

Total sleep time is defined as the duration of scored sleep epochs, expressed in minutes. We defined the AB as the cumulative duration of all arousal events relative to TST:

$$\mathrm{AB} = \frac{\sum_{i=1}^{N} d_i}{\mathrm{TST}} \times 100\ (\%) \tag{C.1}$$

where N is the number of arousals, and d is the duration of arousal i, expressed in minutes. If arousal terminated in a wake epoch, arousal duration was calculated as the interval between arousal onset and end of the related sleep epoch. The AI was calculated as:

$$\mathrm{AI} = \frac{\mathrm{N}}{\mathrm{TST}} \times 60\ (\mathrm{h}^{-1}) \tag{C.2}$$

Other measures

All participants were required to attend a clinical interview and complete an enrolment form that contained a questionnaire on medical history in advance of overnight PSG recordings (see Supplementary material online). Blood pressure was measured during the clinical visit.

From overnight PSG, we derived the mean respiratory rate (Baumert *et al.*, 2019), the time of sleep spent below 90% oxygen saturation (Baumert *et al.*, 2020), average ventricular rate, the apnoea-hypopnoea index (AHI) and the periodic limb movement index (PLMI).

Statistical analysis

Arousal burden data were divided into quartiles for Kaplan–Meier curve survival analysis and log-rank testing. Anthropometric data, lifestyle metrics, and medical history were compared using dichotomized AB (Q4 vs. Q1–Q3) and Student's t-test and χ^2 test, respectively. Cox proportional hazard models were constructed for continuous and categorical AB variables. The proportionality of hazard ratio (HR) was tested using cumulative sums of martingale residuals. Correlations were assessed using Spearman's rank coefficient. Exposure–response relationships were evaluated separately for men and women. We used restricted cubic splines with knots at the 5th, 35th, 65th, and 95th percentile to explore the potential non-linear association of the continuous variables with the outcome. Estimations for the exposure variables were made separately varying over their default range while adjusting the covariates to their median value. The Wald χ^2 test determined the effect of explanatory variables in a multivariable model. A *p*-value of 0.05 was considered statistically significant.

The associations between AB and mortality and AI with mortality were assessed with Kaplan–Meier curves and Cox proportional hazard models. Schenfeld's global test was applied to evaluate the proportionality of hazards. Cumulative incidence function and Fine–Gray subdistribution hazard model were applied to observe associations in the presence of competing risk.

MATLAB (R2019a, MathWorks, Natick, MA, USA) and R statistical software (R Foundation for Statistical Computing, Vienna, Austria) were used for statistical analysis and computing.

Results

Participant characteristics

At the baseline visit, MrOS cohort participants were 76.6 $\pm$ 5.5 years old. Almost half of them were overweight (Table C.1). Half of the men had a history of hypertension, and 13% had diabetes, while 17.3% reported histories of coronary artery disease/myocardial infarction (CAD/MI), 3.7% stroke, and 6.1% heart failure.

Women in the SOF cohort were 82.9 ± 3.2 years old at baseline (Table C.1). Nearly 40% were overweight, and about 60% had a history of hypertension, and 13.7% had diabetes. The histories of stroke, CAD/MI, and heart failure were 13.9%, 12.9%, and 8.5%, respectively.

The SHHS cohort included 2574 women and 2221 men (Table C.2). The average age was 64 years. The prevalence of CAD/MI in men was approximately twice that of women (11.1% vs. 8.4%). About 3.6% of men and 4.2% of women had a history of heart failure, 16.5% of men and 8.3% of women had a history of stroke.

Systolic and diastolic blood pressure recorded at the sleep visit was greater in women of SOF than other cohorts, while the mean heart rate of MrOS participants was higher than SOF and SHHS cohorts.

Polysomnographic assessment

Total sleep time recorded on PSG were 5.9 ± 1.2, 5.8 ± 1.3, 5.9 ± 1.0, and 6.2 ± 1.1 h in the MrOS, SOF, SHHS-men, and women cohorts; the AHI values were 20 ± 12.9, 27.6 ± 18.3, 12.3 ± 14.1, and 7.2 ± 10.5 h^{-1}, respectively. Periodic limb movement index in MrOS and SOF cohorts were 10.6 ± 9.9 and 13.3 ± 19 h^{-1}, respectively (not measured in SHHS).

The AB was significantly higher in the MrOS cohort than in the SOF cohort (6.60 ± 3.34% vs. 5.50 ± 3.05%, $p < 0.001$), and also higher in SHHS-men than in SHHS-women (7.14 ± 3.72% vs. 5.43 ± 2.62%, $p < 0.001$).

The AB correlated only weakly with TST (MrOS: ρ = -0.18, $p < 0.001$; SOF: ρ = -0.20, $p < 0.001$; SHHS: ρ = -0.20, $p < 0.001$; Supplementary material online, Figure C.7) and people with a low AB tended to sleep longer. The AB was also correlated with AHI (> 0.4 in all cohorts except SHHS-women; Supplementary material online, Figure C.8).

Cardiovascular and all-cause mortality

In the MrOS cohort, mortality data were available for 2782 participants (Supplementary material online, Figure C.6). During the follow-up period of 11.2 ± 2.1 years, 665 (23.4%) men died. Causes of deaths included CV disease (n = 236; 35.5%), cancer (n = 146; 22%), pulmonary disease (n = 55; 8.3%), and others (n = 228; 34.3%).

Table C.1. Cohort characteristics of the Osteoporotic Fractures in Men Study and the Study of Osteoporotic Fractures.

	MrOS (Men)				SOF (Women)			
	ABI	ABI $\leq$ 8.5%	ABI >8.5%	*p*-value	ABI	ABI $\leq$ 6.5%	ABI >6.5%	*p*-value
Subjects(n)	2782	2179	603		424	303	121	
Anthropometric and Ethnicity Data								
Age(years)	76.6 $\pm$ 5.5	76.1 $\pm$ 5.5	77.4 $\pm$ 5.4	<0.001	82.9 $\pm$ 3.2	82.8 $\pm$ 3.2	83.4 $\pm$ 3.3	0.082
White(n)	2526(90.8)	1961(90)	565(93.7)	0.002	397(93.6)	281(92.7)	116(95.9)	0.234
African American(n)	92(3.3)	77 (3.5)	15(2.5)	0.209	27(6.4)	22(7.3)	5(4.1)	0.234
Asian(n)	82(2.9)	68(3.1)	14(2.3)	0.312				
Other(n)	82(2.9)	70(3.2)	12(2.0)	0.120				
Body weight								
BMI(kg/m^2)	27.2 $\pm$ 3.8	26.9 $\pm$ 3.7	27.9 $\pm$ 4.1	<0.001	27.7 $\pm$ 4.6	27.7 $\pm$ 4.6	27.5 $\pm$ 4.5	0.596
Overweight(n)	1375(49.4)	1072(49.2)	304(50.4)	0.597	169(39.9)	121(39.9)	48(39.6)	0.951
Obese(n)	557(20.1)	401(18.4)	156(25.9)	<0.001	125(29.5)	92(30.4)	33(27.3)	0.529
Cardiac assessment								
Ventricular rate(bpm)	80 $\pm$ 14.1	79.9 $\pm$ 14.1	80.5 $\pm$ 14	0.437	65.3 $\pm$ 9.7	64.9 $\pm$ 9.1	66.1 $\pm$ 11	0.278
Afib(%)	295(10.6)	229(10.5)	66(11)	0.407				
SBP(mmHg)	126.5 $\pm$ 16.5	126.4 $\pm$ 16.6	127.2 $\pm$ 16.1	0.268	137.9 $\pm$ 17.6	137.7 $\pm$ 17.4	138.3 $\pm$ 18	0.731
DBP(mmHg)	67.5 $\pm$ 9.5	67.4 $\pm$ 9.5	67.6 $\pm$ 9.5	0.690	76.9 $\pm$ 8.6	76.9 $\pm$ 8.6	76.9 $\pm$ 8.6	0.982
Lifestyle								
Never-smokers(n)	1107(39.8)	891(40.9)	219(36.3)	0.043	271(63.9)	197(65.0)	74(61.2)	0.455
Ex-smokers(n)	1619(58.2)	1246(57.2)	373(61.9)	0.039	146(34.4)	99(32.7)	47(38.8)	0.227
Current Smokers(n)	56(2.0)	42(1.9)	14(2.3)	0.542	7(1.7)	7(2.31)	0	0.091
Current alcohol consumers(n)	1825(65.6)	1428(65.5)	398(66.0)	0.830	157(37.0)	108(35.6)	49(40.5)	0.350
Medical History								
Stroke(n)	102(3.67)	86(3.95)	16(2.7)	0.135	59(13.9)	43(14.2)	16(13.2)	0.794
CAD/MI(n)	481(17.3)	366(16.8)	115(19)	0.191	55(12.9)	33(10.9)	22(18.2)	0.044
CHF(n)	170(6.11)	120(5.55)	50(8.3)	0.013	36(8.5)	25(8.25)	11(9.1)	0.779
Asthma(n)	220(7.91)	178(8.17)	42(6.9)	0.332	53(12.5)	45(14.8)	8(6.6)	0.021
COPD(n)	145(5.21)	105(4.8)	40(6.6)	0.076				
HTN(n)	1389(50)	1078(49.5)	311(51.9)	0.361	252(59.4)	181(59.7)	71(58.7)	0.841
Depression(n)					48(11.3)	33(10.9)	15(12.4)	0.659
Parkinson(n)	31(1.11)	27(1.24)	4(0.66)	0.233				
Diabetes (n)	363 (13.0)	281 (12.9)	83 (13.8)	0.576	58 (13.7)	35 (11.5)	23 (19.0)	0.043
Overnight polysomnography								
WASO(min)	114.9 $\pm$ 67	110 $\pm$ 64.7	132.8 $\pm$ 70	<0.001	103.4 $\pm$ 72	91.4 $\pm$ 68	133.6 $\pm$ 76	<0.001
TST(min)	355 $\pm$ 69.4	362.4 $\pm$ 65	329.7 $\pm$ 78.2	<0.001	348.6 $\pm$ 77.5	357.5 $\pm$ 74.6	326.2 $\pm$ 80.3	<0.001
T90(min)	14.5 $\pm$ 33	14.6 $\pm$ 34	13.9 $\pm$ 32	0.680	12.9 $\pm$ 36	14.2 $\pm$ 40	9.7 $\pm$ 21.6	0.240
RR(min^{-1})	14.5 $\pm$ 1.9	14.6 $\pm$ 1.9	14.4 $\pm$ 1.9	0.230	15.3 $\pm$ 1.8	15.3 $\pm$ 1.8	15.4 $\pm$ 1.7	0.445
AHI(h^{-1})	20.1 $\pm$ 12.9	17.1 $\pm$ 10.30	30.8 $\pm$ 15.3	<0.001	27.6 $\pm$ 18.3	23.1 $\pm$ 14.3	38.7 $\pm$ 22.2	<0.001
PLMI(h^{-1})	10.6 $\pm$ 9.9	10.5 $\pm$ 9.8	10.9 $\pm$ 10.1	0.462	13.3 $\pm$ 19	12.7 $\pm$ 19.4	14.8 $\pm$ 18	0.298

Data are presented as mean $\pm$ standard deviation, or n (%).

AB, arousal burden; AF, atrial fibrillation; AHI, apnoea-hypopnoea index; BMI, body mass index; CAD, coronary artery index; CHF, congestive heart failure; COPD, chronic obstructive pulmonary disease; DBP, diastolic blood pressure; HTN, hypertension; MI, myocardial infarction; MrOS, Osteoporotic Fractures in Men Study; PLMI, periodic limb movement index; RR, respiratory rate; SBP, systolic blood pressure; SOF, Study of Osteoporotic Fractures; T90, time of sleep spent below 90% oxygen saturation; TST, total sleep time; WASO, wake after sleep onset.

Appendix C

Table C.2. Cohort characteristics of the Sleep Heart Health Study.

	Men				Women			
	ABI	ABI ≤ 8.5 %	ABI >8.5 %	*p*-Value	ABI	ABI ≤ 6.5 %	ABI >6.5 %	*p*-Value
Subjects (n)	2221	1643	578		2574	1918	656	
Anthropometric and Ethnicity Data								
Age(years)	64 ± 10.9	62.9 ± 11.1	66.9 ± 9.8	<0.001	63.7 ± 11.4	62.3 ± 11.3	67.8 ± 10.6	<0.001
White(n)	1956(88)	1444(87.9)	512(88.6)	0.658	2224(86.4)	586(85.4)	586(89.3)	0.011
African- American(n)	128(5.8)	83(5)	45(7.8)	0.015	180(7)	137(7.1)	43(6.6)	0.612
Other(n)	137(6.2)	116(7.1)	21(3.6)	0.003	170(6.6)	143(7.5)	27(4.1)	0.003
Body weight								
BMI(kg/m^2)	28.5 ± 4.3	28.2 ± 4.1	29.2 ± 4.9	<0.001	28.1 ± 5.6	27.9 ± 5.5	28.4 ± 5.9	0.06
Overweight(n)	1072(48.3)	822(50.0)	250(43.3)	0.005	955(37.1)	725(37.8)	230(35.1)	0.21
Obese(n)	687(30.9)	468(28.5)	219(37.9)	<0.001	792(30.8)	574(29.9)	218(33.2)	0.113
Cardiac assessment								
Ventricular rate(bpm)	65 ± 10	65 ± 10.1	64.4 ± 9.1	0.369	64.3 ± 9.9	64.2 ± 9.9	65.1 ± 10	0.213
Atrial fibrillation(%)	38(1.7)	25(1.5)	13(2.3)	0.246	24(0.9)	18(0.9)	8(1.2)	0.534
SBP(mmHg)	125.4 ± 18.5	125.1 ± 18.5	127.6 ± 19	0.054	125.9 ± 18.6	125.7 ± 18.4	128 ± 20.5	0.022
DBP(mmHg)	72.5 ± 10.7	72.4 ± 10.7	73.2 ± 10.7	0.276	72.3 ± 11.2	72.4 ± 11.2	71.5 ± 11.2	0.241
Lifestyle								
Never- Smokers(n)	768(34.6)	573(34.9)	195(33.7)	0.621	1458(56.6)	1081(56.4)	377(57.5)	0.621
Ex-smokers(n)	1222(55)	892(54.3)	330(57.1)	0.244	882(34.3)	655(34.2)	227(34.6)	0.833
Current Smokers(n)	231(10.4)	178(10.8)	53(9.2)	0.26	234(9.1)	182(9.5)	52(7.9)	0.229
Medical History								
Stroke(n)	367(16.5)	259(15.8)	108(18.7)	0.104	213(8.3)	147(7.7)	66(10.1)	0.05
CAD/MI(n)	246(11.1)	163(9.9)	83(14.4)	0.003	217(8.4)	152(7.9)	65(9.9)	0.114
CHF(n)	79(3.56)	54(3.29)	25(4.33)	0.245	109(4.2)	69(3.6)	40(6.1)	0.006
HTN(n)	924(35.9)	629(38.3)	295(51.0)	<0.001	1018(39.5)	715(37.3)	303(46.2)	<0.001
Diabetes(n)	189(7.3)	129(7.9)	60(10.4)	0.060	159(6.2)	104(5.4)	55(8.4)	0.007
Overnight Polysomnography								
WASO(min)	67.4 ± 46	59.7 ± 42	89.3 ± 52	<0.001	57.4 ± 41	50.6 ± 36.6	77.4 ± 45.5	<0.001
TST(min)	354.6 ± 61.1	361 ± 57.4	336 ± 67.4	<0.001	371.6 ± 64.6	379 ± 60.7	349.8 ± 71	<0.001
T90(min)	11.4 ± 35.2	11.5 ± 35.5	10.9 ± 31.9	0.809	13.4 ± 39	13.3 ± 39.4	14.1 ± 35.7	0.748
RR(min^{-1})	14.7 ± 1.7	14.7 ± 1.7	14.8 ± 1.7	0.924	14.7 ± 1.7	14.7 ± 1.7	14.8 ± 1.7	0.224
AHI(h^{-1})	12.3 ± 14.1	8.7 ± 9.1	22.9 ± 19.5	<0.001	7.2 ± 10.5	5.3 ± 7.1	12.8 ± 15.7	<0.001

Data are presented as mean ± standard deviation, or n (%).
AB, arousal burden; AF, atrial fibrillation; AHI, apnoea-hypopnoea index; BMI, body mass index; CAD, coronary artery disease; CHF, congestive heart failure; DBP, diastolic blood pressure; HTN, hypertension; MI, myocardial infarction; SBP, systolic blood pressure; T90, time of sleep spent below 90% oxygen saturation; TST, total sleep time; RR, respiratory rate; WASO, wake after sleep onset.

In the SOF cohort, mortality data were available for 424 women. During the 6.4 ± 1.6 years of follow-up, 105 (24.8%) women died, including 47 (44.8%) CV deaths, 17 (16.2%) cancer deaths, 27 (25.7%) pulmonary deaths, and 14 (13.3%) deaths due to other reasons.

In the SHHS cohort, mortality data were available for 4795 individuals. During the follow-up period of 10.7 ± 3.1 years, 987 (20.6%) individuals died (525 men, 462 women). Among those, 344 (34.9%) were CV deaths. Kaplan–Meier survival analysis of AB quartiles demonstrates the association between AB and all-cause mortality (Figure C.2). When comparing people in Q4 against all others (men: AB > 8.5%; women: AB > 6.5%), all-cause mortality was greater in all cohorts (MrOS: 3.2%, $p = 0.017$; SOF: 8.9%, $p = 0.039$; SHHS-men: 10.2%, $p < 0.001$; SHHS-women: 8.4%, $p < 0.001$; Figure C.3).

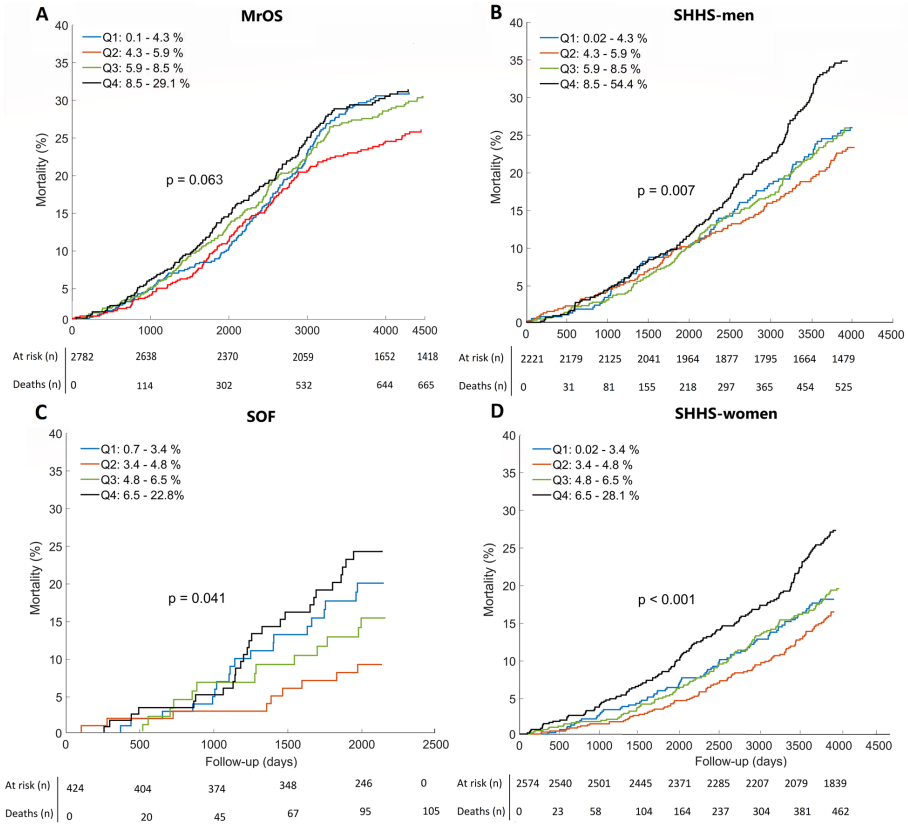


Figure C.2. Arousal burden and all-cause mortality for men and women from MrOS, SOF, and SHHS sleep studies. Arousal burden and all-cause mortality. Kaplan–Meier curves indicate arousal burden quartiles for (A) men from the Osteoporotic Fractures in Men Study (MrOS) Sleep cohort, (B) men from the Sleep Heart Health Study (SHHS), (C) women from the Study of Osteoporotic Fractures (SOF) cohort, and (D) women from the Sleep Heart Health Study. The *p*-values show log-rank test results.Q1–Q4, quartiles 1–4.

When assessing the competing risk of CV vs. non-CV deaths across all four cohorts, the highest AB quartile was associated with relatively higher CV mortality in the MrOS, SOF, and SHHS-women cohorts (4.8%, 8%, and 5.5%, respectively; Figure C.4). Q4 was also associated with a higher probability of non-CV mortality in the SOF and SHHS-men and SHHS-women cohorts (3.5%, 6.9%, and 7.4%).

Characteristics of participants in the highest AB quartiles are summarized in Tables C.1 and C.2. In the MrOS cohort, men in Q4 were more likely to be older, obese, ex-smokers, have a history of heart failure and chronic obstructive pulmonary disease, and have higher AHI values. In the SOF cohort, women in Q4 were more likely to be diabetic, asthmatic, have a history of CAD/MI and suffering from severe SDB. There was no significant association between high AB and history of stroke and prevalence of hypertension either in SOF or MrOS participants. Neither was there an association with chronic obstructive pulmonary

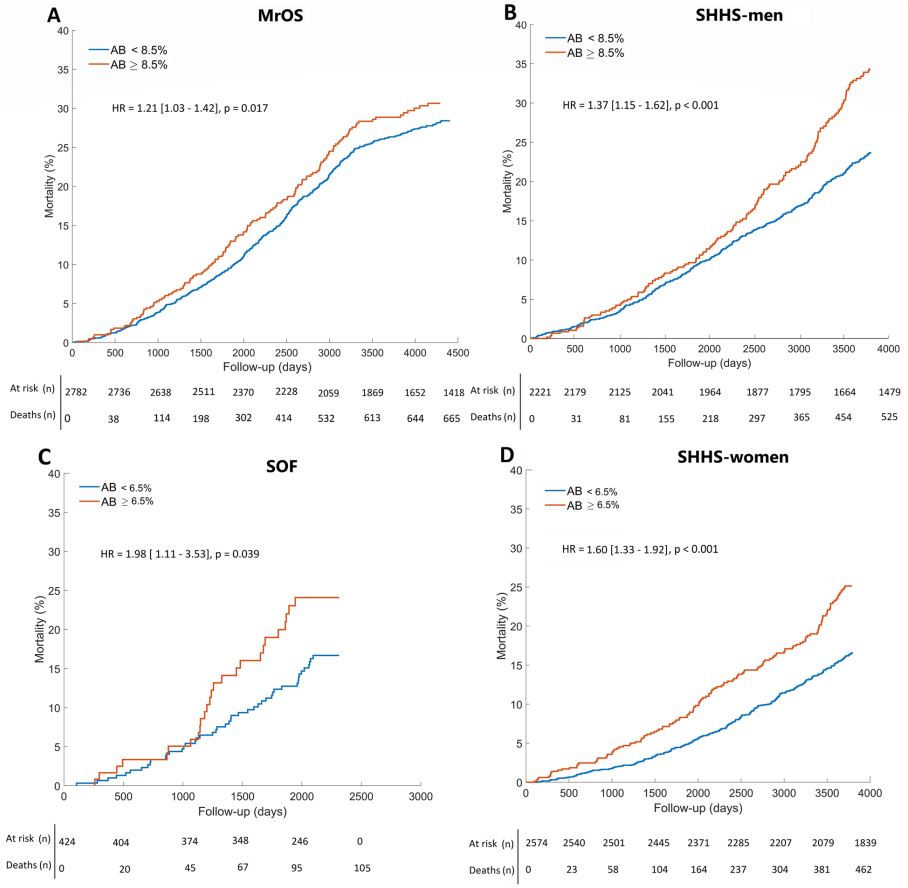


Figure C.3. Dichotomised arousal burden and all-cause mortality for men and women from MrOS, SOF, and SHHS sleep studies. Arousal burden and all-cause mortality. Kaplan–Meier curves indicate dichotomised arousal burden data for (A) men from the Osteoporotic Fractures in Men Study (MrOS) Sleep cohort, (B) men from the Sleep Heart Health Study (SHHS), (C) women from the Study of Osteoporotic Fractures (SOF) cohort, and (D) women from the Sleep Heart Health Study. The *p*-values show log-rank test results. HR, hazard ratio.

disease or depression. In SHHS-men in Q4 were more likely to be older, overweight, or obese, have a history of congestive heart failure, hypertension, and to have a moderate-to-severe degree of sleep apnoea. In SHHS-women in Q4 were likely to be older, white, have a history of CAD/MI, hypertension, diabetes, stroke, and mild-to-moderate sleep apnoea and higher systolic blood pressure.

In univariate Cox proportional hazard analysis, AB > 8.5% and AB > 6.5% were significantly associated with all-cause mortality in men [MrOS: HR 1.21, 95% confidence interval (CI) 1.03–1.42, $p = 0.02$; SHHS-men: HR 1.37 (1.15–1.62), $p < 0.001$] and women [SOF: HR 1.98 (1.11–3.53), $p = 0.02$; SHHS-women: HR 1.60 (1.33–1.92), $p < 0.001$] (Table C.3). Arousal burden > 6.5% was also associated with CV mortality in women [SOF: HR 2.15 (1.08–4.28), $p = 0.031$; SHHS-women: HR 1.95 (1.38–2.75), $p < 0.001$]. In male cohorts,

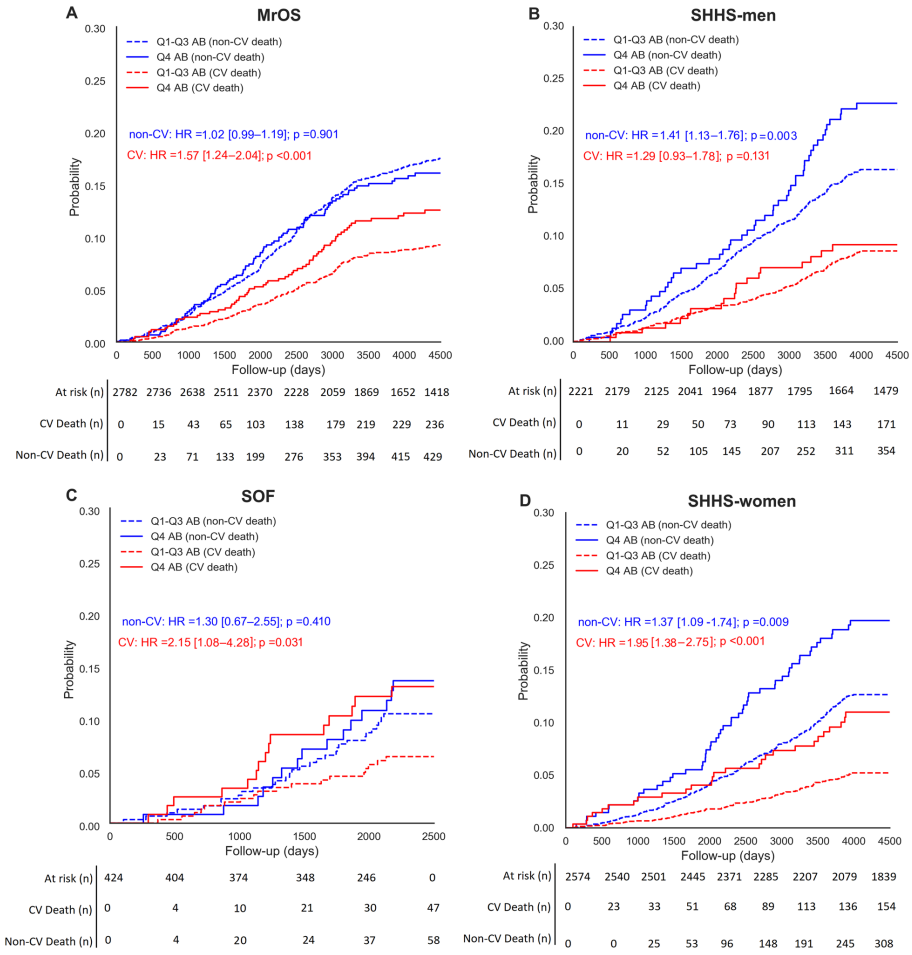


Figure C.4. Comparison of competing risk of arousal burden of cardiovascular, non-cardiovascular and all-cause mortality in older men and women. Commutative incident function curves compare the competing risk of arousal burden of cardiovascular, non-cardiovascular and all-cause mortality in (A) men from the Osteoporotic Fractures in Men Study (MrOS) Sleep cohort, (B) men from the Sleep Heart Health Study (SHHS), (C) women from the Study of Osteoporotic Fractures (SOF) cohort, and (D) women from the Sleep Heart Health Study. Hazard ratios (HR) and *p*-values were estimated through a sub-distributional Fine–Gray hazard model.

AB > 8.5% was associated with CV mortality in MrOS [HR= 1.57 (1.24–2.04), $p < 0.001$] but not in SHHS-men ($p = 0.131$).

After adjusting the regression models for TST, age, average heart rate, respiratory rate, systolic and diastolic blood pressure, time spent with oxygen desaturation below 90%, history of hypertension, stroke, MI, congestive heart failure, diabetes and hypertension, body mass index category, AHI, and smoking habits, AB >8.5% remained associated with CV mortality in the MrOS cohort [HR 1.35 (1.02–1.79), $p = 0.034$] but not for all-cause mortality (Table C.3). After adjusting the SOF cohort model for the same variables, AB > 6.5% remained associated with all-cause [HR 1.58 (1.01–2.42), $p = 0.038$] and CV mortality

Table C.3. Association of arousal burden with cardiovascular and all-cause mortality.

	All-cause mortality				Cardiovascular mortality				Non-cardiovascular Mortality			
	Univariate analysis		Multivariable analysis		Univariate analysis		Multivariable analysis		Univariate analysis		Multivariable analysis	
	HR(95% CI)	*p*	HR(95% CI)	*p*	HR(95% CI)	*p*	HR(95% CI)	*p*	HR(95% CI)	*p*	HR(95% CI)	*p*
MrOS Sleep												
AB(%)	1.03(1.00–1.05)	0.011	1.02(1.00–1.04)	0.049	1.05(1.02–1.09)	0.003	1.04(1.00–1.08)	0.028	1.01(0.99–1.04)	0.301	0.99(0.97–1.02)	0.955
AB>8.5%	1.21(1.03–1.42)	0.020	1.11(0.94–1.32)	0.261	1.57(1.24–2.04)	<0.001	1.35(1.02–1.79)	0.034	1.02(0.99–1.25)	0.897	0.95(0.76–1.19)	0.651
SOF												
AB(%)	1.11(1.03–1.19)	0.002	1.06(1.00–1.13)	0.050	1.14(1.05–1.23)	0.004	1.14(1.05–1.25)	0.002	1.06(0.97–1.15)	0.201	1.05(0.96–1.15)	0.255
AB>6.5%	1.98(1.11–3.53)	0.020	1.58(1.01–2.42)	0.038	2.15(1.08–4.28)	0.031	2.17(1.04–4.50)	0.037	1.30(0.70–2.41)	0.410	1.31(0.67–2.55)	0.431
SHHS-men												
AB(%)	1.04(1.02–1.06)	<0.001	1.02(0.98–1.06)	0.199	1.01(0.96–1.06)	0.661	0.96(0.90–1.03)	0.251	1.05(1.01–1.08)	0.004	1.03(0.98–1.07)	0.194
AB>8.5%	1.37(1.15–1.62)	<0.001	1.31(1.06–1.62)	0.011	1.29(0.93–1.78)	0.131	1.24(0.86–1.79)	0.271	1.41(1.13–1.76)	0.003	1.34(1.04–1.74)	0.005
SHHS-women												
AB(%)	1.10(1.07–1.13)	<0.001	1.07(1.03–1.12)	0.003	1.15(1.08–1.22)	<0.001	1.12(1.06–1.20)	<0.001	1.10(1.06–1.16)	<0.001	1.09(1.04–1.15)	0.001
AB>6.5%	1.60(1.33–1.92)	<0.001	1.21(1.06–1.42)	0.012	1.95(1.38–2.75)	<0.001	1.60(1.12–2.28)	0.009	1.37(1.09–1.74)	0.009	1.28(1.01–1.63)	0.038

HR for AB (%) indicate the risk increment per 1% increase in AB. For categorical risk analysis, AB was dichotomized on the 4th quartile (men: AB > 8.5; women: AB > 6.5). Multivariable analysis was adjusted for total sleep duration, age, history of stroke, coronary artery disease/myocardial infarction, congestive heart failure, diabetes, hypertension, mean heart rate, mean respiratory rate, systolic and diastolic blood pressure, time of sleep spent below 90% oxygen saturation, total wake after sleep onset, categorized body mass index, apnoea-hypopnoea index, and smoking habit. AB, arousal burden; CI, confidence interval; HR, hazard ratio; MrOS, Osteoporotic Fractures in Men Study; SHHS, Sleep Heart Health Study; SOF, Study of Osteoporotic Fractures.

[HR 2.17 (1.04–4.50), p = 0.037]. In the adjusted model, AB was also associated with all-cause mortality in SHHS-men [HR 1.31 (1.06–1.62), p = 0.011] and SHHS-women [HR 1.21 (1.06–1.42), p = 0.012] and CV mortality in SHHS-women [HR 1.60 (1.12–2.28), p = 0.009] (Table C.3). Concerning non-CV mortality, the highest AB quartile was associated with increased risk in both SHHS-men [HR 1.34 (1.04–1.74), p = 0.005] and SHHS-women [HR 1.28 (1.01–1.63), p < 0.038].

For a detailed mortality analysis of AI, see Supplementary material online. Kaplan–Meier analysis of AI (Supplementary material online, Figure C.9) as well as competing risk analysis for CV and non-CV mortality (Supplementary material online, Figure C.10) showed a somewhat weaker association of AI with mortality, although AI was strongly correlated with AB (Supplementary material online, Figure C.11). Kaplan–Meier curves of combined AB and AI did not demonstrate any significant difference between subgroups for all-cause (Supplementary material online, Figure C.12) or CV mortality (Supplementary material online, Figure C.13).

To further explore the non-linear association between AB and mortality in men and women, we collated datasets and assessed exposure–response relationships (Figure C.5), adjusting for the same covariates. We observed a significant gender effect on the prognostic value of AB for mortality (all-cause: p = 0.034; CV: p = 0.030). Overall, the association between AB and mortality appears to be stronger in women than in men (Graphical abstract). Men demonstrated an initial decrease in all-cause mortality up to almost AB ≈6% [logHR 0.08

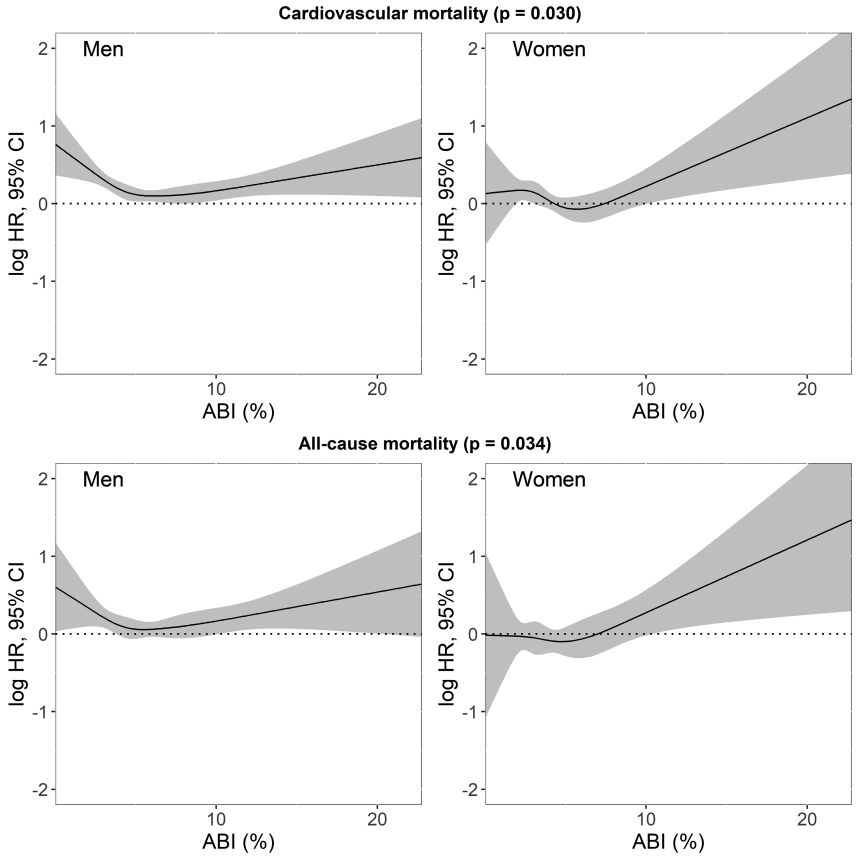


Figure C.5. Exposure–response relationships of arousal burden and all-cause and cardiovascular mortality for older men and women. The exposure–response relationships of arousal burden and all-cause and cardiovascular mortality for men and women regardless of their cohort, adjusted for total sleep duration, age, systolic and diastolic blood pressure, average heart rate, mean respiratory rate, time spent below 90% oxygen desaturation, history of stroke, myocardial infarction/coronary artery disease, congestive heart failure, categorized body mass and apnoea/hypopnoea indices, the total duration of wake after sleep onset, smoking habit, and history of hypertension and diabetes. CI, confidence interval; HR, hazard ratio. The *p*-values show Wald χ^2 test results.

(0.02–0.14)], the cut-off of Q2 in Kaplan–Meier plots for men (Figure C.2), followed by a gradual increase [AB 10%: logHR 0.13 (0.04–0.21), AB 20%: logHR 0.41 (0.11–0.72)]. In women, the all-cause mortality risk similarly reached the minimum value at the Q2 cut-off [AB ≈5%: logHR -0.03 (-0.12–0.07)], while an almost linear increase in all-cause mortality was observed with increasing AB [AB 10%: logHR 0.26 (0.09–0.43), AB 20%: logHR 1.07 (1.45–1.67)]. The exposure–response relationships for CV mortality were similar.

Discussion

This study is the first to evaluate the sex-specific association of sleep AB and long-term CV and all-cause mortality. Data from three large cohort studies, totalling 8001 participants demonstrate that a high AB is associated with significantly increased CV and all-cause mortality. While the frequency of arousals was lower in women than in men, the association with mortality was stronger in women.

We used two metrics to quantify nocturnal AB, AB and AI, quantifying the percentage of sleep time affected by arousal and the frequency of arousal per hour of sleep, respectively. Arousal burden correlated only weakly with total sleep duration. People with low AB tend to sleep longer, presumably due to lower levels of sleep fragmentation. The wake time after sleep onset was associated with AB in our study, but AB in itself appears to carry important prognostic information. While AB was more strongly associated with mortality than AI, both metrics confirm the relationship between AB and longterm mortality as well as the difference between genders. While the severity of SDB, periodic limb movement disorder and other sleep pathologies increase the rate of arousals, mechanisms underlying the arousal duration are less clear. Boselli *et al.* (1998) showed that the mean arousal duration is independent of age or gender. The extent and duration of arousals elicited by apnoeas are typically greater than that of hypopnoeas. Apnoeas exceeding 20 s accompanied by a minimum oxygen desaturation of 86% are likelier to be terminated by arousals longer than 11 s (Nigro and Rhodius, 2005). Generally, arousal duration reflects arousal intensity and the level of concurrent autonomic activation such as post-arousal heart rate (Javaheri *et al.*, 2018). Particularly in SDB, reflex control and chemoreceptor sensitivity, which determine arousal thresholds and contribute to sympathovagal disbalance and haemodynamic responses, may affect arousal duration (Eckert and Younes, 2013).

Comparing AB between the MrOS and SOF cohorts suggests lower AB in women than in men, despite their more advanced age and somewhat higher AHI and PLMI. In the SHHS cohort, AB was greater in men as was AHI, but was not associated with CV mortality, unlike AB in women. This suggests that women are more prone to arousal-related consequences than men.

Interestingly, the lowest mortality was observed among people in the 2nd AB quartile, demonstrating the non-linear relationship between AB and mortality. People experiencing infrequent arousals may suffer from elevated arousal thresholds, which could result in worse outcomes (Kaur and Saper, 2019).

Cardiovascular risk factors are more prevalent in participants with high AB; CV disease and arrhythmias are known to contribute to increased mortality (Cappuccio *et al.*, 2010). Importantly, AB was associated with mortality when adjusted for concomitant risk factors. Arousal-related pathological conditions, involving autonomic nervous system activation, circadian rhythm impairment due to sleep fragmentation (Bonnet, 1989), nocturnal blood pressure and heart rate rises (Bennett *et al.*, 1998; O'Driscoll *et al.*, 2004; Bonnet and Arand, 1997) or concomitant conditions such as SDB have all shown to increase all-cause and CV mortality (Kaur and Saper, 2019; Linz *et al.*, 2019; Azarbarzin *et al.*, 2019). Moreover, irregular sleep duration and timing have been recently shown as risk factors of CV disease, independent of traditional CV disease risk factors and sleep quality and quantity.5 Here, we observed a significant association between history of heart failure and MI and AB in some cohorts, linking AB to CV disease. We also report a higher prevalence of diabetes in women with AB >6.5%, corroborating on the link between short sleep, type 2 diabetes, and glucose metabolism (Gangwisch *et al.*, 2005; Yaggi *et al.*, 2006). However, we did not observe this relationship in men.

Strengths

This study had multiple strengths, including the large sample size and hypothesis testing across three independent cohorts; prospective evaluation of outcomes over significant time periods; the focus on hard endpoints (which are less sensitive to misclassification and are most clinically relevant); rigorously collected PSG data scored blinded to other data; and the ability to adjust for multiple potential confounders.

Limitations

All cohorts comprise predominately white men and women of predominantly middle to older age. Hence, our findings cannot be extrapolated to other races or younger individuals. Baseline exposure to various conditions was self-reported rather than systematically ascertained through medical records or direct measurement. We did not consider the possible confounding effects of medications. We did not distinguish arousal types, assuming that regardless of the cause, arousals disrupt sleep architecture the same way. In line with the American Academy of Sleep Medicine (AASM) scoring rules, subcortical arousals were not considered (Iber *et al.*, 2007). Standard AASM 30 s sleep staging may have resulted in underestimating AB by cutting short arousals that terminated in 'wake' stages. Our findings

were obtained on single-night in-home overnight PSG; night-to-night variability in AB may exist and affect the estimated strength of the observed associations (Linz *et al.*, 2018).

Future directions

We demonstrated a clear association between nocturnal AB and long-term CV and all-cause mortality. Arousal burden may represent a promising marker to identify patients at risk. We quantified AB using EEG recordings from overnight PSG. The clinical implementation of AB assessment in routine risk stratification strategies will require easily scalable, widely accessible, and affordable techniques to estimate the duration and fragmentation of sleep and to detect arousals (e.g. wrist actigraphy or peripheral arterial tonometry) (Yalamanchali *et al.*, 2013). To determine whether a more detailed description of the AB, incorporating alternative arousal characteristics such as the frequency of wake periods or sleep stage transitions from deep to light sleep results in a better risk prediction requires further study. Arousal burden may also represent a modifiable risk factor for CV and all-cause mortality, which warrants future prospective intervention studies. Possible interventions to reduce AB require a multi-modal assessment and may involve pharmacological elevation of the arousal threshold (Eckert and Younes, 2013).

Conclusion

Nocturnal AB is associated with long-term CV and all-cause mortality in women and to a lesser extent in men.

Funding

This study was supported through a grant from the Australian Research Council (DP066334 5). The MrOS Study and the SOF Study are supported by NIH funding. The following institutes provided support: The National Institute on Aging, the National Institute of Arthritis and Musculoskeletal and Skin Diseases, the National Center for Advancing Translational Sciences, and NIH Roadmap for Medical Research (U01 AG027810, U01 AG042124, U01 AG042139, U01 AG042140, U01 AG042143, U01 AG042145, U01 AG042168, U01 AR066160, and UL1 TR000128). The National Heart, Lung, and Blood Institute provided funding for the MrOS Sleep ancillary study (R01 HL071194, R01 HL070848, R01

HL070847, R01 HL070842, R01 HL070841, R01 HL070837, R01 HL070838, and R01 HL070839), and the National Sleep Research Resource (R24-HL-114473). The SOF sleep study was supported by grant AG021918, AG026720, AG05394, AG05407, AG08415, AR35582, AR35583, AR35584, R01 AG005407, R01 AG027576-22, 2 R01 AG005394-22A1, 2 R01 AG027574-22A1, HL40489, and T32 AG000212-14. Dr S.R. was supported in part by NIH R35HL135818.

Role of funding sources: The study sponsors had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Conflict of interest: Dr D.L. reports having served on the advisory board of LivaNova. Dr D.L. reports having received lecture and/or consulting fees from LivaNova, and ResMed.

Data availability

All of the individual participant data generated during this study will be made available at the MrOS Online (https://mrosonline.ucsf.edu/) and SOF Online (https://sof.ucsf.edu/) and sleepdata.org websites.

Supplementary material

Study Populations

MrOS sleep study

The MrOS cohort observational study enrolled 5995 community-dwelling men older than 65 years between March 2000 and April 2002 at six clinical centres in the United States to investigate the epidemiology of osteoporosis in older men and identify the risk factors for fracture and bone loss (Orwoll *et al.*, 2005). Enrolled participants had to be able to walk without any assistance from another person and not have a bilateral hip replacement (Blank *et al.*, 2005). The Outcomes of Sleep Disorders in Older Men (MrOS Sleep) sub-study recruited 3135 participants from MrOS. All men provided written informed consent, and the Institutional Review Board approved the study at each site. All men completed the clinical visit and in-home overnight PSG between December 2003, and March 2005 (Baumert *et al.*, 2019). Of these participants, 2892 (92.2%) had adequate PSG datasets.

SOF

The SOF observational cohort study enrolled 9704 community-dwelling Caucasian female participants aged ≤ 65 years who lived in the US between September 1986 and October 1988 (Cummings *et al.*, 1990). Later, 662 African-American women recruited between February 1997 and 1998 were added to the study. Participants were reassessed biannual follow-up visits. Four hundred sixty-one participants completed overnight in-home PSG between January 2002 and February 2004 (Baumert *et al.*, 2019; Spira *et al.*, 2008). Of these women, 453 had adequate PSG.

SHHS

The SHHS is a prospective multi-centre cohort study implemented by the National Heart Lung & Blood Institute to investigate OSA and other SDB as risk factors for the development of CV disease (Haas *et al.*, 2005). Thus, SHHS participants were recruited from ongoing cohort studies of CV or respiratory disease with no treatment of SDB with continuous positive airway pressure (CPAP), no tracheotomy and no current home oxygen therapy (Haas *et al.*, 2005). Among 11503 eligible individuals in parent cohort studies, 6841 (62%) participants completed the home overnight PSG sleep study between November 1995 and January 1998 (Haas *et al.*, 2005; Redline *et al.*, 1998). In this study, the PSG of 5791 participants were available for analysis (89.9%).

Follow-up

MrOS sleep participants were followed up every four months to survey for new symptoms of CV or clinically relevant arrhythmia by postcards and/or phone with > 99% response rate. A board-certified cardiologist then verified all relevant medical records and supporting documents for centralised adjudication using a pre-specified protocol (Koo *et al.*, 2011). The death certificate and hospital records from the time of death were collected for fatal events. If a fatal event did not occur at the hospital, a proxy interview with next of kin and the participant's most recent hospitalisation documents in the prior 12 months were collected. Only events confirmed by the adjudicator are included for analysis (Baumert *et al.*, 2019).

Deaths of SOF participants were centrally adjudicated using a state-registered certificate of death which was submitted to the coordinating centre. The principal investigator at each of four clinical sites indicated the initial diagnosis for the cause of death. The final classification of cause-specific mortality was centrally adjudicated at the coordinating centre by a

trained physician adjudicator, using the International Classification of Diseases 9th Revision Clinical Modification (ICD-9-CM) (Baumert *et al.*, 2019).

Deaths in the SHHS cohort from any cause were identified and confirmed using multiple concurrent approaches including follow-up interviews, written annual questionnaires or telephone contacts with study participants or next-of-kin, surveillance of local hospital records and community obituaries, and linkage with the Social Security Administration Death Master File (Punjabi *et al.*, 2009).

In-home overnight polysomnography

Sleep recordings were performed using an unattended, portable in-home PSG over one night at the participant's residence using the Compumedics (Abbotsford, Australia) Safiro sleep monitoring system for MrOS Sleep, the Compumedics Siesta system for SOF3 and Compumedics *p*-series for SHHS (Punjabi *et al.*, 2009). Trained staff members visited the participants to attach the sensors and electrodes and conduct overnight PSG. The setup included two central electroencephalograms (EEG), bilateral electrooculograms, bilateral chin electromyogram, a bipolar electrocardiogram, nasal-oral thermistor, nasal flow via pressure transducer and nasal cannula, abdominal and respiratory inductance plethysmography, finger pulse oximetry, bilateral leg movements by piezoelectric sensors and body position (Baumert *et al.*, 2019; Punjabi *et al.*, 2009; Song *et al.*, 2015).

Sleep scoring

Certified sleep technicians scored sleep events according to the standard criteria (Bonnet *et al.*, 1992; Rechtschaffen and Kales, 1968). The apnoea-hypopnoea index (AHI) was calculated as the number of apnoea and hypopnoea episodes per hour of sleep. Apnoea was defined as the complete or near-complete cessation of airflow for more than ten seconds, and hypopnoeas were scored if clear reductions in breathing amplitude (at least 30% below baseline breathing) occurred and lasted more than 10 seconds. Only apnoea and hypopnoea events that were associated with a 3% or greater desaturation were included in the AHI (Song *et al.*, 2015). The severity of sleep apnoea/hypopnoea syndrome can be determined through the AHI as mild those with $5 \leq \text{AHI} < 15$ events per hour, moderate ($15 \leq \text{AHI} < 30$ h^{-1}) and severe ($\text{AHI} \geq 30$ h^{-1}). Sleep duration or total sleep time (TST) was obtained from REM and NREM sleep stages after PSG analysis. Sleep arousals were also

scored according to the American sleep disorders association criteria (Bonnet *et al.*, 1992) and verified with AASM requirements for arousal scoring (Iber *et al.*, 2007).

Periodic limb movement (PLM) events were scored according to AASM criteria which includes individual movements with clear amplitude increase from baseline in legs movement and the duration between 0.5 and 5 seconds. To be considered periodic, at least 4 movements required to occur in succession no less than 5 seconds and no more than 90 seconds apart excluding PLM after respiratory events. The periodic limb movement index (PLMI) was the total number of PLM events per hour of sleep (Koo *et al.*, 2011). PLM events were only scored in MrOS and SOF datasets.

Respiratory rate (RR) during sleep extracted from the thoracic respiratory inductance plethysmography belt signal of the PSG data. During pre-processing, signal offsets were removed and a low-pass forward and reverse Butterworth filter (1 Hz) was applied. Expiratory and inspiratory onsets were determined from the respiratory signal by identifying the peaks and valleys using the first-order derivative. The inspiratory onset of artefact-free breaths was used to compute a breath-by-breath measure of the respiratory interval which were then averaged within each subject (Baumert *et al.*, 2019).

Additional measures

In the MrOS and SOF cohorts, the participants' history of physician diagnosis of diabetes, hypertension (HT), coronary artery disease (CAD), myocardial infarction (MI), congestive heart failure (CHF), asthma, stroke and Parkinson were surveyed. The SOF questionnaires contained additional items on the history of depression. In contrast, MrOS questionnaires contained additional questions on the history of chronic obstructive pulmonary disease (COPD), atrial fibrillation or flutter (Afib), smoking habits and alcohol consumption. The SHHS questionnaire included items on the history of stroke, CHF, MI, diabetes, HT, Afib and smoking habits. The ethnicity of all participants in the three cohorts (MrOS, SOF and SHHS) was also included in the analysis.

Survival analysis using the arousal index

Arousal index data of all cohorts were divided into AI quartiles. To compare participants in the fourth AI quartile against the participants in the lower three quartiles, we used cut-off values of 25 h^{-1} and 30 h^{-1} for women and men, respectively.

The AI was significantly higher in the MrOS cohort than in the SOF cohort (25.2 $\pm$ 12.6 h^{-1} vs. 22.1 $\pm$ 12.6 h^{-1}, $p < 0.001$). Within the SHHS cohort, the AI was also significantly higher in men than in women (25.2 $\pm$ 13 h^{-1} vs. 19.9 $\pm$ 10 h^{-1}, $p < 0.001$).

Kaplan-Meier curves of dichotomised AI data for all-cause mortality is shown in Figure C.9. The competing risk of CV and non-CV mortality in four cohorts is depicted in Figure C.10. In SOF cohort, the probability of CV mortality was significantly increased in women with AI>25 h^{-1} compared with women with an AI $\leq$ 25 h^{-1} (about 10%). Similarly, CV mortality in SHHS-women with AI > 25 h^{-1} was nearly 4% greater. In the MrOS cohort, the CV mortality increased by about 5% in participants with AI > 30 h^{-1} compared with men with AI < 30 h^{-1}. But no significant association between CV mortality and AI > 30 h^{-1} was observed in SHHS-men.

All-cause mortality in SHHS-men with AI > 30 h^{-1} was nearly 7% higher than men with AI < 30. In SHHS-women, all-cause mortality for AI > 25 h^{-1} was 11.5% greater. Log-rank tests showed no significant association between AI distribution and all-cause mortality in MrOS and SOF cohorts (MrOS: $p = 0.131$; SOF: $p = 0.127$).

In Cox proportional hazard analysis of the MrOS cohort, AI > 30 h^{-1} was associated with CV mortality (univariate: HR = 1.38 [1.09-1.76], $p = 0.009$; multivariable: HR = 1.29 [1.01-1.63], $p = 0.048$), but not with all-cause mortality (Table C.4).

Similarly, AI > 25 h^{-1} was significantly associated with CV mortality in SOF (univariate: HR = 2.57 [1.30-5.11], $p = 0.001$; multivariable: HR = 2.68 [1.22-5.82], $p = 0.013$), but not with all-cause mortality.

In women of the SHHS cohort, AI > 25 h^{-1} was significantly associated with CV (univariate: HR = 1.81 [1.27-2.58], $p < 0.001$; multivariable: HR = 1.53 [1.07-2.22], $p = 0.022$) and all-cause mortality (univariate: HR = 1.69 [1.39-2.05], $p < 0.001$; multivariable: HR = 1.47 [1.21-1.80], $p < 0.001$). In men, AI > 30 h^{-1} was only significantly associated with all-cause mortality in univiariate analysis (HR = 1.24 [1.03-1.49], $p = 0.02$).

In all datasets, ABI and AI values are highly correlated (Figure C.11: $\rho > 0.9$, $p < 0.001$). The highest quartile of ABI (ABI > 8.5% in men and ABI > 6.5% in women) may reflect the highest quartile of AI (AI > 30 h^{-1} in men and AI > 25 h^{-1} in women).

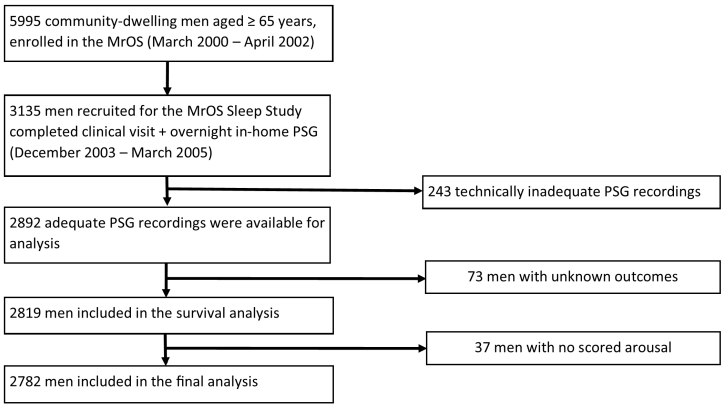
To test whether combining AB and AI yields stronger associations with mortality, we dichotomised AB and AI values using the cohort medians to low/high AB and low/high AI, respectively, resulting in subgroups of participants with either low AB & low AI, high AB & high AI, low AB & high AI or high AB & low AI. We created KM plots for the resultant

Table C.4. Association of arousal index (AI) with cardiovascular and all-cause mortality.

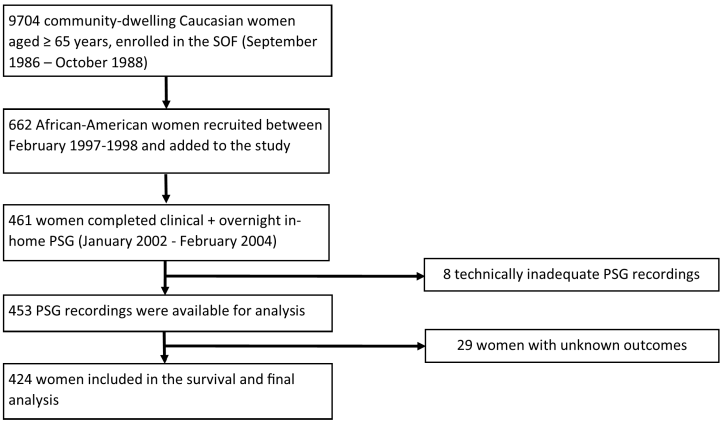
	All-cause mortality				Cardiovascular mortality				Non-cardiovascular Mortality			
	Univariate analysis		Multivariable analysis		Univariate analysis		Multivariable analysis		Univariate analysis		Multivariable analysis	
	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*	HR (95% CI)	*p*
MrOS Sleep												
AI (h^{-1})	1.01 (1.00–1.01)	0.05	1.00 (0.99–1.01)	0.312	1.01 (1.00–1.02)	0.013	1.00 (0.99–1.01)	0.092	1.00 (0.99–1.01)	0.5	1.00 (0.99–1.01)	0.944
AI >30	1.12 (0.96–1.31)	0.126	1.07 (0.91–1.26)	0.384	1.38 (1.09–1.76)	0.009	1.29 (1.01–1.63)	0.048	1.00 (0.82–1.20)	0.9	0.97 (0.80–1.19)	0.841
SOF												
AI (h^{-1})	1.03 (1.01 – 1.04)	0.002	1.03 (1.00–1.05)	0.002	1.04 (1.02 – 1.06)	0.001	1.03 (1.01 – 1.06)	0.001	1.01 (0.99–1.03)	0.352	1.01 (0.99–1.04)	0.287
AI >25	1.42 (0.90 – 2.24)	0.1	1.45 (0.87–2.42)	0.149	2.57 (1.30 – 5.11)	0.001	2.68 (1.22 – 5.82)	0.013	0.88 (0.46–1.67)	0.688	0.89 (0.42–1.85)	0.758
SHHS-men												
AB (h^{-1})	1.01 (1.00 – 1.02)	0.01	1.00(0.99–1.02)	0.192	1.00 (0.99 – 1.01)	0.8	1.00 (0.98 – 1.01)	0.247	1.01 (1.00–1.02)	0.003	1.01(1.00–1.02)	0.015
AB >30	1.24 (1.03 – 1.49)	0.02	1.20 (0.97–1.48)	0.094	1.01 (0.99 – 1.42)	0.9	0.94 (0.63 – 1.41)	0.764	1.36 (1.09–1.71)	0.007	1.34 (1.04–1.72)	0.024
SHHS-women												
AB (h^{-1})	1.02 (1.01 – 1.04)	<0.001	1.02 (1.01–1.03)	<0.001	1.03 (1.02 – 1.04)	<0.001	1.03(1.01–1.04)	<0.001	1.02 (1.01–1.03)	<0.001	1.02 (1.01–1.03)	0.001
AB >25	1.69 (1.39 – 2.05)	<0.001	1.47 (1.21–1.80)	<0.001	1.81 (1.27 – 2.58)	<0.001	1.53 (1.07–2.22)	0.022	1.64 (1.30–2.07)	<0.001	1.44 (1.13–1.83)	0.003

Hazard ratios (HR) for ABI (%) indicate the risk increment per 1% increase in ABI. For categorical risk analysis, ABI was dichotomised on the fourth quartile (men: ABI > 8.5; women: ABI > 6.5). Multivariable analysis was adjusted for total sleep duration, age, history of stroke, myocardial infarction/coronary artery disease, congestive heart failure, diabetes, hypertension, mean heart rate, mean respiratory rate, systolic and diastolic blood pressure, time of sleep spent below 90% oxygen saturation, categorised body mass index, apnoea-hypopnea index and smoking habit, CI: confidence interval.

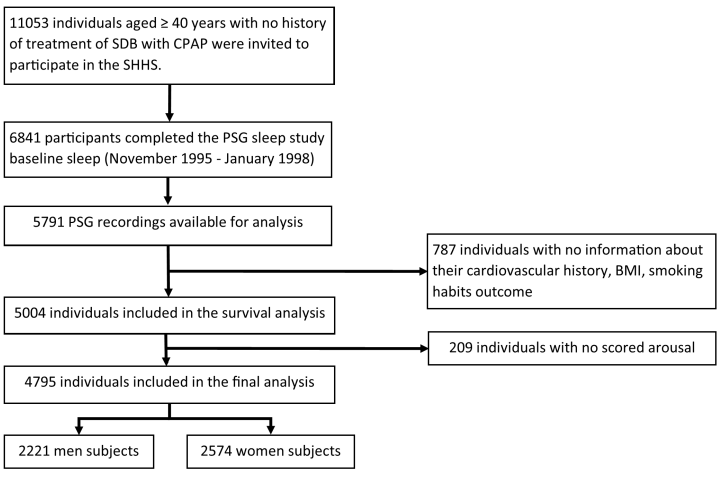
composite index. Neither all-cause mortality (Figure C.12) nor CV mortality (Figure C.13) was significantly different between the composite index subgroups. Owing to the strong the between AB and AI, the vast majority of participants with a high AB also have a high AI, and vice versa.



(a)



(b)



(c)

Figure C.6. Flow charts of recruitment process for MrOS, SOF, and SHHS. Flow charts of participants included in the analysis of arousal burden index for a) the Osteoporotic Fractures in Men Study (MrOS), b) the Study of Osteoporotic Fractures (SOF) and c) the Sleep Heart Health Study (SHHS). PSG: polysomnography, BMI: body mass index, SDB: sleep-disordered breathing. CPAP: continuous positive airway pressure.

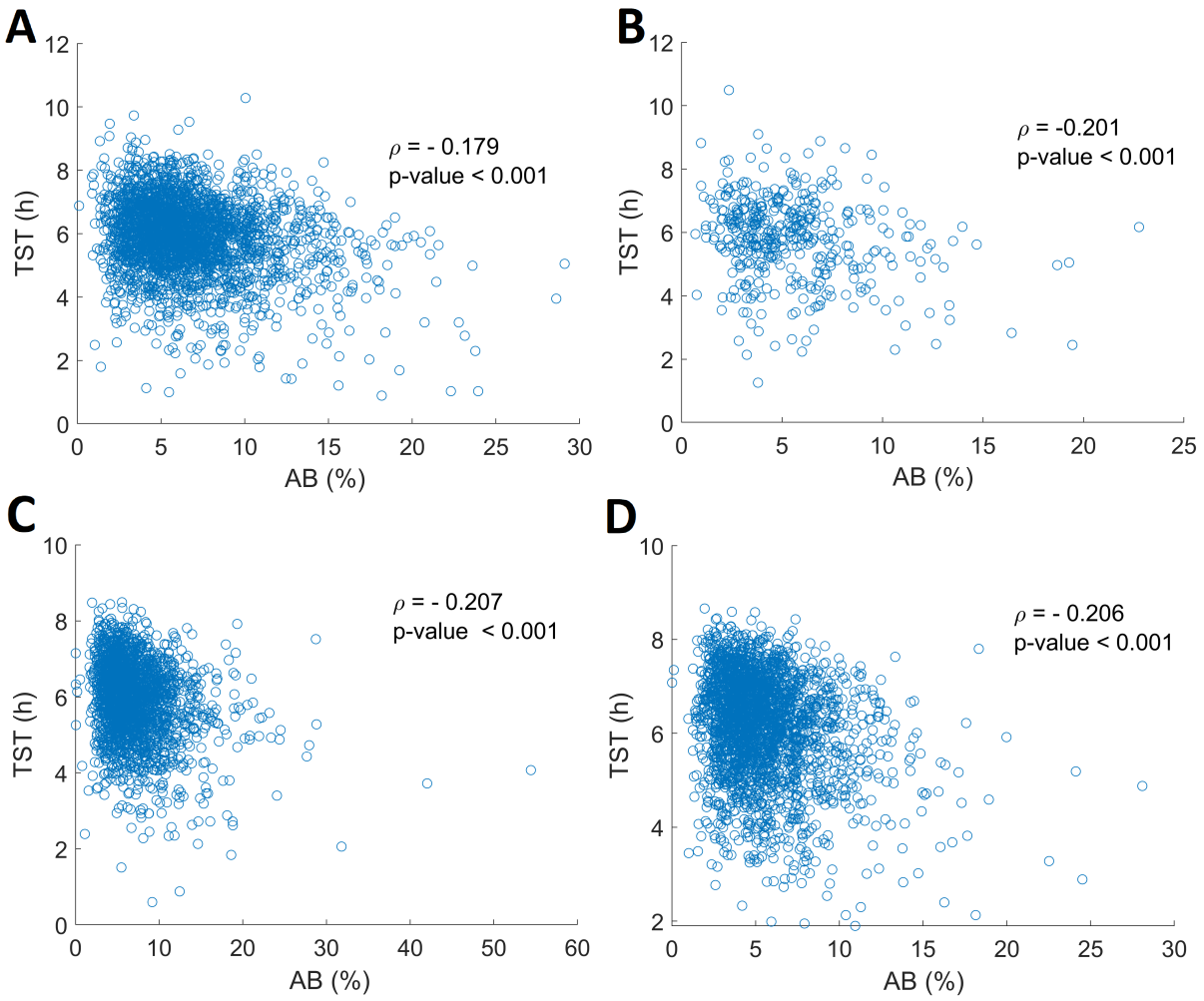


Figure C.7. Illustration of correlation between ABI and TST. Association between arousal burden index (ABI) and total sleep time (TST) in A) the Osteoporotic Fractures in Men Study (MrOS), B) the Study of Osteoporotic Fractures (SOF), C) men in the Sleep Heart Health Study (SHHS) and D) women in SHHS. ρ – Spearman's correlation coefficient.

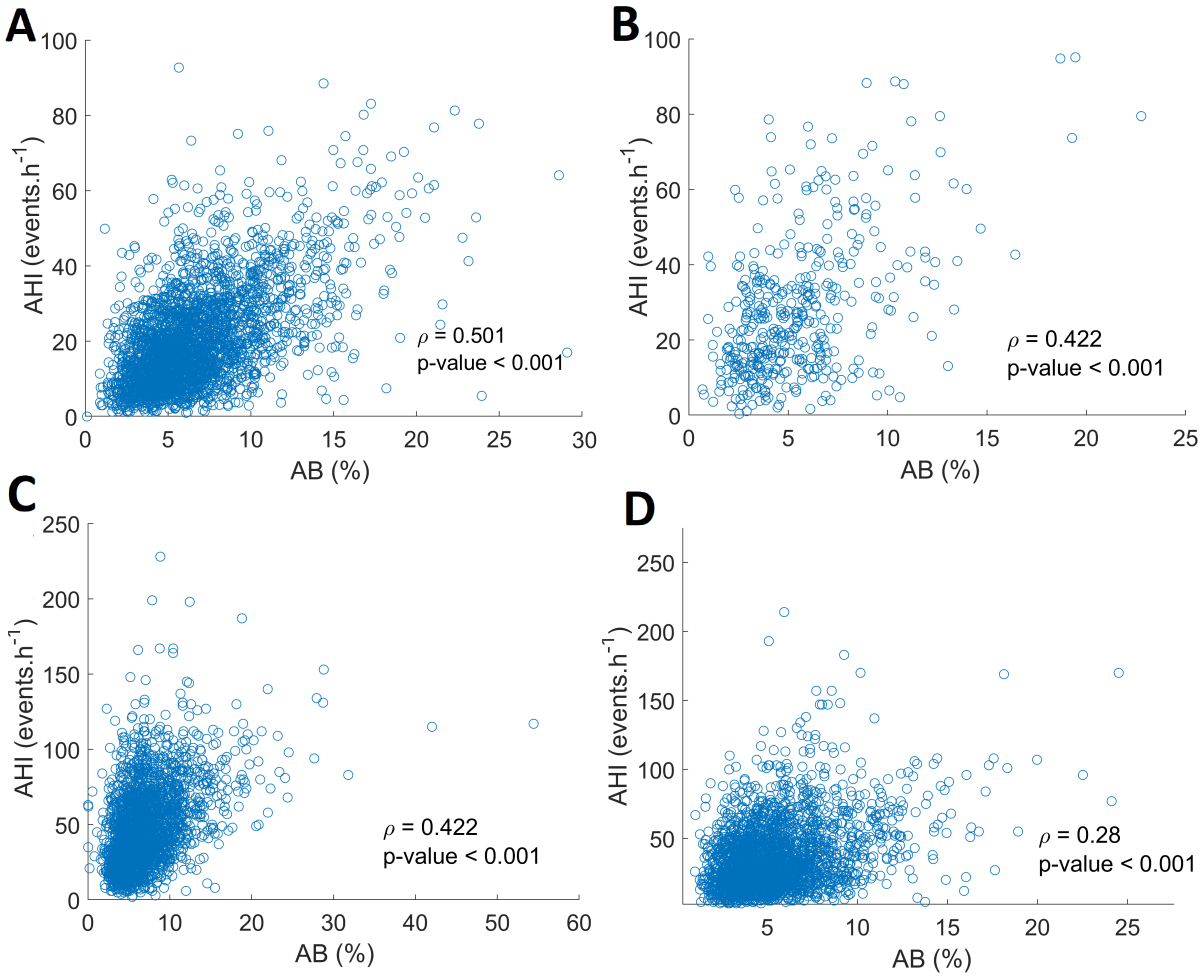


Figure C.8. Illustration of correlation between ABI and AHI. Association between arousal burden index (ABI) and apnoea-hypopnoea index (AHI) in A) the Osteoporotic Fractures in Men Study (MrOS), B) the Study of Osteoporotic Fractures (SOF), C) men in the Sleep Heart Health Study (SHHS) and D) women in SHHS. ρ – Spearman's correlation coefficient.

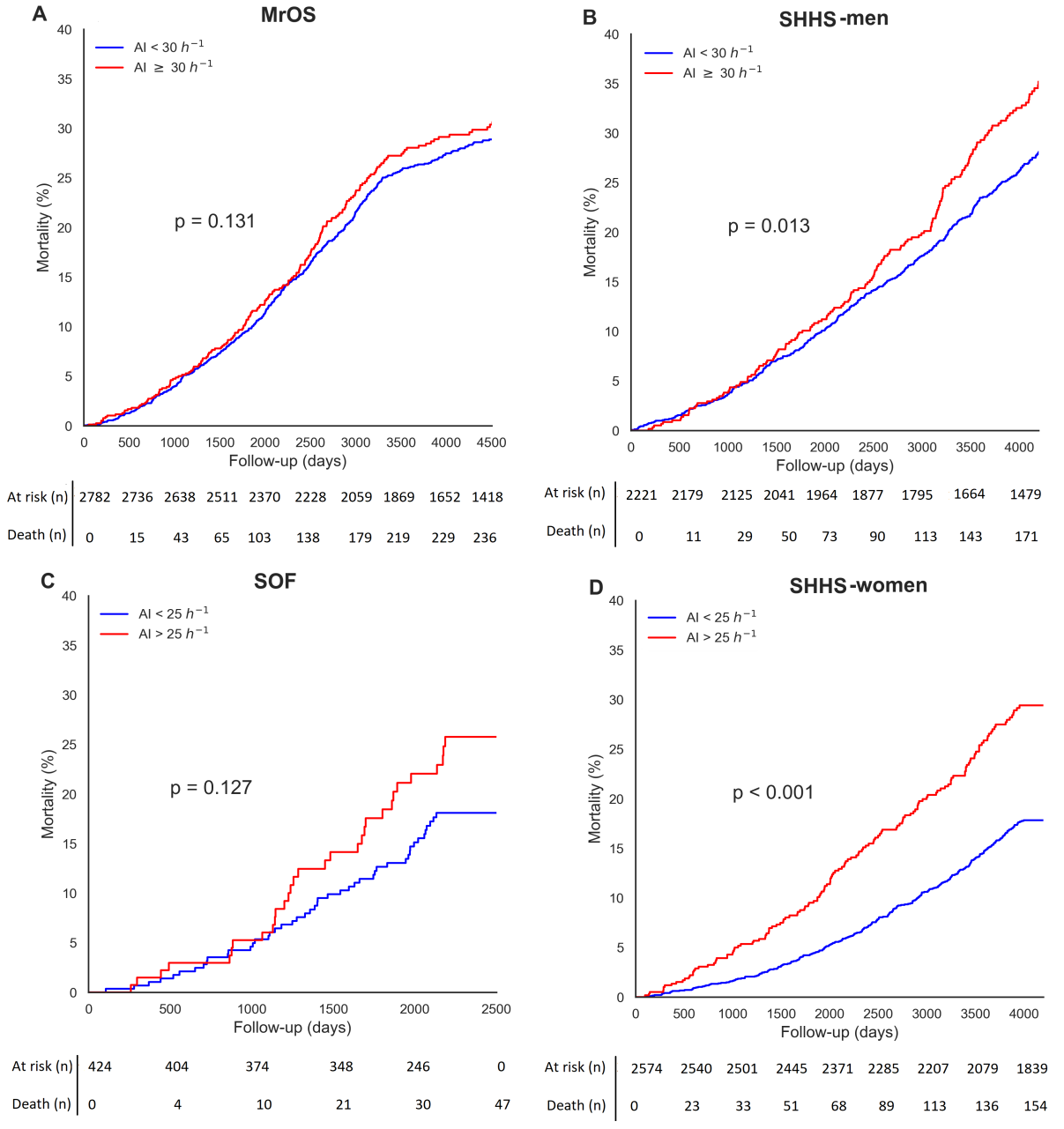


Figure C.9. Kaplan-Meier curves for arousal index (AI) and all-cause mortality. Kaplan-Meier curves for arousal index (AI) and all-cause mortality in (A) the Osteoporotic Fractures in Men Study (MrOS) Sleep cohort, (B) men in Sleep Heart Health Study (SHHS), (C) the Study of Osteoporotic (SOF) cohort and (D) women in the SHHS cohort. AI values were dichotomised at 30 h^{-1} for men and 25 h^{-1} for women. *p*-values refer to log-rank test results.

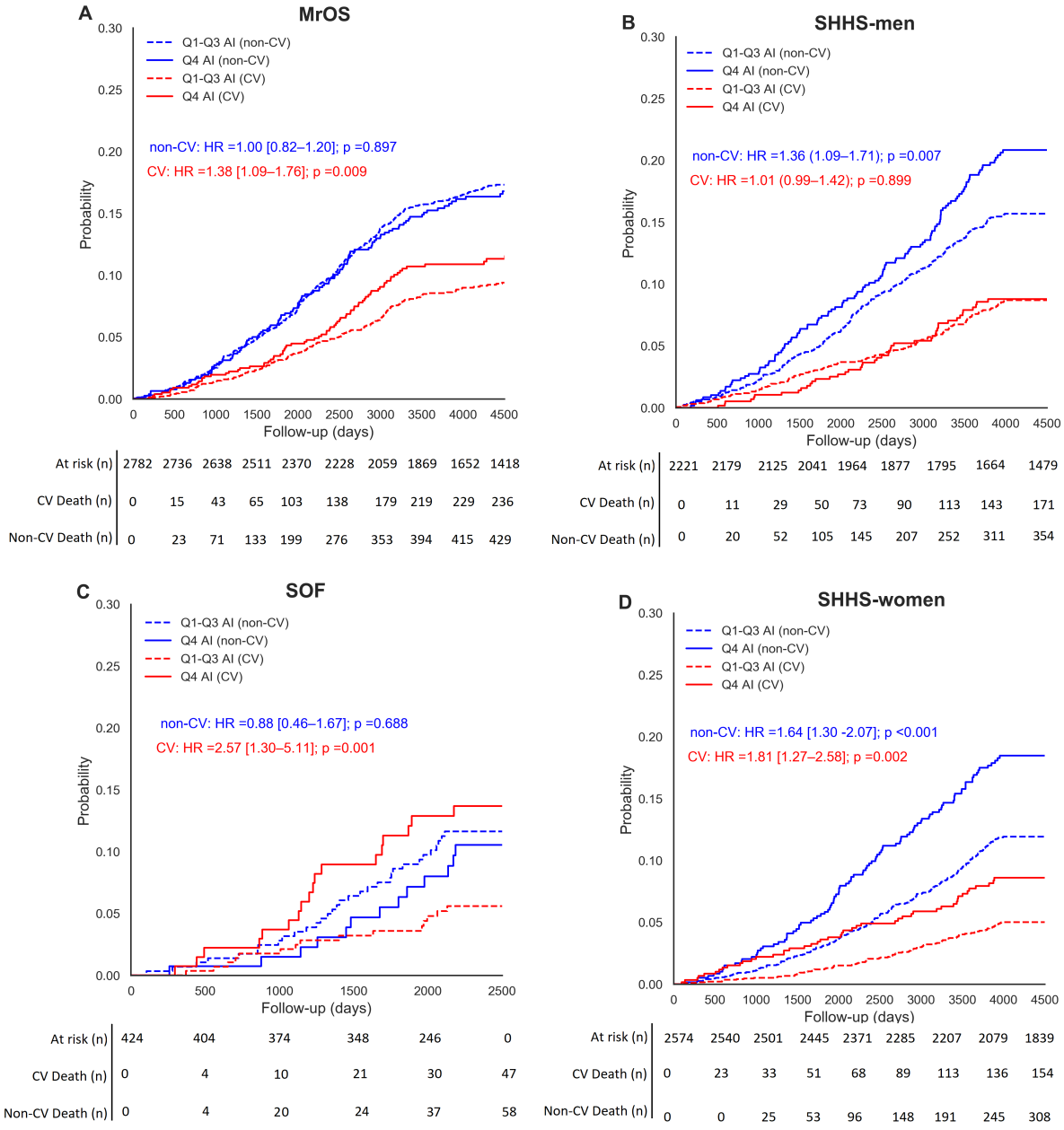


Figure C.10. Cummulative incident function curves compare the competing risk of arousal index (AI) of cardiovascular (CV), non-cardiovascular and all-cause mortality. Cummulative incident function curves compare the competing risk of arousal index (AI) of cardiovascular (CV), non-cardiovascular and all-cause mortality in (A) men from the Osteoporotic Fractures in Men Study (MrOS) Sleep cohort, (B) men from the Sleep Heart Health Study, (C) in women from the Study of Osteoporotic (SOF) cohort and (D) women from the Sleep Heart Health Study. Hazard ratio (HR) and *p*-value were estimated through subdistributional Fine-Gray hazard model.

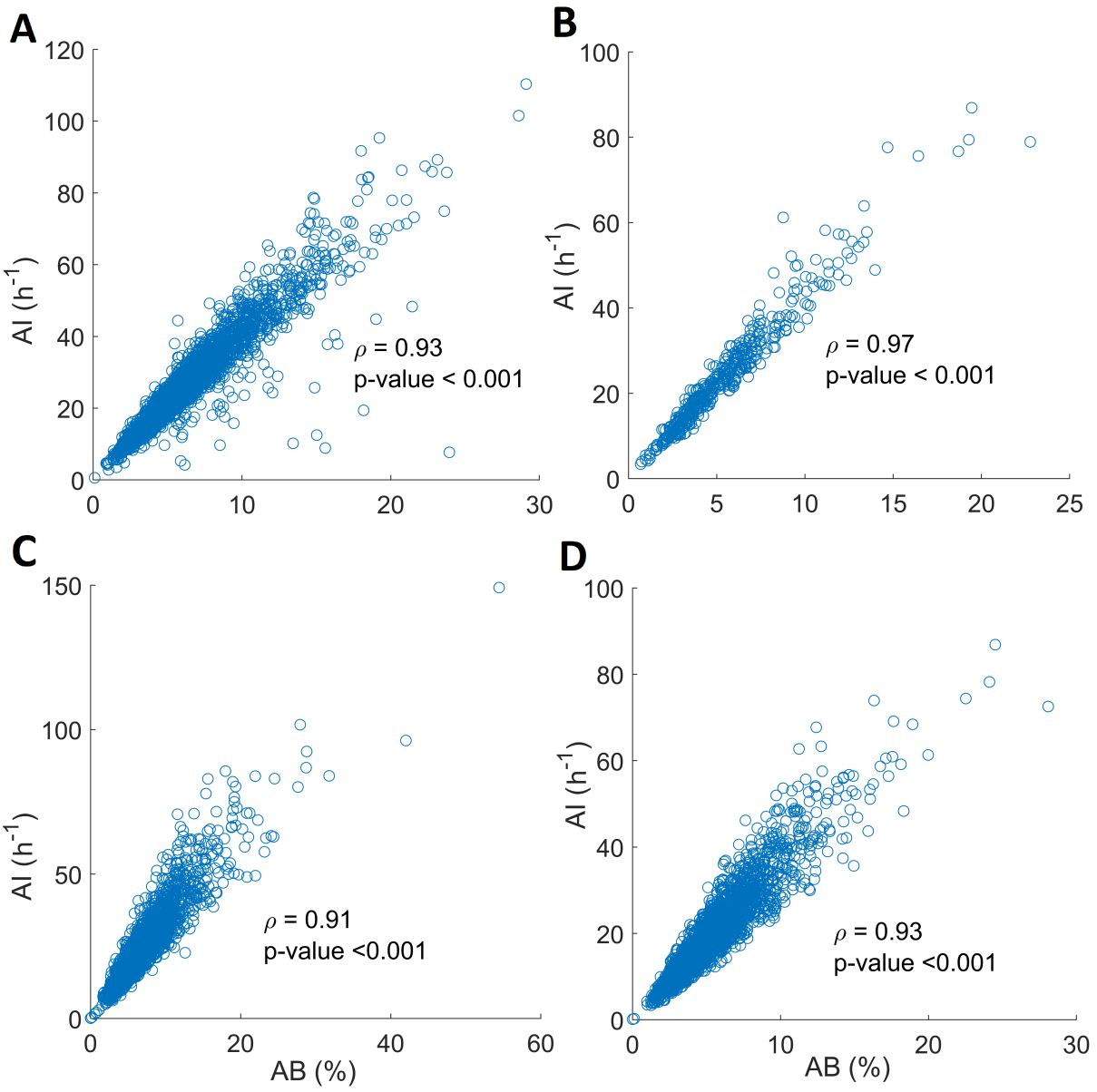


Figure C.11. Illustration of correlation between ABI and AI. Association between arousal burden index (ABI) and arousal index (AI) in A) the Osteoporotic Fractures in Men Study (MrOS), B) the Study of Osteoporotic Fractures (SOF), C) men in the Sleep Heart Health Study (SHHS) and D) women in SHHS. $/rho$ – Spearman's correlation coefficient.

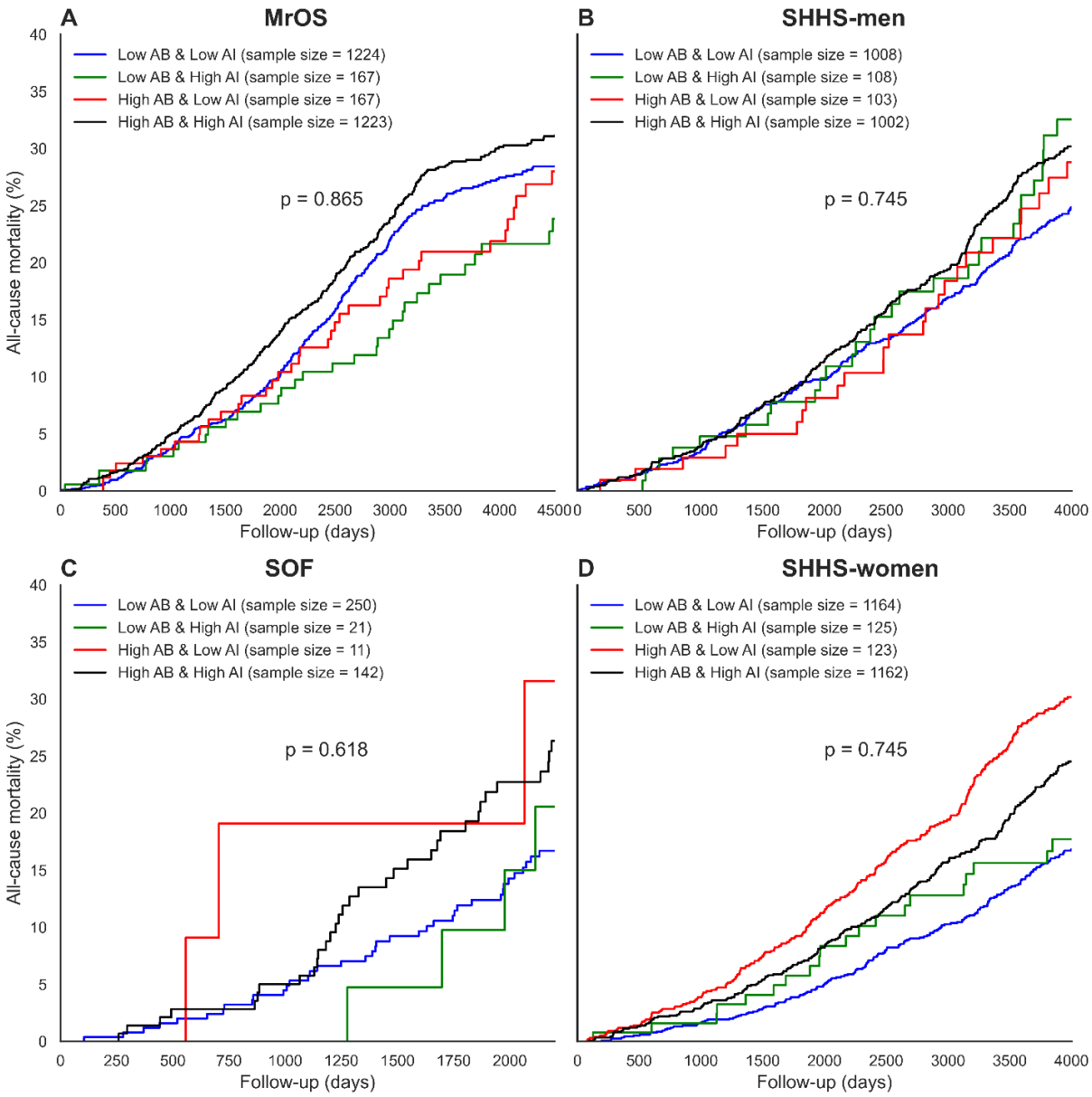


Figure C.12. Kaplan-Meier curves of combined arousal burden (AB) and arousal index (AI) and all-cause mortality in (A). Kaplan-Meier curves of combined arousal burden (AB) and arousal index (AI) and all-cause mortality in (A) the Osteoporotic Fractures in Men Study (MrOS) Sleep cohort, (B) men in the Sleep Heart Health Study (SHHS), (C) the Study of Osteoporotic (SOF) cohort and (D) women in the SHHS cohort. AB and AI values were dichotomised based on cohort median values.

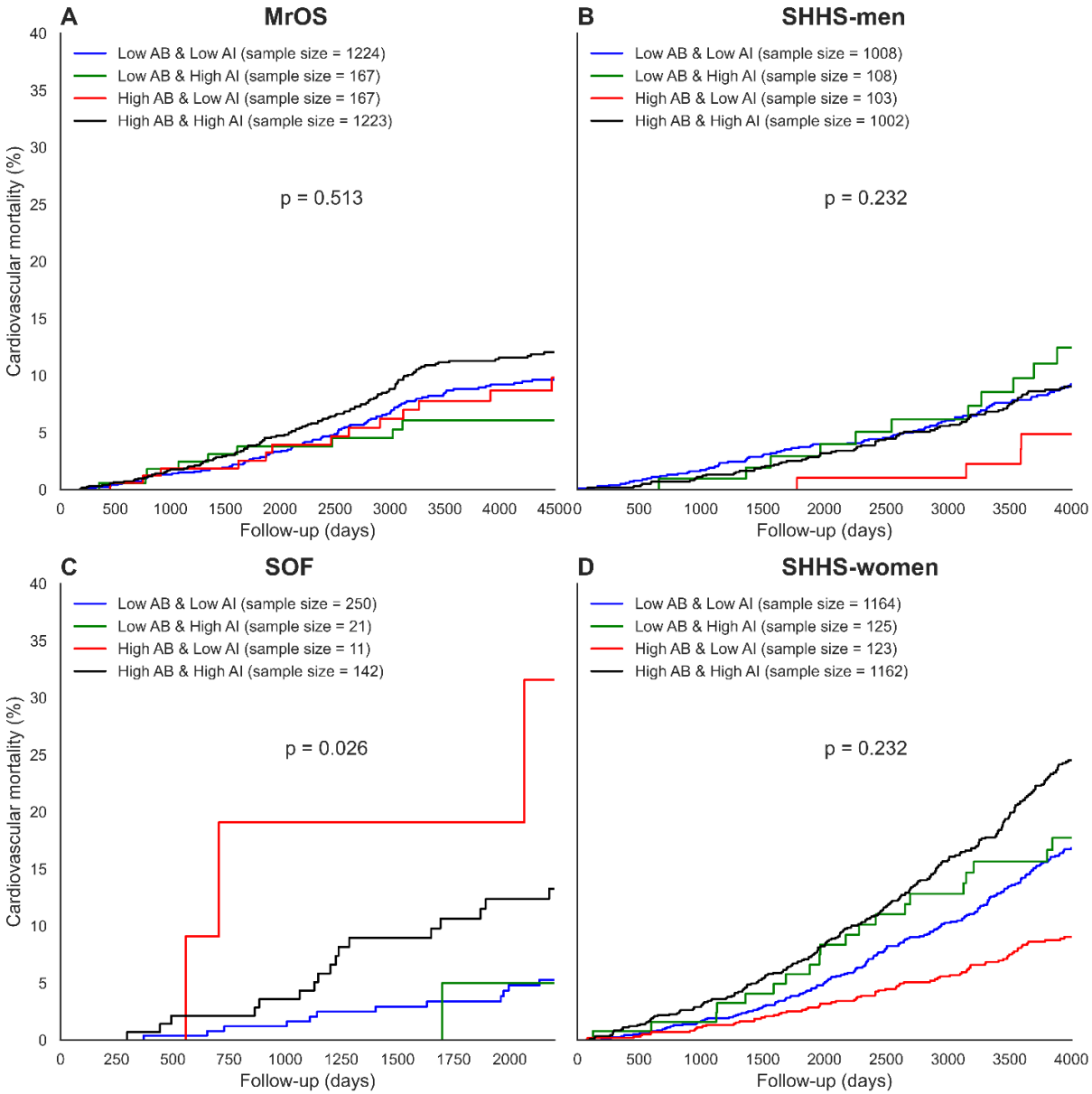


Figure C.13. Kaplan-Meier curves of combined arousal burden (AB) and arousal index (AI) and cardiovascular mortality. Kaplan-Meier curves of combined arousal burden (AB) and arousal index (AI) and cardiovascular mortality in (A) the Osteoporotic Fractures in Men Study (MrOS) Sleep cohort, (B) men in the Sleep Heart Health Study (SHHS), (C) the Study of Osteoporotic (SOF) cohort and (D) women in the SHHS cohort. AB and AI values were dichotomised based on cohort median values.

Bibliography

Ali, N. J., Pitson, D. J., and Stradling, J. R. (1993). 'Snoring, sleep disturbance, and behaviour in 4-5 year olds', *Archives of Disease in Childhood*, **68**(3), pp. 360–366.

Alvarez-Estevez, D., and Fernández-Varela, I. (2020). 'Addressing database variability in learning from medical data: An ensemble-based approach using convolutional neural networks and a case of study applied to automatic sleep scoring', *Computers in Biology and Medicine*, **119**, 103697.

Aly, M. (2005). *Survey on multiclass classification methods*, Technical report, California Institute of Technology.

Amzica, F., and Steriade, M. (2002). 'The functional significance of K-complexes', *Sleep Medicine Reviews*, **6**(2), pp. 139–149.

Arce-Santana, E. R., Alba, A., Mendez, M. O., and Arce-Guevara, V. (2020). 'A-phase classification using convolutional neural networks', *Medical & Biological Engineering & Computing*, **58**(5), pp. 1003–1014.

ASDA. (1992). 'EEG arousals: scoring rules and examples', *Sleep*, **15**(2), 173.

Aserinsky, E., and Kleitman, N. (1953). 'Regularly occurring periods of eye motility, and concomitant phenomena, during sleep', *Science*, **118**(3062), pp. 273 – 274.

Azarbarzin, A., Sands, S. A., Stone, K. L., Taranto-Montemurro, L., Messineo, L., Terrill, P. I., Ancoli-Israel, S., Ensrud, K., Purcell, S., White, D. P., Redline, S., and Wellman, A. (2019). 'The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the Osteoporotic Fractures in Men Study and the Sleep Heart Health Study', *European Heart Journal*, **40**(14), pp. 1149–1157.

Babaeizadeh, S., Zhou, S. H., Pittman, S. D., and White, D. P. (2011). 'Electrocardiogram-derived respiration in screening of sleep-disordered breathing', *Journal of Electrocardiology*, **44**(6), pp. 700–706.

Baharav, A., Kotagal, S., Gibbons, V., Rubin, B. K., Pratt, G., Karin, J., and Akselrod, S. (1995). 'Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability', *Neurology*, **45**(6), pp. 1183 – 1187.

Bahdanau, D., Cho, K., and Bengio, Y. (2014). 'Neural machine translation by jointly learning to align and translate', *arXiv preprint arXiv:1409.0473*, .

Bandarabadi, M., Teixeira, C. A., Rasekhi, J., and Dourado, A. (2015). 'Epileptic seizure prediction using relative spectral power features', *Clinical Neurophysiology*, **126**(2), pp. 237–248.

Barcaro, U., Bonanni, E., Maestri, M., Murri, L., Parrino, L., and Terzano, M. G. (2004). 'A general automatic method for the analysis of NREM sleep microstructure', *Sleep Medicine*, **5**(6), pp. 567–576.

Barcaro, U., Navona, C., Belloli, S., Bonanni, E., Gneri, C., and Murri, L. (1998). 'A simple method for the quantitative description of sleep microstructure', *Electroencephalography and Clinical Neurophysiology*, **106**(5), pp. 429–432.

Barrett, A., and Barnett, L. (2013). 'Granger causality is designed to measure effect, not mechanism', *Frontiers in Neuroinformatics*, **7**, 6.

Bartsch, R. P., Liu, K. K., Bashan, A., and Ivanov, P. C. (2015). 'Network physiology: how organ systems dynamically interact', *PLoS ONE*, **10**(11), e0142143.

Bashan, A., Bartsch, R. P., Kantelhardt, J. W., Havlin, S., and Ivanov, P. C. (2012). 'Network physiology reveals relations between network topology and physiological function', *Nature Communications*, **3**, 702.

Baumert, M., Immanuel, S. A., Stone, K. L., Litwack Harrison, S., Redline, S., Mariani, S., Sanders, P., McEvoy, R. D., and Linz, D. (2020). 'Composition of nocturnal hypoxaemic burden and its prognostic value for cardiovascular mortality in older community-dwelling men', *European Heart Journal*, **41**(4), pp. 533–541.

Baumert, M., Kohler, M., Kabir, M., Kennedy, D., and Pamula, Y. (2010). 'Cardiorespiratory response to spontaneous cortical arousals during stage 2 and rapid eye movement sleep in healthy children', *Journal of Sleep Research*, **19**(3), pp. 415–424.

Baumert, M., Linz, D., Stone, K., McEvoy, R. D., Cummings, S., Redline, S., Mehra, R., and Immanuel, S. (2019). 'Mean nocturnal respiratory rate predicts cardiovascular and all-cause mortality in community-dwelling older men and women', *European Respiratory Journal*, **54**(1), 1802175.

Becker, D. E. (2006). 'Fundamentals of electrocardiography interpretation', *Anesthesia Progress*, **53**(2), pp. 53–64.

Bengio, Y., Simard, P., and Frasconi, P. (1994). 'Learning long-term dependencies with gradient descent is difficult', *IEEE Transactions on Neural Networks*, **5**(2), pp. 157–166.

Benington, J. H. (2000). 'Sleep homeostasis and the function of sleep', *Sleep*, **23**(7), pp. 959–966.

Bennett, L. S., Langford, B. A., Stradling, J. R., and Davies, R. J. O. (1998). 'Sleep fragmentation indices as predictors of daytime sleepiness and nCPAP response in obstructive sleep apnea', *American Journal of Respiratory and Critical Care Medicine*, **158**(3), pp. 778–786.

Berry, R. B., and Gleeson, K. (1997). 'Respiratory arousal from sleep: mechanisms and significance', *Sleep*, **20**(8), pp. 654–675.

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Marcus, C., and Vaughn, B. V. (2015). *The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*, Version 2.2, American Academy of Sleep Medicine, Darien, Illinois.

Bishop, C. M. (2006). '*Pattern recognition and machine learning*', 1 edn, Springer, New York, NY.

Bixler, E. O., Papaliaga, M. N., Vgontzas, A. N., Lin, H.-M., Pejovic, S., Karataraki, M., Vela-Bueno, A., and Chrouso, G. P. (2009). 'Women sleep objectively better than men and the sleep of young women is more resilient to external stressors: effects of age and menopause', *Journal of Sleep Research*, **18**(2), pp. 221–228.

Blackwell, T., Yaffe, K., Ancoli-Israel, S., Redline, S., Ensrud, K. E., Stefanick, M. L., Laffan, A., and Stone, K. L. (2011). 'Associations between sleep architecture and sleep-disordered breathing and cognition in older community-dwelling men: the Osteoporotic Fractures in Men Sleep Study', *Journal of the American Geriatrics Society*, **59**(12), pp. 2217–2225.

Blank, J. B., Cawthon, P. M., Carrion-Petersen, M. L., Harper, L., Johnson, J. P., Mitson, E., and Delay, R. R. (2005). 'Overview of recruitment for the Osteoporotic Fractures in Men Study (MrOS)', *Contemporary Clinical Trials*, **26**(5), pp. 557–568.

Boly, M., Perlbarg, V., Marrelec, G., Schabus, M., Laureys, S., Doyon, J., Pélégrini-Issac, M., Maquet, P., and Benali, H. (2012). 'Hierarchical clustering of brain activity during human nonrapid eye movement sleep', *Proceedings of the National Academy of Sciences*, **109**(15), pp. 5856 – 5861.

Bonnet, M., and Arand, D. (1997). 'Heart rate variability: sleep stage, time of night, and arousal influences', *Electroencephalography and Clinical Neurophysiology*, **102**(5), pp. 390–396.

Bonnet, M., Carley, D., Carskadon, M., and Easton, P. (1992). 'EEG arousals: scoring rules and examples', *Sleep*, **15**(2), pp. 173–184.

Bonnet, M. H. (1989). 'The effect of sleep fragmentation on sleep and performance in younger and older subjects', *Neurobiology of Aging*, **10**(1), pp. 21–25.

Bonnet, M. H., and Arand, D. L. (2003). 'Clinical effects of sleep fragmentation versus sleep deprivation', *Sleep Medicine Reviews*, **7**(4), pp. 297–310.

Bonnet, M. H., and Arand, D. L. (2007). 'EEG arousal norms by age', *Journal of Clinical Sleep Medicine*, **3**(03), pp. 271–274.

Boselli, M., Parrino, L., Smerieri, A., and Terzano, M. G. (1998). 'Effect of age on EEG arousals in normal sleep', *Sleep*, **21**(4), pp. 361–367.

Bosi, M., Milioli, G., Riccardi, S., Melpignano, A., Vaudano, A. E., Cortelli, P., Poletti, V., and Parrino, L. (2018). 'Arousal responses to respiratory events during sleep: the role of pulse wave amplitude', *Journal of Sleep Research*, **27**(2), pp. 261–269.

Brand, S., and Kirov, R. (2011). 'Sleep and its importance in adolescence and in common adolescent somatic and psychiatric conditions', *International Journal of General Medicine*, **4**, pp. 425–442.

Bressler, S. L., and Seth, A. K. (2011). 'Wiener–Granger causality: a well established methodology', *NeuroImage*, **58**(2), pp. 323–329.

Brouillette, R. T., Fernbach, S. K., and Hunt, C. E. (1982). 'Obstructive sleep apnea in infants and children', *The Journal of Pediatrics*, **100**(1), pp. 31–40.

Bruni, O., and Ferri, R. (2009). 'Neurocognitive deficits in pediatric obstructive sleep apnea: a multifaceted pathogenetic model', *Sleep Medicine*, **10**(2), pp. 161–163.

Bruni, O., Ferri, R., Miano, S., Verrillo, E., Vittori, E., Della Marca, G., Farina, B., and Mennuni, G. (2002). 'Sleep cyclic alternating pattern in normal school-age children', *Clinical Neurophysiology*, **113**(11), pp. 1806–1814.

Bruni, O., Ferri, R., Miano, S., Verrillo, E., Vittori, E., Farina, B., Smerieri, A., and Terzano, M. G. (2005). 'Sleep cyclic alternating pattern in normal preschool-aged children', *Sleep*, **28**(2), pp. 220–230.

Bruni, O., Ferri, R., Novelli, L., Finotti, E., Miano, S., and Guilleminault, C. (2008). 'NREM sleep instability in children with sleep terrors: the role of slow wave activity interruptions', *Clinical Neurophysiology*, **119**(5), pp. 985–992.

Bruni, O., Ferri, R., Novelli, L., Finotti, E., Terribili, M., Troianiello, M., Valente, D., Sabatello, U., and Curatolo, P. (2009). 'Slow EEG amplitude oscillations during NREM sleep and reading disabilities in children with dyslexia', *Developmental Neuropsychology*, **34**(5), pp. 539–551.

Bruni, O., Ferri, R., Vittori, E., Novelli, L., Vignati, M., Porfirio, M. C., Aricò, D., Bernabei, P., and Curatolo, P. (2007). 'Sleep architecture and NREM alterations in children and adolescents with Asperger syndrome', *Sleep*, **30**(11), pp. 1577–1585.

Bruni, O., Kohler, M., Novelli, L., Kennedy, D., Lushington, K., Martin, J., and Ferri, R. (2012). 'The role of NREM sleep instability in child cognitive performance', *Sleep*, **35**(5), pp. 649–656.

Bruni, O., Novelli, L., Luchetti, A., Zarowski, M., Meloni, M., Cecili, M., Villa, M., and Ferri, R. (2010a). 'Reduced NREM sleep instability in benign childhood epilepsy with centro-temporal spikes', *Clinical Neurophysiology*, **121**(5), pp. 665–671.

Bruni, O., Novelli, L., Miano, S., Parrino, L., Terzano, M. G., and Ferri, R. (2010b). 'Cyclic alternating pattern: a window into pediatric sleep', *Sleep Medicine*, **11**(7), pp. 628–636.

Burges, C. J. C. (1998). 'A tutorial on support vector machines for pattern recognition', *Data Mining and Knowledge Discovery*, **2**(2), pp. 121–167.

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Hoch Carolyn C.., Yeager, A. L., and Kupfer, D. J. (1991). 'Quantification of subjective sleep quality in healthy elderly men and women using the Pittsburgh Sleep Quality Index (PSQI)', *Sleep*, **14**(4), pp. 331–338.

Cao, Y., Guo, Y., Yu, H., and Yu, X. (2017). 'Epileptic seizure auto-detection using deep learning method', *2017 4th International Conference on Systems and Informatics (ICSAI)*, pp. 1076–1081.

Cappuccio, F. P., D'Elia, L., Strazzullo, P., and Miller, M. A. (2010). 'Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies', *Sleep*, **33**(5), pp. 585–592.

Carskadon, M. A., and Dement, W. C. (2005). 'Normal human sleep: an overview', *Principles and Practice of Sleep Medicine*, **4**, pp. 13–23.

Carskadon, M. A., Brown, E. D., and Dement, W. C. (1982). 'Sleep fragmentation in the elderly: relationship to daytime sleep tendency', *Neurobiology of Aging*, **3**(4), pp. 321–327.

Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C., and Nazeran, H. (2018). 'A review on wearable photoplethysmography sensors and their potential future applications in health care', *International Journal of Biosensors & Bioelectronics*, **4**(4), pp. 195–202.

Catrambone, V., Greco, A., Vanello, N., Scilingo, E. P., and Valenza, G. (2019). 'Time-resolved directional brain–heart interplay measurement through synthetic data generation models', *Annals of Biomedical Engineering*, **47**(6), pp. 1479–1489.

Chambon, S., Galtier, M. N., and Gramfort, A. (2018a). 'Domain adaptation with optimal transport improves EEG sleep stage classifiers', *2018 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, 10.1109/PRNI.2018.8423957.

Chambon, S., Galtier, M. N., Arnal, P. J., Wainrib, G., and Gramfort, A. (2018b). 'A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **26**(4), pp. 758–769.

Cho, S.-H., Hur, J., and Chung, I.-Y. (2014). 'An applicability of Teager Energy Operator and energy separation algorithm for waveform distortion analysis : harmonics, interharmonics and frequency variation', *Journal of Electrical Engineering and Technology*, **9**, pp. 1210–1216.

Čić, M., Šoda, J., and Bonković, M. (2013). 'Automatic classification of infant sleep based on instantaneous frequencies in a single-channel EEG signal', *Computers in Biology and Medicine*, **43**(12), pp. 2110–2117.

Cirelli, C., and Tononi, G. (2008). 'Is sleep essential?', *PLoS Biol*, **6**(8), e216.

Claman, D. M., Redline, S., Blackwell, T., Ancoli-Israel, S., Surovec, S., Scott, N., Cauley, J. A., Ensrud, K. E., and Stone, K. L. (2006). 'Prevalence and correlates of periodic limb movements in older women', *Journal of Clinical Sleep Medicine*, **2**(04), pp. 438–445.

Colrain, I. M. (2005). 'The K-complex: a 7-decade history', *Sleep*, **28**(2), pp. 255–273.

Cummings, S. R., Black, D. M., Nevitt, M. C., Browner, W. S., Cauley, J. A., Genant, H. K., Mascioli, S. R., Scott, J. C., Seeley, D. G., Steiger, P., and Vogt, T. M. (1990). 'Appendicular bone density and age predict hip fracture in women', *JAMA*, **263**(5), pp. 665–668.

Darvishi, S., Gharabaghi, A., Boulay, C. B., Ridding, M. C., Abbott, D., and Baumert, M. (2017). 'Proprioceptive feedback facilitates motor imagery-related operant learning of sensorimotor β-band modulation', *Frontiers in Neuroscience*, **11**, 60.

de Zambotti, M., Trinder, J., Silvani, A., Colrain, I., and Baker, F. C. (2018). 'Dynamic coupling between the central and autonomic nervous systems during sleep: a review', *Neuroscience & Biobehavioral Reviews*, **90**, pp. 84–103.

de Zambotti, M., Willoughby, A. R., Franzen, P. L., Clark, D. B., Baker, F. C., and Colrain, I. M. (2016). 'K-complexes: interaction between the central and autonomic nervous systems during sleep', *Sleep*, **39**(5), pp. 1129–1137.

Deak, M., and Epstein, L. J. (2009). 'The history of polysomnography', *Sleep Medicine Clinics*, **4**(3), pp. 313–321.

Dean II, D. A., Goldberger, A. L., Mueller, R., Kim, M., Rueschman, M., Mobley, D., Sahoo, S. S., Jayapandian, C. P., Cui, L., Morrical, M. G., Surovec, S., Zhang, G.-Q., and Redline, S. (2016). 'Scaling up scientific ciscovery in sleep medicine: The National Sleep Research Resource', *Sleep*, **39**(5), pp. 1151–1164.

Deboer, T. (2015). 'Behavioral and electrophysiological correlates of sleep and sleep homeostasis', *in* P. Meerlo., R. M. Benca., and T. Abel. (eds.), *Sleep, Neuronal Plasticity and Brain Function*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–24.

Della Marca, G., Farina, B., Mennuni, G. F., Mazza, S., Di Giannantonio, M., Spadini, V., De Risio, S., Ciocca, A., and Mazza, M. (2004). 'Microstructure of sleep in eating disorders: preliminary results', *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, **9**(1), pp. 77–80.

DelRosso, L. M., and Ferri, R. (2019). 'The prevalence of restless sleep disorder among a clinical sample of children and adolescents referred to a sleep centre', *Journal of Sleep Research*, **28**(6), e12870.

DelRosso, L. M., Bruni, O., and Ferri, R. (2018). 'Restless sleep disorder in children: a pilot study on a tentative new diagnostic category', *Sleep*, **41**(8), zsy102.

DelRosso, L. M., Jackson, C. V., Trotter, K., Bruni, O., and Ferri, R. (2019). 'Video-polysomnographic characterization of sleep movements in children with restless sleep disorder', *Sleep*, **42**(4), zsy269.

Dhok, S., Pimpalkhute, V., Chandurkar, A., Bhurane, A. A., Sharma, M., and Acharya, U. R. (2020). 'Automated phase classification in cyclic alternating patterns in sleep stages using Wigner–Ville distribution based features', *Computers in Biology and Medicine*, **119**, 103691.

Dijk, D.-J. (2009). 'Regulation and functional correlates of slow wave sleep', *Journal of Clinical Sleep Medicine*, **5**(2 suppl), pp. S6–S15.

Ding, M., Bressler, S. L., Yang, W., and Liang, H. (2000). 'Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment', *Biological Cybernetics*, **83**(1), pp. 35–45.

Diykh, M., Li, Y., and Wen, P. (2016). 'EEG sleep stages classification based on time domain features and structural graph similarity', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **24**(11), pp. 1159–1168.

Dong, H., Supratak, A., Pan, W., Wu, C., Matthews, P. M., and Guo, Y. (2018). 'Mixed neural network approach for temporal sleep stage classification', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **26**(2), pp. 324–333.

Dorantes, G., Méndez, M., Alba, A., González, J. S., Parrino, L., and Milioli, G. (2015). 'Heart rate variability in cyclic alternating pattern during sleep in healthy and Nocturnal Front Lobe Epilepsy patients', *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Vol. 2015-Novem, pp. 5944–5947.

Dorantes-Méndez, G., Mendez, M. O., Alba, A., Parrino, L., and Milioli, G. (2018). 'Time-varying analysis of the heart rate variability during A-phases of sleep: healthy and pathologic conditions', *Biomedical Signal Processing and Control*, **40**, pp. 111–116.

Eberhard, A., Calabrese, P., Baconnier, P., and Benchetrit, G. (2001). 'Comparison between the respiratory inductance plethysmography signal derivative and the airflow signal', *in* C.-S. Poon., and H. Kazemi. (eds.), *Frontiers in Modeling and Control of Breathing: Integration at Molecular, Cellular, and Systems Levels*, Springer US, Boston, MA, pp. 489–494.

Eckert, D. J., and Younes, M. K. (2013). 'Arousal from sleep: implications for obstructive sleep apnea pathogenesis and treatment', *Journal of Applied Physiology*, **116**(3), pp. 302–313.

Elgendi, M., Fletcher, R., Liang, Y., Howard, N., Lovell, N. H., Abbott, D., Lim, K., and Ward, R. (2019). 'The use of photoplethysmography for assessing hypertension', *npj Digital Medicine*, **2**(1), 60.

Erhan, D., Bengio, Y., Courville, A. C., and Vincent, P. (2009). *Visualizing higher-layer features of a deep network*, Technical report, University of Montreal.

Faes, L., Marinazzo, D., Jurysta, F., and Nollo, G. (2014a). 'Granger causality analysis of sleep brain-heart interactions', *2014 8th Conference of the European Study Group on Cardiovascular Oscillations (ESGCO)*, pp. 5–6.

Faes, L., Marinazzo, D., Jurysta, F., and Nollo, G. (2015). 'Linear and non-linear brain-heart and brain-brain interactions during sleep', *Physiological Measurement*, **36**(4), pp. 683–698.

Faes, L., Marinazzo, D., Stramaglia, S., Jurysta, F., Porta, A., and Giandomenico, N. (2016). 'Predictability decomposition detects the impairment of brain–heart dynamical networks during sleep disorders and their recovery with treatment', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **374**(2067), 20150177.

Faes, L., Nollo, G., Jurysta, F., and Marinazzo, D. (2014b). 'Information dynamics of brain-heart physiological networks during sleep', *New Journal of Physics*, **16**(10), 105005.

Fan, M., Sun, D., Zhou, T., Heianza, Y., Lv, J., Li, L., and Qi, L. (2020). 'Sleep patterns, genetic susceptibility, and incident cardiovascular disease: a prospective study of 385 292 UK biobank participants', *European Heart Journal*, **41**(11), pp. 1182–1189.

Farina, B., Della Marca, G., Grochocinski, V. J., Mazza, M., Buysse, D. J., Di Giannantonio, M., Francesco Mennuni, G., De Risio, S., Kupfer, D. J., and Frank, E. (2003). 'Microstructure of sleep in depressed patients according to the cyclic alternating pattern', *Journal of Affective Disorders*, **77**(3), pp. 227–235.

Ferini-Strambi, L., Bianchi, A., Zucconi, M., Oldani, A., Castronovo, V., and Smirne, S. (2000). 'The impact of cyclic alternating pattern on heart rate variability during sleep in healthy young adults', *Clinical Neurophysiology*, **111**(1), pp. 99–101.

Ferri, R., Bruni, O., Miano, S., and Terzano, M. G. (2005a). 'Topographic mapping of the spectral components of the cyclic alternating pattern (CAP)', *Sleep Medicine*, **6**(1), pp. 29–36.

Ferri, R., Bruni, O., Miano, S., Plazzi, G., Spruyt, K., Gozal, D., and Terzano, M. G. (2006). 'The time structure of the cyclic alternating pattern during sleep', *Sleep*, **29**(5), pp. 693–699.

Ferri, R., Bruni, O., Miano, S., Smerieri, A., Spruyt, K., and Terzano, M. G. (2005b). 'Inter-rater reliability of sleep cyclic alternating pattern (CAP) scoring and validation of a new computer-assisted CAP scoring method', *Clinical Neurophysiology*, **116**(3), pp. 696–707.

Ferri, R., Franceschini, C., Zucconi, M., Drago, V., Manconi, M., Vandi, S., Poli, F., Bruni, O., and Plazzi, G. (2009). 'Sleep polygraphic study of children and adolescents with narcolepsy/cataplexy', *Developmental Neuropsychology*, **34**(5), pp. 523–538.

Ferri, R., Manconi, M., Aricò, D., Sagrada, C., Zucconi, M., Bruni, O., Oldani, A., and Ferini-Strambi, L. (2010). 'Acute dopamine-agonist treatment in restless legs syndrome: effects on sleep architecture and NREM sleep instability', *Sleep*, **33**(6), pp. 793–800.

Ferri, R., Miano, S., Bruni, O., Vankova, J., Nevsimalova, S., Vandi, S., Montagna, P., Ferini-Strambi, L., and Plazzi, G. (2005c). 'NREM sleep alterations in narcolepsy/cataplexy', *Clinical Neurophysiology*, **116**(11), pp. 2675–2684.

Ferri, R., Parrino, L., Smerieri, A., Terzano, M. G., Elia, M., Musumeci, S. A., and Pettinato, S. (2000). 'Cyclic alternating pattern and spectral analysis of heart rate variability during normal sleep', *Journal of Sleep Research*, **9**(1), pp. 13–18.

Ferri, R., Rundo, F., Novelli, L., Terzano, M. G., Parrino, L., and Bruni, O. (2012). 'A new quantitative automatic method for the measurement of non-rapid eye movement

sleep electroencephalographic amplitude variability', *Journal of Sleep Research*, **21**(2), pp. 212–220.

Fiorillo, L., Puiatti, A., Papandrea, M., Ratti, P.-L., Favaro, P., Roth, C., Bargiotas, P., Bassetti, C. L., and Faraci, F. D. (2019). 'Automated sleep scoring: a review of the latest approaches', *Sleep Medicine Reviews*, **48**, 101204.

Frank, Y., Kravath, R. E., Pollak, C. P., and Weitzman, E. D. (1983). 'Obstructive sleep apnea and its therapy: clinical and polysomnographic manifestations', *Pediatrics*, **71**(5), pp. 737 – 742.

Franken, P., and Dijk, D.-J. (2009). 'Circadian clock genes and sleep homeostasis', *European Journal of Neuroscience*, **29**(9), pp. 1820–1829.

Gaillard, J. M., and Tissot, R. (1973). 'Principles of automatic analysis of sleep records with a hybrid system', *Computers and Biomedical Research*, **6**(1), pp. 1–13.

Gangwisch, J. E., Malaspina, D., Boden-Albala, B., and Heymsfield, S. B. (2005). 'Inadequate sleep as a risk factor for obesity: analyses of the NHANES I', *Sleep*, **28**(10), pp. 1289–1296.

Garrett, D., Peterson, D. A., Anderson, C. W., and Thaut, M. H. (2003). 'Comparison of linear, nonlinear, and feature selection methods for EEG signal classification', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **11**(2), pp. 141–144.

Gers, F. A., Schraudolph, N. N., and Schmidhuber, J. (2002). 'Learning precise timing with LSTM recurrent networks', *Journal of Machine Learning Research*, **3**(Aug), pp. 115–143.

Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., and Stanley, H. E. (2000). 'PhysioBank, PhysioToolkit, and PhysioNet', *Circulation*, **101**(23), pp. e215 – e220.

Goldberger, A. L., Goldberger, Z. D., and Shvilkin, A. (2017). '*Clinical electrocardiography: a simplified approach e-book*', 9 edn, Elsevier Health Sciences, Philadelphia, PA.

Gonzalez-Salazar, J. S., Alba, A., Mendez, M. O., Luna-Rivera, J. M., Parrino, L., Grassi, A., Terzano, M., and Milioli, G. (2014). 'Characterization of the autonomic system during the cyclic alternating pattern of sleep', *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, **2014**, pp. 3805–3808.

Gozal, D., Crabtree, V. M., Sans Capdevila, O., Witcher, L. A., and Kheirandish-Gozal, L. (2007). 'C-reactive protein, obstructive sleep apnea, and cognitive dysfunction in school-aged children', *American Journal of Respiratory and Critical Care Medicine*, **176**(2), pp. 188–193.

Granger, C. W. J. (1969). 'Investigating causal relations by econometric models and cross-spectral methods', *Econometrica*, **37**(3), pp. 424–438.

Graves, A., and Schmidhuber, J. (2005). 'Framewise phoneme classification with bidirectional LSTM and other neural network architectures', *Neural Networks*, **18**(5-6), pp. 602–610.

Graves, A., Jaitly, N., and Mohamed, A. r. (2013). 'Hybrid speech recognition with deep bidirectional LSTM', *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, pp. 273–278.

Greco, A., Faes, L., Catrambone, V., Barbieri, R., Scilingo, E. P., and Valenza, G. (2019). 'Lateralization of directional brain-heart information transfer during visual emotional elicitation', *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **317**(1), pp. R25–R38.

Greenhill, L., Puig-Antich, J., Goetz, R., Hanlon, C., and Davies, M. (1983). 'Sleep architecture and REM sleep measures in prepubertal children with attention deficit disorder with hyperactivity', *Sleep*, **6**(2), pp. 91–101.

Grigg-Damberger, M., Gozal, D., Marcus, C. L., Quan, S. F., Rosen, C. L., Chervin, R. D., Wise, M., Picchietti, D. L., Sheldon, S. H., and Iber, C. (2007). 'The visual scoring of sleep and arousal in infants and children', *Journal of Clinical Sleep Medicine*, **3**(2), pp. 201–240.

Grözinger, M., Röschke, J., and Klöppel, B. (1995). 'Automatic recognition of rapid eye movement (REM) sleep by artificial neural networks', *Journal of Sleep Research*, **4**(2), pp. 86–91.

Guilleminault, C., da Rosa, A., Hagen, C. C., and Prilipko, O. (2006a). 'Cyclic alternating pattern (CAP), sleep disordered breathing, and automatic analysis', *Sleep Medicine Clinics*, **1**(4), pp. 483–489.

Guilleminault, C., Kirisoglu, C., da Rosa, A. C., Lopes, C., and Chan, A. (2006b). 'Sleepwalking, a disorder of NREM sleep instability', *Sleep Medicine*, **7**(2), pp. 163–170.

Guilleminault, C., Lopes, M. C., Hagen, C. C., and da Rosa, A. (2007). 'The cyclic alternating pattern demonstrates increased sleep instability and correlates with fatigue and sleepiness in adults with upper airway resistance syndrome', *Sleep*, **30**(5), pp. 641–647.

Haas, D. C., Foster, G. L., Nieto, F. J., Redline, S., Resnick, H. E., Robbins, J. A., Young, T., and Pickering, T. G. (2005). 'Age-dependent associations between sleep-disordered breathing and hypertension: importance of discriminating between systolic/diastolic hypertension and isolated systolic hypertension in the Sleep Heart Health Study', *Circulation*, **111**(5), pp. 614–621.

Halász, P., Terzano, M., Parrino, L., and Bódizs, R. (2004). 'The nature of arousal in sleep', *Journal of Sleep Research*, **13**(1), pp. 1–23.

Hartmann, S., and Baumert, M. (2019). 'Automatic A-Phase detection of cyclic alternating patterns in sleep using dynamic temporal information', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **27**(9), pp. 1695–1703.

Hartmann, S., Bruni, O., Ferri, R., Redline, S., and Baumert, M. (2020). 'Characterization of cyclic alternating pattern during sleep in older men and women using large population studies', *Sleep*, **43**(7), zsaa016.

Hartmann, S., Bruni, O., Ferri, R., Redline, S., and Baumert, M. (2021). 'Cyclic alternating pattern in children with obstructive sleep apnea and its relationship with adenotonsillectomy, behavior, cognition, and quality of life', *Sleep*, **44**(1), zsaa145.

He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., and Zhang, K. (2019). 'The practical implementation of artificial intelligence technologies in medicine', *Nature Medicine*, **25**(1), pp. 30–36.

Hecht-Nielsen. (1989). 'Theory of the backpropagation neural network', *International 1989 Joint Conference on Neural Networks*, pp. 593–605.

Hirshkowitz, M. (2002). 'Arousals and anti-arousals', *Sleep Medicine*, **3**(3), pp. 203–204.

Hirshkowitz, M. (2015). 'The history of polysomnography: tool of scientific discovery', *in* S. Chokroverty., and M. Billiard. (eds.), *Sleep Medicine: A Comprehensive Guide to Its Development, Clinical Milestones, and Advances in Treatment*, Springer, New York, NY, pp. 91–100.

Hjorth, B. (1970). 'EEG analysis based on time domain properties', *Electroencephalography and Clinical Neurophysiology*, **29**(3), pp. 306–310.

Hobson, J. A., Spagna, T., and Malenka, R. (1978). 'Ethology of sleep studied with time-lapse photography: postural immobility and sleep-cycle phase in humans', *Science*, **201**(4362), pp. 1251 – 1253.

Hochreiter, S., and Schmidhuber, J. (1997). 'Long short-term memory', *Neural Computation*, **9**(8), pp. 1735–1780.

Horne, J. (1992). 'Human slow wave sleep: a review and appraisal of recent findings, with implications for sleep functions, and psychiatric illness', *Experientia*, **48**(10), pp. 941–954.

Huang, T., Mariani, S., and Redline, S. (2020). 'Sleep irregularity and risk of cardiovascular events: The Multi-Ethnic Study of Atherosclerosis', *Journal of the American College of Cardiology*, **75**(9), pp. 991–999.

Hussein, R., Palangi, H., Ward, R. K., and Wang, Z. J. (2019). 'Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals', *Clinical Neurophysiology*, **130**(1), pp. 25–37.

Iber, C., Ancoli-Israel, S., Chesson, A., and Quan, S. (2007). '*The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications*', 1st edn, American Academy of Sleep Medicine, Westchester, IL.

Imai, K., Keele, L., and Tingley, D. (2010). 'A general approach to causal mediation analysis.', *Psychological Methods*, **15**(4), pp. 309–334.

Isaacson, W. (2011). '*Steve Jobs: A biography*', 1 edn, Simon & Schuster, New York, NY, US.

Itowi, N., Yamatodani, A., Kiyono, S., Hiraiwa, M., and Wada, H. (1990). 'Development of a computer program classifing rat sleep stages', *Journal of Neuroscience Methods*, **31**(2), pp. 137–143.

Ivanov, P. C., Liu, K. K. L., and Bartsch, R. P. (2016). 'Focus on the emerging new fields of network physiology and network medicine', *New Journal of Physics*, **18**(10), 100201.

Jang, H.-J., and Cho, K.-O. (2019). 'Dual deep neural network-based classifiers to detect experimental seizures', *The Korean Journal of Physiology & Pharmacology*, **23**(2), pp. 131–139.

Javaheri, S., Zhao, Y. Y., Punjabi, N. M., Quan, S. F., Gottlieb, D. J., and Redline, S. (2018). 'Slow-wave sleep is associated with incident hypertension: The Sleep Heart Health Study', *Sleep*, **41**(1), zsx179.

Jurysta, F., van de Borne, P., Migeotte, P.-F., Dumont, M., Lanquart, J.-P., Degaute, J.-P., and Linkowski, P. (2003). 'A study of the dynamic interactions between sleep EEG and heart rate variability in healthy young men', *Clinical Neurophysiology*, **114**(11), pp. 2146–2155.

Kaplan, K. A., Hirshman, J., Hernandez, B., Stefanick, M. L., Hoffman, A. R., Redline, S., Ancoli-Israel, S., Stone, K., Friedman, L., and Zeitzer, J. M. (2017). 'When a gold standard isn't so golden: lack of prediction of subjective sleep quality from sleep polysomnography', *Biological Psychology*, **123**, pp. 37–46.

Karimzadeh, F., Seraj, E., Boostani, R., and Torabi-Nami, M. (2015). 'Presenting efficient features for automatic CAP detection in sleep EEG signals', *2015 38th International Conference on Telecommunications and Signal Processing (TSP)*, pp. 448–452.

Kaur, S., and Saper, C. B. (2019). 'Neural circuitry underlying waking up to hypercapnia', *Frontiers in Neuroscience*, **13**, 401.

Keller, J. M., Gray, M. R., and Givens, J. A. (1985). 'A fuzzy k-nearest neighbor algorithm', *IEEE Transactions on Systems, Man, and Cybernetics*, **SMC-15**(4), pp. 580–585.

Kern, M., Aertsen, A., Schulze-Bonhage, A., and Ball, T. (2013). 'Heart cycle-related effects on event-related potentials, spectral power changes, and connectivity patterns in the human ECoG', *NeuroImage*, **81**, pp. 178–190.

Kheirandish-Gozal, L., McManus, C. J. T., Kellermann, G. H., Samiei, A., and Gozal, D. (2013). 'Urinary neurotransmitters are selectively altered in children with obstructive sleep apnea and predict cognitive morbidity', *Chest*, **143**(6), pp. 1576–1583.

Kheirandish-Gozal, L., Miano, S., Bruni, O., Ferri, R., Pagani, J., Villa, M. P., and Gozal, D. (2007). 'Reduced NREM sleep instability in children with sleep disordered breathing', *Sleep*, **30**(4), pp. 450–457.

Kohavi, R. (1995). 'A study of cross-validation and bootstrap for accuracy estimation and model selection', *Ijcai*, **14**(2), pp. 1137–1145.

Kohler, M. J., Lushington, K., van den Heuvel, C. J., Martin, J., Pamula, Y., and Kennedy, D. (2009). 'Adenotonsillectomy and neurocognitive deficits in children with Sleep Disordered Breathing', *PLoS ONE*, **4**(10), pp. e7343–e7343.

Kondo, H., Ozone, M., Ohki, N., Sagawa, Y., Yamamichi, K., Fukuju, M., Yoshida, T., Nishi, C., Kawasaki, A., Mori, K., Kanbayashi, T., Izumi, M., Hishikawa, Y., Nishino, S., and Shimizu, T. (2014). 'Association between heart rate variability, blood pressure and autonomic activity in cyclic alternating pattern during sleep', *Sleep*, **37**(1), pp. 187–194.

Kononenko, I. (2001). 'Machine learning for medical diagnosis: history, state of the art and perspective', *Artificial Intelligence in Medicine*, **23**(1), pp. 89–109.

Koo, B. B., Blackwell, T., Ancoli-Israel, S., Stone, K. L., Stefanick, M. L., and Redline, S. (2011). 'Association of incident cardiovascular disease with periodic limb movements during sleep in older men: outcomes of sleep disorders in older men (MrOS) study', *Circulation*, **124**(11), pp. 1223–1231.

Kröller-Schön, S., Daiber, A., Steven, S., Oelze, M., Frenis, K., Kalinovic, S., Heimann, A., Schmidt, F. P., Pinto, A., Kvandova, M., Vujacic-Mirski, K., Filippou, K., Dudek, M., Bosmann, M., Klein, M., Bopp, T., Hahad, O., Wild, P. S., Frauenknecht, K., Methner, A., Schmidt, E. R., Rapp, S., Mollnau, H., and Münzel, T. (2018). 'Crucial role for Nox2 and sleep deprivation in aircraft noise-induced vascular and cerebral oxidative stress, inflammation, and gene regulation', *European Heart Journal*, **39**(38), pp. 3528–3539.

Kuwahara, H., Higashi, H., Mizuki, Y., Matsunari, S., Tanaka, M., and Inanaga, K. (1988). 'Automatic real-time analysis of human sleep stages by an interval histogram method', *Electroencephalography and Clinical Neurophysiology*, **70**(3), pp. 220–229.

Largo, R., Lopes, M. C., Spruyt, K., Guilleminault, C., Wang, Y. P., and Rosa, A. C. (2019). 'Visual and automatic classification of the cyclic alternating pattern in electroencephalography during sleep', *Brazilian Journal of Medical and Biological Research*, **52**(3).

Largo, R., Munteanu, C., and Rosa, A. (2005). 'CAP event detection by wavelets and GA tuning', *IEEE International Workshop on Intelligent Signal Processing, 2005.*, pp. 44–48.

Larkin, E. K., Rosen, C. L., Kirchner, H. L., Storfer-Isser, A., Emancipator, J. L., Johnson, N. L., Zambito, A. M. V., Tracy, R. P., Jenny, N. S., and Redline, S. (2005). 'Variation of C-reactive protein levels in adolescents', *Circulation*, **111**(15), pp. 1978–1984.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). 'Deep learning', *Nature*, **521**(7553), pp. 436–444.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). 'Gradient-based learning applied to document recognition', *Proceedings of the IEEE*, **86**(11), pp. 2278–2324.

Leon-Lomeli, R., Murguia, J., Chouvarda, I., Mendez, M., González-Galván, E., Alba, A., Milioli, G., Grassi, A., Terzano, M., and Parrino, L. (2014). 'Relation between heart beat fluctuations and cyclic alternating pattern during sleep in insomnia patients', *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, Vol. 2014.

Li, Y., Zhang, X., Winkelman, J. W., Redline, S., Hu, F. B., Stampfer, M., Ma, J., and Gao, X. (2014). 'Association between insomnia symptoms and mortality: a prospective study of U.S. men', *Circulation*, **129**(7), pp. 737–746.

Liang, Y., Abbott, D., Howard, N., Lim, K., Ward, R., and Elgendi, M. (2019). 'How effective is pulse arrival time for evaluating blood pressure? Challenges and recommendations from a study using the mimic database', *Journal of Clinical Medicine*, **8**(3), 337.

Lim, A. J., and Winters, W. D. (1980). 'A practical method for automatic real-time EEG sleep state analysis', *IEEE Transactions on Biomedical Engineering*, **BME-27**(4), pp. 212–220.

Lima, J. A., and Rosa, A. (1997). 'Maximum likelihood based classification for the microstructure of human sleep', *SIGBIO Newsletter*, **17**(3), pp. 2–6.

Linz, D., Baumert, M., Catcheside, P., Floras, J., Sanders, P., Lévy, P., Cowie, M. R., and Doug McEvoy, R. (2018). 'Assessment and interpretation of sleep disordered breathing severity in cardiology: Clinical implications and perspectives', *International Journal of Cardiology*, **271**, pp. 281–288.

Linz, D., Baumert, M., Desteghe, L., Kadhim, K., Vernooy, K., Kalman, J. M., Dobrev, D., Arzt, M., Sastry, M., Crijns, H. J. G. M., Schotten, U., Cowie, M. R., McEvoy, R. D., Heidbuchel, H., Hendriks, J., Sanders, P., and Lau, D. H. (2019). 'Nightly sleep apnea severity in patients with atrial fibrillation: potential applications of long-term sleep apnea monitoring', *International Journal of Cardiology. Heart & Vasculature*, **24**, 100424.

Liu, X., Pamula, Y., Kohler, M., and Baumert, M. (2019). 'A method for estimating pulse wave amplitude variability in children with sleep disordered breathing', *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2289–2292.

Lopes, M. C., and Guilleminault, C. (2006). 'Chronic snoring and sleep in children: a demonstration of sleep disruption', *Pediatrics*, **118**(3), pp. e741 – e746.

Lopes, M.-C., and Marcus, C. L. (2007). 'The significance of ASDA arousals in children', *Sleep Medicine*, **9**(1), pp. 3–8.

Lopes, M. C., Rosa, A., Roizenblatt, S., Guilleminault, C., Passarelli, C., Tufik, S., and Poyares, D. (2005). 'Cyclic alternating pattern in peripubertal children', *Sleep*, **28**(2), pp. 215–219.

Lorena, A. C., Jacintho, L. F. O., Siqueira, M. F., Giovanni, R. D., Lohmann, L. G., de Carvalho, A. C., and Yamamoto, M. (2011). 'Comparing machine learning classifiers in potential distribution modelling', *Expert Systems with Applications*, **38**(5), pp. 5268–5275.

Lumeng, J. C., and Chervin, R. D. (2008). 'Epidemiology of pediatric obstructive sleep apnea', *Proceedings of the American Thoracic Society*, **5**(2), pp. 242–252.

Machado, F., Sales, F., Bento, C., Dourado, A., and Teixeira, C. (2015). 'Automatic identification of cyclic alternating pattern (CAP) sequences based on the Teager Energy Operator', *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5420–5423.

Machado, F., Sales, F., Santos, C., Dourado, A., and Teixeira, C. A. (2018). 'A knowledge discovery methodology from EEG data for cyclic alternating pattern detection', *BioMedical Engineering OnLine*, **17**(1), 185.

Maestri, M., Carnicelli, L., Tognoni, G., Di Coscio, E., Giorgi, F. S., Volpi, L., Economou, N.-T., Ktonas, P., Ferri, R., Bonuccelli, U., and Bonanni, E. (2015). 'Non-rapid eye movement sleep instability in mild cognitive impairment: a pilot study', *Sleep Medicine*, **16**(9), pp. 1139–1145.

Malhotra, R. K., and Avidan, A. Y. (2013). 'Sleep stages and scoring technique', *Atlas of Sleep Medicine*, pp. 77–99.

Mancia, G. (1993). 'Autonomic modulation of the cardiovascular system during sleep', *New England Journal of Medicine*, **328**(5), pp. 347–349.

Manconi, M., Ferri, R., Zucconi, M., Bassetti, C. L., Fulda, S., Aricò, D., and Ferini-Strambi, L. (2012). 'Dissociation of periodic leg movements from arousals in restless legs syndrome', *Annals of Neurology*, **71**(6), pp. 834–844.

Manni, R., Terzaghi, M., Sartori, I., Veggiotti, P., and Parrino, L. (2004). 'Rhythmic movement disorder and cyclic alternating pattern during sleep: a video-polysomnographic study in a 9-year-old boy', *Movement Disorders*, **19**(10), pp. 1186–1190.

Marcus, C. L., Greene, M. G., and Carroll, J. L. (1998). 'Blood pressure in children with obstructive sleep apnea', *American Journal of Respiratory and Critical Care Medicine*, **157**(4), pp. 1098–1103.

Marcus, C. L., Moore, R. H., Rosen, C. L., Giordani, B., Garetz, S. L., Taylor, H. G., Mitchell, R. B., Amin, R., Katz, E. S., Arens, R., Paruthi, S., Muzumdar, H., Gozal, D., Thomas, N. H., Ware, J., Beebe, D., Snyder, K., Elden, L., Sprecher, R. C., Willging, P., Jones, D., Bent, J. P., Hoban, T., Chervin, R. D., Ellenberg, S. S., and Redline, S. (2013). 'A randomized trial of adenotonsillectomy for childhood sleep apnea', *New England Journal of Medicine*, **368**(25), pp. 2366–2376.

Mariani, S., Bianchi, A. M., Manfredini, E., Rosso, V., Mendez, M. O., Parrino, L., Matteucci, M., Grassi, A., Cerutti, S., and Terzano, M. G. (2010). 'Automatic detection of A phases of the cyclic alternating pattern during sleep', *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 5085–5088.

Mariani, S., Grassi, A., Mendez, M. O., Milioli, G., Parrino, L., Terzano, M. G., and Bianchi, A. M. (2013). 'EEG segmentation for improving automatic CAP detection', *Clinical Neurophysiology*, **124**(9), pp. 1815–1823.

Mariani, S., Grassi, A., Mendez, M. O., Parrino, L., Terzano, M. G., and Bianchi, A. M. (2011a). 'Automatic detection of CAP on central and fronto-central EEG leads via support vector machines', *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 1491–1494.

Mariani, S., Manfredini, E., Rosso, V., Grassi, A., Mendez, M. O., Alba, A., Matteucci, M., Parrino, L., Terzano, M. G., Cerutti, S., and Bianchi, A. M. (2012). 'Efficient automatic classifiers for the detection of A phases of the cyclic alternating pattern in sleep', *Medical and Biological Engineering and Computing*, **50**(4), pp. 359–372.

Mariani, S., Manfredini, E., Rosso, V., Mendez, M. O., Bianchi, A. M., Matteucci, M., Terzano, M. G., Cerutti, S., and Parrino, L. (2011b). 'Characterization of A phases during the Cyclic Alternating Pattern of sleep', *Clinical Neurophysiology*, **122**(10), pp. 2016–2024.

Marinazzo, D., Pellicoro, M., and Stramaglia, S. (2008). 'Kernel method for nonlinear Granger causality', *Physical Review Letters*, **100**(14), 144103.

McIntosh, A. R., and Gonzalez-Lima, F. (1994). 'Structural equation modeling and its application to network analysis in functional brain imaging', *Human Brain Mapping*, **2**(1-2), pp. 2–22.

McNamara, F., Issa, F. G., and Sullivan, C. E. (1996). 'Arousal pattern following central and obstructive breathing abnormalities in infants and children', *Journal of Applied Physiology*, **81**(6), pp. 2651–2657.

Medsker, L. R., and Jain, L. C. (2001). 'Recurrent neural networks', *Design and Applications*, **5**, pp. 64–67.

Memar, P., and Faradji, F. (2018). 'A novel multi-class EEG-based sleep stage classification system', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **26**(1), pp. 84–95.

Mendez, M. O., Chouvarda, I., Alba, A., Bianchi, A. M., Grassi, A., Arce-Santana, E., Milioli, G., Terzano, M. G., and Parrino, L. (2016). 'Analysis of A-phase transitions during the cyclic alternating pattern under normal sleep', *Medical & Biological Engineering & Computing*, **54**(1), pp. 133–148.

Mendonça, F., Fred, A., Mostafa, S. S., Morgado-Dias, F., and Ravelo-García, A. G. (2018a). 'Automatic detection of cyclic alternating pattern', *Neural Computing and Applications*, .

Mendonça, F., Mostafa, S. S., Morgado-Dias, F., and Ravelo-García, A. G. (2018b). 'Sleep quality estimation by cardiopulmonary coupling analysis', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **26**(12), pp. 2233–2239.

Mendonça, F., Mostafa, S. S., Morgado-Dias, F., and Ravelo-García, A. G. (2020a). 'Cyclic alternating pattern estimation based on a probabilistic model over an EEG signal', *Biomedical Signal Processing and Control*, **62**, 102063.

Mendonça, F., Mostafa, S. S., Morgado-Dias, F., and Ravelo-García, A. G. (2020b). 'Matrix of lags: a tool for analysis of multiple dependent time series applied for CAP scoring', *Computer Methods and Programs in Biomedicine*, **189**, 105314.

Mendonça, F., Mostafa, S. S., Morgado-Dias, F., and Ravelo-García, A. G. (2020c). 'On the use of patterns obtained from LSTM and feature-based methods for time series analysis: application in automatic classification of the CAP A phase subtypes', *Journal of Neural Engineering*, **18**(3), 036004.

Mendonça, F., Mostafa, S. S., Morgado-Dias, F., and Ravelo-García, A. G. (2021). 'A method based on cardiopulmonary coupling analysis for sleep quality assessment with FPGA implementation', *Artificial Intelligence in Medicine*, **112**, 102019.

Mendonça, F., Mostafa, S. S., Morgado-Dias, F., and Ravelo-García, G. A. (2019). 'A portable wireless device for cyclic alternating pattern estimation from an EEG monopolar derivation', *Entropy*, **21**(12), 1203.

Mendonça, F., Mostafa, S. S., Morgado-Dias, F., Juliá-Serdá, G., and Ravelo-García, A. G. (2020d). 'A method for sleep quality analysis based on CNN ensemble with implementation in a portable wireless device', *IEEE Access*, **8**, pp. 158523–158537.

Miano, S., Bruni, O., Elia, M., Scifo, L., Smerieri, A., Trovato, A., Verrillo, E., Terzano, M. G., and Ferri, R. (2008). 'Sleep phenotypes of intellectual disability: a polysomnographic evaluation in subjects with Down syndrome and Fragile-X syndrome', *Clinical Neurophysiology*, **119**(6), pp. 1242–1247.

Miano, S., Bruni, O., Elia, M., Trovato, A., Smerieri, A., Verrillo, E., Roccella, M., Terzano, M. G., and Ferri, R. (2007). 'Sleep in children with autistic spectrum disorder: a questionnaire and polysomnographic study', *Sleep Medicine*, **9**(1), pp. 64–70.

Miano, S., Donfrancesco, R., Bruni, O., Ferri, R., Galiffa, S., Pagani, J., Montemitro, E., Kheirandish, L., Gozal, D., and Villa, M. P. (2006). 'NREM sleep instability is reduced in children with attention-deficit/hyperactivity disorder', *Sleep*, **29**(6), pp. 797–803.

Miano, S., Paolino, M. C., Urbano, A., Parisi, P., Massolo, A. C., Castaldo, R., and Villa, M. P. (2011). 'Neurocognitive assessment and sleep analysis in children with sleep-disordered breathing', *Clinical Neurophysiology*, **122**(2), pp. 311–319.

Miano, S., Rizzoli, A., Evangelisti, M., Bruni, O., Ferri, R., Pagani, J., and Villa, M. P. (2009). 'NREM sleep instability changes following rapid maxillary expansion in children with obstructive apnea sleep syndrome', *Sleep Medicine*, **10**(4), pp. 471–478.

Mijovic, B., Vos, M. D., Gligorijevic, I., Taelman, J., and Huffel, S. V. (2010). 'Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis', *IEEE Transactions on Biomedical Engineering*, **57**(9), pp. 2188–2196.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., and Hassabis, D. (2015). 'Human-level control through deep reinforcement learning', *Nature*, **518**(7540), pp. 529–533.

Montgomery-Downs, H. E., Crabtree, V. M., and Gozal, D. (2005). 'Cognition, sleep and respiration in at-risk children treated for obstructive sleep apnoea', *European Respiratory Journal*, **25**(2), pp. 336 – 342.

Mukkamala, R., Hahn, J., Inan, O. T., Mestha, L. K., Kim, C., Töreyin, H., and Kyal, S. (2015). 'Toward ubiquitous blood pressure monitoring via pulse transit time: theory and practice', *IEEE Transactions on Biomedical Engineering*, **62**(8), pp. 1879–1901.

Münzel, T., Kröller-Schön, S., Oelze, M., Gori, T., Schmidt, F. P., Steven, S., Hahad, O., Röösli, M., Wunderli, J.-M., Daiber, A., and Sørensen, M. (2020). 'Adverse cardiovascular effects of traffic noise with a focus on nighttime noise and the new WHO noise guidelines', *Annual Review of Public Health*, **41**(1), pp. 309–328.

Nalivaiko, E., Catcheside, P. G., Adams, A., Jordan, A. S., Eckert, D. J., and McEvoy, R. D. (2007). 'Cardiac changes during arousals from non-REM sleep in healthy volunteers', *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **292**(3), pp. R1320–R1327.

Nan, Y., Chai, K. M., Lee, W. S., and Chieu, H. L. (2012). 'Optimizing F-measure: a tale of two approaches', *arXiv preprint arXiv:1206.4625*, .

Navona, C., Barcaro, U., Bonanni, E., Di Martino, F., Maestri, M., and Murri, L. (2002). 'An automatic method for the recognition and classification of the A-phases of the cyclic alternating pattern.', *Clinical Neurophysiology*, **113**(11), pp. 1826–31.

Nigro, C. A., and Rhodius, E. E. (2005). 'Variation in the duration of arousal in obstructive sleep apnea', *Medical Science Monitor*, **11**(4), pp. CR188–CR192.

Niknazar, H., Seifpour, S., Mikaili, M., Nasrabadi, A. M., and Banaraki, A. K. (2015). 'A novel method to detect the A phases of cyclic alternating pattern (CAP) using similarity index', *2015 23rd Iranian Conference on Electrical Engineering*, IEEE, pp. 67–71.

O'Driscoll, D. M., Meadows, G. E., Corfield, D. R., Simonds, A. K., and Morrell, M. J. (2004). 'Cardiovascular response to arousal from sleep under controlled conditions of central and peripheral chemoreceptor stimulation in humans', *Journal of Applied Physiology*, **96**(3), pp. 865–870.

Ohayon, M. M. (2005). 'Prevalence and correlates of nonrestorative sleep complaints', *Archives of Internal Medicine*, **165**(1), pp. 35–41.

Orwoll, E., Blank, J. B., Barrett-Connor, E., Cauley, J., Cummings, S., Ensrud, K., Lewis, C., Cawthon, P. M., Marcus, R., Marshall, L. M., McGowan, J., Phipps, K., Sherman, S., Stefanick, M. L., and Stone, K. (2005). 'Design and baseline characteristics of the Osteoporotic Fractures in Men (MrOS) study — A large observational study of the determinants of fracture in older men', *Contemporary Clinical Trials*, **26**(5), pp. 569–585.

Ozone, M., Yagi, T., Itoh, H., Tamura, Y., Inoue, Y., Uchimura, N., Sasaki, M., Shimizu, T., Terzano, M. G., and Parrino, L. (2008). 'Effects of zolpidem on cyclic alternating pattern, an objective marker of sleep instability, in Japanese patients with psychophysiological insomnia: a randomized crossover comparative study with placebo', *Pharmacopsychiatry*, **41**(03), pp. 106–114.

Pan, J., and Tompkins, W. J. (1985). 'A real-time QRS detection algorithm', *IEEE Transactions on Biomedical Engineering*, **BME-32**(3), pp. 230–236.

Park, C., Took, C. C., and Seong, J.-K. (2018). 'Machine learning in biomedical engineering', *Biomedical Engineering Letters*, **8**(1), pp. 1–3.

Parrino, L. (2021). 'In Memoriam: Mario Giovanni Terzano, MD (1944–2020): father of the cyclic alternating pattern in sleep', *Sleep and Vigilance*, **5**(1), pp. 167–168.

Parrino, L., Boselli, M., Buccino, G. P., Spaggiari, M. C., Di Giovanni, G., and Terzano, M. G. (1996). 'The cyclic alternating pattern plays a gate-control on periodic limb movements during non-rapid eye movement sleep', *Journal of Clinical Neurophysiology*, **13**(4), pp. 314–323.

Parrino, L., Boselli, M., Spaggiari, M. C., Smerieri, A., and Terzano, M. G. (1998). 'Cyclic alternating pattern (CAP) in normal sleep: polysomnographic parameters in different age groups', *Electroencephalography and Clinical Neurophysiology*, **107**(6), pp. 439–450.

Parrino, L., Ferri, R., Bruni, O., and Terzano, M. G. (2012). 'Cyclic alternating pattern (CAP): the marker of sleep instability.', *Sleep Medicine Reviews*, **16**(1), pp. 27–45.

Parrino, L., Halasz, P., Tassinari, C. A., and Terzano, M. G. (2006). 'CAP, epilepsy and motor events during sleep: the unifying role of arousal', *Sleep Medicine Reviews*, **10**(4), pp. 267–285.

Parrino, L., Milioli, G., Melpignano, A., and Trippi, I. (2016). 'The cyclic alternating pattern and the brain-body-coupling during sleep', *Epileptologie*, **33**(1), pp. 150–160.

Parrino, L., Smerieri, A., Rossi, M., and Terzano, M. G. (2001). 'Relationship of slow and rapid EEG components of CAP to ASDA arousals in normal sleep', *Sleep*, **24**(8), pp. 881–885.

Parrino, L., Spaggiari, M. C., Boselli, M., Barusi, R., and Terzano, M. G. (1993). 'Effects of prolonged wakefulness on cyclic alternating pattern (CAP) during sleep recovery at different circadian phases', *Journal of Sleep Research*, **2**(2), pp. 91–95.

Parrino, L., Thomas, R. J., Smerieri, A., Spaggiari, M. C., Felice, A. D., and Terzano, M. G. (2005). 'Reorganization of sleep patterns in severe OSAS under prolonged CPAP treatment', *Clinical Neurophysiology*, **116**(9), pp. 2228–2239.

Parvez, M. Z., and Paul, M. (2016). 'Epileptic seizure prediction by exploiting spatiotemporal relationship of EEG signals using phase correlation', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **24**(1), pp. 158–168.

Pastor-Pellicer, J., Zamora-Martínez, F., Espana-Boquera, S., and Castro-Bleda, M. J. (2013). 'F-measure as the error function to train neural networks', *International Work-Conference on Artificial Neural Networks*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 376–384.

Penzel, T., and Conradt, R. (2000). 'Computer based sleep recording and analysis', *Sleep Medicine Reviews*, **4**(2), pp. 131–148.

Pereda, E., Quiroga, R. Q., and Bhattacharya, J. (2005). 'Nonlinear multivariate analysis of neurophysiological signals', *Progress in Neurobiology*, **77**(1), pp. 1–37.

Pereira, A. M., Bruni, O., Ferri, R., and Nunes, M. L. (2012a). 'Sleep instability and cognitive status in drug-resistant epilepsies', *Sleep Medicine*, **13**(5), pp. 536–541.

Pereira, A. M., Bruni, O., Ferri, R., Palmini, A., and Nunes, M. L. (2012b). 'The impact of epilepsy on sleep architecture during childhood', *Epilepsia*, **53**(9), pp. 1519–1525.

Phan, H., Andreotti, F., Cooray, N., Chén, O. Y., and Vos, M. D. (2019). 'SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **27**(3), pp. 400–410.

Phan, H., Chen, O. Y., Koch, P., Lu, Z., McLoughlin, I., Mertins, A., and Vos, M. D. (2020). 'Towards more accurate automatic sleep staging via deep transfer learning', *IEEE Transactions on Biomedical Engineering*, **68**(6), pp. 1787 – 1798.

Pilcher, J. J., and Huffcutt, A. I. (1996). 'Effects of sleep deprivation on performance: a meta-analysis', *Sleep*, **19**(4), pp. 318–326.

Porta, A., and Faes, L. (2016). 'Wiener–Granger causality in network physiology with applications to cardiovascular control and neuroscience', *Proceedings of the IEEE*, **104**(2), pp. 282–309.

Porta, A., Bassani, T., Bari, V., Pinna, G. D., Maestri, R., and Guzzetti, S. (2012). 'Accounting for respiration is necessary to reliably infer Granger causality from cardiovascular variability series', *IEEE Transactions on Biomedical Engineering*, **59**(3), pp. 832–841.

Prendergast, K. B., Schofield, G. M., and Mackay, L. M. (2016). 'Associations between lifestyle behaviours and optimal wellbeing in a diverse sample of New Zealand adults', *BMC Public Health*, **16**(1), 62.

Priano, L., Grugni, G., Miscio, G., Guastamacchia, G., Toffolet, L., Sartorio, A., and Mauro, A. (2006). 'Sleep cycling alternating pattern (CAP) expression is associated with hypersomnia and GH secretory pattern in Prader–Willi syndrome', *Sleep Medicine*, **7**(8), pp. 627–633.

Principe, J. C., and Tome, A. M. P. (1989). 'Performance and training strategies in feedforward neural networks: an application to sleep scoring', *International 1989 Joint Conference on Neural Networks*, pp. 341–346.

Proekt, A. (2018). 'Chapter fifteen - Brief introduction to electroencephalography', *in* R. G. Eckenhoff., and I. Dmochowski. (eds.), *Chemical and Biochemical Approaches for the Study of Anesthetic Function Part B*, Vol. 603, Academic Press, pp. 257–277.

Punjabi, N. M., Caffo, B. S., Goodwin, J. L., Gottlieb, D. J., Newman, A. B., O'Connor, G. T., Rapoport, D. M., Redline, S., Resnick, H. E., Robbins, J. A., Shahar, E., Unruh, M. L., and Samet, J. M. (2009). 'Sleep-Disordered breathing and mortality: a prospective cohort study', *PLOS Medicine*, **6**(8), e1000132.

Qureshi, S. S., Qureshi, M. A., and Zohra, R. R. (2014). 'The relationship between sleep and cognitive functioning in adult people.', *Pakistan Journal of Pharmaceutical Sciences*, **27**(6), pp. 2153–2156.

Ray, S. R., Lee, W. D., Morgan, C. D., and Airth-Kindree, W. (1986). 'Computer sleep stage scoring - an expert system approach', *International Journal of Bio-Medical Computing*, **19**(1), pp. 43–61.

Rechtschaffen, A., and Kales, A. (1968). '*A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects*', Brain information service/brain research institute, Los Angeles.

Redline, S., Amin, R., Beebe, D., Chervin, R. D., Garetz, S. L., Giordani, B., Marcus, C. L., Moore, R. H., Rosen, C. L., Arens, R., Gozal, D., Katz, E. S., Mitchell, R. B., Muzumdar, H., Taylor, H. G., Thomas, N., and Ellenberg, S. (2011). 'The Childhood Adenotonsillectomy Trial (CHAT): Rationale, design, and challenges of a randomized controlled trial evaluating a standard surgical procedure in a pediatric population', *Sleep*, **34**(11), pp. 1509–1517.

Redline, S., Budhiraja, R., Kapur, V., Marcus, C. L., Mateika, J. H., Mehra, R., Parthasarthy, S., Somers, V. K., Strohl, K. P., Sulit, L. G., Gozal, D., Wise, M. S., and Quan, S. F. (2007). 'The scoring of respiratory events in sleep: reliability and validity', *Journal of Clinical Sleep Medicine*, **03**(02), pp. 169–200.

Redline, S., Kirchner, H. L., Quan, S. F., Gottlieb, D. J., Kapur, V., and Newman, A. (2004). 'The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture', *JAMA Internal Medicine*, **164**(4), pp. 406–418.

Redline, S., Sanders, M. H., Lind, B. K., Quan, S. F., Iber, C., Gottlieb, D. J., Bonekat, W. H., Rapoport, D. M., Smith, P. L., and Kiley, J. P. (1998). 'Methods for obtaining and analyzing

unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group.', *Sleep*, **21**(7), pp. 759–767.

Reinhard, W., Plappert, N., Zeman, F., Hengstenberg, C., Riegger, G., Novack, V., Maimon, N., Pfeifer, M., and Arzt, M. (2013). 'Prognostic impact of sleep duration and sleep efficiency on mortality in patients with chronic heart failure', *Sleep Medicine*, **14**(6), pp. 502–509.

Roberts, V. C. (1982). 'Photoplethysmography - fundamental aspects of the optical properties of blood in motion', *Transactions of the Institute of Measurement and Control*, **4**(2), pp. 101–106.

Rosa, A. C., and Allen, L. J. (1996). 'Fuzzy classification of microstructural dynamics of human sleep', *1996 IEEE International Conference on Systems, Man and Cybernetics. Information Intelligence and Systems (Cat. No.96CH35929)*, Vol. 2, pp. 1108–1113.

Rosa, A. C., Parrino, L., and Terzano, M. G. (1999). 'Automatic detection of cyclic alternating pattern (CAP) sequences in sleep: preliminary results', *Clinical Neurophysiology*, **110**(4), pp. 585–592.

Rosenberg, R. S., and Van Hout, S. (2021). 'The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring', *Journal of Clinical Sleep Medicine*, **9**(1), pp. 81–87.

Schaltenbrand, N., Lengelle, R., and Macher, J.-P. (1993). 'Neural network model: application to automatic analysis of human sleep', *Computers and Biomedical Research*, **26**(2), pp. 157–171.

Scherzinger, A., Klemm, S., Berh, D., and Jiang, X. (2017). 'CNN-based background subtraction for long-term in-vial FIM imaging', *in* M. Felsberg., A. Heyden., and N. Krüger. (eds.), *International Conference on Computer Analysis of Images and Patterns*, Springer International Publishing, Cham, pp. 359–371.

Schiatti, L., Nollo, G., Rossato, G., and Faes, L. (2015). 'Extended Granger causality: a new tool to identify the structure of physiological networks', *Physiological Measurement*, **36**(4), pp. 827–843.

Schiecke, K., Schumann, A., Benninger, F., Feucht, M., Baer, K.-J., and Schlattmann, P. (2019). 'Brain–heart interactions considering complex physiological data: processing

schemes for time-variant, frequency-dependent, topographical and statistical examination of directed interactions by convergent cross mapping', *Physiological Measurement*, **40**(11), 114001.

Scholle, S., and Zwacka, G. (2001). 'Arousals and obstructive sleep apnea syndrome in children', *Clinical Neurophysiology*, **112**(6), pp. 984–991.

Seo, H., Back, S., Lee, S., Park, D., Kim, T., and Lee, K. (2020). 'Intra- and inter-epoch temporal context network (IITNet) using sub-epoch features for automatic sleep scoring on raw single-channel EEG', *Biomedical Signal Processing and Control*, **61**, 102037.

Sforza, E., Jouny, C., and Ibanez, V. (2000). 'Cardiac activation during arousal in humans: further evidence for hierarchy in the arousal response', *Clinical Neurophysiology*, **111**(9), pp. 1611–1619.

Shahrbabaki, S. S., Linz, D., Hartmann, S., Redline, S., and Baumert, M. (2021). 'Sleep arousal burden is associated with long-term all-cause and cardiovascular mortality in 8001 community-dwelling older men and women', *European Heart Journal*, **42**(21), pp. 2088–2099.

Siegel, S. (1956). '*Nonparametric statistics for the behavioral sciences.*', McGraw-Hill, New York, NY, US.

Silvani, A., and Dampney, R. A. L. (2013). 'Central control of cardiovascular function during sleep', *American Journal of Physiology-Heart and Circulatory Physiology*, **305**(12), pp. H1683–H1692.

Silvani, A., Calandra-Buonaura, G., Benarroch, E. E., Dampney, R. A. L., and Cortelli, P. (2015). 'Bidirectional interactions between the baroreceptor reflex and arousal: an update', *Sleep Medicine*, **16**(2), pp. 210–216.

Silvani, A., Grimaldi, D., Vandi, S., Barletta, G., Vetrugno, R., Provini, F., Pierangeli, G., Berteotti, C., Montagna, P., Zoccoli, G., and Cortelli, P. (2008). 'Sleep-dependent changes in the coupling between heart period and blood pressure in human subjects', *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, **294**(5), pp. R1686–R1692.

Simonds, J. F., and Parraga, H. (1984). 'Sleep behaviors and disorders in children and adolescents evaluated at psychiatric clinics', *Journal of Developmental & Behavioral Pediatrics*, **5**(1), pp. 6–10.

Singh, A., Thakur, N., and Sharma, A. (2016). 'A review of supervised machine learning algorithms', *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1310–1315.

Smerieri, A., Parrino, L., Agosti, M., Ferri, R., and Terzano, M. G. (2007). 'Cyclic alternating pattern sequences and non-cyclic alternating pattern periods in human sleep', *Clinical Neurophysiology*, **118**(10), pp. 2305–2313.

Smurra, M. V., Dury, M., Aubert, G., Rodenstein, D. O., and Liistro, G. (2001). 'Sleep fragmentation: comparison of two definitions of short arousals during sleep in OSAS patients', *European Respiratory Journal*, **17**(4), pp. 723 – 727.

Sokolova, M., and Lapalme, G. (2009). 'A systematic analysis of performance measures for classification tasks', *Information Processing & Management*, **45**(4), pp. 427–437.

Somers, V. K., Dyken, M. E., Mark, A. L., and Abboud, F. M. (1993). 'Sympathetic-nerve activity during sleep in normal subjects', *New England Journal of Medicine*, **328**(5), pp. 303–307.

Song, Y., Blackwell, T., Yaffe, K., Ancoli-Israel, S., Redline, S., and Stone, K. L. (2015). 'Relationships between sleep stages and changes in cognitive function in older men: the MrOS Sleep Study', *Sleep*, **38**(3), pp. 411–421.

Soster, L. A., Alves, R. C., Fagundes, S. N., Lebl, A., Garzon, E., Koch, V. H., Ferri, R., and Bruni, O. (2017). 'Non-REM sleep instability in children with primary monosymptomatic sleep enuresis', *Journal of Cinical Sleep Medicine*, **13**(10), pp. 1163–1170.

Spira, A. P., Blackwell, T., Stone, K. L., Redline, S., Cauley, J. A., Ancoli-Israel, S., and Yaffe, K. (2008). 'Sleep-disordered breathing and cognition in older women', *Journal of the American Geriatrics Society*, **56**(1), pp. 45–50.

Stanus, E., Lacroix, B., Kerkhofs, M., and Mendlewicz, J. (1987). 'Automated sleep scoring: a comparative reliability study of two algorithms', *Electroencephalography and Clinical Neurophysiology*, **66**(4), pp. 448–456.

Stephansen, J. B., Olesen, A. N., Olsen, M., Ambati, A., Leary, E. B., Moore, H. E., Carrillo, O., Lin, L., Han, F., Yan, H., Sun, Y. L., Dauvilliers, Y., Scholz, S., Barateau, L., Hogl, B., Stefani, A., Hong, S. C., Kim, T. W., Pizza, F., Plazzi, G., Vandi, S., Antelmi, E., Perrin, D., Kuna, S. T., Schweitzer, P. K., Kushida, C., Peppard, P. E., Sorensen, H. B. D., Jennum,

P., and Mignot, E. (2018). 'Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy', *Nature Communications*, **9**(1), 5229.

Stickgold, R., and Walker, M. P. (2005). 'Memory consolidation and reconsolidation: what is the role of sleep?', *Trends in Neurosciences*, **28**(8), pp. 408–415.

Suen, J. S., Arnold, J. E., and Brooks, L. J. (1995). 'Adenotonsillectomy for treatment of obstructive sleep apnea in children', *Archives of Otolaryngology–Head & Neck Surgery*, **121**(5), pp. 525–530.

Sundararajan, M., Taly, A., and Yan, Q. (2017). 'Axiomatic attribution for deep networks', *International Conference on Machine Learning*, PMLR, pp. 3319–3328.

Supratak, A., Dong, H., Wu, C., and Guo, Y. (2017). 'DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, **25**(11), pp. 1998–2008.

Svetnik, V., Ferri, R., Ray, S., Ma, J., Walsh, J. K., Snyder, E., Ebert, B., and Deacon, S. (2010). 'Alterations in cyclic alternating pattern associated with phase advanced sleep are differentially modulated by gaboxadol and zolpidem', *Sleep*, **33**(11), pp. 1562–1570.

Tamura, T., Maeda, Y., Sekine, M., and Yoshida, M. (2014). 'Wearable photoplethysmographic sensors—past and present', *Electronics*, **3**(2).

Teixeira, C. A., Direito, B., Bandarabadi, M., Quyen, M. L. V., Valderrama, M., Schelter, B., Schulze-Bonhage, A., Navarro, V., Sales, F., and Dourado, A. (2014). 'Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients', *Computer Methods and Programs in Biomedicine*, **114**(3), pp. 324–336.

Terzano, M. G., and Parrino, L. (1993). 'Clinical applications of cyclic alternating pattern', *Physiology & Behavior*, **54**(4), pp. 807–813.

Terzano, M. G., and Parrino, L. (2000). 'Origin and significance of the cyclic alternating pattern (CAP)', *Sleep Medicine Reviews*, **4**(1), pp. 101–123.

Terzano, M. G., and Parrino, L. (2005). 'Chapter 8 The cyclic alternating pattern (CAP) in human sleep', *in* C. Guilleminault. (ed.), *Handbook of Clinical Neurophysiology*, Vol. 6, Elsevier, pp. 79–93.

Terzano, M. G., Mancia, D., Salati, M. R., Costani, G., Decembrino, A., and Parrino, L. (1985). 'The cyclic alternating pattern as a physiologic component of normal NREM sleep', *Sleep*, **8**(2), pp. 137–145.

Terzano, M. G., Parrino, L., and Spaggiari, M. C. (1988). 'The cyclic alternating pattern sequences in the dynamic organization of sleep', *Electroencephalography and Clinical Neurophysiology*, **69**(5), pp. 437–447.

Terzano, M. G., Parrino, L., Boselli, M., Smerieri, A., and Spaggiari, M. C. (2000). 'CAP components and EEG synchronization in the first 3 sleep cycles', *Clinical Neurophysiology*, **111**(2), pp. 283–290.

Terzano, M. G., Parrino, L., Boselli, M., Spaggiari, M. C., and Di Giovanni, G. (1996). 'Polysomnographic analysis of arousal responses in obstructive sleep apnea syndrome by means of the cyclic alternating pattern.', *Journal of Clinical Neurophysiology*, **13**, pp. 145 – 155.

Terzano, M. G., Parrino, L., Fioriti, G., Orofiamma, B., and Depoortere, H. (1990). 'Modifications of sleep structure induced by increasing levels of acoustic perturbation in normal subjects', *Electroencephalography and Clinical Neurophysiology*, **76**(1), pp. 29–38.

Terzano, M. G., Parrino, L., Rosa, A., Palomba, V., and Smerieri, A. (2002). 'CAP and arousals in the structural development of sleep: an integrative perspective', *Sleep Medicine*, **3**(3), pp. 221–229.

Terzano, M. G., Parrino, L., Sherieri, A., Chervin, R., Chokroverty, S., Guilleminault, C., Hirshkowitz, M., Mahowald, M., Moldofsky, H., Rosa, A., Thomas, R., and Walters, A. (2001). 'Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep.', *Sleep Medicine*, **2**(6), pp. 537–553.

Terzano, M. G., Parrino, L., Spaggiari, M. C., Palomba, V., Rossi, M., and Smerieri, A. (2003). 'CAP variables and arousals as sleep electroencephalogram markers for primary insomnia', *Clinical Neurophysiology*, **114**(9), pp. 1715–1723.

Terzano, M. G., Smerieri, A., Felice, A. D., Giglia, F., Palomba, V., and Parrino, L. (2006). 'Cyclic alternating pattern (CAP) alterations in narcolepsy', *Sleep Medicine*, **7**(8), pp. 619–626.

Thomas, R. J., Mietus, J. E., Peng, C.-K., and Goldberger, A. L. (2005). 'An electrocardiogram-based technique to assess cardiopulmonary coupling during sleep', *Sleep*, **28**(9), pp. 1151–1161.

Thompson, N. C., Greenewald, K., Lee, K., and Manso, G. F. (2020). 'The computational limits of deep learning', *arXiv preprint arXiv:2007.05558*, .

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). 'Mediation: R package for causal mediation analysis', *Journal of Statistical Software*, **59**(5).

Trinder, J., Allen, N., Kleiman, J., Kralevski, V., Taylor, D., Anson, K., and Kim, Y. (2003). 'On the nature of cardiovascular activation at an arousal from sleep', *Sleep*, **26**, pp. 543–551.

Trinder, J., Kleiman, J., Carrington, M., Smith, S., Breen, S., Tan, N., and Kim, Y. (2001a). 'Autonomic activity during human sleep as a function of time and sleep stage', *Journal of Sleep Research*, **10**(4), pp. 253–264.

Trinder, J., Padula, M., Berlowitz, D., Kleiman, J., Breen, S., Rochford, P., Worsnop, C., Thompson, B., and Pierce, R. (2001b). 'Cardiac and respiratory activity at arousal from sleep under controlled ventilation conditions', *Journal of Applied Physiology*, **90**(4), pp. 1455–1463.

Trinder, J., Waloszek, J., Woods, M. J., and Jordan, A. S. (2012). 'Sleep and cardiovascular regulation', *Pflügers Archiv - European Journal of Physiology*, **463**(1), pp. 161–168.

Tsinalis, O., Matthews, P. M., Guo, Y., and Zafeiriou, S. (2016). 'Automatic sleep stage scoring with single-channel EEG using convolutional neural networks', *arXiv preprint arXiv:1610.01683*, .

Ulate-Campos, A., Coughlin, F., Gaínza-Lein, M., Fernández, I. S., Pearl, P. L., and Loddenkemper, T. (2016). 'Automated seizure detection systems and their effectiveness for each type of seizure', *Seizure*, **40**, pp. 88–101.

Van Dongen, H. P. A., Maislin, G., Mullington, J. M., and Dinges, D. F. (2003). 'The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation', *Sleep*, **26**(2), pp. 117–126.

Verrillo, E., Bruni, O., Franco, P., Ferri, R., Thiriez, G., Pavone, M., Petrone, A., Paglietti, M. G., Crinò, A., and Cutrera, R. (2009). 'Analysis of NREM sleep in children with

Prader–Willi syndrome and the effect of growth hormone treatment', *Sleep Medicine*, **10**(6), pp. 646–650.

Vilamala, A., Madsen, K. H., and Hansen, L. K. (2017). 'Deep convolutional neural networks for interpretable analysis of EEG sleep stage scoring', *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*, 10.1109/MLSP.2017.8168133.

Wang, C., Bangdiwala, S. I., Rangarajan, S., Lear, S. A., AlHabib, K. F., Mohan, V., Teo, K., Poirier, P., TSE, L. A., Liu, Z., Rosengren, A., Kumar, R., Lopez-Jaramillo, P., Yusoff, K., Monsef, N., Krishnapillai, V., Ismail, N., Seron, P., Dans, A. L., Kruger, L., Yeates, K., Leach, L., Yusuf, R., Orlandini, A., Wolyniec, M., Bahonar, A., Mohan, I., Khatib, R., Temizhan, A., Li, W., and Yusuf, S. (2019). 'Association of estimated sleep duration and naps with mortality and cardiovascular events: a study of 116 632 people from 21 countries', *European Heart Journal*, **40**(20), pp. 1620–1629.

Wei, X., Zhou, L., Chen, Z., Zhang, L., and Zhou, Y. (2018). 'Automatic seizure detection using three-dimensional CNN based on multi-channel EEG', *BMC Medical Informatics and Decision Making*, **18**(5), 111.

Winterhalder, M., Schelter, B., Hesse, W., Schwab, K., Leistritz, L., Klan, D., Bauer, R., Timmer, J., and Witte, H. (2005). 'Comparison of linear signal processing techniques to infer directed interactions in multivariate neural systems', *Signal Processing*, **85**(11), pp. 2137–2160.

Witting, W., van der Werf, D., and Mirmiran, M. (1996). 'An on-line automated sleep-wake classification system for laboratory animals', *Journal of Neuroscience Methods*, **66**(2), pp. 109–112.

Wolf, G. K., and Arnold, J. H. (2005). 'Noninvasive assessment of lung volume: Respiratory inductance plethysmography and electrical impedance tomography', *Critical Care Medicine*, **33**(3), pp. S163–S169.

Worrell, G., and Gotman, J. (2011). 'High-frequency oscillations and other electrophysiological biomarkers of epilepsy: clinical studies', *Biomarkers in Medicine*, **5**(5), pp. 557–566.

Xie, L., Kang, H., Xu, Q., Chen, M. J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D. J., Nicholson, C., Iliff, J. J., Takano, T., Deane, R., and Nedergaard, M. (2013). 'Sleep drives metabolite clearance from the adult brain', *Science*, **342**(6156), pp. 373–377.

Yaggi, H. K., Araujo, A. B., and McKinlay, J. B. (2006). 'Sleep duration as a risk factor for the development of type 2 diabetes', *Diabetes Care*, **29**(3), pp. 657–661.

Yalamanchali, S., Farajian, V., Hamilton, C., Pott, T. R., Samuelson, C. G., and Friedman, M. (2013). 'Diagnosis of obstructive sleep apnea by peripheral arterial tonometry: meta-analysis', *JAMA Otolaryngology–Head & Neck Surgery*, **139**(12), pp. 1343–1350.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). 'Large-scale automated synthesis of human functional neuroimaging data', *Nature Methods*, **8**, pp. 665–670.

Zhang, G.-Q., Cui, L., Mueller, R., Tao, S., Kim, M., Rueschman, M., Mariani, S., Mobley, D., and Redline, S. (2018). 'The National Sleep Research Resource: towards a sleep data commons', *Journal of the American Medical Informatics Association*, **25**(10), pp. 1351–1358.

Zielinski, M. R., McKenna, J. T., and McCarley, R. W. (2016). 'Functions and mechanisms of sleep', *AIMS Neuroscience*, **3**(1), pp. 67–104.

Zucconi, M., Oldani, A., Smirne, S., and Ferini-Strambi, L. (2000). 'The macrostructure and microstructure of sleep in patients with autosomal dominant nocturnal frontal lobe epilepsy', *Journal of Clinical Neurophysiology*, **17**(1), pp. 77–86.

Acronyms

AASM	American Academy of Sleep Medicine
AI	artificial intelligence
ANN	artificial neural network
ANS	autonomous nervous system
AT	adenotonsillectomy
BP	blood pressure
CAP	cyclic alternating pattern
CFA	cardiac field artefact
CHAT	Childhood Adenotonsillectomy Trial
CNN	convolutional neural network
CNS	central nervous system
eAT	early adenotonsillectomy
ECG	electrocardiography
EEG	electroencephalography
EMG	electromyography
EOG	electrooculogram
GC	Granger causality
HR	heart rate
HRV	heart rate variability
kNN	k-nearest neighbour
LDA	linear discriminant analysis
LSTM	long-short term memory
MrOS	Osteoporotic Fractures of Men
NFLE	nocturnal frontal lobe epilepsy
NREM	non-rapid eye-movement
NSRR	National Sleep Research Resource

OSA	obstructive sleep apnea
PAT	pulse arrival time
PPG	photoplethysmogram
PSG	polysomnography
PTT	pulse transit time
PWA	pulse wave amplitude
PWV	pulse wave velocity
REM	rapid eye movement
RIP	respiratory inductance plethysmography
RLS	restless limb syndrome
RNN	recurrent neural network
RSD	restless sleep disorder
SHHS	Sleep Heart Health Study
SOF	Study of Osteoporotic Fractures
SVM	support vector machine
SWS	slow wave sleep
WWSC	watchful waiting with supportive care

Biography

Simon Hartmann graduated with a B.Sc. degree in Electrical and Electronic Engineering from the University of Ulm, Germany, in 2014. For his Bachelor thesis, he developed a microcontroller-controlled measurement setup to investigate inductive power transmission as power supply for a retina prosthesis. In 2017, he received the M.Sc. degree in Electrical and Electronic Engineering from the Karlsruhe Institute of Technology, Germany. For his Master thesis, he developed a chest-worn sensor-patch to enable convenient screening for sleep-disordered breathing. Following his master studies, he started his Ph.D. degree with the School of Electrical and Electronic Engineering at the University of Adelaide, Australia, under the supervision of M. Baumert, and D. Abbott. Simon's personal interests and skills lie at the exciting intersection of advanced signal processing and human physiology. He is passionate about developing tools to enhance the understanding of the role of the human brain.

He was granted the GOstralia! PhD scholarship from The University of Adelaide, Australia, to study toward his Ph.D. degree in 2017. He was also a member of an international collaboration that received a Lab2lab grant by the Graduate Academy of Technical University Dresden, Germany. Furthermore, he has served as a reviewer for a number of recognised journals and conferences including IEEE Access, Sleep, and Biomedical Signal Processing and Controls. He is currently a NHMRC Grant Funded Researcher at the School of Psychiatry, The University of Adelaide, Australia, working on multi-modal risk prediction of early psychosis onset as part of the Centre of Research Excellence (CRE) in PREdiction of Early Mental Disorder and Preventive Treatment (PRE-EMPT).