SLEEP AND DEATH:
THE RELATIONSHIP BETWEEN REM SLEEP AND MORTALITY

A DISSERTATION
SUBMITTED TO THE DEPARTMENT
OF EPIDEMIOLOGY AND POPULATION HEALTH
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

EILEEN B. LEARY, MS, RPSGT

MAY, 2020

This dissertation is online at: http://purl.stanford.edu/bn040xk9167

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Steven Goodman, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Emmanuel Mignot

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

James Zou

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

Sleep is a non-negotiable requirement for a happy, healthy life. In the last 70 years, our understanding of sleep has grown exponentially. However, in our busy society, sleep is often overlooked and undervalued. This is surprising given that sleep disorders and sleep dysregulation have been linked to multiple systemic and brain-based diseases, including cardiovascular disease, type 2 diabetes, dementia, and major depressive disorder. Additionally, sleep disorders and sleep characteristics (e.g. sleep duration) have been linked to higher rates of mortality. Despite the emerging evidence of a sleep-mortality association, the mechanisms underlying the relationship are not well understood. Little is known about how the proportion of time spent in each sleep stage relate to timing or cause of death.

This dissertation is an in-depth investigation of the relationship between rapid eye movement (REM) sleep and risk of mortality. Specific aim one combines traditional and machine learning analytic approaches to evaluate whether lower levels of REM sleep would be associated with an increased rate of mortality.

Sleep is comprised of multiple sleep stages that by nature are highly correlated. Therefore, it is necessary to tease apart whether another sleep stage could be a better predictor of mortality. Aim two used supervised machine learning to rank the four sleep stages from most to least predictive in context of one another.

The hypotheses were increased mortality rates would be associated with lower quantities of REM sleep and that compared to other sleep stages, REM would be the best predictor of mortality.

Specific aim three was to evaluate the validity, consistency, and generalizability of the findings. to do this, the final models were validated in two independent cohorts and the results from all three cohorts were combined in a meta-analysis.

Materials and Methods

Three longitudinal, population-based cohorts were used in this project. The Osteoporotic Fractures in Men (MrOS) sample included 2,675 older men (mean age 76.3 years ± 5.5 years) recruited from 2003 to 2005 and followed for a median of 12.1 years. The Wisconsin Sleep Cohort (WSC) started in 1988 and followed 1,386 participants (45.7% women, mean age 51.5 years ± 8.5) for a median of 20.8 years. The Sleep Heart Health Study (SHHS) was comprised of 5,550 participants (52.4% women, mean age 63.0 years ± 11.2) recruited between 1995 and 1998 and monitored for a median of 11.9 years.

The exposure was percent of total sleep time spent in REM sleep and was evaluated at baseline using polysomnography. The main outcomes included all-cause and cause-specific (cardiovascular, cancer, other) mortality confirmed with death certificates.

Cox proportional hazards regression models were used to evaluate the association between percent REM and mortality. The first model contained a core set of covariates selected a priori based on existing literature and clinical experience. Additional covariates commonly associated with sleep architecture were evaluated using 6-fold cross-validation with a forward step-wise feature selection algorithm to obtain the best candidates for the final multivariate regression models. A threshold effect was suspected based on Kaplan-Meier curves, so separate models were run with percent REM as a binary variable using 15% as the cut-point.

Conditional inference survival tree and random survival forest analyses were conducted to identify which sleep stage(s) were driving the significance of the finding and to evaluate relevant cut-points. Several sensitivity analyses were completed to rule out alternative explanations for the findings. The findings were replicated using data from the Wisconsin Sleep Cohort (WSC) and Sleep Heart Health Study (SHHS). A meta-analysis pooled and weighted the results from all three studies to provide a global quantification of the hazard ratio.

Results

MrOS participants had a 13% higher mortality rate for every 5% reduction in REM sleep (percent REM standard deviation = 6.6%) after adjusting for multiple demographic, sleep, and health covariates including study site, age at sleep visit, race, education, medication use, smoking status, caffeine intake, respiratory disturbance index, and actigraphy measures (age adjusted hazard ratio [HR] = 1.12, fully adjusted HR = 1.13, 95% CI, 1.08–1.19). The association was also present for cardiovascular disease-related mortality (CVD) (HR = 1.18, 95% CI, 1.09–1.28), cancer related mortality (HR = 1.14, 95% CI, 1.03–1.26), and non-cardiovascular, non-cancer related mortality (HR = 1.19, 95% CI, 1.10–1.28). Individuals with <15% REM had a higher mortality rate relative to individuals with ≥15% for each mortality outcome with odds ratios ranging from 1.20 to 1.35. The random forest model identified REM as the most important sleep stage for predicting survival.

In the WSC, the effect size for 5% reduction in REM on risk of all-cause mortality was similar despite the younger age, inclusion of women, and longer follow-up period (HR = 1.17, 95% CI, 1.03–1.34). When stratified by gender, lower percent REM was associated with all-cause mortality in women (HR = 1.34, 95% CI, 1.07–1.68) but was not statistically significant in men

(HR = 1.09, 95% CI, 0.92–1.30). In the SHHS results were consistent with the other cohorts with a 13% increase in all-cause mortality rate for every 5% reduction in REM (HR = 1.13, 95% CI, 1.07–1.18) and a 7% increase in cardiovascular mortality rate (HR = 1.07, 95% CI, 0.97–1.17). Unlike WSC, when stratified by gender, the hazard ratio was higher in men (HR = 1.16, 95% CI, 1.08–1.17) than women (HR = 1.09, 95% CI, 1.02–1.17). Meta-analysis of the three cohorts yielded an overall hazard ratio of 1.13 (95% CI 1.10–1.17) for all-cause mortality and 1.10 (95% CI 1.03–1.16) for cardiovascular mortality.

Conclusion and Relevance

There was a robust association between lower levels of REM sleep and mortality in three independent cohorts, which persisted across different causes of death and multiple sensitivity analyses. Given the complex underlying biological functions, further studies are required to understand whether the relationship is causal. Accelerated brain aging may result in reduced REM sleep, making it a disease, frailty or biologic aging marker rather than a direct mortality risk factor. Mechanistic studies are needed and strategies to preserve REM may influence clinical therapies and reduce mortality risk, particularly for adults with <15% REM.

Preface

Dreams have always fascinated me; it blows my mind that every night we slip into an alternate world where the rules of space and time don't apply. I was thrilled when I discovered at the age of 16 that I could study sleep for a living. I started my career as a sleep technologist to gain clinical insight and practical, hands-on experience at the patient level. I spent my nights interpreting squiggly lines and making sense of rich physiologic waveforms such as EEG and breathing patterns. I would get to know each patient and wondered why sleep patterns varied so much one person to another. Could high REM density be a marker of higher IQ? Could we predict who would live longer based on their sleep architecture? After more than two decades working in sleep medicine, I have finally explored some of the questions I used to contemplate while working nights.

I want to thank the many colleagues and mentors who provided advice and support on this project and throughout my career: Emmanuel Mignot, Steven Goodman, Katie Stone, Kathleen Watson, Sonia Ancoli-Israel, Susan Redline, Kristine Yaffe, Terry Blackwell, Laurel Ravelo, Paul Peppard, James Zou, Michelle Odden, Victor Henderson, Rita Popat, William C. Dement, James Walsh, Kimberly Trotter, and countless others.

And of course I would like to thank my friends and family for their unending support and for never complaining (to me) even when I bored them to tears by talking endlessly about this project. Thank you Travis Neal, Margaret Wylder, Arthur Leary, Alice McLaughlin, Cate Leary, Rita Leary, Carly Message, Anna Castorena, Jules Saskin, and many others. Travis deserves a special thanks AND a medal for his unfailing support and keeping me sane during this process.

This effort is dedicated to my parents, David and Jean Leary, and my late husband Steven H. Sykes Jr. Words cannot express how much I miss them every day.

Eileen B. Leary
Stanford University
May 2020

Table of Contents

List of Tables

List of Figures

Introduction

History of the Sleep Field

Humans spend nearly a third of their lives asleep and sleep is fundamental to animal life, yet the science of sleep is a relatively new discipline and the function of sleep remains largely unknown[1]. The first textbook on sleep was published in 1913 when the French scientist W. H. Howell reported on the physiology of sleep in "Le probleme physiologique du sommeil[2]." Shortly after, Dr. Nathanial Kleitman began studying the regulation of sleep and wake at the University of Chicago in the 1920s. At that time, most people assumed that sleep was a passive activity where both the body and brain were dormant.

That belief changed in the 1950's when Eugene Aserinsky, Kleitman's student, identified a marked increase in frequency and speed of eye movements during periods in the night. Today that activity is known as rapid eye movements (REM). William C. Dement (see Figure 1a), another student of Kleitman, was the first to make the connection between REM and dreaming. Throughout the 1950s, Dement studied and described sleep patterns in human beings and documented the occurrence of REM sleep in newborn babies and cats (one of the first animal models in sleep)[3]. Michel Jouvet recognized that the brainwaves during REM sleep were similar to wake; classifying REM as an independent state of alertness he termed "paradoxical sleep." In the early 1960's Jouvet mapped out the brain structures that generate REM at the University Claude Bernard in Lyon[4].

In 1963, Dement left Chicago and joined the Psychiatry Department at Stanford University where he continued to develop and publicize the new field of sleep science. He founded the Stanford Narcolepsy Clinic to study and provide clinical care to individuals with narcolepsy, a sleep disordered marked by disordered REM sleep that we now know affects 0.05% of the population in the US[5]. Together with Dr. Christian Guilleminault, Dement founded the first sleep disorders clinic at Stanford in the 1970s and used all-night polysomnography (PSG) to objectively examine and treat patients with sleep-related complaints. Numerous clinical advances in sleep have taken place at Stanford University to support patient care. Dr. Mary Carskadon invented the Multiple Sleep Latency Test to formally diagnose narcolepsy and idiopathic hypersomnia[6]. Dr. Rachel Manber launched of the first group therapy program for insomnia in 1999.

In addition, many groundbreaking scientific advances and discoveries took place at Stanford, many by Dr. Emmanuel Mignot (see Figure 1b) whose research focuses on the neurobiology, genetics, and immunology of narcolepsy. Mignot used positional cloning to identify an autosomal

recessive mutation responsible for narcolepsy (hypocretin receptor 2) in the canine model[7]. He also discovered that narcolepsy in humans is caused by an immune-mediated destruction of hypocretin/orexin neurons in the hypothalamus[8-10]. In addition, he isolated the gene causing DNMT1, the methylopathy Autosomal Dominant Cerebellar Ataxia, Deafness and Narcolepsy (ADCA-DN)[11] and identified a genetic marker for sleep apnea, which is also associated with Alzheimer's and cardiovascular diseases[12].

Overview of Sleep Science and Medicine

Today, sleep is a rapidly growing, international field that is separated into two main disciplines: sleep science and sleep medicine. The science of sleep can be broadly divided into basic and clinical research. Basic research includes the study of genetics, proteomics, immunology, neurobiology, neurophysiology, pharmacology, animal models (from drosophila and zebra fish to primates), and countless other sub-specialties. Clinical research includes observational studies that systematically collect data on the causes, consequences, and natural progression of sleep disorders. It also covers clinical trials used to gauge the effectiveness, tolerability, and safety of medications or medical devices in the treatment of sleep disorders.

Sleep medicine encompasses the diagnosis and treatment of sleep disorders. There are more than 800 different sleep disorders that are broken down into eight main categories: (a) insomnias, (b) sleep-related breathing disorders, (c) hypersomnias of central origin, (d) circadian rhythm disorders, (e) parasomnias, (f) sleep-related movement disorders, (g) isolated symptoms, apparently normal variants, and unresolved issues, (h) others[13]. Common sleep disorders include obstructive sleep apnea, insomnia, restless legs syndrome, shift work, and narcolepsy.

One major issue in the field of sleep medicine is the inability to effectively scale the sleep care model. Sleep disorders and sleep dysregulation are a global issue with over 100 million affected in the US alone[14]. Impaired sleep is a substantial contributor to contributing to multisystemic medical consequences[15] including cardiovascular disease (arrhythmia, hypertension, stroke)[16,17], metabolic disorders (diabetes, obesity)[18], psychiatric conditions (depression, addictive behaviors, suicide)[19,20], impaired cognition (Alzheimer's, dementia)[21,22], quality of life outcomes (daytime sleepiness, irritability, concentration)[23], and all-cause mortality[16]. The field of sleep medicine is simply too small and ill equipped to support the sheer volume of people suffering with sleep issues.

The current clinical model starts with a referral for a consultation appointment with a sleep clinician. Many clinics have a long waiting list for these visits. After collecting information about

the patient's clinical and sleep history, usually the physician will order an overnight sleep study, called a polysomnogram (PSG), to monitors brain waves, heart rate, leg movements, and breathing during sleep. For most patients, this results in yet another long wait, sometimes several months.

The night of their study, the patient arrives in the early evening and spends the next two to three hours completing paperwork and having wired sensors applied from head to toe by a trained polysomnographic technologist. Most in-lab PSGs collect between 16 and 25 channels of digital information throughout the night to acquire a comprehensive assessment of the individual's sleep. standard PSG channels includes: electroencephalogram (EEG), electrooculogram (EEG), electromyogram (EMG) of the submentalis muscle as well as the anterior tibialis muscle of each leg, electrocardiogram (ECG), snoring, breathing effort using inductance plethysmography summation systems, both oral and nasal airflow, as well oxygen hemoglobin. A 2-minute PSG recording is presented in Figure 2.

The night is divided into five discrete stages: wake, non-REM stage 1 (N1), stage 2 (N2) stage 3 (N3), and REM. Currently, a trained technician manually assigns a sleep stage to every 30-second epoch of the 8-hour study through visual inspection. The stage is selected based on a series of rules originally published in by Rechtschaffen & Kales in 1968[24]. Updated consensus rules are regularly published by the American Academy of Sleep Medicine[8,9]. It takes about 2 hours for a technologist to score an 8-hour recording.

This manual approach has multiple problems. It is time consuming, expensive, inconsistent, and subjective. In addition, the 30-second window (epoch) was chosen before digital recording, when sleep studies were collected using paper and ink. Thirty seconds of data was recorded on each page, so it was convenient when flipping through the paper records. However, biologically it does not make sense to assign a single stage for such a long period because it is extremely common to have two or more stages present in a single epoch. Therefore, the technician has to make a judgement call about which sleep stage represents the majority of the epoch. A study performed by Rosenberg and colleagues[20] found inter-scorer reliability for sleep stage scoring to be 82.6% on average, a result consistently found by other teams[21–24]. N1 and N3 are particularly problematic with agreements as low as 63% and 67%, which significantly reduce their clinical usefulness[20].

In the US, there are only 2,600 American Academy of Sleep Medicine accredited sleep disorders centers. These clinics are supported by 6,000 board-certified sleep medicine physicians[25] and 20,000 Registered Polysomnographic Technologists[26] and perform 1-2 million sleep studies each

year. This is simply not enough to provide adequate care to those in need. As a result, the field of sleep medicine is being forced to change to accommodate the growing need for care.

Advances in Sleep Medicine

The landscape of sleep medicine has been able to change rapidly due to the ongoing advancements in technology, machine learning, data storage, and wearable devices. While it may be tempting to replace old systems with innovative approaches, care must be taken to ensure critical knowledge is carried forward, particularly because sleep is a relatively new discipline. A more effective strategy is to combine clinical knowledge with advances in machine learning and statistical methods to carry the field forward while breaking the bottleneck of care created by traditional sleep center workflows.

The analysis of overnight PSGs is an obvious area that can be transformed with the thoughtful application of machine learning and artificial neural networks. PSGs are an incredibly rich phenotype that have been underutilized due to the size and complexity of the studies. Currently, eight hours of biological data are condensed into roughly a dozen summary variables, such as how long it took to fall asleep (sleep latency), total sleep time, and number of abnormal breathing events per hour of sleep (apnea hypopnea index). However, the historical limitations around studying PSGs are being eliminated as securely transferring and storing large files becomes cheaper and easier. Additionally, data is now being shared through mechanisms like the National Sleep Research Resource[27] (supported by the National Heart, Lung, and Blood Institute) that offers free access to de-identified PSGs and clinical data elements from multiple well-characterized research cohorts and clinical trials[28]. As these constraints related to access and computational power are reduced, automated scoring algorithms have become a viable solution to change how PSGs are analyzed. Automation will reduce cost and improve the reproducibility of the data used to make clinical decisions[29]. Additionally, a deep learning, quantitative analysis of sleep will lead to the development of new biomarkers and diagnostic criteria[30].

Over the last 30 years, there have been numerous attempts to create software to assist sleep scoring[31], but the resulting systems have not been well received by the sleep community because they are cumbersome and expensive with variable utility. The Academic Alliance for Sleep Research (AASR) conducted a study to investigate the inter-scorer variability of computer-assisted manual PSG scoring between and across sites. The AASR was a consortium of academic sleep centers established to perform multicenter research studies. For this study, ten experienced scoring techs from five institutions scored the same 70 studies using the American Academy of Sleep Medicine standards for scoring sleep stages, arousals, and apneas[32]. Hypopneas were scored

using three different criteria: recommended, alternate, and research (Chicago)[32]. The intraclass correlation coefficients (ICC) ranged considerably depending on the metric evaluated. The ICCs across sites ranged from 0.402 (95% CI, 0.352–0.452) for arousals during REM sleep to 0.870 (95% CI, 0.847–0.889) for total sleep time. Agreement across techs within-site was higher but had more variance with ICCs ranging from 0.557 (95% CI, 0.267–0.827) for arousals during REM sleep to 0.892 (95% CI, 0.779–0.980) for total sleep time[33]. What makes this project particularly noteworthy is that the same studies were analyzed using an automatic scoring software developed by Magdy Younes at YST Limited (Winnipeg, Canada). Overall, the automated scoring results were comparable to those obtained by experienced technologists. The ICC for automatic versus average manual scoring for arousals during REM sleep was 0.39 (95% CI, 0.02–0.63) and for total sleep time the ICC was 0.87 (95%, CI 0.79–0.91)[29].

It is not surprising that software created with the goal of mimicking how humans evaluate sleep studies has intrinsic limitations given the rules used by humans are fairly arbitrary. Mignot and colleagues[30] adopted a new strategy by transcending the current rules developed decades ago by a consensus of sleep experts. The team explored automated sleep staging from different perspectives by applying multiple deep learning algorithms to data from numerous independent cohorts, including data scored by multiple technologists. The best model performed better than any individual scorer did (Cohen's kappa of $87.0 \pm 4.5\%$ versus consensus) and surpassed the performance reported from the Younes software[29].

The Mignot model produces a hynodensity graph, which is a probability distribution of sleep stages across the night. This is novel measure is similar to the classic hypnogram (see Figure 2), but instead of assigning a single sleep stage to each epoch of sleep, the system provides a membership function to each of the sleep stages across the night. This approach breaks the dependency on a 30-second epoch to categorize sleep, which does not have a physiological basis. It also conveys more information about sleep trends than traditional human scoring. For example, there were fundamental differences in the hypnodensity graphs of narcolepsy patients compared to other individuals. Specifically, there was a greater than normal overlap between sleep stages. Type-1 narcolepsy is an auto-immune disease[34] characterized by sleepiness, cataplexy (episodes of muscle weakness triggered by emotions), poor nocturnal sleep (insomnia) and symptoms of dissociated REM sleep. This dissociation is demonstrated when REM sleep is intermingled with wakefulness, producing reports of dreams that interrupt wakefulness and seem real (hypnagogic hallucinations), or episodes where the sleeper is awake but paralyzed (sleep paralysis)[35-39]. The hypnodenisty graph made this dissociation and mixing of sleep stages very apparent. When

combined with PSG features identified from prior work to aid in the diagnosis of narcolepsy[40], the Mignot model achieved a sensitivity of 93% and a specificity of 91% in the replication sample.

Zhang and colleagues[41] also used deep neural networks to build a multi-model integration scoring strategy with multiple signal channels as input. The network's performance was comparable to humans with an average accuracy of 0.8181 and Cohen's Kappa of 0.7276. Their approach was less successful because the performance of the model was significantly impacted by arousals. This is a serious problem as arousals are a common phenomenon in sleep clinic patients.

Machine learning has also proved useful in the characterization of sleep clinic patients. The ability to better articulate traits of clinic patients is a major breakthrough as the heterogeneity of patients with any given sleep disorder is still largely unknown. As a result, it is nearly impossible for physicians to develop a thoughtful treatment plan that is tailored to a patient's needs. Recently there has been considerable expansion in this area, particularly with respect to obstructive sleep apnea (OSA). Predicting who will be adherent to positive airway pressure (PAP) has proven to be an incredibly complex task. Patient characteristics, disease characteristics, initial CPAP exposure factors, and psychosocial factors have all been extensively examined as possible predictors with little success[42]. In general, patient reported symptoms and adherence to treatment are poorly correlated with disease severity in both adults[42] and children[43,44]. Dr. Allan Pack and colleagues tackled this problem by using unsupervised clustering on self-reported symptoms at diagnosis to identify relevant clinical subtypes of OSA related to (a) insomnia/disturbed sleep, (b) minimal symptoms, and (c) excessive sleepiness. These subtypes were first identified using the Icelandic Sleep Apnea Cohort[45]. They have now been validated in several other independent cohorts including: the Korean Genomic Cohort[46], the Sleep Apnea Global Interdisciplinary Consortium (SAGIC)[47] and the Sleep Heart Health Study (SHHS)[48]. This work led to the discovery that the cardiovascular risk of OSA appears to be driven by the subset of patients in the excessively sleepy subtype[49]. From a clinical perspective, this type of breakthrough is critical because now we know that treatment of these sub-groups should be addressed differently and there needs to a major focus on patients who are excessively sleepy.

The idea that self-reported symptoms can guide clinical therapy is far from novel. In sleep, there are hundreds of validated sleep surveys, but most focus on a single disorder and strive for brevity. However, sleep and sleep disorders are far too complex for tunnel vision to be an effective strategy. As a result, these measures fail to gain enough context to provide a useful diagnosis or treatment plan. The Alliance Sleep Questionnaire (ASQ) changes that paradigm. The ASQ is an

innovative, on-line questionnaire designed to address the multifaceted nature of sleep and was originally created by the same Academic Alliance for Sleep Research (AASR) that conducted the sleep scoring study described above. Stanford University led the multidisciplinary team of domain experts from Harvard University, University of Pennsylvania, University of Wisconsin-Madison, and Stanford University to design the questionnaire's content using novel questions and validated measures[50]. The questionnaire uses complex branching logic to lead users through a comprehensive set of questions covering medical history, current medications, previous treatments for sleep disorders, sleep habits/schedule, daytime fatigue, as well as symptoms of insomnia, obstructive sleep apnea (OSA), restless legs syndrome (RLS), narcolepsy, and other sleep related disorders.

The ASQ has been standard of care since 2012 at the Stanford Sleep Medicine Center[51,52]. The questionnaire takes a median of 35 minutes for sleep disorder patients to complete, and 20 minutes for a young volunteer without sleep issues. The completion rate for clinical patients is 88.4%[51]. The ASQ's ability to identify key disorders has been evaluated using clinical diagnosis as the gold standard. Type 1 narcolepsy and RLS were evaluated first on a small scale using simple rules to classify patients based on responses to the corresponding module. The type 1 narcolepsy algorithm had a sensitivity of 80.0% and a specificity of 99.0%[53]. The RLS algorithm had a sensitivity of 68.9% and a specificity of 91.2%[54]. More advanced techniques were used to evaluate sleep apnea and insomnia. Instead of restricting the model to the validated measure used in the questionnaire, the algorithms contained variables from other ASQ modules. In both cases, the ASQ algorithm had better sensitivity and specificity compared to the validated measure. The ASQ's OSA model had a sensitivity and specificity of 95.5% and 43.3% compared to the multi-variate apnea predictor (MAP)'s[55] sensitivity of 95.1% and specificity of 38.4%[56]. The ASQ insomnia model had a sensitivity of 75.1% and a specificity of 73.0% compared to the insomnia severity index (ISI)'s[57] sensitivity of 71.5%, and specificity of 62.0%[58].

To date more than 15,000 new and returning patients have taken the ASQ before their consultation or sleep study appointment. Combining this volume of standardized subjective data with the final diagnosis could be an incredibly valuable resource for the field of sleep if widely shared. However, extracting a definitive diagnosis from the electronic health record (EHR) is an extremely complex task because the data is disorganized and often embedded in clinical notes. Cleveland Clinic was an early adopter of the EHR. In 1995, they began transitioning to Epic, an EHR software system. By 2018, less than 5% of their EHR data are codified variables[59] which speaks to the complexity of this data. To provide a clean dataset for statistical analysis, the data

must be pre-processed using numerous statistical techniques to parse, map and validate the raw EHR data. The Observational Health Data Sciences and Informatics (OHDSI)[60] group is using large-scale analytics to develop an open-sourced pipelines to map data to a detailed ontology. The Observational Medical Outcomes Partnership (OMOP) common data model provides resources to convert a wide variety of datasets into a format that can shared across institutions. This standardization of EHR data is a critical step in evaluating generalizability of findings. Researchers can access EHR data from standard encounter tables as well as clinical notes, use the provided natural language processing tools to engineer features needed to answer scientific questions. This common data model also provides a mechanism to validate findings with data from other institutions. Although working with EHR data is still a complex problem that requires a high level of knowledge and expertise, these tools will revolutionize how health outcomes are monitored.

When taken together, these types of advances are coming together to create new opportunities in sleep medicine for large-scale analysis of real-world data. There have already been noteworthy progress to automate and standardize overnight sleep studies. In addition, complex datasets can now be assembled using OHDSI and OMOPs to combine EHR and comprehensive subjective data from surveys like the Alliance Sleep Questionnaire. These datasets can be posted to the National Sleep Research Resource[28] for access by researchers interested in building prediction models that will result in customized clinical workflows and ultimately improved patient outcomes.

This type of data sharing is critical to combat the replication crisis described by Dr. John Ioannidis in 2005[61]. In this controversial paper, Ioannidis explained how most published studies are false, with most findings at best accurately reflecting the prevailing bias and at worst incorrect interpretations of faulty methodology. The “Reproducibility Project: Psychology” was developed in 2011 to evaluate reproducibility of previously published work[62]. Using a strict protocol[63], more than 250 scientists from around the world attempted to reproduce 100 key 2008 articles published in three leading psychology journals. In the original analyses, 97% of the studies had significant results ($p < 0.05$). However, only 36% of the replication studies had a significant finding Correlational tests performed by the team found that replication success was better predicted by the strength of original evidence than by characteristics of the analyses performed by either the original and replication teams[64].

Now that we know how pervasive the lack of reproducibility is in existing literature, we must set a higher standard for future publications. Many journals are now requiring authors share their raw

data as part of their publication. Some journals are even providing opportunity to publish datasets as a citable scientific publication[65]. These systemic changes are improving the quality of science in multiple ways. Providing an easy mechanism for peers to reproduce findings is setting a new level of accountability. In addition, publicly available datasets allow researchers to validate their own findings to ensure they are valid before publication. Finally, scientists are able to fill in knowledge gaps using existing datasets rather than generating new data, which is both expensive and time consuming.

This dissertation leverages many of the recent advances in science by using a combination of traditional and machine learning analytic approaches to answer a scientific question using existing data. The finding was replicated in two independent cohorts to prove reproducibility and strengthen the validity and generalizability of the results.

Thesis Specific Aims

The goal of this study was to investigate the relationship between REM sleep and mortality. A combination of traditional statistics and machine learning were used to conduct a thorough analysis on the association and explore whether another sleep stage could be a better predictor of mortality. The hypotheses were: (1) lower quantities of REM sleep would be associated with increased mortality rates and (2) compared to other sleep stages, REM would be the best predictor of mortality. To evaluate generalizability of the findings, the final models were replicated in two independent cohorts. The results from all three cohorts were combined in a meta-analysis.

The dissertation was separated into the following three aims:

Specific Aim 1. Examine the relationship between REM sleep and mortality.

REM sleep has been linked with multiple aspects of mental and physical health[66-74]. After total sleep deprivation, REM is the first sleep stage to rebound[75]. Additionally, there is evidence of a U-shaped relationship between sleep duration and mortality[17,76,77] therefore it is logical that a reduction in REM would be associated with adverse outcomes.

Data from the Osteoporotic Fractures in Men (MrOS) Sleep Study[78] were used to exhaustively evaluate the relationship between REM sleep and mortality. MrOS is an observational, longitudinal cohort study that collected baseline sleep data, including overnight polysomnography (PSG), on 3,135 community-dwelling men ≥65 years of age between December 2003 and March 2005. Mortality data were last updated in 2018 and 2,675 men ages met criteria for inclusion in

this analysis (mean age 76.3 years ± 5.5 years) and were followed a median of 12.1 years. Crude and multi-variable Cox proportional hazards or logistic regression models were used to evaluate the effect of a 5% reduction in REM sleep on the rate of mortality. Mortality data were updated in 2018 and five outcome definitions were used (1) all-cause mortality, (2) cardiovascular mortality, (3) cancer mortality, (4) other morality, and (5) longevity (survival $\geq$ 90 years of age). A combination of clinical knowledge and 6-fold cross-validation was used to control for potential confounders. Multiple sensitivity analyses were conducted to evaluate whether other factors could explain the findings.

My hypothesis was that lower levels of REM sleep would be associated with an increased rate of mortality.

Specific Aim 2. Identify which sleep stage best predicts mortality.

Sleep is comprised of multiple sleep stages and by nature are highly correlated. Little is known about how the proportion of time spent in each sleep stage relate to timing or cause of death. Therefore, the role of each sleep stage was explored using supervised machine learning.

Random survival forests[79,80] were used because they are indifferent to highly correlated data as well as outliers and non-linear data. Additionally, due to the ensemble nature of the model building, there is minimal risk of overfitting the data. Therefore, it was possible to include all sleep stages (non-REM N1, N2, N3, and REM) were included in the models to predict all-cause mortality using data from the MrOS study. How much each individual sleep stage contributed to the model was evaluated using mean decreased accuracy. This method ranks the sleep stages from most to least predictive in context of one another.

An individual conditional survival inference trees was created to visualize an example tree from the random forest model. The algorithm empirically selects the best sleep stages and cut-points to divide the sample based on all-cause mortality risk. The data are iteratively split into subsamples until only subsamples with different risks remain.

My hypothesis was that REM sleep would be the most important sleep stage for predicting mortality.

Specific Aim 3. Assess generalizability by replicating findings using external datasets.

To improve the validity, consistency, and generalizability of the finding, the analyses were replicated using two independent, well-characterized sleep cohorts. Although the data available in

the replication datasets did not exactly match MrOS multi-variate models, the exposure and outcomes were equivalent.

The Wisconsin Sleep Cohort (WSC) is an ongoing, longitudinal, population-based study of the causes, consequences, and natural history of sleep disorders[81]. Participants were aged 30 to 60 years when enrolled and 1,546 agreed to participate in a sub-study that objectively measured sleep every four years[82,83]. Of these individuals, 1,386 met criteria for inclusion for analysis (45.7% women, mean age 51.5 years ± 8.5) and were followed a median of 20.8 years. Deaths were updated through 2018.

The Sleep Heart Health Study (SHHS) is a prospective cohort study of cardiovascular consequences of sleep-disordered breathing[84]. The population was recruited between 1995 and 1998 and included 6,441 individuals who were at least 40 years of age and were not being treated for sleep-disordered breathing. 5,550 met criteria for inclusion for analysis (52.4% women, mean age 63.0 years ± 11.2) and monitored for 11.9 years. Mortality data were last updated in 2011.

A meta-analysis was performed using the results from all three studies to create a pooled, weighted measure of association and quantitatively provide a more precise estimate of effect magnitude.

Impact: The long-term goal of this study is to improve our understanding of the underlying biological pathways linking sleep and mortality. Results from this analysis will inform future studies examine the importance and function of REM sleep. A better understanding of this complex biological process will ultimately inform clinical decisions related to medical and pharmacological treatments for at risk populations.

My overall career goal is to improve the quality of life for millions of people that struggle nightly to obtain refreshing, restorative sleep by developing new diagnostic criteria and treatment options using the latest computational solutions in phenotype assessment.

Background and Literature Review

Sleep is a regulated, reversible, and recurring loss of consciousness. It is a non-negotiable requirement for a happy, healthy life. In our busy society, sleep is often overlooked and undervalued even though sleep disorders and sleep dysregulation impact an estimated 50 to 70 million Americans[14]. Sleep issues contribute to multisystemic medical consequences including cardiovascular issues (arrhythmia, hypertension, stroke)[16,17], metabolic disorders (diabetes, obesity)[18], psychiatric problems (depression, suicide)[19,20], impaired cognition (Alzheimer's, dementia)[21,22], quality of life outcomes (daytime sleepiness, irritability, concentration)[23], and all-cause mortality[16].

During sleep, the central nervous system undergoes several dynamic changes that can be tracked and monitored throughout the night. Based on these physiologic changes, sleep is split into two main states: non-rapid eye movement (NREM) and rapid eye movement (REM) sleep. NREM is further broken down into three stages (N1, N2, N3) based on frequency, amplitude and morphology of the brain waves. N1 is the transition from wake to sleep, characterized by slowing of the brain waves, disappearance of occipital alpha waves, decreased EMG and slow rolling eye movements. During N2, the heart rate slows and the body's core temperature begins to drop. Brain waves continue to slow with occasional bursts of faster waves called sleep spindles and slower waves called K-complexes. N3 is characterized by a dominance of slow, high amplitude brain waves (>30%). These three NREM stages roughly correspond with the depth of sleep, with N1 having the lowest arousal threshold (easiest to waken) and N3 the highest (hardest to waken).

REM sleep is a highly active mental state defined by rapid eye movements, mixed low-voltage, low voltage, desynchronized brain waves in the beta or theta range with occasional saw tooth patterned brain waves, and loss of muscle tone. REM sleep is associated with dreaming[85] and homeostatically regulated[86]. REM is homeostatically regulated and selective REM sleep deprivation results in an extended time spent in REM sleep during recovery, known as REM rebound[86].

A normal adult spends approximately 5% of the night in N1, 50% in N2, and 20% in N3. Twenty to 25% of the night is spent in REM sleep (about 2 hours). A normal sleeper will cycle through the four sleep stages in approximately 90 minutes and will have between four to six sleep cycles per night with REM propensity highest in the last third of the sleep period[87]. Sleep architecture is the term used to describe sleep's cyclical pattern as it shifts between the different stages. Figure 3 is an example of a normal sleeper's hypnogram, which graphs sleep stages as a function of time.

Numerous studies have shown the critical role REM sleep plays with respect to learning[88], memory[89], and mood[90]. Decreased REM has been linked with poor mental and physical health outcomes[66-74]. NREM sleep is associated with a synchronized EEG pattern in which sleep spindles and/or K complexes found in N2 and high-voltage slow wave activity found in N3. NREM is thought to maintain cell homeostasis by reducing the number of synaptic connections to a basic level[91]. This process is necessary for normal physical and intellectual performance and behavior.

Many older adults experience sleep issues even though sleep disturbances are not an inherent part of the aging process[92]. While it's true the natural aging process has been shown to significantly change sleep structure and duration in the general population[93]. Most changes to sleep architecture occur during childhood and early adulthood. After that, REM sleep declines ~0.6% per decade of life; stabilizing ~75 years of age[94]. Nevertheless, the increasing prevalence of co-morbidities and primary sleep disorders associated with aging or polypharmacy factors have been found to make older adults more vulnerable to sleep disturbances[93,95,96]. Sleep complaints, whether related to disorders or sleepiness, have important consequences in older adults as these symptoms predict poor physical and mental health-related quality of life[97]. Spending proportionally less time in REM and more time in N1 has been associated with lower levels of cognitive function[98].

Although sleep is increasingly recognized as an important component of health and well-being, maintaining healthy sleep habits is complicated due to the multiple dimensions impacting sleep quality including timing, duration, and the absence of sleep disorders[99]. There are guidelines on the recommended amount of sleep needed for optimal health, however sleep needs vary from person to person as well as across the lifespan[100]. As a result, the importance of sleep is regularly undervalued.

Numerous studies have reported on the association between sleep and mortality, most focusing on the U-shaped relationship between self-reported sleep duration and mortality[17,76,77]. Despite emerging evidence of a sleep-mortality association, the mechanisms underlying the relationship are not well understood. Specifically, we have not identified which component of sleep is the most clearly associated with mortality. A recent meta-analysis[76] found that sleep duration may also be associated with obesity[101], hypertension[102], and cardiovascular outcomes[103], in addition to all-cause mortality[16] which may provide insight into a possible biological pathway. Alternatively, obstructive sleep apnea (OSA) has been linked to cardiovascular outcomes such as early mortality, stroke, and cardiovascular disease (CVD)[104-107]. Therefore, OSA may play a role in the

association between sleep duration and mortality. However, there is very limited data showing how sleep architecture relates to timing and cause of death[108].

Determining which aspects of sleep contribute to the mortality association has been a challenging topic to study given the multi-dimensionality of both sleep and mortality. The goal of this study is to use a combination of traditional regression and supervised machine learning techniques to evaluate the relationships between sleep architecture and mortality. Given the evidence of the importance of REM sleep, I postulated that lower levels of REM would be associated with increased rates of mortality. As described in the Introduction, I tested this hypothesis using data from the Osteoporotic Fractures in Men Study (MrOS)[78]. Five mortality outcome definitions were evaluated: (1) all-cause mortality, (2) cardiovascular mortality, (3) cancer mortality, and (4) other morality (non-cardiovascular and non-cancer related), and (5) longevity (survival > 90 years of age).

One major challenge in designing this study was the sheer volume of potential confounders. I felt it was important to carefully consider this problem from a clinical perspective rather than relying on machine learning to select the parameters. Therefore, I created a directed acyclic graph (DAGs) to identify the presence of confounding and other sources of bias. In epidemiologic research, a DAG is a common method used to evaluate the causal relationships between the exposure and outcome while accounting for multiple covariates[109,110]. Given scarcity of data proving casual relationships in sleep medicine, I created a modified DAG to graphically organize and cluster potential covariates available in the MrOS cohort (See Figure 4).

The next challenge was to evaluate which clinically relevant covariates identified in the DAG should be included in the final regression model. Stepwise variable selection methods regularly used in regression are an outdated approach that violates the assumption that a single hypothesis is tested. The creation of multiple models with different combinations of covariates will bias standard errors and p-values toward zero and parameter estimates away from zero[111]. Therefore, I used a mixture of clinical knowledge and k-fold cross-validation to build the final model. I opted to separate the data by recruitment site instead of randomly splitting the data into 10 groups (10-fold cross-validation) or performing a stratified cross-validation, both of which are common approaches. This strategy allowed me to evaluate heterogeneity across sites and was less computationally intensive than leave-one-out cross-validation.

Cox proportional hazard's regression[112] was a natural choice for statistical test given it is one of the most popular regression techniques for survival analysis. Logistic or linear regression would not be appropriate since the scientific question involved estimating the cumulative hazard using

both time-to-event (time a subject was followed in the trial) and whether the event occurred (died / survived).

To gauge reproducibility, consistency, and generalizability of the findings, the final models were validated using data from the Wisconsin Sleep Cohort (WSC)[82,83] and the Sleep Heart Health Study (SHHS)[84]. To improve precision of the results, a meta-analysis was performed.

This multi-dimensional approach provided a solid framework to evaluate the scientific question. Reliable, accurate results were achieved by reducing threats to the study's internal validity. Replication proved the findings were generalizable to other age groups and populations.

Methods

Participants

The Stanford University Institutional Review Board (IRB) determined that this research does not involve human subjects as defined in 45 CFR 46.102(f) or 21 CFR 50.3(g).

MrOS Sleep Study

The Osteoporotic Fractures in Men (MrOS) parent study[113] is an observational, longitudinal cohort of older community-dwelling men. The study enrolled 5,994 men for a baseline study (2000–2002) across six clinical centers throughout the United States (Birmingham, Alabama; Minneapolis, Minnesota; Palo Alto, California; Monongahela Valley, Pennsylvania; Portland, Oregon; and San Diego, California)[78,114]. Subjects were ≥65, able to walk without assistance, and without bilateral hip replacements[114].

The MrOS Sleep Study is an ancillary project that assessed sleep using a combination of subjective and objective measures to understand the relationships between sleep disorders and falls, fractures, mortality, and vascular disease. This sub-study included 3,135 (59.2%) men from the parent study recruited between December 2003 and March 2005[115]. Participants were not eligible for inclusion in the MrOS Sleep Study if they were being treated for sleep apnea with positive pressure or oral appliances, or if they used overnight nocturnal oxygen therapy (n = 150). Other reasons for non-participation included: death prior to the sleep study (n = 349), study participation terminated (n = 39), declined participation in the sleep study (n = 1,997), or because MrOS Sleep Study recruitment goals were met prior to enrollment (n = 324)[78].

Of the 3,135 men enrolled in the MrOS Sleep Study, 207 did not complete the overnight sleep study and 56 were missing sleep staging data. An additional 197 individuals did not have complete mortality data, leaving 2,675 participants who contributed data to this analysis. See Figure 5 for an inclusion/exclusion flow chart of MrOS participants. All men provided written informed consent; each site received IRB approval.

Wisconsin Sleep Cohort (WSC)

Data from the Wisconsin Sleep Cohort (WSC) were used to replicate the main findings. The WSC is an ongoing, longitudinal, population-based study of the causes, consequences, and natural history of sleep disorders. It was established in 1988 from a sample of employees from four state agencies in south central Wisconsin. There are 2,940 participants in the main study aged 30 to 60 years at enrollment[82,83]. Participants were mailed questionnaires every five years

asking about the participant's sleep and medical history. A subsample of 1,546 (53%) agreed to undergo in-lab nocturnal polysomnography (PSG) every four years. Rationale and design of the WSC study were published in 2009[82].

This study used participants who had data on as many of the covariates included in the MrOS models as possible. Of the 1,546 who agreed to have sleep studies, 22 were excluded because they had inadequate sleep data and two died the same year as the sleep study. Another 136 were excluded from analysis because they were either missing mortality or questionnaire data. See Figure 6 for a flowchart of the 1,386 WSC participants analyzed. Informed consent was obtained from all subjects under a University of Wisconsin-Madison Health Sciences approved IRB protocol.

Sleep Heart Health Study (SHHS)

The Sleep Heart Health Study (SHHS) is a prospective cohort study of cardiovascular consequences of sleep-disordered breathing. Participants were recruited between 1995 and 1998 from prospective cohort studies including the Framingham Offspring and Omni Study, the Atherosclerosis Risk in Communities Study, the Cardiovascular Health Study, the Strong Heart Study, and the cohort studies of respiratory disease in Tucson and of hypertension in New York. Details of the SHHS study design have been previously reported[84].

Eligible individuals were at least 40 years of age and were not being treated for sleep-disordered breathing with positive airway pressure, oral appliance, oxygen, or tracheostomy at the time of enrollment. A total of 6,441 participants completed the baseline examination. Of those, 637 withdrew consent, 252 were excluded due to sleep staging issues (85 had unreliable REM data, 143 were scored as sleep/wake only, and 24 were scored as REM/NREM/wake only), and two were missing all-cause mortality data. See Figure 7 for a flowchart of the 5,550 SHHS participants included in the final dataset. Every participant provided written consent and the study protocol was approved by the local institutional review board of the participating field site.

Measures

Polysomnography (PSG) and Subjective Sleep Measures

MrOS: A full, unattended, portable in-home baseline PSG was conducted at Sleep Visit 1 (Safiro, Compumedics, Inc., Melbourne, Australia)[116]. A standard sleep study includes: electroencephalogram (EEG), electrooculogram (EEG), electromyogram (EMG) of the

submentalis muscle as well as the anterior tibialis muscles of each leg, electrocardiogram (ECG), snoring, breathing effort using respiratory inductance plethysmography summation, airflow from a nasal-oral thermistor and a nasal pressure transducer, and oxygen hemoglobin from a finger oximeter probe using a 3-second averaging rate. Studies were scored visually by centrally trained research technologists to evaluate sleep stages (REM, and non-REM stages N1, N2, and N3) using current consensus rules published in the American Academy of Sleep Medicine (AASM) Scoring Manual[32].

The primary exposure was REM sleep during a single night, evaluated as percent of total sleep time (TST) and total number of minutes spent in REM.

Standard PSG characteristics were considered for inclusion in the final model using previously published definitions[117]. These include:

1. Time in Bed (min)
2. Total Sleep Time (number of minutes spent in non-REM and REM during the sleep period)
3. Time Spent in Each Sleep non-REM Stage (percent time and total minutes)
4. Wake After Sleep Onset (WASO) (min)
5. Sleep Efficiency (%)
6. Sleep Latency (min) and REM Latency (min)
7. Overall Arousal Index (number of awakenings between 3 and 14 seconds per hour of sleep
8. Apnea Hypopnea Index (number of apneas or hypopneas per hour of sleep
9. Desaturation Index (number of desaturations per hour of sleep)
10. Periodic Limb Movements Index (PLMI – number of leg movements per hour of sleep)
11. Sleep Time with Saturated Oxygen Below 80% (% time)
12. Number of N2 to N1 and Number of N3 to N1/N2 Sleep Stage Shifts/hour

The variables used to evaluate sleep related breathing disorder included apnea hypopnea index (number of apneas and hypopneas per hour of sleep), desaturation index (number of oxygen desaturations per hour of sleep) as well as percent sleep time with saturated oxygen below 80%. Apnea was defined as a ≥ 90% decrease in airflow for ≥ 10 seconds. Hypopnea was defined as a reduction in airflow (at least 30% below baseline) lasting ≥ 10 seconds and associated with either a 3% reduction in oxygen saturation or an arousal.

Sleep fragmentation was assessed using sleep efficiency (percentage of time in bed spent asleep), arousal index (number of awakenings between 3 and 14 seconds per hour of sleep), total number of minutes awake after sleep onset, number of N2 to N1 and number of N3 to N1/N2 sleep stage shifts per hour of sleep. Other PSG variables included time in bed (min), total sleep time, sleep latency (min), REM latency, and periodic limb movement index (with and without arousals).

Self-reported sleep data included Epworth Sleepiness Scale (ESS)[118], Pittsburgh Sleep Quality Index[119], Functional Outcomes of Sleep Questionnaire[120], and positive airway pressure use >3 months.

WSC: Full, nocturnal in-lab PSGs were collected using Grass-Telefactor Heritage digital sleep systems. The majority of studies used a previously published conventional scoring approach[121]; however, PSG interpretation protocols changed slightly over time to stay current with evolving definitions. Trained technicians scored and reviewed sleep stage and apnea/hypopnea events using standard criteria[24,122]. The usual sleep study variables were analyzed (see Table 4). Hypopnea was defined as a discernible reduction in breathing (sum of chest and abdominal excursions) with a ≥ 4% oxyhemoglobin saturation reduction. Since the current analysis included the ESS (which was collected beginning in 1993), the PSG data used were the earliest available with ESS data (collected via mailed survey).

SHHS: A full-montage, unattended home PSG was collected on each participant at baseline using a portable monitor (P-Series, Compumedics, Abbotsville, AU). Details of polysomnographic equipment, hook-up procedures, failure rates, scoring, and quality assurance and control have been previously published[123]. Standard sleep study variables were analyzed (see Table 6). The SHHS PSG protocol was based on the MrOS protocol and all studies from both cohorts were scored at the same central sleep reading center s (Case Western Reserve University, directed by S Redline). The main difference was that SHHS studies only used thermistors and inductance plethysmography to monitor breathing while MrOS measured nasal pressure, thermistors, and inductance plethysmography. Hypopnea was defined as a 30% amplitude reduction in breathing effort lasting ≥ 10 seconds and associated with either a ≥ 3% desaturation or arousal. Desaturation index was defined as a ≥ 4% oxyhemoglobin saturation reduction. SpO2 signals were captured by fingertip pulse oximeter (Nonin, Minneapolis, MN, USA) sampled at 1 Hz.

Actigraphy Measures

MrOS: Home sleep-wake patterns were estimated using wrist actigraphy (SleepWatch-O, Ambulatory Monitoring, Inc., Ardsley, NY). Participants were asked to wear the devices

continuously on their non-dominant wrist for a minimum of four consecutive 24 hour periods and complete sleep logs to track time into and time out of bed, as well as times the device was removed. The sleep logs were used to edit data and set sleep intervals. The devices recorded data in three modes: proportional integration mode (PIM), time above threshold (TAT) and zero crossings mode (ZCM). In a previous study, the PIM mode was shown to corresponded better to PSG[124], so that is the mode used for this analysis. Actigraphy scoring algorithms used in this study have been published[125].

Actigraphy measures included: mean TST (min) while in bed, mean sleep latency (min), mean WASO while in bed (min), mean nighttime sleep efficiency (% time), and mean TST (min) outside sleep interval. Data were averaged over the entire period the device was worn to best reflect usual sleep patterns.

WSC and SHHS: Actigraphy was not collected.

Mortality Outcome

MrOS: After baseline, participants were contacted every four months to document vital status. Next of kin were contacted in cases of non-response. Reported deaths through 2018 were confirmed by centralized review of death certificates. Of the 2,872 men with sleep data, 31 were missing information on mortality data and 166 (5.8%) ended participation during the follow-up period and were excluded from analyses. Cause of death was categorized by ICD-9 codes as cardiovascular (396.9-442, 966.71, 785.51), cancer (141.9-208.0), and other (codes not in previous categories). For each analysis of cause-specific mortality, individuals who died of a different cause were censored. In addition, longevity was defined as survival > 90 years old compared with death before 90 years old. Individuals who were less than 90 years old and still alive were omitted from the analysis for this outcome.

WSC: Deaths through 2018 were identified by matching social security numbers with two death record sources: National Death Index and Wisconsin State Bureau of Health Information and Policy, Vital Records Section. Matches on social security number were verified with participants' age and sex. Cause of death was categorized using MrOS ICD-9 codes. Longevity was not evaluated for this cohort due to the younger age.

SHHS: Multiple concurrent approaches were used to identify and confirm deaths through 2011. These included 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 as previously described[126].

Only cardiovascular related cause of death were tracked and ICD-9 codes were not available. Therefore, cause of death is not available on all individuals and it was not possible to use the same categories used for MrOS and WSC. Instead, cardiovascular related death was defined as death caused by cardiovascular disease, coronary heart disease, myocardial infarct, or stroke. Longevity was not evaluated for this cohort due to the younger age.

Other Mental and Physical Measures

MrOS: Education, race, body mass index (BMI), neck and hip circumference, smoking status, weekly alcohol use, and daily caffeine intake were collected at baseline along with the Geriatric Depression Scale-15 (GDS)[127], Modified Mini-Mental State Examination[128] (evaluated as continuous and binary variable with a score <77 indicating impairment[129]), and Physical Activity Scale for the Elderly (PASE) Scale[130].

Data on prescription and non-prescription medications were collected at baseline and entered into an electronic database managed by the San Francisco Coordinating Center. Each medication was matched to its ingredient(s) using the Iowa Drug Information Service Drug Vocabulary[131]. Current use of medications known to affect sleep architecture were included (antidepressants, benzodiazepines), as well as sleep medications (nonbenzodiazepines, nonbarbiturate sedative hypnotics).

Self-reported history of physician diagnosis was used to identify chronic medical conditions including hypertension, angina, stroke, heart attack, transient ischemic attack (TIA), congestive heart failure, type 2 diabetes, chronic obstructive pulmonary disease, osteoarthritis, and rheumatoid arthritis.

WSC: BMI was assessed during the PSG visit. Data on education, race, smoking habits, weekly alcohol, and daily caffeine use were collected by questionnaire. Participants reported current use of anti-depressants or sedatives and physician diagnosis of hypertension, coronary artery disease, or heart attack.

SHHS: A baseline examination was performed on all participants that included a detailed health interview to collect data on history of hypertension, angina, heart attack, congestive heart failure, type 2 diabetes, chronic obstructive pulmonary disease, and stroke as well as current prescription medication use (anti-depressants and benzodiazepine). Short Form 36 Health Survey (SF-36) was also collected which is comprised of self-report quality-of-life measures created as part of the Medical Outcomes Study[132-134]. The Mental Component Scale Standardized Score was used in place of the 3MS to evaluate cognition, the Mental Health Index Standardized Score was used in

lieu of the GAD to estimate mental health, and the Physical Component Scale Standardized Score was used instead of the PASE to account for physical activity. Information on education, race, smoking status, caffeine and alcohol use was obtained by self-report. Anthropometric measures including weight, height, neck circumference, and waist girth were obtained at the baseline sleep study by trained and certified technicians. Although SHHS was a multi-site study, information on enrollment site was not included in the shared dataset.

Statistical Analysis

Characterization of Study Populations

Descriptive statistics and frequency distributions were performed on demographic, behavioral and clinical characteristics in the overall cohort and by quartiles of percent REM. Similarly, characteristics of sleep were examined by overall cohort and by REM quartile. Means and standard deviations are reported for continuous variables, and number and percent for categorical variables. Analyses were performed using SAS 9.4 (Cary, NC)[135], R Studio Version 1.1.463[136], and RevMan 5.3 software[137].

Data were evaluated for missingness and extreme or implausible values. Missing data were considered missing at random and imputed using multivariate imputation by chained equations operationalized using the R 3.2.5 MICE package[138]. Rate of loss to follow-up was <10%. Collinearity was evaluated using Pearson's correlation coefficient[139] with > 0.70 as the cut-point. Based on actigraphy data, individuals with mean TST>8h were categorized as long sleepers and TST<5h as short sleepers.

Survival Models

Cox proportional hazards models[112] were used to assess associations between percent REM and all-cause, cardiovascular, cancer, and other (non-cardiovascular, non-cancer) mortality. The Cox functional formula is

$$\ln(\, h(t)/h_0(t)) = (\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

where h(t) is the expected hazard at time t. The baseline hazard is $h_0(t)$ is and represents the hazard when all independent variables (X_1, X_2 , X_k) equal zero. Since there are no assumptions about the shape of the baseline hazard function, Cox is a semi-parametric model. Results are reported as hazard ratios (HR) with 95% confidence intervals (95% CI) for each Cox model.

Percent REM was evaluated as a continuous variable with decrements of 5% as the unit rather than 1% to aid with interpretation. Five percent was selected because it is similar to the 6.6% standard deviation, but provides a universal unit because it is not dataset dependent. Minutes in REM were evaluated using a 25 minutes decrease, which was similar to the 28.6 minute standard deviation.

Cox models were built in two steps using a combination of clinical knowledge and empirical covariate selection using data from the MrOS Cohort. Model 1, included covariates selected based on known associations including: age, race, education, BMI, smoking status, alcohol, caffeine, medication use (sleep, antidepressant, or benzodiazepine), and study site. Next, more than 60 variables commonly associated with sleep architecture or mortality were evaluated using six-fold cross-validation to obtain the best features to add to Model 1 for the final regression models (Model 2)[140]. At each fold, data from one site was withheld and an optimal model was identified using the My.stepwise R package. Criteria for entry into the final model was inclusion in at least three of six optimal models. I chose not to withhold a random sample to gauge whether site affected variable selection. The REM sleep beta co-efficient from the final model was systematically compared to the REM beta co-efficient from each of the six folds to identify whether the effect sizes were similar.

Potential interactions were explored using random survival forests[79,80]. Models were tested for the proportional hazards assumption, influential observations, and linearity assumption using Schoenfeld residuals, graphical methods, and martingale residuals. The Schoenfeld residuals test for independence between residuals and time for each covariate and for the model. Percent REM met the proportional hazards assumption with a Shoenfeld residuals of $p = 0.45$, however a threshold effect was suspected based on Kaplan-Meier survival curves[141]. Therefore, separate models were run with percent REM as a binary variable using 15% as the cut-point, which is similar to the first cut-point identified by the conditional inference survival tree. Martingale residuals were used to assess nonlinearity for each covariate by evaluating the functional form of each covariate. Linearity assumptions were met for all models. Deviance residuals were used to evaluate the potential effect of outliers using a normalized transform of the martingale residual. There was no indication of influential observations.

Logistic regression was used to estimate the odds ratio and 95% CI for the association between percent REM and longevity. The functional formula for logistic regression is

$$\ln({}^{P}/_{1-P}) = (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

where the output is logit, or the natural logarithm of the odds where p is probability. The y intercept is β_0, representing the expected mean value when all independent variables (X_1, X_2 , X_k) are equal to zero. The longevity models were adjusted for the same set of covariates used in the Cox regression models. Assumptions of logistic regression were checked and met, including independence, collinearity, and linearity in the association between continuous predictors or covariates and the logit.

Kaplan-Meier survival curves[141] were used to graphically present the probability of surviving to any point in time. The horizontal axis (x-axis) represents time in years, and the vertical axis (y-axis) shows the probability of surviving. The lines represent the different survival curve by group. A vertical drop in the curves indicates an event. A vertical tick mark means that a patient was censored at that time point. Poor survival rates are reflected by a curve that drops relatively rapidly towards zero. The Log-Rank Test was used to compare the total number of events observed in each REM quartile with the number of events that would be expected if there were no group effect. The Log-Rank statistic is approximately distributed as a chi-square test statistic.

Sensitivity Analyses

Multiple sensitivity analyses were conducted to rule out alternative explanations. First, to attenuate the possibility of REM sleep being a marker of ill health/undiagnosed health issue at baseline, or reverse causation, analyses were run after excluding individuals censored in the first two years. Then, to evaluate whether the presence of sleep apnea or medication use influenced the association, anyone with an AHI>30 and/or using antidepressants, benzodiazepines, or sleep medications were excluded. Another analysis excluded depressed individuals (GDS score >4 or antidepressant use). To assess residual confounding from sleep duration, I performed two analyses excluding individuals using different TST definitions (TST<5h or >8h from either PSG or actigraphy based on previous MrOS publication[142] and TST<6h or >8h). Finally, models were run using total REM minutes and dichotomized percent REM (cut-point = 15%; similar to lowest quartile threshold).

Sleep Architecture and Mortality

Sleep stages are inter-dependent (they add to 100%) so conditional inference survival tree and random survival forest methods to identify which sleep stage(s) were driving the significance of the finding. Conditional inference survival tree and random survival forest analyses were performed using the party package ctree and cforest functions in R[136,143]. All sleep stages (non-

REM N1, N2, N3, and REM) were included in the models to predict all-cause mortality using data from the MrOS study.

Conditional inference tree models were used to empirically select the best sleep stages and cut-points to divide the sample based on all-cause mortality risk. To create a tree, data are iteratively split into subsamples until only subsamples with different risks remain. The minsplit (minimum sum of weights in a node in order to be considered for splitting) was set to four and I did not limit tree depth (successive splits).

Random survival forests[79,80] is a supervised machine learning technique that uses an ensemble approach to classify data. Simply, random survival forests summarize the results from multiple different decision tree models. For this analysis, a series of survival trees were built by splitting on a random subset of features using bootstrap samples (sampled with replacement from the original training set). The control parameter setting cforest_unbiased was used to build the forest and reduce bias in the variable selection and variable importance. The ntree argument (controls the overall number of trees in the forest) was set to 1,000, and the mtry argument (controls the number of randomly preselected predictor variables for each split) was set to eight. To control the tree sizes, a minimum of 400 subjects was required in a node. This classification technique was selected because random forests are indifferent to highly correlated data as well as outliers and non-linear data and there is minimal risk of overfitting the data.

After fitting this model, I calculated the variable importance for each individual sleep stage using mean decreased accuracy to rank the variables from most to least predictive in context of one another. This style of permutation importance was employed because the model uses uncorrelated predictors sub-sampled without replacement and unbiased trees will be used to construct the forest. To obtain the mean decrease in accuracy, each predictor variable is randomly permutated in order to break the original association between the predictor and the outcome. Variables were ranked from most to least predictive in context of one another based on the resulting difference in prediction accuracy before and after this process. The resulting difference in prediction accuracy before and after this process is known as mean decrease in accuracy.

Validation

Cox models were repeated using the WSC and SHHS dataset after matching all possible covariates across both datasets. All-cause mortality model was stratified by gender in both cohorts.

Given the smaller samples size, a power calculation was performed for the WSC before running the survival models. The Power and Sample Size Program version 3.1.6 was used to evaluate the study's power to reject the null hypothesis that the experimental and control survival curves are equal given the sample size. The existing WSC sample of 1,386 includes 345 individuals in the lowest quartile of REM sleep. The median survival time for those in the highest quartile was 20.9 years compared to 20.8 in the lowest. The Type I error probability associated with this test of this null hypothesis is 0.05. Assuming the true hazard ratio is 1.10[144], the power to reject the null hypothesis was less than .20 indicating the study is underpowered.

Meta-analysis

To improve precision of the findings and evaluate heterogeneity across studies, a meta-analysis was performed using the data from each of the cohorts. Separate models were created for all-cause and cardiovascular related mortality. All-cause mortality was also stratified by gender. Summary estimates (HR, 95% CI) were calculated using a fixed-effect model. Since estimates of log hazard ratios and standard errors from the Cox proportional hazards regression models were available, the inverse variance-weighted method was utilized to calculate confidence intervals. Heterogeneity across studies was estimated by the Cochran Q test (Chi^2, $p \leq 0.1$ to be indicative of statistically significant heterogeneity) and I^2 statistic[145]. The I^2 statistic describes the percentage of variation across studies that is due to heterogeneity rather than chance. Analyses were performed with RevMan 5.3 software[137].

Results

Characterization of Study Populations

MrOS Cohort: 2,675 males (91.5% white) with a mean age of 76.3 ± 5.5 years at baseline and followed for a median of 12.1 years were included. Average age was 86.6 ± 5.2 years at follow-up. Percent REM ranged from 0 to 43.9% and was normally distributed. Overall mean was 19.2 ± 6.6% (69.7 ± 28.6 min) with values increasing from 14.8% in the lowest quartile to 23.6% in the highest. Table 1 reports demographic, lifestyle, and health characteristics varying across REM quartile. Those in the lowest quartile tended to be older with higher rates of anti-depressant use, hypertension, heart attack, and TIA, and lower PASE scores. As expected, most objective sleep variables differed across quartiles. Subjective sleepiness as measured by total Epworth Sleepiness Scale did not have notable variation across quartile. Sleep variables for the MrOS cohort are presented in Table 2. Of note was the high amount of wake after sleep onset which was both elevated compared to the other groups and significantly higher in the lowest REM quartile (134.5 ± 73.4 min) compared to the highest (102.5 ± 61.7 min).

WSC: 1,386 subjects (54.3% men, 94.6% white) with a mean age of 51.5 ± 8.5 years at baseline were followed for a median of 20.8 years (70.2 ± 7.7 years at follow-up). Table 3 describes demographic, lifestyle, and health characteristics for the overall cohort and by REM quartile.

WSC subjects were younger, mixed-gender, more obese (30.7 vs 27.2 kg/m2), consumed more alcohol (3.6 vs 1.9 drinks/week), more likely to be current or never smokers, had fewer co-morbidities, and had higher anti-depressant and/or sedative use compared to MrOS. Despite these differences, most measures had similar distributions across REM quartiles compared to MrOS. Demographic and lifestyle variables are compared graphically to MrOS by REM quartile in Figure 8 and sleep variables are compared in Figure 9.

Minutes in REM (67.8 ± 28.9 min) and percent REM range (0-43.0%) were similar, but overall mean percent REM (17.6 ± 6.5%) was lower compared to MrOS and SHHS data, possibly due to in-lab vs in-home studies. Objective and subjective sleep measures for the WSC cohort are provided in Table 4 including variation by REM quartile. While the Epworth Sleepiness Scale did not change significantly across REM quartiles, the overall totals were higher than the other cohorts indicating the population overall was sleepier.

SHHS: 5,550 participants (47.6% men, 84.6% white) with a mean age 63.0 ± 11.2 years at baseline and monitored for a median of 11.9 years (74.0 ± 10.5 years at follow-up). Demographic,

lifestyle, and health characteristics for the SHHS cohort are shown in Table 5. SHHS participants were older than WSC and younger than MrOS. There was more ethnic diversity (>15% non-white) in SHHS compared to the other cohorts, both of which were almost exclusively white. There was a higher percentage of individuals who did not drink alcohol (65.1%) and fewer participants took medications (anti-depressants or benzodiazepine) compared to the other groups.

The overall mean (19.4 ± 6.9) and range (0–43.3%) for percent REM was nearly identical to the MrOS cohort. TST (390.1 ± 64.2) was longer in SHHS than both MrOS (356.3 ± 68.4) and WSC (378.3 ± 58.8). All objective and subjective sleep measures are presented in Table 6.

Mortality Data

More than half of the MrOS cohort (52.5%) had died compared to only 13.3% of the WSC and 22.3% from SHHS. These differences are expected given the underlying age differences at baseline. The percentage of men that died was higher than women in both WSC (15.1 vs 11.1%) and SHHS (25.2 vs 19.7%). Longevity was only evaluated in the MrOS cohort. Of the 1,759 individuals analyzed, 737 (41.9%) lived to be at least 90 years old. There were more people in the lowest REM quartile met criteria for the longevity analysis (29.3%) compared to the other quartiles (22.3–25.1%). However the lowest REM quartile also had the lowest percent of people who actually lived to aged 90 (39.6%) compared to the other quartiles (41.6–44.1%).

There were fewer deaths in the WSC sample (184, 13.3%) which was expected given younger ages. The SHHS had 1,237 (22.3%) deaths. As with MrOS, those in the lowest REM quartile had the highest percentages of death. Mortality statistics for all three cohorts are provided in Table 7.

Survival Models

Evaluation of the MrOS Cohort

In MrOS, 1,404 deaths (52.5%) were reported over a median follow-up of 12.1 years (interquartile range [IQR] = 5.4). Regression analysis of percent REM as a continuous variable showed a downward trend, reflected in the lowest quartile of percent REM having the highest percentage of deaths for each mortality category. A 13% higher all-cause mortality rate for every 5% reduction in REM (HR = 1.13, 95% CI, 1.08–1.19) was observed after adjusting for covariates identified using cross-validation. The multi-variate model controlled for the following measures: age, race (white versus non–white), education, BMI, smoking status, weekly alcohol,

daily caffeine, antidepressants, benzodiazepines, sleep medications, site, overall arousal index, sleep time with saturated oxygen below 80%, percent stage N2 sleep, actigraphy mean scored sleep while outside of sleep interval, actigraphy wake after sleep onset, Epworth Sleepiness Scale Score, Teng Mini Mental State Examination Score, Physical Activity Scale for the Elderly Score, Depression, congestive heart failure, chronic obstructive pulmonary disease, type 2 diabetes, heart attack, and stroke. There were no noteworthy interactions based on the random forest analysis for interaction terms.

The REM/mortality association persisted for cardiovascular-related mortality (HR = 1.11, 95% CI, 1.02–1.20), and other (non-cardiovascular, non-cancer) mortality (HR = 1.19, 95% CI, 1.11–1.28) but was not statistically significant for cancer-related mortality (HR = 1.06, 95% CI, 0.96–1.17). The odds of living past 90 years old (longevity) was reduced by 11% with every 5% reduction in REM in the final model (odds ratio [OR] = 0.89, 95% CI, 0.81–0.92, Table 8).

Kaplan-Meier curves of the MrOS dataset showed a possible threshold effect (see Figures 10–13). This was particularly notable for cancer mortality (Figure 12). Individuals with <15% REM had a significantly higher mortality rate relative to individuals with ≥15% REM for all mortality definitions except cancer in the WSC (HRs 1.15–1.78, Table 9).

Sensitivity analyses of the MrOS cohort found the 1,776 individuals who slept between 5 and 8 and the 850 individuals who slept between 6 and 8 hours had larger effect sizes on all outcomes except cardiovascular-related mortality in the 6 to 8 hour group (HR = 1.0, 95% CI 0.85–1.19). No substantial differences were found analyzing the 2,591 individuals who survived the first two years, the 1,684 without severe sleep apnea or medication use, the 2,291 individuals without depression (Table 10), or when using absolute time in REM (data not shown).

Validation in the WSC

The effect size for 5% reduction in REM on risk of all-cause mortality (HR = 1.17, 95% CI, 1.03–1.34) and non-CVD, non-cancer related mortality (HR = 1.26, 95% CI, 1.01–1.58) were significant despite younger age, inclusion of both men and women, longer follow-up period, reduced samples size, and event frequency. Effect sizes and direction for CVD-related mortality (HR = 1.13, 95% CI, 0.87–1.45) and cancer-related mortality (HR = 1.13, 95% CI, 0.91–1.40) were similar to MrOS, although the smaller sample size widened CIs (Table 11).

When stratified by gender, REM was associated with a 34% increased rate of all-cause mortality in women for every 5% REM reduction (HR = 1.34, 95% CI, 1.07–1.68); but was not statistically significant in men (HR = 1.09, 95% CI, 0.92–1.30). The possibility of an interaction between age

and gender were evaluated; the estimates provided modest statistical evidence for a difference (test for interaction p-value = 0.08).

Individuals with <15% REM had a higher mortality rate relative to individuals with ≥15% with odds ratios ranging from 1.36 to 1.78 for all mortality definitions except cardiovascular (HR = 1.00, Table 12).

Validation in the SHHS

The all-cause mortality effect size in the SHHS cohort was almost identical to MrOS with a 13% increase in all-cause mortality rate for every 5% reduction in REM (HR = 1.13, 95% CI, 1.07–1.18). There was a 7% increase in cardiovascular mortality rate (HR = 1.07, 95% CI, 0.97–1.17), which was lower than the other two cohorts were, possibly because a different definition of cardiovascular death was used. Unlike WSC, when stratified by gender, the hazard ratio was higher in men (HR = 1.16, 95% CI, 1.08–1.17) than women (HR = 1.09, 95% CI, 1.02–1.17, Table 13). When REM was analyzed as a binary variable, individuals with <15% REM had a higher mortality rate relative to individuals with ≥15% with odds ratios ranging from 1.20 to 1.36 for all-cause and cardiovascular mortality and when stratified by gender (Table 14).

Sleep Architecture and Mortality

REM sleep was identified as the most important sleep stage for predicting all-cause mortality based on both conditional survival tree and random forest modeling. The first and second nodes in the conditional survival tree were percent REM with cut-points of 15.4% (similar to the lowest quartile cut-point) and 10.9%. Percentage N1 was the fifth node with a cut-point of 13.6% (Figure 14). The random survival forest model identified percent REM as the most important sleep stage for predicting survival (mean decrease in accuracy = 0.058). Percent N1 was a distant second at 0.001 (Figure 15). Both techniques found percent REM overwhelmingly important compared to other sleep stages, implying that contributions from other stages were inconsequential. This is consistent with the Cox results where the REM beta coefficient (0.17) was substantially higher than the N2 beta (0.06).

Meta-analysis

The meta-analysis of the three cohorts yielded an overall hazard ratio of 1.13 (95% CI 1.10–1.17) for all-cause mortality with very little variation between trials ($Chi^2 = 0.31$, $I^2 = 0\%$). The overall rate of mortality was significant for both genders with a 14% increased rate of mortality for men (HR = 1.14, 95% CI, 1.10–1.18) for each 5% reduction in REM and an $I^2 = 0\%$. There was an 11% increase for women (HR = 1.11, 95% CI, 1.04–1.19), however there was more variation

between trials ($Chi^2 = 2.93$, I^2 of 66%) using the fixed effects model. A random effects model was evaluated ($Tau^2 = 0.01$) for women due to this heterogeneity which yielded a HR = 1.18 (95% CI 1.04–1.19). For cardiovascular mortality, there was a 10% increase in mortality rate (95% CI 1.03–1.16) and the studies had little variation ($Chi^2 = 0.45$, I^2 of 0%). Table 15 contains all data used for the meta-analysis. Forest plots and summary of results are presented in Figures 16 to19.

Discussion

Survival analysis of older, community-based men found an association between lower percentages of REM sleep and increased mortality. This finding replicated in two independent datasets of middle-aged men and women. Similar effect sizes (HR ranging from 1.13–1.19 for a 5% decrease in REM) were observed in MrOS for all-cause, cardiovascular, cancer, and other mortality, even after adjusting for confounding demographic, sleep, and health-related covariates. These effect sizes are slightly larger than the mortality risk resulting from aging one year based on MrOS data (HR for one-year increase in age ranging from 1.11–1.16). Sensitivity analyses using MrOS data showed findings persisted in subgroups that survived the first two years of the study and those with sleep duration between 5–8 and 6–8 hours (except cardiovascular), without depression, severe sleep apnea, and in persons not using medications that may affect REM sleep.

The findings replicated in two independent, well-characterized sleep cohorts (WSC and SHHS) which is a critical step showing the validity, consistency, and generalizability of the results. Specifically, reproducing in the WSC generalized the finding from older white men to middle-aged, white men and women. The SHHS further expanded the generalizability as the cohort had a higher proportion of non-white participants. When stratified by gender, there was a higher rate of all-cause mortality in women compared to men in the WSC (1.34 vs 1.09). The SHHS data had a higher rate of all-cause mortality in men than women (1.16 vs 1.09).

A meta-analysis of the three cohorts produced an overall HR of 1.13 for all-cause mortality and 1.10 for cardio-vascular mortality. When stratified by gender, the rate of mortality was similar for men and women (HR= 1.14 vs 1.11). In sum, decreased REM was an indicator of increased mortality rate across both genders and a broad age range.

Since sleep is comprised of multiple sleep stages, it was critical to evaluate the role the other sleep stages may play in the association. Therefore, all sleep stages (not just REM) were included in the 6-fold cross-validation process used to empirically identify which covariates to include in the final models. Only percent N2 was selected for inclusion. In addition, I performed a random survival forest analysis to evaluate whether any of the other sleep stages could be driving the finding. As explained in the results section, REM sleep was found to be the most important sleep stage for predicting mortality based on results from the random forest model, the conditional inference trees, and the Cox REM beta coefficient compared to the N2 beta. Therefore, it is unlikely the association was caused by a different sleep stage.

This finding is supported by a recent publication evaluating REM and mortality in the Sleep Heart Health Study[108]. In their analysis, Zhang and colleagues found the proportion of REM sleep to be negatively associated with all-cause mortality even after adjustments for age, gender, race, BMI, smoking status, total cholesterol, triglycerides, high-density lipoprotein, history of diabetes, history of hypertension, and AHI (HR=0.972, 95% CI: 0.963–0.981). For this analysis, a 1% increase in percent REM was used compared to our analysis that used a 5% decrease.

In MrOS, mean TST actigraphy (in bed) was lower than expected. When combined with mean TST out of bed, total daily TST was similar to reports from a different, similar aged cohort[146]. A meta-analysis[76] suggested sleep duration is associated with obesity[101], hypertension[102], cardiovascular outcomes[103], and all-cause mortality[16]. Yin and colleagues[147] conducted a meta-analysis on sleep duration and all-cause mortality and cardiovascular events and observed a curvilinear U-shaped association with the lowest risk of all-cause mortality at a sleep duration of about 7 hours per day. Quantity of REM is impacted by sleep duration given the known circadian component to REM sleep (most of REM occurs in last third of the sleep period)[87]. Therefore, I controlled for total sleep time in multiple ways to ensure our finding was not a byproduct of the known association between sleep duration and mortality. First, I used percent time in REM rather than total minutes in REM, which factors in total sleep time from the overnight sleep study. Next, I included TST from both the overnight sleep study and home actigraphy in the 6-fold cross-validation process in the event of residual confounding. Only actigraphy TST was selected for inclusion. Finally, I performed sensitivity analyses limiting the population to participants who slept 5 to 8 and 6 to 8 hours/night and found no change in effect sizes except for cardiovascular related mortality in the 6 to 8 hour group. Therefore, I felt confident that TST was adequately addressed as a potential confounder.

Further, obstructive sleep apnea has been linked to mortality, stroke, and cardiovascular disease[104-107]. I found the REM and all-cause mortality association remained significant after excluding individuals with AHI>30 and/or using antidepressants, benzodiazepines, or sleep medications.

Dew and colleagues found an association between mortality and sleep latency >30m and sleep efficiency <80%[148]. Wallace and colleagues[149] used several statistical approaches to evaluate sleep health characteristics and found that rhythmicity and continuity were the strongest predictors of time to all-cause mortality. Individual sleep stages were not evaluated.

Although different outcomes were used, my findings were consistent with reports linking REM to other age-related diseases and conditions. Song and colleagues found that increased time in N1

sleep and less time in REM were associated with worsening cognitive performance in MrOS[98]. Smagula and colleagues evaluated the relationship between sleep stages and comorbid depression and anxiety in MrOS and found men with clinically significant depressive symptoms spent more time in N2 and less time in REM[150]. Suh and colleagues found a short average cycle length (sequence of NREM and REM sleep stages, both >2 min, and not interrupted by >2 min of wake) were significantly associated with cognitive decline[151 152]. While not directly related to mortality, these studies support the value of REM sleep and the importance of evaluating sleep stages independently from sleep duration.

Other researchers focused on the association between REM sleep and neurological disease and cardiovascular diseases. For example, Pase and colleagues[153] evaluated whether REM sleep could be linked to declarative memory impairment on the premise that REM abnormalities may be suggest deterioration of cholinergic pathways. They found that relatively shorter REM sleep is associated with a higher risk of developing dementia. Giubilei and colleagues[154] found the number and duration of REM phases were significantly reduced in the acute phase of ischemic stroke. This reduction correlated with both the severity of neurological deficit and with the anatomical site of the lesion. Another study found that patients with vascular cognitive impairment-no dementia had reduced REM sleep, among other polysomnographic differences, compared to controls[155].

The HypnoLaus study evaluated the relationship between different sleep stages and hypertension, diabetes, overweight/obesity, and metabolic syndrome. Although initially a higher prevalence of metabolic syndrome was observed in individuals with decreased REM, after multivariate adjustment the study concluded that normal variations in sleep stages contribute little to metabolic syndrome and associated disorders[156].

Sun and colleagues used machine learning to predict brain age from sleep EEG[157]. The resulting brain age for healthy individuals correlated well with their chronological age. However, there was evidence that individuals with significant neurological or psychiatric disease, hypertension or diabetes had a significantly higher brain age compared to their true age. The study extracted 102 features related to sleep staging and then averaged the features in each of the five sleep stages over time to yield 510 features per EEG. Most features related to relative bands of power spectra and macro-structure sleep features. Percent REM sleep may be another important, easy to interpret, aging biomarker.

Strengths and Limitations

This study has many strengths. The most important being replication of the original finding in not one, but two independent datasets. This provided strong evidence for the validity, consistency, and generalizability of the findings. In addition, the study used robust methods to evaluate the association through a combination of machine learning and traditional statistics. Numerous sensitivity analyses were conducted to control for potential biases. Finally, three instrumental, well-characterized, population-based sleep cohorts were used.

Limitations include the possibility of unmeasured and residual confounding. To address this concern, I used clinical knowledge and empirical model building to select covariates for the final models. MrOS did not include women and the population's mean age at baseline was 76.4 ± 5.5 years. However, validation in the WSC comprised of men and women with a lower mean age (51.5 ± 8.5 years) reduces the likelihood these factors influenced the findings. All three cohorts are comprised of community dwelling volunteers and therefore may be healthier than the general population, however I adjusted for comorbidities and do not believe this impacts the associations presented.

While replication provides generalizability across a larger age group and reduces, it does not eliminate the likelihood of reverse causality. Although >90% of the MrOS and WSC cohorts were Caucasian, replication in the SHHS broadens the generalizability somewhat to other races given >15% of the population was non-white.

For all three datasets, REM was quantified based on one night of polysomnography at baseline. Although it is possible the "first-night" effect biased our results, it is unlikely the effect would be differential with respect to mortality. In addition, a previous publication of the SHHS found no evidence of "first-night" effect[158].

Conclusion and Future Directions

In this analysis, I found a robust association between percent REM sleep and mortality, which persisted across different causes of death and multiple sensitivity analyses. The findings replicated in two independent cohorts comprised of different populations providing strong evidence for the validity, consistency, and generalizability of the findings. Based on this study, REM sleep appears to be a reliable predictor of mortality and may have other predictive health values. Strategies to preserve REM may influence clinical therapies and reduce mortality risk, particularly for adults with <15% REM. However, mechanistic studies are needed to better understand the relationship.

Given the complex underlying biological functions, further studies are required to understand whether the relationship is causal. Before expanding this research into other areas, it would be useful to improve internal reliability of the exposure by running all studies through an automated scoring algorithm. This would reduce variation in the exposure introduced by using manually annotated sleep stages.

Understanding how much within-subject variation is present over time and whether that change is associated with mortality will be an important addition to the REM / mortality story. All three cohorts used in this analysis have at least one additional visit with PSG data. Therefore, the relative change in REM over time can easily be determined. Even if the finding shows there is no association between change in REM and mortality, the results will be informative.

It is possible that accelerated brain aging is linked with lower levels of REM sleep, making it a marker rather than a direct mortality risk factor. The Mignot Lab already has a deep learning technique to predict age based on PSG that will soon be submitted for publication. This algorithm can be modified to explore the relationship between PSG derived brain age, quantity of REM sleep, and mortality rate. I would seek additional advice from Dr. M. Brandon Westover at Harvard University, who is also working in this area.

Finally, it would be worth exploring the genetics of REM given it is a robust phenotype known to have a genetic component in both animal and twin studies. I plan to perform a genome wide association study (GWAS) on quantity of REM sleep using a collection of international cohorts previously assembled by the Mignot Lab to evaluate the genetics of periodic leg movements during sleep. Then I would replicate any GWAS findings in the Sleep Apnea Global Interdisciplinary Consortium (SAGIC) dataset[47].

Currently measuring REM sleep currently requires EEG, making it difficult to effectively monitor REM sleep outside of a sleep clinic and reducing the utility of REM as a biomarker. However, there have been significant advances in wearable devices (i.e., actigraphy). This rapid expansion and consumer acceptance of wearable devices offers a low burden mechanism to monitor sleep over time. A recent study I was involved in found that when compared to PSG, sleep and wake estimates generated by a consumer-grade wearable were comparable to those generated by a clinical-grade actigraph[159]. Newer devices include heart rate data, making it possible to differentiate sleep stages[160]. The ability to use consumer grade actigraphy devices to accurately evaluate sleep stages over multiple nights will reduce barriers for easily evaluating REM in the general population.

Cohort Acknowledgements

The Osteoporotic Fractures in Men (MrOS) Study is supported by National Institutes of Health funding. The following institutes provide support: the National Institute on Aging (NIA), the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), the National Center for Advancing Translational Sciences (NCATS), and NIH Roadmap for Medical Research under the following grant numbers: 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 (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 Wisconsin Sleep Cohort was supported by grants R01HL62252, RR03186, and R01AG14124 from the National Institutes of Health.

The Sleep Heart Health Study (SHHS) acknowledges the Atherosclerosis Risk in Communities Study, the Cardiovascular Health Study, the Framingham Heart Study, the Cornell/Mt. Sinai Worksite and Hypertension Studies, the Strong Heart Study, the Tucson Epidemiologic Study of Airways Obstructive Diseases, and the Tucson Health and Environment Study for allowing their cohort members to be part of the SHHS and for permitting the use of the data acquired by them. SHHS is particularly grateful to the members of these cohorts who agreed to participate in SHHS as well. SHHS further recognizes all of the investigators and staff who have contributed to its success. A list of SHHS investigators, staff, and their participating institutions is available on the SHHS website, https://sleepdata.org/datasets/shhs. The SHHS supported by grants (U01HL53916, U01HL53931, U01HL53934, U01HL53937, U01HL53938, U01HL53940, U01HL53941, U01HL64360) from the National Institutes of Health.

References

1. Keene AC, Duboue ER. The origins and evolution of sleep. *The Journal of experimental biology.* Jun 12 2018;221(Pt 11).

2. Howell WH. Le problème physiologique du Sommeil. *Science* 04/04/1913 1913;Vol. 37(953):525-526.

3. Dement WC. A brief history of sleep research. 1999; https://web.stanford.edu/~dement/history.html. Accessed 03/17, 2020.

4. Michel Jouvet, pioneering sleep scientist – obituary https://www.telegraph.co.uk/obituaries/2017/10/12/michel-jouvet-pioneering-sleep-scientist-obituary/. Accessed 03/17, 2020.

5. Hallmayer J, Faraco J, Lin L, et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. *Nature genetics.* Jun 2009;41(6):708-711.

6. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep.* Dec 1986;9(4):519-524.

7. Lin L, Faraco J, Li R, et al. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell.* Aug 6 1999;98(3):365-376.

8. Chemelli RM, Willie JT, Sinton CM, et al. Narcolepsy in orexin knockout mice: molecular genetics of sleep regulation. *Cell.* Aug 20 1999;98(4):437-451.

9. Han F, Lin L, Warby SC, et al. Narcolepsy onset is seasonal and increased following the 2009 H1N1 pandemic in China. *Annals of neurology.* Sep 2011;70(3):410-417.

10. Kornum BR, Kawashima M, Faraco J, et al. Common variants in P2RY11 are associated with narcolepsy. *Nature genetics.* Jan 2011;43(1):66-71.

11. Winkelmann J, Lin L, Schormair B, et al. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Human molecular genetics.* May 15 2012;21(10):2205-2210.

12. Kadotani H, Kadotani T, Young T, et al. Association between apolipoprotein E epsilon4 and sleep-disordered breathing in adults. *Jama.* Jun 13 2001;285(22):2888-2890.

13. Ito E, Inoue Y. [The International Classification of Sleep Disorders, third edition. American Academy of Sleep Medicine. Includes bibliographies and index]. *Nihon rinsho. Japanese journal of clinical medicine.* Jun 2015;73(6):916-923.

14. Medicine Io. Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem. Washington, DC: The National Academies Press; 2006.

15. *Social and Economic Dimensions of Sleep Disorders.* Vol 12-1 Elsevier; February 2017

16. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep.* May 2010;33(5):585-592.

17. Khan H, Kella D, Kunutsor SK, Savonen K, Laukkanen JA. Sleep Duration and Risk of Fatal Coronary Heart Disease, Sudden Cardiac Death, Cancer Death, and All-Cause Mortality. *The American journal of medicine.* Dec 2018;131(12):1499-1505 e1492.

18. Kay DB, Karim HT, Soehner AM, et al. Sleep-Wake Differences in Relative Regional Cerebral Metabolic Rate for Glucose among Patients with Insomnia Compared with Good Sleepers. *Sleep.* Oct 1 2016;39(10):1779-1794.

19. Benca RM, Obermeyer WH, Thisted RA, Gillin JC. Sleep and psychiatric disorders. A meta-analysis. *Archives of general psychiatry.* Aug 1992;49(8):651-668; discussion 669-670.

20. Wetter TC, Zils E, Fulda S. Sleep and psychiatric disorders. *Sleep medicine.* 2008:170-185.

21. Brzecka A, Leszek J, Ashraf GM, et al. Sleep Disorders Associated With Alzheimer's Disease: A Perspective. *Frontiers in neuroscience.* 2018;12:330.

22. Ancoli-Israel S, Clopton P, Klauber MR, Fell R, Mason W. Use of wrist activity for monitoring sleep/wake in demented nursing-home patients. *Sleep.* Jan 1997;20(1):24-27.

23. Sariarslan HA, Gulhan YB, Unalan D, Basturk M, Delibas S. The relationship of sleep problems to life quality and depression. *Neurosciences.* Jul 2015;20(3):236-242.

24. Rechtschaffen A KA, editors. Los Angeles: Brain Information Service/Brain Research Institute. A manual of standardized terminology, techniques and scoring system of sleep stages in human subjects. University of California; 1968.

25. (AASM) AAoSM. Report of the 2017 – 2018 AASM president. 2017; Dr. Ilene Rosen's address at the AASM General Membership Meeting. Available at: https://aasm.org/report-of-the-2017-2018-aasm-president/. Accessed 03/18, 2020.

26. Technologists BoRP. About the RPSGT Credential. 2019; https://www.brpt.org/rpsgt/about-the-rpsgt-credential/ Accessed 03/18, 2020.

27. Zhang GQ, Cui L, Mueller R, et al. The National Sleep Research Resource: towards a sleep data commons. *Journal of the American Medical Informatics Association : JAMIA.* Oct 1 2018;25(10):1351-1358.

28. Dean DA, 2nd, Goldberger AL, Mueller R, et al. Scaling Up Scientific Discovery in Sleep Medicine: The National Sleep Research Resource. *Sleep.* May 1 2016;39(5):1151-1164.

29. Malhotra A, Younes M, Kuna ST, et al. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep.* Apr 1 2013;36(4):573-582.

30. Stephansen JB, Olesen AN, Olsen M, et al. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. *Nature communications.* Dec 6 2018;9(1):5229.

31. Penzel T, Hirshkowitz M, Harsh J, et al. Digital analysis and technical specifications. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine.* Mar 15 2007;3(2):109-120.

32. Berry RB BR, Gamaldo CE, et al; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications. Version 2.4.* Darien, IL American Academy of Sleep Medicine; 2017.

33. Kuna ST, Benca R, Kushida CA, et al. Agreement in computer-assisted manual scoring of polysomnograms across sleep centers. *Sleep.* Apr 1 2013;36(4):583-589.

34. Kornum BR, Jennum P. The case for narcolepsy as an autoimmune disease. *Expert review of clinical immunology.* Jan 27 2020:1-3.
35. Zawilska JB, Woldan-Tambor A, Plocka A, Kuzajska K, Wojcieszak J. [Narcolepsy: etiology, clinical features, diagnosis and treatment]. *Postepy higieny i medycyny doswiadczalnej.* Oct 24 2012;66:771-786.
36. Ahmed I, Thorpy M. Clinical features, diagnosis and treatment of narcolepsy. *Clinics in chest medicine.* Jun 2010;31(2):371-381.
37. Scammell TE. The neurobiology, diagnosis, and treatment of narcolepsy. *Annals of neurology.* Feb 2003;53(2):154-166.
38. Thorpy M. Current concepts in the etiology, diagnosis and treatment of narcolepsy. *Sleep medicine.* Jan 2001;2(1):5-17.
39. Dement WC, Carskadon MA, Guilleminault C, Zarcone VP. Narcolepsy. Diagnosis and treatment. *Primary care.* Dec 1976;3(4):609-623.
40. Andlauer O, Moore H, Jouhier L, et al. Nocturnal rapid eye movement sleep latency for identifying patients with narcolepsy/hypocretin deficiency. *JAMA neurology.* Jul 2013;70(7):891-902.
41. Zhang X, Xu M, Li Y, et al. Automated multi-model deep neural network for sleep stage scoring with unfiltered clinical data. *Sleep & breathing = Schlaf & Atmung.* Jan 14 2020.
42. Weaver TE, Sawyer AM. Adherence to continuous positive airway pressure treatment for obstructive sleep apnoea: implications for future interventions. *The Indian journal of medical research.* Feb 2010;131:245-258.
43. Blinder H, Momoli F, Bokhaut J, et al. Predictors of adherence to positive airway pressure therapy in children: a systematic review and meta-analysis. *Sleep medicine.* Jan 11 2020;69:19-33.
44. DiFeo N, Meltzer LJ, Beck SE, et al. Predictors of positive airway pressure therapy adherence in children: a prospective study. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine.* Jun 15 2012;8(3):279-286.
45. Ye L, Pien GW, Ratcliffe SJ, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. *The European respiratory journal.* Dec 2014;44(6):1600-1607.
46. Kim J, Keenan BT, Lim DC, Lee SK, Pack AI, Shin C. Symptom-Based Subgroups of Koreans With Obstructive Sleep Apnea. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine.* Mar 15 2018;14(3):437-443.
47. Keenan BT, Kim J, Singh B, et al. Recognizable clinical subtypes of obstructive sleep apnea across international sleep centers: a cluster analysis. *Sleep.* Mar 1 2018;41(3).
48. Pack AI. Further Development of P4 Approach to Obstructive Sleep Apnea. *Sleep medicine clinics.* Sep 2019;14(3):379-389.
49. Mazzotti DR, Keenan BT, Lim DC, Gottlieb DJ, Kim J, Pack AI. Symptom Subtypes of Obstructive Sleep Apnea Predict Incidence of Cardiovascular Outcomes. *American journal of respiratory and critical care medicine.* Aug 15 2019;200(4):493-506.

50. Leary EB, Barger L, Hall-Porter JM, et al. Alliance Sleep Questionnaire (ASQ): A Collaborative Online Sleep Assessment Questionnaire. *Sleep.* 2012;35(Abstract Supplement):A432.
51. Leary EB, Griffin KS, Malunjkar S, et al. Alliance Sleep Questionnaire (ASQ) Feasibility Pilot Study. *Sleep.* 2012;35(Abstract Supplement):A432.
52. Leary EB, Malunjkar S, Qadri S, et al. Deployment of the Alliance Sleep Questionnaire (ASQ) at the Stanford Sleep Disorders Clinic (SSDC) as Standard of Care. *Sleep.* 2013;36(Abstract Supplement):A413.
53. Leary EB, Einen M, Malunjkar S, Ruoff CM, Walsh JK, Mignot E. Validation of the Alliance Sleep Questionnaire (ASQ) Narcolepsy Module in Sleep Disordered Patients. *Sleep.* 2014;37(Abstract Supplement):A360.
54. Leary EB, Malunjkar S, Ruoff CM, Walsh JK, Mignot E. Validation of the Alliance Sleep Questionnaire (ASQ) Restless Legs Syndrome Module Module in Sleep Disordered Patients. *Sleep.* 2015;38(Abstract Supplement).
55. Maislin G, Pack AI, Kribbs NB, et al. A survey screen for prediction of apnea. *Sleep.* Apr 1995;18(3):158-166.
56. Leary EB, Malunjkar S, Cheung J, Schneider L, Walsh JK, Mignot E. Validation of the Alliance Sleep Questionnaire (ASQ) Obstructive Sleep Apnea (OSA) Module in Sleep Disordered Patients. *Sleep.* 2016((Abstract suppl.)):A124.
57. Bastien CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep medicine.* Jul 2001;2(4):297-307.
58. Leary EB, Joergensen SH, Malunjkar S, Kim SJ, Mignot E, Barwick FH. Insomnia Module in Sleep Disordered Patients. *Sleep.* 2017(Abstract suppl.):A117.
59. Milinovich A, Kattan MW. Extracting and utilizing electronic health data from Epic for research. *Annals of translational medicine.* Feb 2018;6(3):42.
60. OHDSI. Observational Health Data Sciences and Informatics. 2020; https://ohdsi.org/. Accessed 03/21, 2020.
61. Ioannidis JP. Why most published research findings are false. *PLoS medicine.* Aug 2005;2(8):e124.
62. Collaboration OS. Estimating the Reproducibility of Psychological Science. 2011; https://osf.io/ezcuj/wiki/home/. Accessed 05-03-2020.
63. Pablos-Mendez A, Klag M, Goldman L. Open-source collaboration for Global Health: Science and Practice. *Global health, science and practice.* Mar 2013;1(1):1-2.
64. Open Science C. PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science.* Aug 28 2015;349(6251):aac4716.
65. Gorgolewski KJ, Margulies DS, Milham MP. Making data sharing count: a publication-based solution. *Frontiers in neuroscience.* 2013;7:9.
66. Pesonen AK, Gradisar M, Kuula L, et al. REM sleep fragmentation associated with depressive symptoms and genetic risk for depression in a community-based sample of adolescents. *Journal of affective disorders.* Feb 15 2019;245:757-763.
67. Hou J, Shen Q, Wan X, Zhao B, Wu Y, Xia Z. REM sleep deprivation-induced circadian clock gene abnormalities participate in hippocampal-dependent memory

impairment by enhancing inflammation in rats undergoing sevoflurane inhalation. *Behavioural brain research.* Feb 16 2019;364:167-176.

68. Yang L, Huang Z, Zhong J, Qiu S, Liu D. The characteristics and clinical significance of REM<10% in children with sleep-disordered breathing. *Int J Pediatr Otorhinolaryngol.* Jul 2018;110:53-56.

69. Targa ADS, Noseda ACD, Rodrigues LS, Aurich MF, Lima MMS. REM sleep deprivation and dopaminergic D2 receptors modulation increase recognition memory in an animal model of Parkinson's disease. *Behavioural brain research.* Feb 26 2018;339:239-248.

70. Basheer R, Magner M, McCarley RW, Shiromani PJ. REM sleep deprivation increases the levels of tyrosine hydroxylase and norepinephrine transporter mRNA in the locus coeruleus. *Brain research. Molecular brain research.* Jun 15 1998;57(2):235-240.

71. Peter JH, Grote L, Fus E, Ploch T, Stammnitz A. REM-sleep-hypertension in obstructive sleep apnea. *European journal of medical research.* Dec 18 1995;1(3):132-136.

72. Siegel JM. Phylogeny and the function of REM sleep. *Behavioural brain research.* Jul-Aug 1995;69(1-2):29-34.

73. Crick F, Mitchison G. REM sleep and neural nets. *Behavioural brain research.* Jul-Aug 1995;69(1-2):147-155.

74. Marks GA, Shaffery JP, Oksenberg A, Speciale SG, Roffwarg HP. A functional role for REM sleep in brain maturation. *Behavioural brain research.* Jul-Aug 1995;69(1-2):1-11.

75. Borbely AA, Baumann F, Brandeis D, Strauch I, Lehmann D. Sleep deprivation: effect on sleep stages and EEG power density in man. *Electroencephalography and clinical neurophysiology.* May 1981;51(5):483-495.

76. da Silva AA, de Mello RG, Schaan CW, Fuchs FD, Redline S, Fuchs SC. Sleep duration and mortality in the elderly: a systematic review with meta-analysis. *BMJ open.* Feb 17 2016;6(2):e008119.

77. Kronholm E, Laatikainen T, Peltonen M, Sippola R, Partonen T. Self-reported sleep duration, all-cause mortality, cardiovascular mortality and morbidity in Finland. *Sleep medicine.* Mar 2011;12(3):215-221.

78. Orwoll E, Blank JB, Barrett-Connor E, et al. 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.* Oct 2005;26(5):569-585.

79. Breiman L. Bagging predictors. *Mach Learn.* Aug 1996;24(2):123-140.

80. Breiman L. Random forests. *Mach Learn.* Oct 2001;45(1):5-32.

81. Young T. Rationale, design and findings from the Wisconsin Sleep Cohort Study: Toward understanding the total societal burden of sleep disordered breathing. *Sleep medicine clinics.* Mar 1 2009;4(1):37-46.

82. Young T, Palta M, Dempsey J, Peppard PE, Nieto FJ, Hla KM. Burden of sleep apnea: rationale, design, and major findings of the Wisconsin Sleep Cohort study. *WMJ : official publication of the State Medical Society of Wisconsin.* Aug 2009;108(5):246-249.

83. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *NEJM.* Apr 29 1993;328(17):1230-1235.
84. Quan SF, Howard BV, Iber C, et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep.* Dec 1997;20(12):1077-1085.
85. Kalia M. Neurobiology of sleep. *Metabolism: clinical and experimental.* Oct 2006;55(10 Suppl 2):S2-6.
86. Dement W. The effect of dream deprivation. *Science.* Jun 10 1960;131(3415):1705-1707.
87. Endo S, Kobayashi T, Yamamoto T, Fukuda H, Sasaki M, Ohta T. Persistence of the circadian rhythm of REM sleep: a variety of experimental manipulations of the sleep-wake cycle. *Sleep.* Sep 1981;4(3):319-328.
88. Smith C, Rose GM. Evidence for a paradoxical sleep window for place learning in the Morris water maze. *Physiology & behavior.* Jan 1996;59(1):93-97.
89. Menz MM, Rihm JS, Salari N, et al. The role of sleep and sleep deprivation in consolidating fear memories. *NeuroImage.* Jul 15 2013;75:87-96.
90. Mueller AD, Meerlo P, McGinty D, Mistlberger RE. Sleep and adult neurogenesis: implications for cognition and mood. *Current topics in behavioral neurosciences.* 2015;25:151-181.
91. de Andres I, Garzon M, Reinoso-Suarez F. Functional Anatomy of Non-REM Sleep. *Frontiers in neurology.* 2011;2:70.
92. Moraes W, Piovezan R, Poyares D, Bittencourt LR, Santos-Silva R, Tufik S. Effects of aging on sleep structure throughout adulthood: a population-based study. *Sleep medicine.* Apr 2014;15(4):401-409.
93. Miner B, Kryger MH. Sleep in the Aging Population. *Sleep medicine clinics.* Mar 2017;12(1):31-38.
94. Floyd JA, Janisse JJ, Jenuwine ES, Ager JW. Changes in REM-sleep percentage over the adult lifespan. *Sleep.* Jul 2007;30(7):829-836.
95. Vaz Fragoso CA, Gill TM. Sleep complaints in community-living older persons: a multifactorial geriatric syndrome. *Journal of the American Geriatrics Society.* Nov 2007;55(11):1853-1866.
96. Marengoni A, Angleman S, Melis R, et al. Aging with multimorbidity: a systematic review of the literature. *Ageing research reviews.* Sep 2011;10(4):430-439.
97. Reid KJ, Martinovich Z, Finkel S, et al. Sleep: a marker of physical and mental health in the elderly. *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry.* Oct 2006;14(10):860-866.
98. Song Y, Blackwell T, Yaffe K, Ancoli-Israel S, Redline S, Stone KL. Relationships between sleep stages and changes in cognitive function in older men: the MrOS Sleep Study. *Sleep.* Mar 1 2015;38(3):411-421.
99. Buysse DJ. Sleep health: can we define it? Does it matter? *Sleep.* Jan 1 2014;37(1):9-17.
100. Chaput JP, Dutil C, Sampasa-Kanyinga H. Sleeping hours: what is the ideal number and how does age impact this? *Nature and science of sleep.* 2018;10:421-430.

101. Cappuccio FP, Taggart FM, Kandala NB, et al. Meta-analysis of short sleep duration and obesity in children and adults. *Sleep.* May 2008;31(5):619-626.
102. Wang Q, Xi B, Liu M, Zhang Y, Fu M. Short sleep duration is associated with hypertension risk among adults: a systematic review and meta-analysis. *Hypertension research : official journal of the Japanese Society of Hypertension.* Oct 2012;35(10):1012-1018.
103. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. *European heart journal.* Jun 2011;32(12):1484-1492.
104. Redline S, Yenokyan G, Gottlieb DJ, et al. Obstructive sleep apnea-hypopnea and incident stroke: the sleep heart health study. *American journal of respiratory and critical care medicine.* Jul 15 2010;182(2):269-277.
105. Ramos AR, Figueredo P, Shafazand S, et al. Obstructive Sleep Apnea Phenotypes and Markers of Vascular Disease: A Review. *Frontiers in neurology.* 2017;8:659.
106. Arzt M, Young T, Finn L, Skatrud JB, Bradley TD. Association of sleep-disordered breathing and the occurrence of stroke. *American journal of respiratory and critical care medicine.* Dec 1 2005;172(11):1447-1451.
107. Yaggi HK, Concato J, Kernan WN, Lichtman JH, Brass LM, Mohsenin V. Obstructive sleep apnea as a risk factor for stroke and death. *The New England journal of medicine.* Nov 10 2005;353(19):2034-2041.
108. Zhang J, Jin X, Li R, Gao Y, Li J, Wang G. Influence of rapid eye movement sleep on all-cause mortality: a community-based cohort study. *Aging.* Mar 13 2019;11(5):1580-1588.
109. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* Jan 1999;10(1):37-48.
110. Schipf S, Knuppel S, Hardt J, Stang A. [Directed acyclic graphs (DAGs) - the application of causal diagrams in epidemiology]. *Gesundheitswesen.* Dec 2011;73(12):888-892.
111. Harrell FE. *Regression modeling strategies: With applications to linear models, logistic regression, and survival*
analysis. New York: Springer-Verlag; 2001.
112. Annesi I, Moreau T, Lellouch J. Efficiency of the logistic regression and Cox proportional hazards models in longitudinal studies. *Statistics in medicine.* Dec 1989;8(12):1515-1521.
113. http://mrosdata.sfcc-cpmc.net.
114. Blank JB, Cawthon PM, Carrion-Petersen ML, et al. Overview of recruitment for the osteoporotic fractures in men study (MrOS). *Contemporary clinical trials.* Oct 2005;26(5):557-568.
115. Blackwell T, Yaffe K, Ancoli-Israel S, et al. Association of sleep characteristics and cognition in older community-dwelling men: the MrOS sleep study. *Sleep.* Oct 1 2011;34(10):1347-1356.
116. Mehra R, Stone KL, Blackwell T, et al. Prevalence and correlates of sleep-disordered breathing in older men: osteoporotic fractures in men sleep study. *Journal of the American Geriatrics Society.* Sep 2007;55(9):1356-1364.
117. Winkelman JW, Blackwell T, Stone K, Ancoli-Israel S, Redline S. Associations of Incident Cardiovascular Events With Restless Legs Syndrome and Periodic

Leg Movements of Sleep in Older Men, for the Outcomes of Sleep Disorders in Older Men Study (MrOS Sleep Study). *Sleep.* Apr 1 2017;40(4).
118. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep.* Aug 1992;15(4):376-381.
119. Smyth C. The Pittsburgh Sleep Quality Index (PSQI). *Journal of gerontological nursing.* Dec 1999;25(12):10-11.
120. Weaver TE, Laizner AM, Evans LK, et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. *Sleep.* Oct 1997;20(10):835-843.
121. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *The New England journal of medicine.* Apr 29 1993;328(17):1230-1235.
122. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The Report of an American Academy of Sleep Medicine Task Force. *Sleep.* Aug 1 1999;22(5):667-689.
123. Redline S, Sanders MH, Lind BK, et al. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. Sleep Heart Health Research Group. *Sleep.* Nov 1 1998;21(7):759-767.
124. Blackwell T, Redline S, Ancoli-Israel S, et al. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. *Sleep.* Feb 2008;31(2):283-291.
125. Blackwell T, Yaffe K, Ancoli-Israel S, et al. 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.* Dec 2011;59(12):2217-2225.
126. Punjabi NM, Caffo BS, Goodwin JL, et al. Sleep-disordered breathing and mortality: a prospective cohort study. *PLoS medicine.* Aug 2009;6(8):e1000132.
127. Yesavage JA. Geriatric Depression Scale. *Psychopharmacology bulletin.* 1988;24(4):709-711.
128. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *The Journal of clinical psychiatry.* Aug 1987;48(8):314-318.
129. Canadian study of health and aging: study methods and prevalence of dementia. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* Mar 15 1994;150(6):899-913.
130. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *Journal of clinical epidemiology.* Feb 1993;46(2):153-162.
131. Pahor M, Chrischilles EA, Guralnik JM, Brown SL, Wallace RB, Carbonin P. Drug data coding and analysis in epidemiologic studies. *European journal of epidemiology.* Aug 1994;10(4):405-411.
132. McHorney CA, Ware JE, Jr., Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical care.* Jan 1994;32(1):40-66.

133. McHorney CA, Ware JE, Jr., Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical care.* Mar 1993;31(3):247-263.
134. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care.* Jun 1992;30(6):473-483.
135. Liu WW, Hu LP. [Survival analysis using SAS software package]. *Zhong xi yi jie he xue bao = Journal of Chinese integrative medicine.* May 2010;8(5):489-493.
136. *R: A language and environment for statistical computing.* [computer program]. Vienna, Austria2014.
137. Version Version 5.3. Copenhagen: The Nordic Cochrane Centre: The Cochrane Collaboration; 2014.
138. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Annals of translational medicine.* Jan 2016;4(2):30.
139. Kirwan JR. Reporting of significance levels versus Pearson's correlation coefficient. *British journal of rheumatology.* Aug 1984;23(3):232-233.
140. Arlot S, Celisse A. A survey of cross-validation procedures for model selection. *Stat Surv.* 2010;4:40-79.
141. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. *International journal of Ayurveda research.* Oct 2010;1(4):274-278.
142. Blackwell T, Yaffe K, Ancoli-Israel S, et al. Association of sleep characteristics and cognition in older community-dwelling men: the MrOS sleep study. *Sleep.* Oct 1 2011;34(10):1347-1356.
143. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods.* Dec 2009;14(4):323-348.
144. Smagula SF, Stone KL, Redline S, et al. Actigraphy- and Polysomnography-Measured Sleep Disturbances, Inflammation, and Mortality Among Older Men. *Psychosom Med.* Jul-Aug 2016;78(6):686-696.
145. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ.* Sep 6 2003;327(7414):557-560.
146. Kurina LM, Thisted RA, Chen JH, McClintock MK, Waite LJ, Lauderdale DS. Actigraphic sleep characteristics among older Americans. *Sleep health.* Dec 2015;1(4):285-292.
147. Yin J, Jin X, Shan Z, et al. Relationship of Sleep Duration With All-Cause Mortality and Cardiovascular Events: A Systematic Review and Dose-Response Meta-Analysis of Prospective Cohort Studies. *Journal of the American Heart Association.* Sep 9 2017;6(9).
148. Dew MA, Hoch CC, Buysse DJ, et al. Healthy older adults' sleep predicts all-cause mortality at 4 to 19 years of follow-up. *Psychosom Med.* Jan-Feb 2003;65(1):63-73.
149. Wallace ML, Stone K, Smagula SF, et al. Which Sleep Health Characteristics Predict All-Cause Mortality in Older Men? An Application of Flexible Multivariable Approaches. *Sleep.* Jan 1 2018;41(1).
150. Smagula SF, Reynolds CF, 3rd, Ancoli-Israel S, et al. Sleep Architecture and Mental Health Among Community-Dwelling Older Men. *The journals of*

gerontology. Series B, Psychological sciences and social sciences. Sep 2015;70(5):673-681.

151. Suh SW, Han JW, Lee JR, et al. Short Average Duration of NREM/REM Cycle Is Related to Cognitive Decline in an Elderly Cohort: An Exploratory Investigation. *Journal of Alzheimer's disease : JAD.* Jul 15 2019.

152. Han JW, Kim TH, Kwak KP, et al. Overview of the Korean Longitudinal Study on Cognitive Aging and Dementia. *Psychiatry investigation.* Aug 2018;15(8):767-774.

153. Pase MP, Himali JJ, Grima NA, et al. Sleep architecture and the risk of incident dementia in the community. *Neurology.* Sep 19 2017;89(12):1244-1250.

154. Giubilei F, Iannilli M, Vitale A, et al. Sleep patterns in acute ischemic stroke. *Acta Neurol Scand.* Dec 1992;86(6):567-571.

155. Jiang B, Ding C, Yao G, et al. Polysomnographic abnormalities in patients with vascular cognitive impairment-no dementia. *Sleep medicine.* Nov 2013;14(11):1071-1075.

156. Haba-Rubio J, Marques-Vidal P, Andries D, et al. Objective sleep structure and cardiovascular risk factors in the general population: the HypnoLaus Study. *Sleep.* Mar 1 2015;38(3):391-400.

157. Sun H, Paixao L, Oliva JT, et al. Brain age from the electroencephalogram of sleep. *Neurobiology of aging.* Feb 2019;74:112-120.

158. Quan SF, Griswold ME, Iber C, et al. Short-term variability of respiration and sleep during unattended nonlaboratory polysomnography--the Sleep Heart Health Study. [corrected]. *Sleep.* Dec 2002;25(8):843-849.

159. Cheung J, EB. L, H. L, JM. Z, E. M. PSG Validation of minute-to-minute scoring for sleep and wake periods in a consumer wearable device. *Sleep medicine.* 2020;Under Review.

160. Walch O, Huang Y, Forger D, Goldstein C. Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep.* Dec 24 2019;42(12).

Tables

Table 1: Baseline Demographic, Lifestyle, and Health Characteristics of the MrOS Cohort, Overall and by Percent Time in REM Quartiles (Q1–Q4)[a]

Characteristic	Overall (n=2,675)	Q1: <14.8% (n=677)	Q2: 14.8–19.5% (n=662)	Q3: 19.5–23.6% (n=667)	Q4: >23.6% (n=669)
Age at sleep visit (years)	76.3 ± 5.5	77.5 ± 5.7	76.4 ± 5.5	75.7 ± 5.5	75.7± 5.2
Age at last follow up (years)	86.6 ± 5.2	86.8 ± 5.6	86.7 ± 5.2	86.5 ± 5.2	86.4 ± 4.9
Race					
White	2,448 (91.5%)	626 (92.5%)	616 (93.1%)	607 (91.0%)	599 (89.5%)
non-White	227 (8.5%)	51 (7.5%)	46 (6.9%)	60 (9.0%)	70 (10.5 %)
Education					
Less than high school degree	130 (4.9%)	46 (6.8%)	33 (5.0%)	23 (3.4%)	28 (4.2%)
High school degree, some college	1,035 (38.7%)	263 (38.8%)	268 (40.5%)	255 (38.2%)	249 (37.2%)
College degree or higher	1510 (56.4%)	368 (54.4%)	361 (54.5%)	389 (58.3%)	392 (58.6%)
Body Mass Index (kg/m^2)	27.2 ± 3.8	27.4 ± 4.1	27.2 ± 3.9	27.0 ± 3.6	27.2 ± 3.6
Average neck circumference (cm)	39.4 ± 2.8	39.5 ± 3.0	39.5 ± 2.9	39.4 ± 2.7	39.4 ± 2.7
Average hip circumference (cm)	102.9 ± 8.5	102.9 ± 8.9	103.0 ± 8.4	102.7 ± 8.2	103.1 ± 8.6
Smoking Status					
Never	1,056 (39.5%)	272 (40.2%)	247 (37.3%)	283 (42.4%)	254 (38.0%)
Past	1,563 (58.5%)	387 (57.2%)	403 (60.9%)	371 (55.6%)	402 (60.1%)
Current	55 (2.1%)	17 (2.5%)	12 (1.8%)	13 (1.9%)	13 (1.9%)
Daily caffeine intake (mg/day)	235.7 ± 246.6	227.0 ± 239.9	235.0 ± 232.9	230.0 ± 249.9	250.8 ± 262.7
Alcohol Use (drinks/week) (n=2,664)	1.9 ± 1.7	1.7 ± 1.6	2.0 ± 1.8	2.0 ± 1.7	1.9 ± 1.7
Weekly Alcohol Use (n=2,664)					
0 Drinks	1,246 (46.8%)	345 (51.3%)	299 (45.3%)	288 (43.2%)	314 (47.1%)
1-4 Drinks	1,271 (47.7%)	301 (44.8%)	319 (48.3%)	335 (50.3%)	316 (47.4%)
>5 Drinks	147 (5.5%)	26 (3.9%)	42 (6.4%)	43 (6.5%)	36 (5.4%)
Medication Use					

Current use of antidepressant	201 (7.5%)	83 (12.3%)	40 (6.0%)	44 (6.6%)	34 (5.1%)
Current use of benzodiazepine	125 (4.7%)	40 (5.9%)	33 (5.0%)	20 (3.0%)	32 (4.8%)
Current use of sleep medications	321 (1.2%)	84 (12.4%)	88 (13.3%)	70 (10.5%)	79 (11.8%)
Hypertension	1,343 (50.2%)	381 (56.4%)	321 (48.5%)	326 (48.9%)	315 (47.1%)
Angina	402 (15.0%)	119 (17.6%)	101 (15.3%)	88 (13.2%)	94 (14.1%)
Stroke	97 (3.6%)	29 (4.3%)	26 (3.9%)	19 (2.8%)	23 (3.4%)
Heart attack	464 (17.4%)	141 (20.9%)	108 (16.3%)	105 (15.7%)	110 (16.4%)
Transient ischemic attack	254 (9.5%)	82 (12.1%)	72 (10.9%)	47 (7.1%)	53 (7.9%)
Congestive heart failure	161 (6.0%)	56 (8.3%)	33 (5.0%)	32 (4.8%)	40 (6.0%)
Type 2 diabetes	353 (13.2%)	102 (15.1%)	83 (12.5%)	84 (12.6%)	84 (12.6%)
Chronic obstructive pulmonary disease	139 (5.2%)	46 (6.8%)	37 (5.6%)	23 (3.4%)	33 (4.9%)
Osteoarthritis	649 (24.3%)	174 (25.7%)	158 (23.9%)	162 (24.3%)	155 (23.2%)
Rheumatoid arthritis	217 (8.1%)	69 (10.2%)	49 (7.4%)	45 (6.7%)	54 (8.1%)
Teng Mini Mental State Examination (3TMS) Score (0 – 100)	92.9 ± 6.1	92.0 ± 7.2	92.9 ± 6.1	93.4 ± 5.2	93.1 ± 5.4
Cognitive impairment based on 3TMS (score <= 77)	57 (2.1%)	21 (3.1%)	15 (2.3%)	10 (1.5%)	11 (1.6%)
Physical Activity Scale for the Elderly Score (0 – 400)	145.4 ± 70.4	134.5 ± 72.4	148.7 ± 69.8	148.2 ± 68.3	150.5 ± 69.9
Geriatric Depression Score (0 – 15)	1.7 ± 2.1	2.1 ± 2.4	1.7 ± 2.1	1.5 ± 1.8	1.6 ± 2.0

[a] Entries are mean ± standard deviation or n (%)

Table 2: Baseline Sleep Data from the MrOS Cohort, Overall and by Percent Time in REM Quartiles (Q1–Q4)[a]

Characteristic	Overall (n=2,675)	Q1: <14.8% (n=677)	Q2: 14.8–19.5% (n=662)	Q3: 19.5–23.6% (n=667)	Q4: >23.6% (n=669)
Objective Sleep Measures					
Stage N1 sleep (% time)	6.8 ± 4.3	7.8 ± 5.3	7.0 ± 4.3	6.3 ± 3.5	6.0 ± 3.4
Stage N1 sleep (min)	23.6 ± 13.5	24.7 ± 14.7	24.7 ± 13.8	23.1 ± 12.5	22.0 ± 12.6
Stage N2 sleep (% time)	62.8 ± 9.6	69.2 ± 9.6	63.8 ± 8.8	61.5 ± 7.5	56.7 ± 7.8
Stage N2 sleep (min)	223.4 ± 53.4	228.7 ± 63.6	229.5 ± 53.3	226.6 ± 46.4	209.0 ± 45.7
Stage N3 sleep (% time)	11.2 ± 8.9	12.3 ± 10.0	11.8 ± 9.4	10.7 ± 7.9	9.9 ± 7.8
Stage N3 sleep (min)	40.3 ± 32.4	41.8 ± 35.5	43.0 ± 34.8	39.6 ± 30.1	36.7 ± 28.5
Stage REM sleep (% time)	19.2 ± 6.6	10.7 ± 3.5	17.4 ± 1.4	21.5 ± 1.2	27.4 ± 3.2
Stage REM sleep (min)	69.7 ± 28.6	36.2 ± 15.0	62.5 ± 12.4	79.3 ± 13.2	101.2 ± 21.2
Time in bed (min)	484.4 ± 74.9	481.1 ± 84.8	488.6 ± 72.3	484.9 ± 69.0	483.0 ± 72.3
Total sleep time (min)	356.3 ± 68.4	330.6 ± 79.8	359.0 ± 64.5	368.0 ± 58.2	368.1 ± 62.1
Long sleeper based on actigraphy (>8h)	190 (7.2%)	41 (6.1%)	50 (7.6%)	38 (5.7%)	61 (9.1%)
Short sleeper based on actigraphy (<5 h)	307 (11.7%)	113 (16.7%)	65 (9.8%)	64 (9.6%)	65 (9.7%)
Wake after sleep onset (WASO) (min)	114.5 ± 65.9	134.5 ± 73.4	115.3 ± 65.7	105.3 ± 57.1	102.5 ± 61.7
Sleep efficiency (%)	76.2 ± 11.9	71.6 ± 13.5	76.3 ± 11.4	78.3 ± 10.2	78.8 ± 10.9
Sleep latency (min)	23.1 ± 27.3	26.6 ± 33.2	24.8 ± 30.0	19.4 ± 21.1	21.6 ± 22.5
REM latency (min)	108.7 ± 77.0	151.9 ± 98.4	107.8 ± 70.3	91.0 ± 53.8	84.0 ± 58.2
Overall arousal Index	23.7 ± 11.8	28.4 ± 13.9	24.2 ± 11.5	22.6 ± 10.5	19.5 ± 8.9
Apnea hypopnea index (3% desat/arousal)	20.7 ± 16.0	25.9 ± 18.4	20.5 ± 15.7	19.3 ± 14.9	17.1 ± 13.3
Desaturation index (desaturations/h)	20.7 ± 16.3	26.3 ± 19.9	20.1 ± 15.3	19.3 ± 14.9	17.0 ± 13.1
Periodic limb movements index	35.8 ± 37.8	40.1 ± 42.0	35.0 ± 37.3	35.4 ± 36.4	32.7 ± 34.7
Periodic limb movements with arousals Index	4.1 ± 5.8	5.3 ± 6.9	4.2 ± 2.8	4.0 ± 5.6	3.0 ± 4.1
Sleep time with saturated oxygen below 80% (% time)	0.09 ± 0.51	0.14 ± 0.62	0.08 ± 0.37	1.09 ± 0.56	0.05 ± 0.42

Characteristic	Overall (n=2,675)	Q1: <14.8% (n=677)	Q2: 14.8–19.5% (n=662)	Q3: 19.5–23.6% (n=667)	Q4: >23.6% (n=669)
Number of N2 to N1 sleep stage shifts/h	0.01 ± 0.05	0.02 ± 0.07	0.01 ± 0.04	0.01 ± 0.04	0.02 ± 0.05
Number of N3 to N1/N2 Sleep Stage Shifts/h	2.6 ± 1.8	2.7 ± 2.0	2.7 ± 1.9	2.7 ± 1.7	2.4 ± 1.5
Actigraphy: mean total sleep time, in bed (min)	385.4 ± 73.1	375.9 ± 79.8	387.1 ± 73.0	385.7 ± 69.3	393.1 ± 68.8
Actigraphy: mean WASO, in bed (min)	78.1 ± 44.3	89.7 ± 51.3	77.4 ± 42.4	74.3 ± 41.1	70.7 ± 38.9
Actigraphy: mean sleep latency, in bed (min)	30.6 ± 31.5	33.7 ± 32.8	32.4 ± 36.4	29.3 ± 29.4	27.0 ± 26.5
Actigraphy: mean sleep efficiency, in bed (% time)	82.6 ± 10.3	80.1 ± 11.9	82.7 ± 10.2	83.4 ± 9.5	84.3 ± 8.9
Actigraphy: mean scored sleep while outside of sleep interval (min)	68.8 ± 59.6	77.0 ± 64.0	68.8 ± 61.0	63.7 ± 55.7	65.8 ± 56.6
Self-reported Sleep Measures					
Epworth Sleepiness Scale Score (ESS) (0–18)	6.1 ± 3.7	6.2 ± 3.8	6.1 ± 3.8	6.4 ± 3.6	5.9 ± 3.5
Excessive Daytime Sleepiness (ESS > 10)	348 (13.0%)	89 (13.1%)	88 (13.3%)	97 (14.5%)	74 (11.1%)
Positive Airway Pressure Use >3 Months	14 (0.5%)	10 (1.5%)	4 (0.6%)	0 (0.0%)	0 (0.0%)
Pittsburgh Sleep Quality Index Total Score	5.6 ± 3.3	6.1 ± 3.6	5.4 ± 3.1	5.3 ± 3.0	5.5 ± 3.4
Functional Outcomes of Sleep Questionnaire Total Score	18.7 ± 1.6	18.6 ± 1.8	18.6 ± 1.6	18.8 ± 1.3	18.7 ± 1.5

[a] Entries are mean ± standard deviation or n (%)

Table 3: Baseline Demographic, Lifestyle, and Health Characteristics of the Wisconsin Sleep Cohort (WSC), Overall and by Percent Time in REM Quartiles (Q1–Q4))[a]

Characteristic	Overall (n=1,386)	Q1: <13.3% (n=345)	Q2: 13.3–17.6% (n=348)	Q3: 17.6–21.8% (n=347)	Q4: >21.8% (n=346)
Age at sleep visit (years)	51.5 ± 8.5	53.7 ± 9.0	51.6 ± 8.5	50.7 ± 8.3	50.0 ± 7.8
Age at last follow up	70.3 ± 7.7	71.6 ± 8.0	70.0 ± 7.7	69.7 ± 7.5	69.8 ± 7.4
Sex					
Male	753 (54.3%)	202 (58.6%)	189 (54.3%)	187 (53.9%)	175 (50.6%)
Female	633 (45.7%)	143 (41.4%)	159 (45.7%)	160 (46.1%)	171 (49.4%)
Race					
White	1,311 (94.6%)	323 (93.6%)	328 (94.3%)	330 (95.1%)	330 (95.4%)
non-White	75 (5.4%)	22 (6.4%)	20 (5.7%)	17 (4.9%)	16 (4.6%)
Education (years) (n=1,375)	14.7 ± 2.3	14.8 ± 2.3	14.8 ± 2.3	14.6 ± 2.4	14.5 ± 2.3
Body Mass Index (kg/m2)	30.7 ± 6.8	32.6 ± 8.2	29.6 ± 6.0	30.0 ± 5.6	30.4 ± 6.6
Average neck circumference (cm)	38.0 ± 4.3	39.1 ± 4.4	37.6 ± 4.3	37.7 ± 4.0	37.5 ± 4.2
Average hip circumference (cm)	108.2 ± 13.5	111.6 ± 15.7	106.5 ± 12.1	106.8 ± 12.0	107.8 ± 13.5
Smoking Status					
Never	668 (48.1%)	168 (48.7%)	166 (47.7%)	172 (49.6%)	162 (46.8%)
Past	510 (36.8%)	133 (38.6%)	121 (34.8%)	131 (37.8%)	125 (36.1%)
Current	208 (15.0%)	44 (12.8%)	61 (17.5%)	44 (12.7%)	59 (17.1%)
Daily caffeine intake (number of servings/day)	3.0 ± 2.5	3.0 ± 2.7	3.0 ± 2.5	3.1 ± 2.5	2.9 ± 2.6
Alcohol use (drinks/week)	3.6 ± 5.2	3.0 ± 4.4	3.6 ± 5.0	3.8 ± 5.0	4.0 ± 6.1
Weekly Alcohol Use					
0 Drinks	662 (47.8%)	184 (53.3%)	159 (45.7%)	156 (45.0%)	163 (47.1%)
1-4 Drinks	349 (25.2%)	74 (21.4%)	96 (27.6%)	94 (27.1%)	85 (24.6%)
>5 Drinks	375 (27.1%)	87 (25.2%)	93 (26.7%)	97 (28.0%)	98 (28.3%)
Medication Use					
Current use of antidepressant	191 (13.8%)	65 (18.8%)	50 (14.4%)	30 (8.6%)	46 (13.3%)
Current use of sedatives	63 (4.5%)	19 (5.5%)	18 (5.2%)	17 (4.9%)	9 (2.6%)

Characteristic	Overall (n=1,386)	Q1: <13.3% (n=345)	Q2: 13.3–17.6% (n=348)	Q3: 17.6–21.8% (n=347)	Q4: >21.8% (n=346)
Hypertension	339 (24.5%)	110 (31.9%)	78 (22.4%)	76 (21.9%)	75 (21.7%)
Angina	29 (2.1%)	11 (3.2%)	4 (1.1%)	10 (2.9%)	4 (1.2%)
Stroke	11 (0.8%)	3 (0.9%)	2 (0.6%)	3 (0.9%)	3 (0.9%)
Heart attack	37 (2.7%)	13 (3.8%)	8 (2.3%)	9 (2.6%)	7 (2.0%)
Type 2 diabetes	85 (6.1%)	39 (11.3%)	22 (6..3%)	8 (2.3%)	16 (4.6%)
Emphysema	17 (1.2%)	2 (0.6%)	3 (0.9%)	5 (1.4%)	7 (2.0%)

[a] Entries are mean ± standard deviation or n (%)

Table 4: Baseline Sleep Data from the Wisconsin Sleep Cohort (WSC), Overall and by Percent Time in REM Quartiles (Q1–Q4)[a]

Characteristic	Overall (n=1,386)	Q1: <13.3% (n=345)	Q2: 13.3–17.6% (n=348)	Q3: 17.6–21.8% (n=347)	Q4: >21.8% (n=346)
Objective Sleep Measures					
Stage N1 sleep (% time)	9.8 ± 6.3	12.4 ± 8.3	10.0 ± 6.0	8.6 ± 4.7	8.2 ± 4.8
Stage N1 sleep (min)	36.2 ± 21.7	42.5 ± 26.0	37.0 ± 22.0	33.2 ± 18.3	32.0 ± 18.1
Stage N2 sleep (% time)	59.9 ± 10.1	64.9 ± 10.4	61.8 ± 9.0	58.8 ± 9.0	54.1 ± 8.8
Stage N2 sleep (min)	226.6 ± 51.8	230.0 ± 56.1	232.7 ± 51.3	228.4 ± 50.3	215.1 ± 47.5
Stage N3 sleep (% time)	12.6 ± 9.7	13.2 ± 10.5	12.6 ± 9.9	12.9 ± 9.5	11.7 ± 8.7
Stage N3 sleep (min)	47.5 ± 36.7	46.9 ± 38.7	47.0 ± 37.1	49.8 ± 37.0	46.2 ± 34.0
Stage REM sleep (% time)	17.7 ± 6.5	9.5 ± 3.0	15.6 ± 1.3	19.6 ± 1.2	25.9 ± 3.7
Stage REM sleep (min)	67.8 ± 28.9	33.8 ± 12.4	58.6 ± 10.1	76.2 ± 12.0	102.6 ± 20.7
Time in bed (min)	452.3 ± 48.4	448.7 ± 45.9	451.1 ± 48.2	453.5 ± 51.7	455.8 ± 47.3
Total sleep time (min)	378.3 ± 58.8	353.5 ± 60.1	375.6 ± 56.8	387.9 ± 57.3	396.3 ± 51.8
Wake after sleep onset (WASO) (min)	60.1 ± 39.4	78.7 ± 44.4	61.7 ± 38.3	52.9 ± 33.6	47.1 ± 32.8
Sleep Efficiency (%)	83.7 ± 9.8	78.8 ± 11.0	83.3 ± 9.4	85.6 ± 8.4	87.1 ± 8.2
Sleep Latency (min)	12.2 ± 14.3	14.7 ± 16.7	12.0 ± 13.5	10.8 ± 12.6	11.2 ± 13.8
REM Latency (min)	124.8 ± 70.6	180.0 ± 84.0	124.3 ± 61.7	104.6 ± 49.9	91.0 ± 45.5
Apnea hypopnea index (3% desat/arousal)	13.7 ± 20.9	23.2 ± 32.2	12.0 ± 18.0	10.0 ± 12.0	9.6 ± 11.3
Limb movements index	26.7 ± 23.0	33.1 ± 25.9	26.6 ± 23.8	23.1 ± 21.0	24.2 ± 19.6
Self-reported Sleep Measures					
Epworth Sleepiness Scale (ESS) (0–18)	8.8 ± 4.1	8.7 ± 4.1	8.7 ± 4.3	8.9 ± 4.0	8.9 ± 3.9
Excessive daytime sleepiness (ESS>10)	452 (32.6%)	103 (29.9%)	102 (29.3%)	130 (37.5%)	117 (33.8%)
Self-reported sleepiness	308 (22.2%)	77 (22.3%)	75 (21.6%)	75 (21.6%)	81 (23.4%)

[a] Entries are mean ± standard deviation or n (%)

Table 5: Baseline Demographic, Lifestyle, and Health Characteristics of the Sleep Heart Health Study (SHHS) Cohort, Overall and by Percent Time in REM Quartiles (Q1–Q4)[a]

Characteristic	Overall (n=5,550)	Q1: <15.6 (n=1,386)	Q2: 15.6–20.0 (n=1,390)	Q3: 20.0–23.9 (n=1,386)	Q4: >23.9 (n=1,388)
Age at Sleep Visit (years)	63.0 ± 11.2	66.0 ± 11.3	63.1 ± 11.1	62.1 ± 10.9	60.8 ± 10.9
Age at Last Follow up	74.0 ± 10.5	76.6 ± 10.2	74.2 ± 10.5	73.3 ± 10.3	72.1 ± 10.5
Sex					
Male	2640 (47.6%)	703 (50.7%)	703 (50.6%)	643 (46.4%)	591 (42.6%)
Female	2910 (52.4%)	683 (49.3%)	687 (49.4%)	743 (53.6%)	797 (57.4%)
Race					
White	4,696 (84.6%)	1,154 (83.3%)	1,186 (85.3%)	1,189 (85.8%)	1,167 (84.1%)
non-White	854 (15.4%)	232(16.7%)	204 (14.7%)	197 (14.2%)	221 (15.9%)
Education (n=5,069)					
Less than 10 years of education	416 (8.2%)	132 (10.3%)	92 (7.2%)	104 (8.3%)	88 (7.0%)
11-15 years of education	2,624 (51.6%)	646 (50.4%)	665 (51.8%)	659 (52.4%)	654 (51.8%)
16-20 years of education	1,826 (35.9%)	450 (35.1%)	466 (36.3%)	452 (35.9%)	458 (36.3%)
More than 20 years of education	220 (4.3%)	55 (4.3%)	60 (4.7%)	43 (3.4%)	62 (4.9%)
Body Mass Index (kg/m2)	28.1 ± 5.0	28.3 ± 5.4	28.1 ± 5.0	28.1 ± 4.9	27.9 ± 4.8
Average Neck Circumference (cm)	37.8 ± 4.2	38.1 ± 4.3	37.9 ± 4.1	37.7 ± 4.2	37.4 ± 4.2
Average Hip Circumference (cm)	106.2 ± 20.3	106.6 ± 17.4	106.1 ± 14.9	106.3 ± 30.2	105.7 ± 14.7
Smoking Status (n=5,510)					
Never	2,593 (47.1%)	636 (46.2%)	629 (45.5%)	681 (49.6%)	647 (46.9%)
Past	2,384 (43.3%)	630 (45.8%)	613 (44.4%)	563 (41.0%)	578 (41.9%)
Current	533 (9.7%)	110 (8.0%)	139 (10.1%)	130 (9.5%)	154 (11.2%)
Daily Caffeine Use (number of servings/day)	2.7 ± 2.7	2.4 ± 2.5	2.7 ± 2.8	2.8 ± 2.7	2.7 ± 2.7
Alcohol Use (drinks/week)	2.8 ± 5.7	2.8 ± 6.2	3.0 ± 6.0	2.6 ± 4.9	2.9 ± 5.7
Weekly Alcohol Use					
0 Drinks	3,353 (65.1%)	872 (67.9%)	824 (63.9%)	845 (65.5%)	812 (63.1%)
1-4 Drinks	758 (14.7%)	158 (12.3%)	194 (15.0%)	194 (15.0%)	212 (16.5%)
>5 Drinks	1,042 (20.2%)	255 (19.8%)	272 (21.1%)	252 (19.5%)	263 (20.4%)

Characteristic	Overall (n=5,550)	Q1: <15.6 (n=1,386)	Q2: 15.6–20.0 (n=1,390)	Q3: 20.0–23.9 (n=1,386)	Q4: >23.9 (n=1,388)
Medication Use					
Current Use of Antidepressant (n=5,533)	152 (2.7%)	62 (4.5%)	44 (3.2%)	21 (1.5%)	25 (1.8%)
Current Use of Benzodiazepine (n=5,533)	282 (5.1%)	120 (8.7%)	61 (4.4%)	53 (3.8%)	48 (3.5%)
Hypertension	2,341 (42.2%)	710 (51.2%)	579 (41.7%)	549 (39.6%)	503 (36.2%)
Angina (n=5,396)	401 (7.4%)	119 (8.9%)	96 (7.1%)	105 (7.7%)	81 (6.0%)
Stroke (n=5,412)	180 (3.3%)	60 (4.5%)	53 (3.9%)	39 (2.9%)	28 (2.1%)
Heart Attack (n=5,386)	333 (6.2%)	86 (6.4%)	90 (6.7%)	86 (6.4%)	71 (5.2%)
Congestive Heart Failure (n=4,810)	129 (2.7%)	42 (3.6%)	33 (2.7%)	33 2.7%)	21 (1.7%)
Type 2 Diabetes (n=5,289)	383 (7.2%)	137 (10.4%)	92 (6.9%)	81 (6.1%)	73 (5.5%)
Chronic Obstructive Pulmonary Disease (n=5,450)	59 (1.1%)	26 (1.9%)	14 (1.0%)	9 (0.7%)	10 (0.7%)
Short Form 36 Health Survey (SF-36) Scores					
Mental Component Scale Standardized Score (n=5,069)	53.3 ± 8.2	52.7 ± 8.6	53.6 ± 8.3	53.6 ± 7.9	53.3 ± 8.0
Mental Health Index Standardized Score (n=5,133)	79.0 ± 15.2	77.7 ± 15.6	79.4 ± 15.4	79.7 ± 14.7	79.1 ± 15.2
Physical Component Scale Standardized Score (n=5,069)	47.5 ± 9.8	45.8 ± 10.7	47.7 ± 9.6	48.1 ± 9.5	48.6 ± 9.3

[a] Entries are mean ± standard deviation or n (%)

Table 6: Baseline Sleep Data from the Sleep Heart Health Study (SHHS) Cohort, Overall and by Percent Time in REM Quartiles (Q1–Q4)[a]

Characteristic	Overall (n=5,550)	Q1: <15.6 (n=1,386)	Q2: 15.6–20.0 (n=1,390)	Q3: 20.0–23.9 (n=1,386)	Q4: >23.9 (n=1,388)
Objective Sleep Measures					
Stage N1 Sleep (% time)	5.2 ± 3.9	5.3 ± 4.8	5.5 ± 3.9	5.1 ± 3.4	5.0 ± 3.3
Stage N2 Sleep (% time)	57.6 ± 13.3	66.6 ± 16.2	57.7 ± 10.8	55.1 ± 10.3	51.1 ± 9.8
Stage N3 Sleep (% time)	17.8 ± 12.0	17.7 ± 14.1	18.9 ± 11.7	17.9 ± 11.1	16.5 ± 10.4
Stage REM Sleep (% time)	19.4 ± 6.9	10.4 ± 4.8	17.9 ± 1.3	21.9 ± 1.1	27.4 ± 3.1
Time in Bed (min)	434.7 ± 59.1	425.3 ± 64.8	436.4 ± 58.4	439.3 ± 55.2	437.8 ± 56.7
Total Sleep Time (min)	390.1 ± 64.2	336.7 ± 70.3	359.0 ± 61.7	372.3 ± 55.7	372.4 ± 61.7
Wake After Sleep Onset (WASO) (min)	60.6 ± 43.6	73.4 ± 50.8	63.8 ± 44.3	53.4 ± 35.3	51.7 ± 38.8
Sleep Efficiency (%)	83.0 ± 10.5	79.3 ± 12.4	82.4 ± 10.1	84.9 ± 8.3	85.2 ± 9.6
Sleep Latency (min)	14.0 ± 20.4	15.2 ± 23.2	13.6 ± 18.4	13.6 ± 18.8	13.7 ± 21.0
REM Latency (min)	86.3 ± 57.8	111.4 ± 84.2	89.9 ± 50.3	77.3 ± 38.6	66.8 ± 34.4
Number of N2 to N1 Sleep Stage Shifts/hour	0.03 ± 0.08	0.04 ± 0.09	0.03 ± 0.09	0.03 ± 0.07	0.02 ± 0.08
Number of N3 to N1/N2 Sleep Stage Shifts/hour	2.9 ± 2.0	2.6 ± 2.3	3.1 ± 2.0	3.1 ± 1.8	3.0 ± 1.8
Apnea Hypopnea Index (3% desat/arousal)	17.8 ± 15.9	21.5 ± 19.1	18.8 ± 16.2	16.2 ± 13.6	14.6 ± 12.9
Desaturation Index (4% desat/hour)	7.6 ± 10.6	10.2 ± 13.9	8.1 ± 10.8	6.5 ± 8.2	5.8 ± 7.7
Sleep Time with Saturated Oxygen Below 80% (% time)	0.2 ± 1.9	0.3 ± 2.1	0.2 ± 1.7	0.2 ± 2.5	0.1 ± 0.7
Self-reported Sleep Measures					
Epworth Sleepiness Scale Score (ESS) (0 - 18) (n=5,335)	7.8 ± 4.4	7.8 ± 4.5	7.9 ± 4.3	7.6 ± 4.4	7.9 ± 4.4
Excessive Daytime Sleepiness (ESS Score > 10) (n=5,335)	1,346 (25.2%)	343 (26.0%)	337 (25.2%)	321 (24.0%)	345 (25.7%)

[a] Entries are mean ± standard deviation or n (%)

Table 7: Mortality Data for the Osteoporotic Fractures in Men Study (MrOS) and Wisconsin Sleep Cohort (WSC) and Sleep Heart Health Study (SHHS), Overall and by Percent Time in REM Quartiles (Q1–Q4)[a]

Mortality of MrOS Cohort, Overall Cohort and by Percent Time in REM Quartiles (Q1–Q4)					
Characteristic	**Overall (n=2,675)**	**Q1: <14.8% (n=677)**	**Q2: 14.8–19.5% (n=662)**	**Q3: 19.5–23.6% (n=667)**	**Q4: >23.6% (n=669)**
All-cause Mortality	1,404 (52.5%)	434 (64.1%)	354 (53.5%)	310 (46.5%)	306 (45.7%)
Cardiovascular Mortality	490 (18.3%)	154 (22.7%)	128 (19.3%)	102 (15.3%)	106 (15.8%)
Cancer Mortality	310 (11.6%)	90 (13.3%)	68 (10.3%)	71 (10.6%)	81 (12.1%)
Other Mortality	604 (22.6%)	190 (28.1%)	158 (23.9%)	137 (20.5%)	119 (17.8%)
Characteristic	**Overall (n=1,759)**	**Q1: <14.8% (n=515)**	**Q2: 14.8–19.5% (n=442)**	**Q3: 19.5–23.6% (n=410)**	**Q4: >23.6% (n=392)**
Longevity (>= 90 years old)	737 (41.9%)	204 (39.6%)	189 (42.8%)	181 (44.1%)	163 (41.6%)

Mortality of Wisconsin Sleep Cohort, Overall Cohort and by Percent Time in REM Quartiles (Q1–Q4)					
Characteristic	**Overall (n=1,386)**	**Q1: <13.3% (n=345)**	**Q2: 13.3–17.6% (n=348)**	**Q3: 17.6–21.8% (n=347)**	**Q4: >21.8% (n=346)**
All-cause Mortality	184 (13.3%)	62 (18.0%)	50 (14.4%)	41 (11.8%)	31 (9.0%)
Females (n=633)	70 (11.1%)	27 (18.9%)	18 (11.3%)	17 (10.6%)	8 (4.7%)
Males (n=753)	114 (15.1%)	35 (17.3%)	32 (16.9%)	24 (12.8%)	23 (13.1%)
Cardiovascular Mortality	50 (3.6%)	17 (4.9%)	17 (4.9%)	9 (2.6%)	7 (2.0%)
Cancer Mortality	71 (5.1%)	22 (6.4%)	19 (5.5%)	18 (5.2%)	12 (3.5%)
Other Mortality	63 (4.5%)	23 (6.7%)	14 (4.0%)	14 (4.0%)	12 (3.5%)

[a] Entries are n (%)

Mortality of Sleep Heart Health Study Cohort, Overall Cohort and by Percent Time in REM Quartiles (Q1–Q4)					
Characteristic	**Overall (n=5,550)**	**Q1: <15.6 (n=1,386)**	**Q2: 15.6–20.0 (n=1,390)**	**Q3: 20.0–23.9 (n=1,386)**	**Q4: >23.9 (n=1,388)**
All-cause Mortality	1,237 (22.3%)	446 (32.2%)	304 (21.9%)	276 (19.9%)	211 (15.2%)
Females (n=2,910)	572 (19.7%)	204 (29.9%)	125 (18.2%)	139 (18.7%)	104 (13.0%)
Males (n=2,640)	665 (25.2%)	242 (34.4%)	179 (25.5%)	137 (21.3%)	107 (18.1%)
Cardiovascular Mortality (n=4,810)	343 (7.1%)	122 (10.5%)	81 (6.7%)	77 (6.2%)	63 (5.2%)

[a] Entries are n (%)

Table 8: Mortality Risk Ratios from Cox Regression for 5% reduction in Percent REM in the Osteoporotic Fractures in Men Study (MrOS)[a]

Mortality Risk Ratios for the MrOS Cohort Using Percent REM as a Continuous Variable (5% Decrease)					
Outcome (n=2,675)	Overall Deaths	Unadjusted HR (95% CI)	Age Adjusted HR (95% CI)	Model 1 HR (95% CI)	Model 2 HR (95% CI)
All–cause Mortality	1,404 (52.5%)	1.19 (1.14–1.24)	1.12 (1.07–1.16)	1.11 (1.07–1.15)	1.13 (1.08–1.19)
Cardiovascular Mortality	490 (18.3%)	1.24 (1.16–1.33)	1.12 (1.05–1.20)	1.11 (1.03–1.18)	1.11 (1.02–1.20)
Cancer Mortality	310 (11.6%)	1.16 (1.06–1.26)	1.03 (0.95–1.13)	1.03 (0.94–1.12)	1.06 (0.96–1.17)
Other Mortality	604 (22.6%)	1.26 (1.19–1.34)	1.16 (1.09–1.23)	1.16 (1.09–1.23)	1.19 (1.11–1.28)
Outcome (n=1,759)	Overall >=90	Unadjusted OR (95% CI)	Age Adjusted OR (95% CI)	Model 1 OR (95% CI)	Model 2 OR (95% CI)
Longevity (>=90 years old)	737 (41.9%)	0.97 (0.90-1.03)	0.85 (0.78-0.93)	0.86 (0.79-0.94)	0.89 (0.81-0.92)

[a] Entries are n (%) Abbreviations: Hazard Ratio (HR), Odds Ratio (OR), Confidence Interval (CI)

Model 1: Age, Race (white versus non–white), Education, Body Mass Index (BMI), Smoking Status, Weekly Alcohol, Daily Caffeine, Antidepressants, Benzodiazepines, Sleep Medications, and Site

Model 2: Model 1 + Overall Arousal Index, Sleep Time with Saturated Oxygen Below 80%, Percent Stage N2 Sleep, Actigraphy Mean Scored Sleep While Outside of Sleep Interval, Actigraphy Wake After Sleep Onset, Epworth Sleepiness Scale Score, Teng Mini Mental State Examination Score, Physical Activity Scale for the Elderly Score, Depression, Congestive Heart Failure, Chronic Obstructive Pulmonary Disease, Type 2 Diabetes, Heart Attack, and Stroke

Table 9: Mortality Risk Ratios from Cox Regression Using Percent REM as Binary Variable (< 15%) in the Osteoporotic Fractures in Men Study (MrOS)[a]

Mortality Risk Ratios for the MrOS Cohort using Percent REM as Binary Variable (< 15%)					
Outcome (n=2,675)	**Overall Deaths**	**Unadjusted HR (95% CI)**	**Age Adjusted HR (95% CI)**	**Model 1 HR (95% CI)**	**Model 2 HR (95% CI)**
All–cause mortality	1,404 (52.5%)	1.59 (1.42–1.78)	1.34 (1.20–1.50)	1.33 (1.18–1.49)	1.28 (1.12–1.46)
Cardiovascular mortality	490 (18.3%)	1.62 (1.34–1.96)	1.34 (1.11–1.62)	1.30 (1.07–1.56)	1.20 (0.96–1.49)
Cancer mortality	310 (11.6%)	1.37 (1.08–1.76)	1.22 (0.95–1.56)	1.21 (0.94–1.55)	1.31 (0.99–1.73)
Other mortality	604 (22.6%)	1.68 (1.42–2.00)	1.41 (1.19–1.68)	1.42 (1.19–1.69)	1.35 (1.10–1.64)
Outcome (n=1,759)	**Overall >=90**	**Unadjusted OR (95% CI)**	**Age Adjusted OR (95% CI)**	**Model 1 OR (95% CI)**	**Model 2 OR (95% CI)**
Longevity (>=90 years old)	737 (41.9%)	0.85 (0.69-1.05)	0.58 (0.44-0.75)	0.59 (0.45-0.77)	0.72 (0.53-0.98)

[a] Entries are n (%) Abbreviations: Hazard Ratio (HR), Odds Ratio (OR), Confidence Interval (CI)

Model 1: Age, Race (white versus non-white), Education, Body Mass Index (BMI), Smoking Status, Weekly Alcohol, Daily Caffeine, Antidepressants, Benzodiazepines, Sleep Medications, and Site

Model 2: Model 1 + Overall Arousal Index, Sleep Time with Saturated Oxygen Below 80%, Percent Stage N2 Sleep, Actigraphy Mean Scored Sleep While Outside of Sleep Interval, Actigraphy Wake After Sleep Onset, Epworth Sleepiness Scale Score, Teng Mini Mental State Examination Score, Physical Activity Scale for the Elderly Score, Depression, Congestive Heart Failure, Chronic Obstructive Pulmonary Disease, Type 2 Diabetes, Heart Attack, and Stroke

Table 10: Sensitivity Analyses of Mortality Hazard Ratios from Cox Regression for the Osteoporotic Fractures in Men Study (MrOS)[a]

Sensitivity Analysis Mortality Hazard Ratios for the MrOS Cohort Using Percent REM as a Continuous Variable (5% Decrease)						
	Full Dataset (n=2,675)		Survived First 2 Years of Study (n=2,591)		No OSA / Medication (n=1,684)	
Outcome	Overall Number Deaths n (%)	Model 2 HR (95% CI)	Overall Number Deaths n (%)	Model 2 HR (95% CI)	Overall Number Deaths n (%)	Model 2 HR (95% CI)
All–cause mortality	1,404 (52.5%)	1.13 (1.08–1.19)	1,320 (50.9%)	1.12 (1.07–1.18)	821 (48.8%)	1.13 (1.06–1.20)
Cardiovascular mortality	490 (18.3%)	1.11 (1.02–1.20)	460 (17.8%)	1.10 (1.02–1.20)	269 (16.0%)	1.10 (0.99–1.22)
Cancer mortality	310 (11.6%)	1.06 (0.96–1.17)	291 (11.2%)	1.06 (0.95–1.17)	184 (10.9%)	1.01 (0.89–1.15)
Other mortality	604 (22.6%)	1.19 (1.11–1.28)	569 (22.0%)	1.18 (1.09–1.27)	368 (21.9%)	1.23 (1.12–1.35)

Survived First 2 Years of Study group excluded individuals who were censored during the first two years

No OSA / Medication group excluded individuals with an AHI > 30 or current use of antidepressants, benzodiazepines, or sleep medications

Sensitivity Analysis Mortality Hazard Ratios for the MrOS Cohort Using Percent REM as a Continuous Variable (5% Decrease)						
	No Depression (n=2,291)		Normal Sleepers (5-8 Hours) (n=1,776)		Normal Sleepers (6-8 Hours) (n=859)	
Outcome	Overall Number Deaths n (%)	Model 2 HR (95% CI)	Overall Number Deaths n (%)	Model 2 HR (95% CI)	Overall Number Deaths n (%)	Model 2 HR (95% CI)
All–cause mortality	1,153 (50.3%)	1.14 (1.08–1.20)	872 (49.1%)	1.17 (1.09–1.25)	408 (47.5%)	1.19 (1.08–1.31)
Cardiovascular mortality	399 (17.4%)	1.13 (1.03–1.24)	310 (17.5%)	1.10 (0.98–1.23)	145 (16.9%)	1.00 (0.85–1.19)
Cancer mortality	271 (11.8%)	1.04 (0.93–1.17)	200 (11.3%)	1.12 (0.98–1.27)	93 (10.8%)	1.13 (0.92–1.37)
Other mortality	483 (21.1%)	1.20 (1.11–1.30)	362 (20.4%)	1.26 (1.14–1.40)	170 (19.8%)	1.38 (1.19–1.61)

[a] Entries are n (%)

No depression group excluded individuals with Geriatric Depression Score >4 or current antidepressant use

Model 2: Age, Race (white versus non–white), Education, Body Mass Index (BMI), Smoking Status, Weekly Alcohol, Daily Caffeine, Antidepressants, Benzodiazepines, Sleep Medications, Site, Overall Arousal Index, Sleep Time with Saturated Oxygen Below 80%, Percent Stage N2 Sleep, Actigraphy Mean Scored Sleep While Outside of Sleep Interval, Actigraphy Wake After Sleep Onset, Epworth Sleepiness Scale Score, Teng Mini Mental State Examination Score, Physical Activity Scale for the Elderly Score, Depression, Congestive Heart Failure, Chronic Obstructive Pulmonary Disease, Type 2 Diabetes, Heart Attack, and Stroke

Table 11: Mortality Hazard Ratios from Cox Regression for 5% reduction in Percent REM in the Wisconsin Sleep Cohort (WSC)[a]

Mortality Hazard Ratios for WSC Using Percent REM as a Continuous Variable (5% Decrease)					
Outcome (n=1,386)	**Overall Deaths**	**Unadjusted HR (95% CI)**	**Age/Gender Adjusted HR (95% CI)**	**Model 1 HR (95% CI)**	**Model 2 HR (95% CI)**
All–cause Mortality	184 (13.3%)	1.22 (1.08–1.36)	1.15 (1.03–1.29)	1.14 (1.01–1.28)	1.17 (1.03–1.34)
Females (n=633)	70 (11.1%)	1.38 (1.15–1.66)	1.36 (1.13–1.64)	1.29 (1.06–1.57)	1.34 (1.07–1.68)
Males (n=753)	114 (15.1%)	1.11 (0.96–1.28)	1.02 (0.88–1.18)	1.02 (0.89–1.19)	1.09 (0.92–1.30)
Cardiovascular Mortality	50 (3.6%)	1.33 (1.07–1.66)	1.21 (0.97–1.51)	1.18 (0.94–1.47)	1.13 (0.87–1.45)
Cancer Mortality	71 (5.1%)	1.16 (0.97–1.40)	1.11 (0.92–1.33)	1.12 (0.93–1.36)	1.13 (0.91–1.40)
Other Mortality	63 (4.5%)	1.20 (0.98–1.46)	1.16 (0.95–1.41)	1.13 (0.93–1.37)	1.26 (1.01–1.58)

[a] Entries are n (%) Abbreviations: Hazard Ratio (HR), Confidence Interval (CI)

Model 1: Age, Sex, Race (white versus non–white), Education, Body Mass Index (BMI), Smoking Status, Weekly Alcohol, Daily Caffeine, Antidepressants, and Sedatives

Model 2: Model 1 + Percent Stage N2 Sleep, Wake After Sleep Onset, Epworth Sleepiness Scale Score, Emphysema, Type 2 Diabetes, Heart Attack, and Stroke

Table 12: Mortality Hazard Ratios from Cox Regression Using Percent REM as Binary Variable (< 15%) in the Wisconsin Sleep Cohort (WSC)[a]

Mortality Risk Ratios for WSC Using Percent REM as Binary Variable (< 15%)					
Outcome (n=1,386)	Overall Deaths	Unadjusted HR (95% CI)	Age/Gender Adjusted HR (95% CI)	Model 1 HR (95% CI)	Model 2 HR (95% CI)
All–cause mortality	184 (13.3%)	1.66 (1.24–2.22)	1.36 (1.01–1.83)	1.34 (0.99–1.81)	1.36 (0.98–1.89)
Females (n=633)	70 (11.1%)	2.11 (1.32–3.34)	1.95 (1.22–3.14)	1.70 (1.03–2.80)	1.70 (0.99–2.91)
Males (n=753)	114 (15.1%)	1.40 (0.96–2.02)	1.06 (0.72–1.55)	1.08 (0.74–1.58)	1.15 (0.75–1.76)
Cardiovascular mortality	50 (3.6%)	1.80 (1.03–3.14)	1.29 (0.73–2.28)	1.26 (0.71–2.23)	1.00 (0.52–1.90)
Cancer mortality	71 (5.1%)	1.54 (0.96–2.47)	1.28 (0.79–2.07)	1.35 (0.83–2.20)	1.36 (0.81–2.30)
Other mortality	63 (4.5%)	1.74 (1.06–2.87)	1.54 (0.93–2.56)	1.46 (0.86–2.44)	1.78 (1.01–3.15)

[a] Entries are n (%), Abbreviations: Hazard Ratio (HR), Confidence Interval (CI)

Model 1: Age, Sex, Race (white versus non–white), Education, Body Mass Index (BMI), Smoking Status, Weekly Alcohol, Daily Caffeine, Antidepressants, and Sedatives

Model 2: Model 1 + Percent Stage N2 Sleep, Wake After Sleep Onset, Epworth Sleepiness Scale Score, Emphysema, Type 2 Diabetes, Heart Attack, and Stroke

Table 13: Mortality Hazard Ratios from Cox Regression for 5% Reduction in Percent REM in the Sleep Heart Health Study (SHHS)[a]

Mortality Hazard Ratios for SHHS Using Percent REM as a Continuous Variable (5% Decrease)					
Outcome (n=5,550)	**Overall Deaths**	**Unadjusted HR (95% CI)**	**Age/Gender Adjusted HR (95% CI)**	**Model 1 HR (95% CI)**	**Model 2 HR (95% CI)**
All-cause Mortality	1,237 (22.3%)	1.27 (1.22 - 1.32)	1.13 (1.09 - 1.18)	1.13 (1.09 - 1.18)	1.13 (1.07 - 1.18)
Females (n=2,910)	572 (19.7%)	1.28 (1.22 - 1.36)	1.11 (1.05 - 1.17)	1.11 (1.05 - 1.17)	1.09 (1.02 - 1.17)
Males (n=2,640)	665 (25.2%)	1.25 (1.18 - 1.31)	1.15 (1.09 - 1.22)	1.15 (1.09 - 1.21)	1.16 (1.08 - 1.24)
Cardiovascular Mortality (n=4,810)	343 (7.1%)	1.25 (1.16 - 1.34)	1.05 (0.97 - 1.13)	1.05 (0.97 - 1.13)	1.07 (0.97 - 1.17)

[a] Entries are n (%) Abbreviations: Hazard Ratio (HR), Confidence Interval (CI)

Model 1: Age, Sex, Race (white versus non-white), Education, Body Mass Index (BMI), Smoking Status, Weekly Alcohol, Daily Caffeine, Antidepressants, and Benzodiazepines

Model 2: Model 1 + Overall Arousal Index, Sleep Time with Saturated Oxygen Below 80%, Percent Stage N2 Sleep, Epworth Sleepiness Scale Score, SF-36 Mental Component Scale Standardized Score, SF-36 Mental Health Index Standardized Score, SF-36 Physical Component Scale Standardized Score, Congestive Heart Failure, Chronic Obstructive Pulmonary Disease, Type 2 Diabetes, Heart Attack, and Stroke

Table 14: Mortality Hazard Ratios from Cox Regression Using Percent REM as Binary Variable (< 15%) in the Sleep Heart Health Study (SHHS)[a]

Mortality Risk Ratios for SHHS Using Percent REM as Binary Variable (< 15%)					
Outcome (n=5,550)	Overall Deaths	Unadjusted HR (95% CI)	Age/Gender Adjusted HR (95% CI)	Model 1 HR (95% CI)	Model 2 HR (95% CI)
All-cause Mortality	1,237 (22.3%)	1.86 (1.65 - 2.09)	1.35 (1.20 - 1.52)	1.35 (1.20 - 1.53)	1.28 (1.12 - 1.47)
Females (n=2,910)	572 (19.7%)	1.97 (1.66 - 2.35)	1.27 (1.07 - 1.52)	1.27 (1.06 - 1.53)	1.20 (0.98 - 1.47)
Males (n=2,640)	665 (25.2%)	1.72 (1.47 - 2.02)	1.41 (1.20 - 1.66)	1.41 (1.20 - 1.66)	1.35 (1.13- 1.63)
Cardiovascular Mortality (n=4,810)	343 (7.1%)	1.93 (1.54 - 2.41)	1.19 (0.94 - 1.49)	1.19 (0.95 - 1.50)	1.26 (0.97 - 1.63)

[a] Entries are n (%),Abbreviations: Hazard Ratio (HR), Confidence Interval (CI)

Model 1: Age, Sex, Race (white versus non-white), Education, Body Mass Index (BMI), Smoking Status, Weekly Alcohol, Daily Caffeine, Antidepressants, and Benzodiazepines

Model 2: Model 1 + Overall Arousal Index, Sleep Time with Saturated Oxygen Below 80%, Percent Stage N2 Sleep, Epworth Sleepiness Scale Score, SF-36 Mental Component Scale Standardized Score, SF-36 Mental Health Index Standardized Score, SF-36 Physical Component Scale Standardized Score, Congestive Heart Failure, Chronic Obstructive Pulmonary Disease, Type 2 Diabetes, Heart Attack, and Stroke

Table 15: Data Used for Meta-analysis

Outcome / Study	Sample Size	Confidence Level	Hazard Ratio (HR)	Log of HR	Standard Error	95% CI Lower Limit	95% CI Upper Limit
All-cause Mortality							
MrOS All-cause	2,675	0.95	1.132695	0.124599749	0.024	1.0806351	1.1872419
WSC All-cause	1,386	0.95	1.1710681	0.157916238	0.066962	1.0270326	1.3353037
SHHS All-cause	5,550	0.95	1.1253946	0.11813373	0.024721	1.0721677	1.181264
All-cause Mortality (Men)							
MrOS All-cause Men	2,675	0.95	1.132695	0.124599749	0.024	1.0806351	1.1872419
WSC All-cause Men	753	0.95	1.094439	0.090241903	0.085866	0.92491593	1.2950331
SHHS All-cause Men	2,640	0.95	1.1574158	0.146189761	0.0346	1.0815274	1.238629
All-cause Mortality (Women)							
WSC All-cause Women	633	0.95	1.343441	0.295234233	0.114315	1.07377455	1.6808312
SHHS All-cause Women	2,910	0.95	1.0943363	0.090148061	0.0360616	1.0196594	1.1744823
Cardiovascular Mortality							
MrOS Cardiovascular	2,675	0.95	1.112	0.106160196	0.0408842	1.0263483	1.2047479
WSC Cardiovascular	1,386	0.95	1.125265	0.118018563	0.129000	0.87395209	1.4488452
SHHS Cardiovascular	4,810	0.95	1.0681594	0.06593698	0.048215	0.9718411	1.1740237

Abbreviations: Hazard Ratio (HR), Confidence Interval (CI)

Figures

Figure 1a-b: Dr. William C. Dement and Dr. Emmanuel Mignot

Figure 1a. Dr. William C. Dement studying rapid eye movement (REM) sleep in cats.

Figure 1b. Dr. Emmanuel Mignot holding his narcoleptic dog, Bear.

Figure 2: Example Polysomnographic Sleep Recording

Two minute recording of rapid eye movement (REM) sleep with the green box highlighting a 30 second epoch of data. The channels are labeled on the left. LOC and ROC are abbreviations for the left and right electrooculogram. Chin represents the electromyogram measured under the chin. Brainwaves are labeled with their derivation (C3-A1, C4-A2, O1-A2, and O2-A1) and measure data from the central and occipital lobes. RAT and LAT are the right and left anterior tibialis leg muscle. ECG is the electrocardiogram. The R-R channel is displaying the time elapsed between two successive R-waves of the QRS signal on the electrocardiogram. Finally, SAO2 is the oxygen hemoglobin from a finger oximeter probe using a 3-second averaging rate. The green box indicates a 30 second epoch.

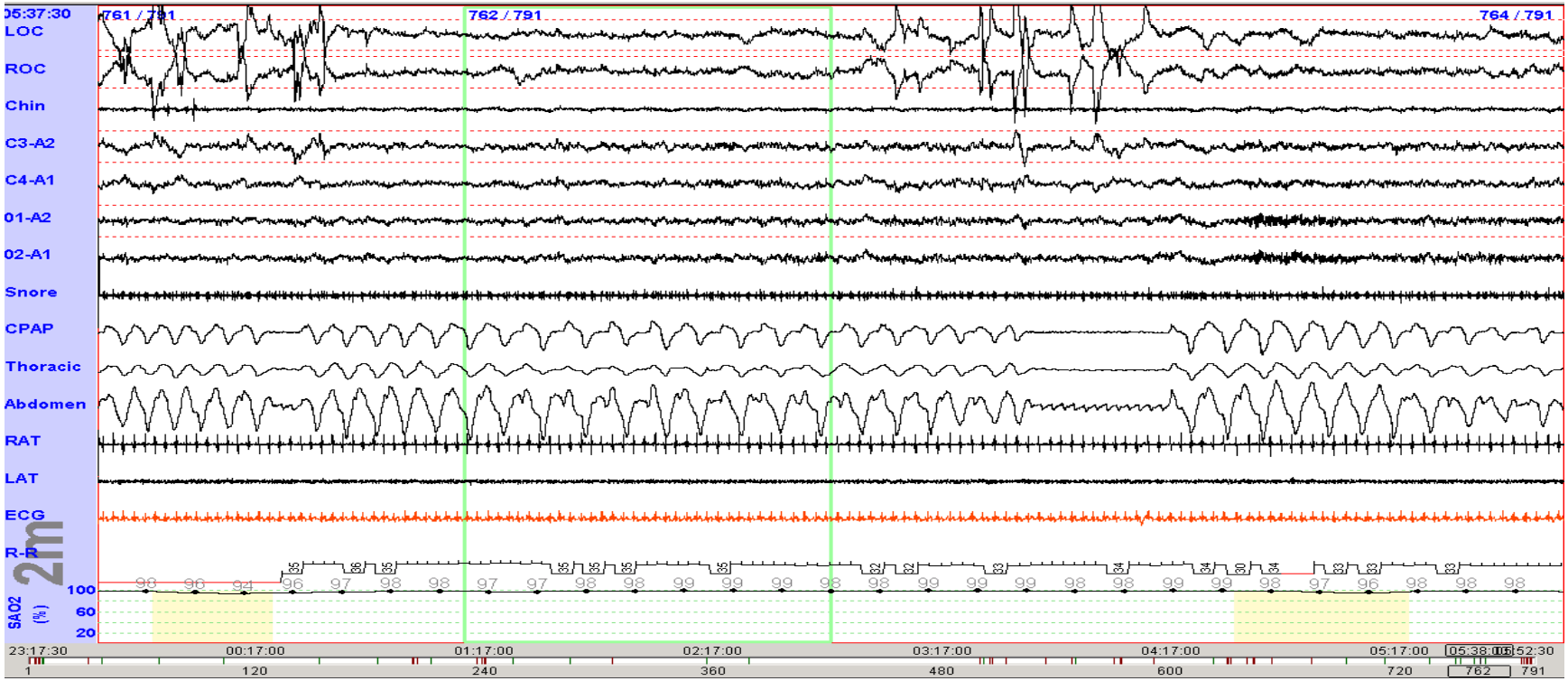


Figure 3: Example Hypnogram for a Healthy Adult

A graphical representation of sleep stages as a function of time. A typical night's sleep consists of 4-6 cycles of a NREM sleep period followed by REM sleep period.

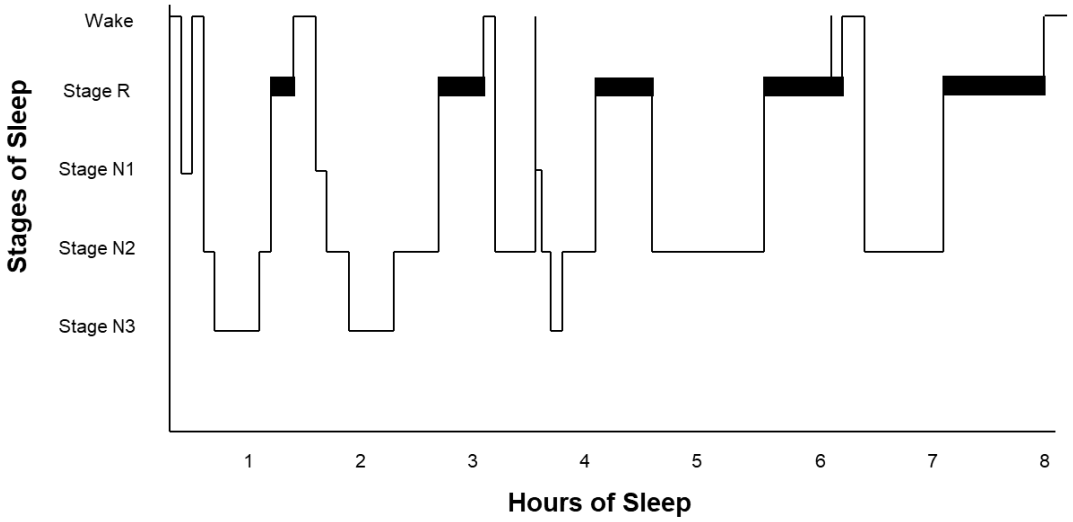


Figure 4: Directed Acyclic Graph (DAG)

A visual representation of causal assumptions used to help identify the presence of confounding. Given lack of information on casual relationships, this diagram was used to graphically organize and cluster available data in the MrOS cohort.

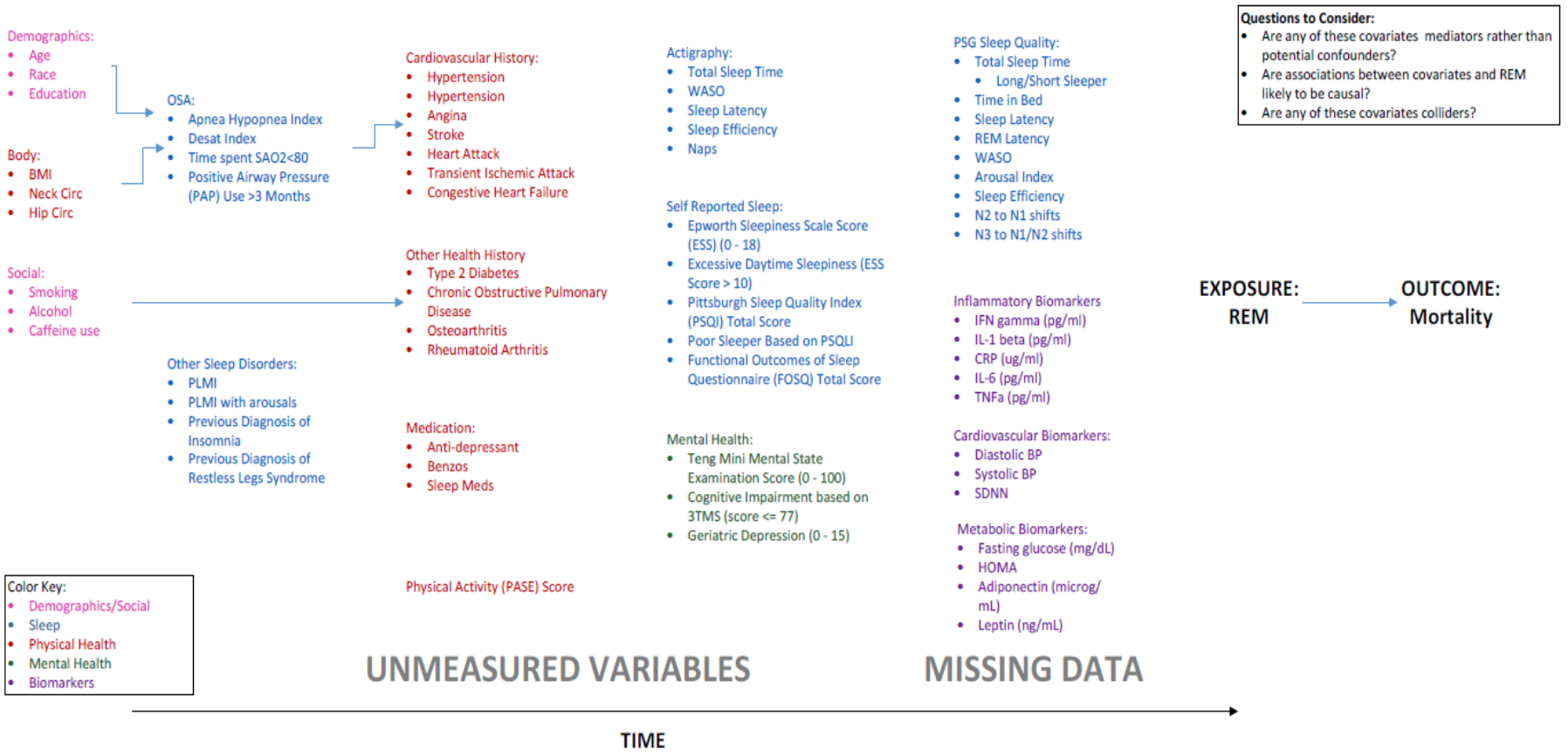


Figure 5: Flowchart Illustrating Participant Selection for Osteoporotic Fractures in Men Study (MrOS) Cohort

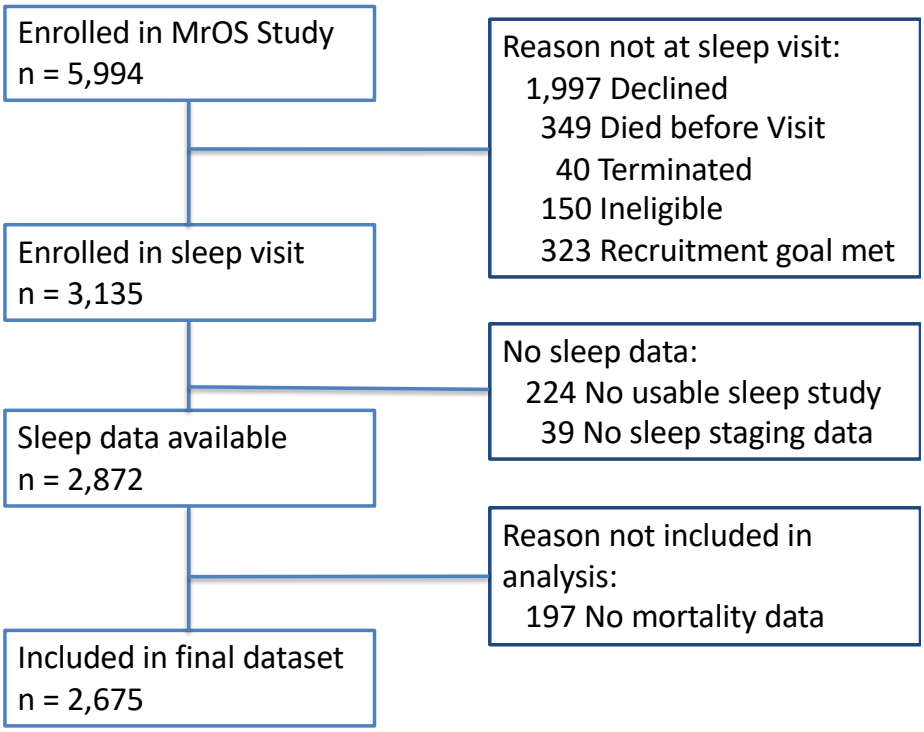


Figure 6: Flowchart Illustrating Participant Selection for the Wisconsin Sleep Cohort (WSC)

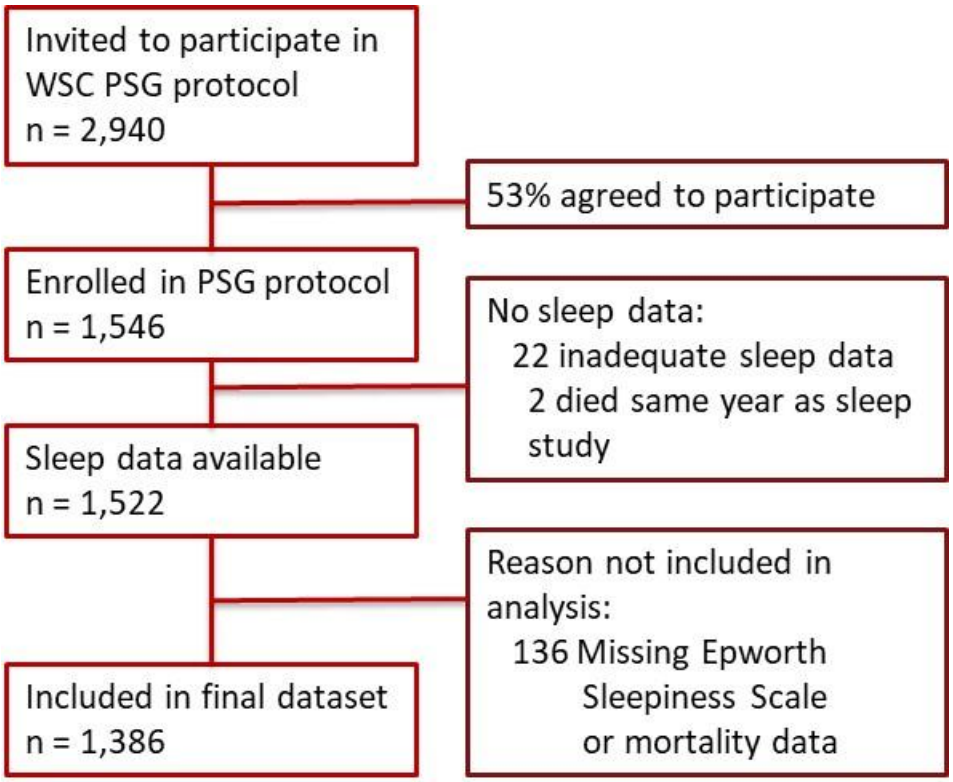


Figure 7: Flowchart Illustrating Participant Selection for the Sleep Heart Health (SHHS) Cohort

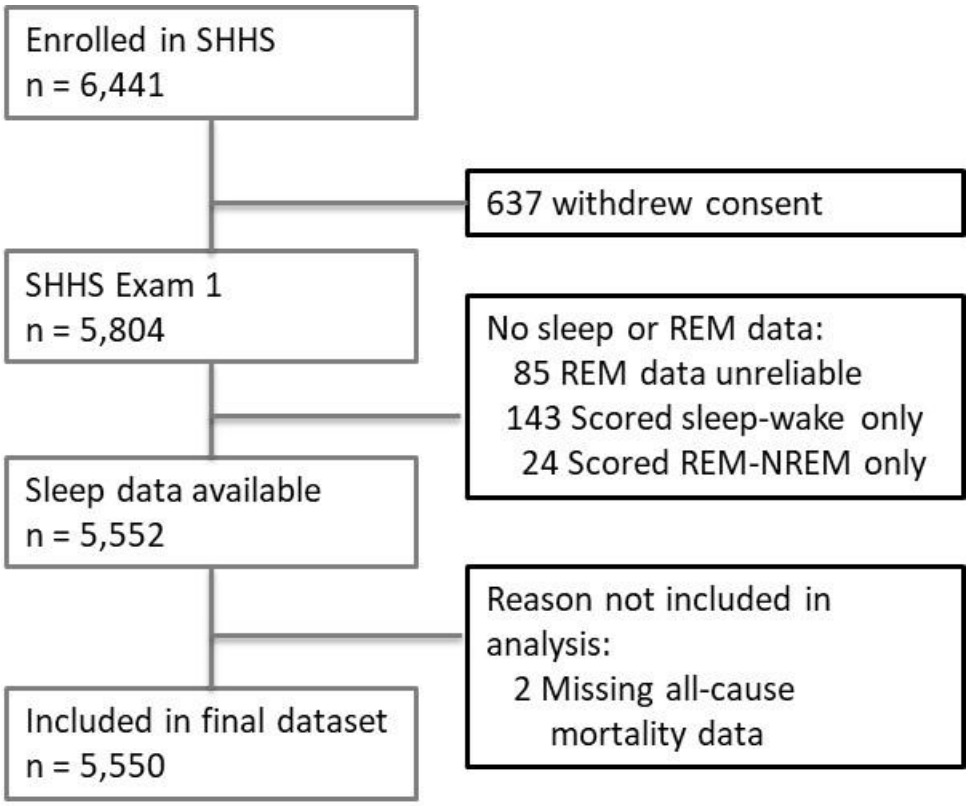
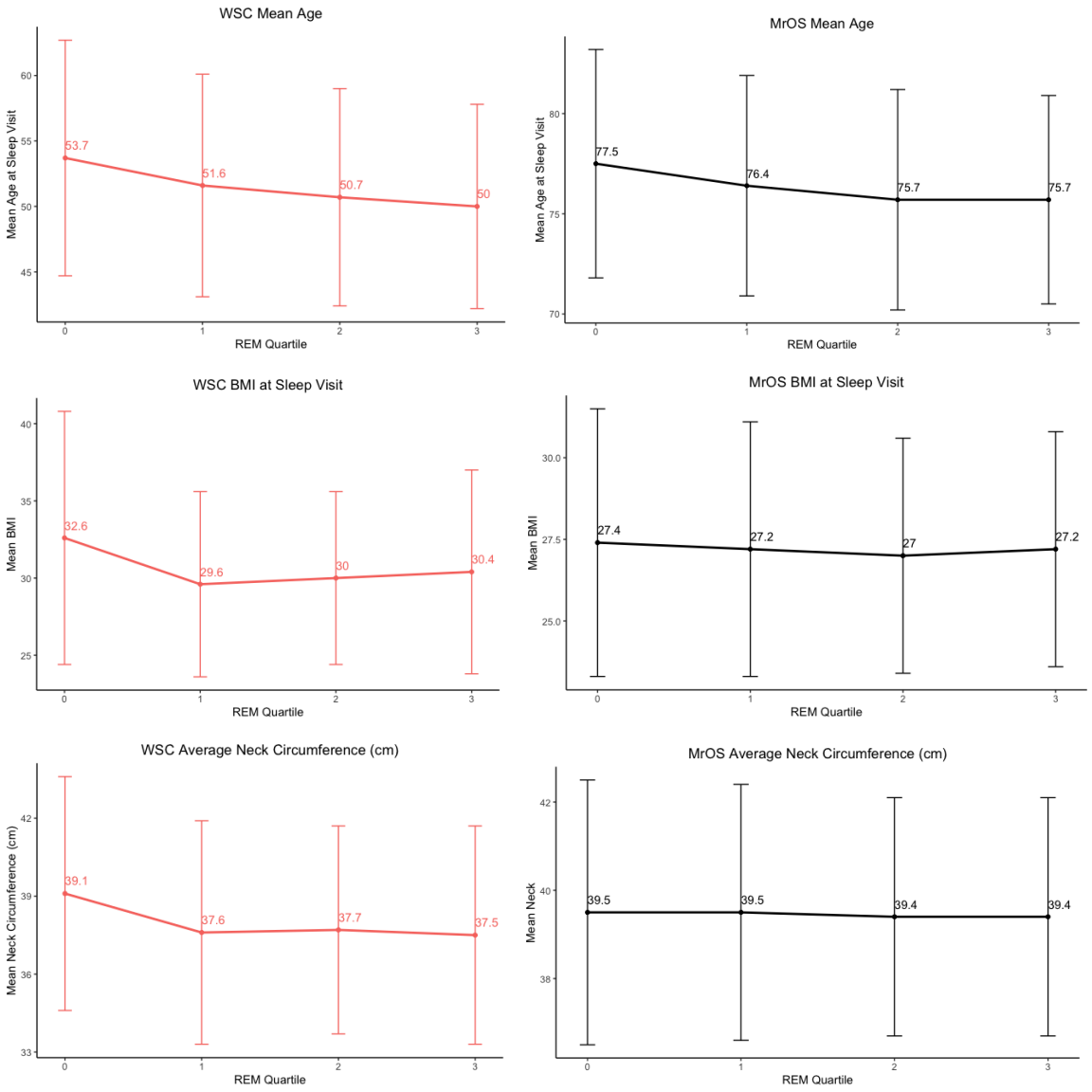
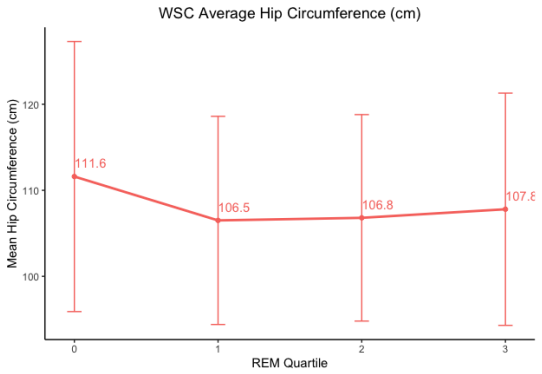
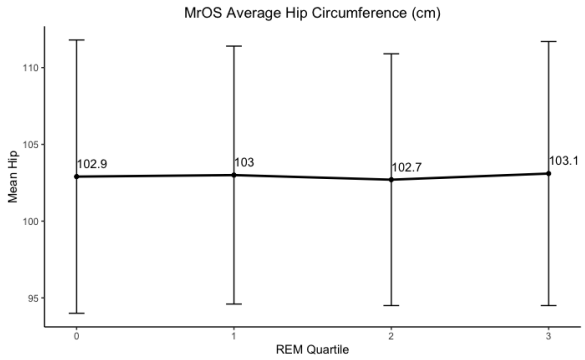


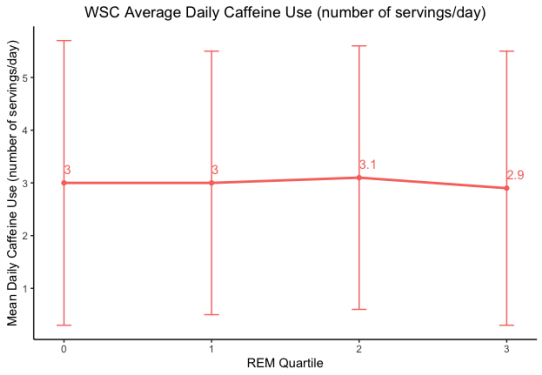
Figure 8: Wisconsin Sleep Cohort (WSC) Demographic and Lifestyle Variables Compared to the Osteoporotic Fractures in Men Study (MrOS) Cohort by REM Quartile

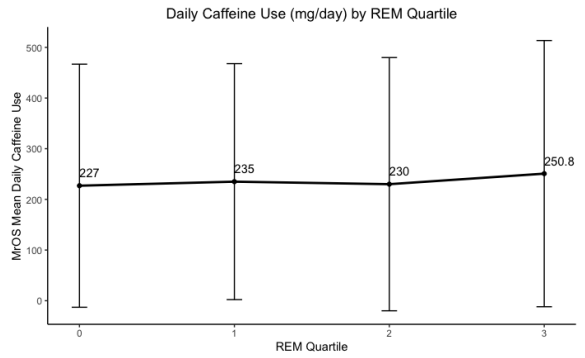
Mean value with standard error by REM quartile. The left column in orange contains data from WSC. The right column in black are values from MrOS.

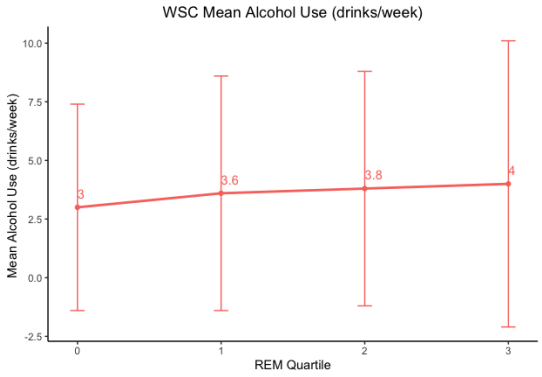



WSC Average Hip Circumference (cm)
Mean Hip Circumference (cm)
120
110
100
111.6
106.5
106.8
107.8
0
1
2
3
REM Quartile


MrOS Average Hip Circumference (cm)
Mean Hip
110
105
100
95
102.9
103
102.7
103.1
0
1
2
3
REM Quartile


WSC Average Daily Caffeine Use (number of servings/day)
Mean Daily Caffeine Use (number of servings/day)
5
4
3
2
1
3
3
3.1
2.9
0
1
2
3
REM Quartile


Daily Caffeine Use (mg/day) by REM Quartile
MrOS Mean Daily Caffeine Use
500
400
300
200
100
0
227
235
230
250.8
0
1
2
3
REM Quartile


WSC Mean Alcohol Use (drinks/week)
Mean Alcohol Use (drinks/week)
10.0
7.5
5.0
2.5
0.0
-2.5
3
3.6
3.8
4
0
1
2
3
REM Quartile

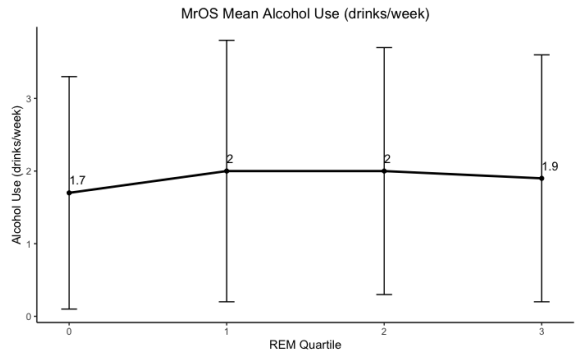
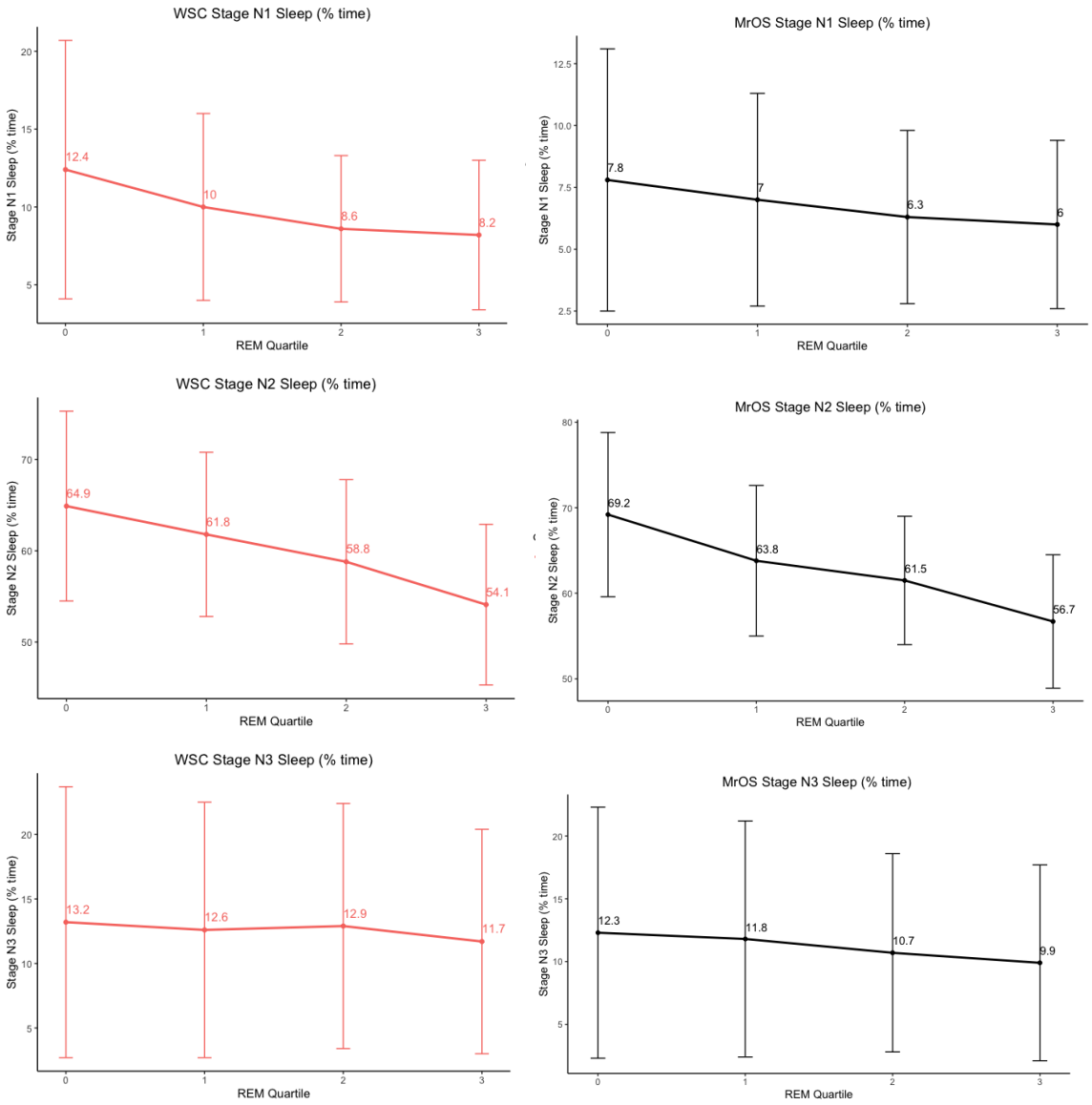
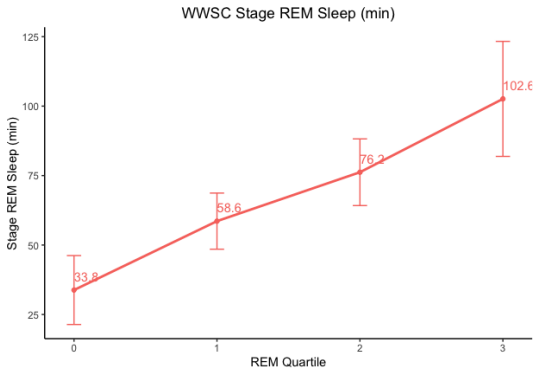
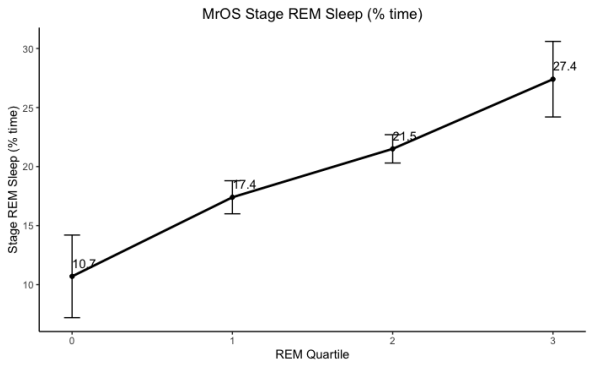

MrOS Mean Alcohol Use (drinks/week)
Alcohol Use (drinks/week)
3
2
1
0
1.7
2
2
1.9
0
1
2
3
REM Quartile

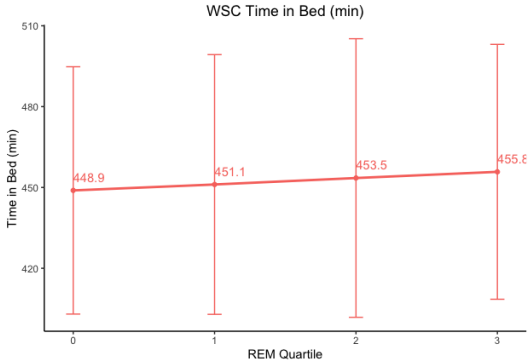
Figure 9: Wisconsin Sleep Cohort (WSC) Sleep Variables Compared to the Osteoporotic Fractures in Men Study (MrOS) Cohort by REM Quartile

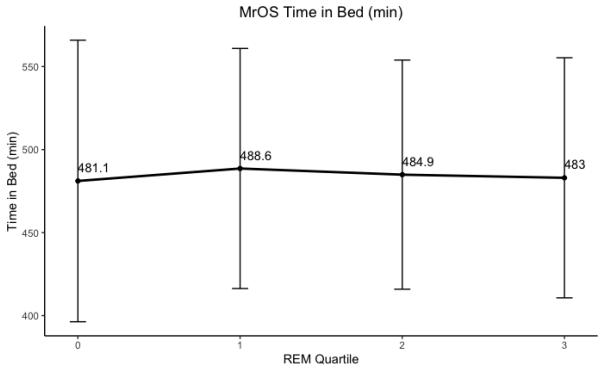
Mean value with standard error by REM quartile. The left column in orange contains data from WSC. The right column in black are values from MrOS.

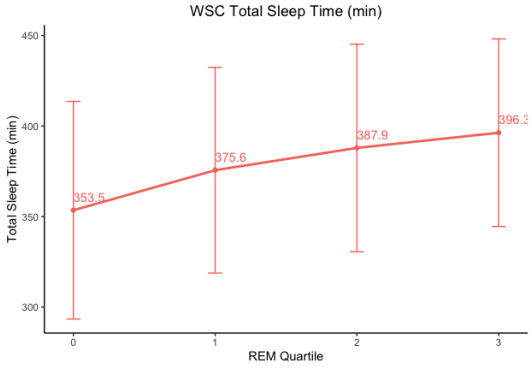


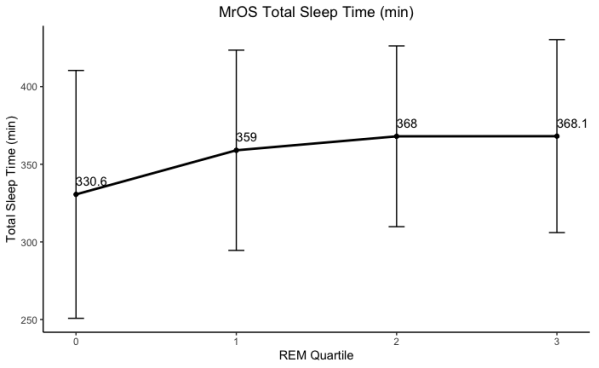

WWSC Stage REM Sleep (min)
Stage REM Sleep (min)
REM Quartile
33.8
58.6
76.2
102.6

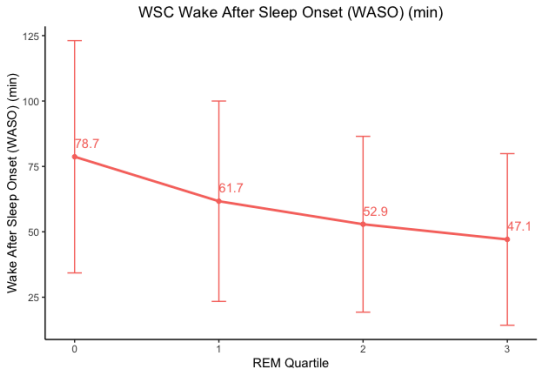

MrOS Stage REM Sleep (% time)
Stage REM Sleep (% time)
REM Quartile
10.7
17.4
21.5
27.4

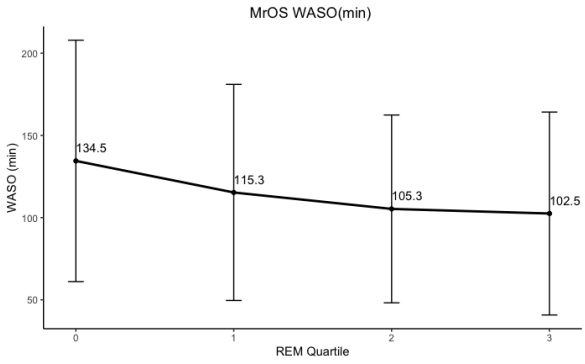

WSC Time in Bed (min)
Time in Bed (min)
REM Quartile
448.9
451.1
453.5
455.8

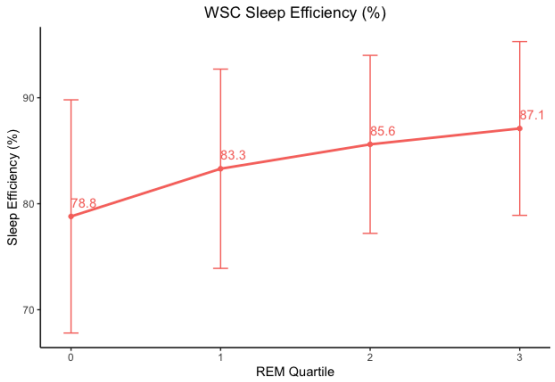

MrOS Time in Bed (min)
Time in Bed (min)
REM Quartile
481.1
488.6
484.9
483

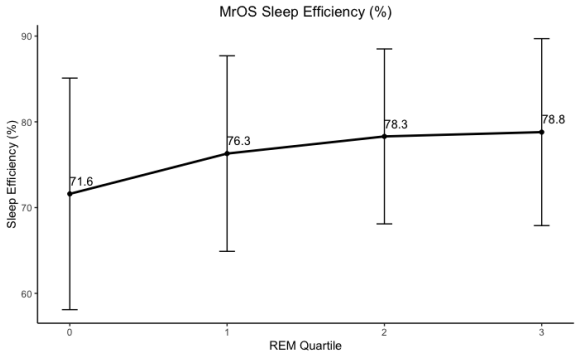

WSC Total Sleep Time (min)
Total Sleep Time (min)
REM Quartile
353.5
375.6
387.9
396.3

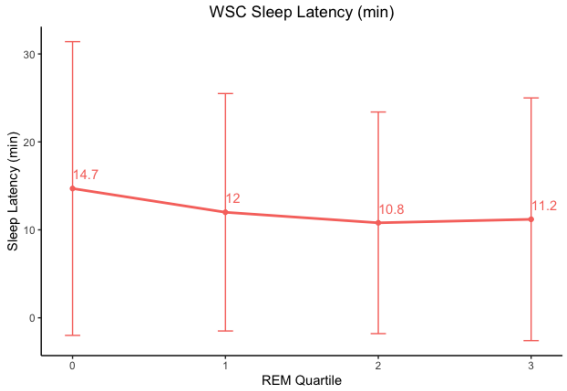

MrOS Total Sleep Time (min)
Total Sleep Time (min)
REM Quartile
330.6
359
368
368.1

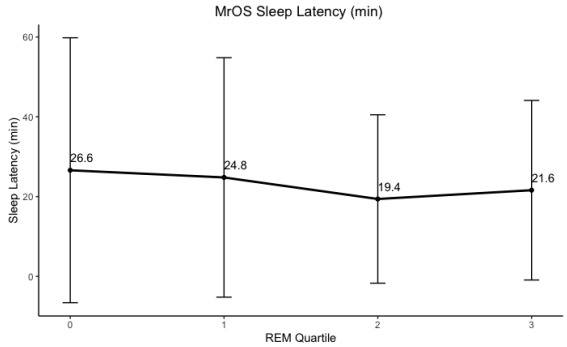

WSC Wake After Sleep Onset (WASO) (min)
Wake After Sleep Onset (WASO) (min)
REM Quartile
78.7
61.7
52.9
47.1

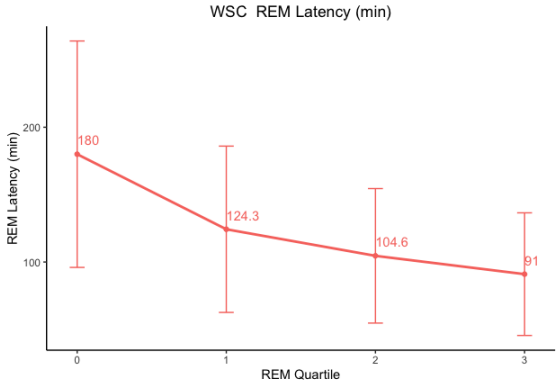

MrOS WASO(min)
WASO (min)
REM Quartile
134.5
115.3
105.3
102.5

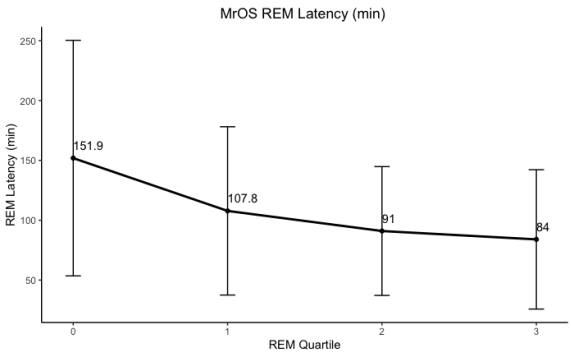

WSC Sleep Efficiency (%)
Sleep Efficiency (%)
REM Quartile
78.8
83.3
85.6
87.1

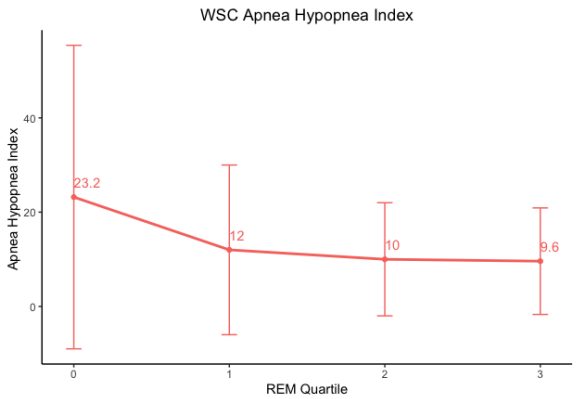

MrOS Sleep Efficiency (%)
Sleep Efficiency (%)
REM Quartile
71.6
76.3
78.3
78.8

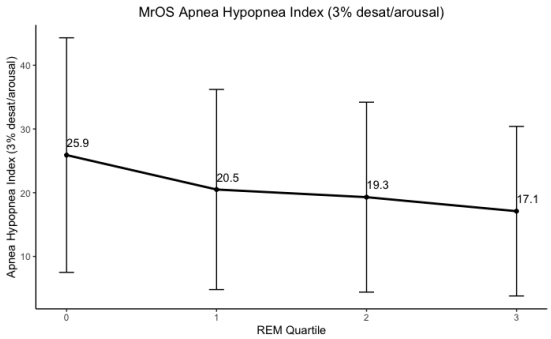

WSC Sleep Latency (min)
Sleep Latency (min)
REM Quartile
14.7
12
10.8
11.2

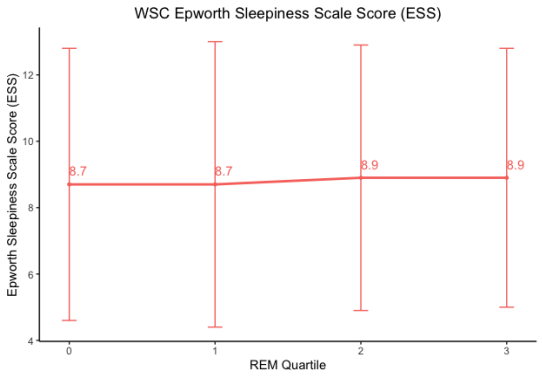

MrOS Sleep Latency (min)
Sleep Latency (min)
REM Quartile
26.6
24.8
19.4
21.6


WSC REM Latency (min)
REM Latency (min)
REM Quartile
180
124.3
104.6
91


MrOS REM Latency (min)
REM Latency (min)
REM Quartile
151.9
107.8
91
84


WSC Apnea Hypopnea Index
Apnea Hypopnea Index
REM Quartile
23.2
12
10
9.6


MrOS Apnea Hypopnea Index (3% desat/arousal)
Apnea Hypopnea Index (3% desat/arousal)
REM Quartile
25.9
20.5
19.3
17.1


WSC Epworth Sleepiness Scale Score (ESS)
Epworth Sleepiness Scale Score (ESS)
REM Quartile
8.7
8.7
8.9
8.9

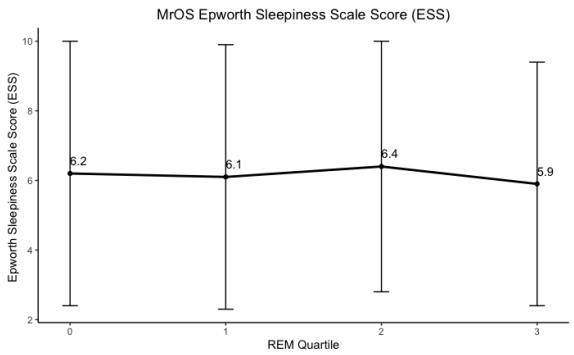

MrOS Epworth Sleepiness Scale Score (ESS)
Epworth Sleepiness Scale Score (ESS)
REM Quartile
6.2
6.1
6.4
5.9

Figure 10: Unadjusted All-cause Mortality Kaplan-Meier Plots by Percent REM Quartile (Q1 through Q4, Top) and 15% REM Threshold (Bottom) in the MrOS Cohort

The horizontal axis (x-axis) represents time in years, and the vertical axis (y-axis) shows the probability of surviving. The lines represent the different survival curve by group. A vertical drop in the curves indicates an event. A vertical tick mark means that a patient was censored at this time.

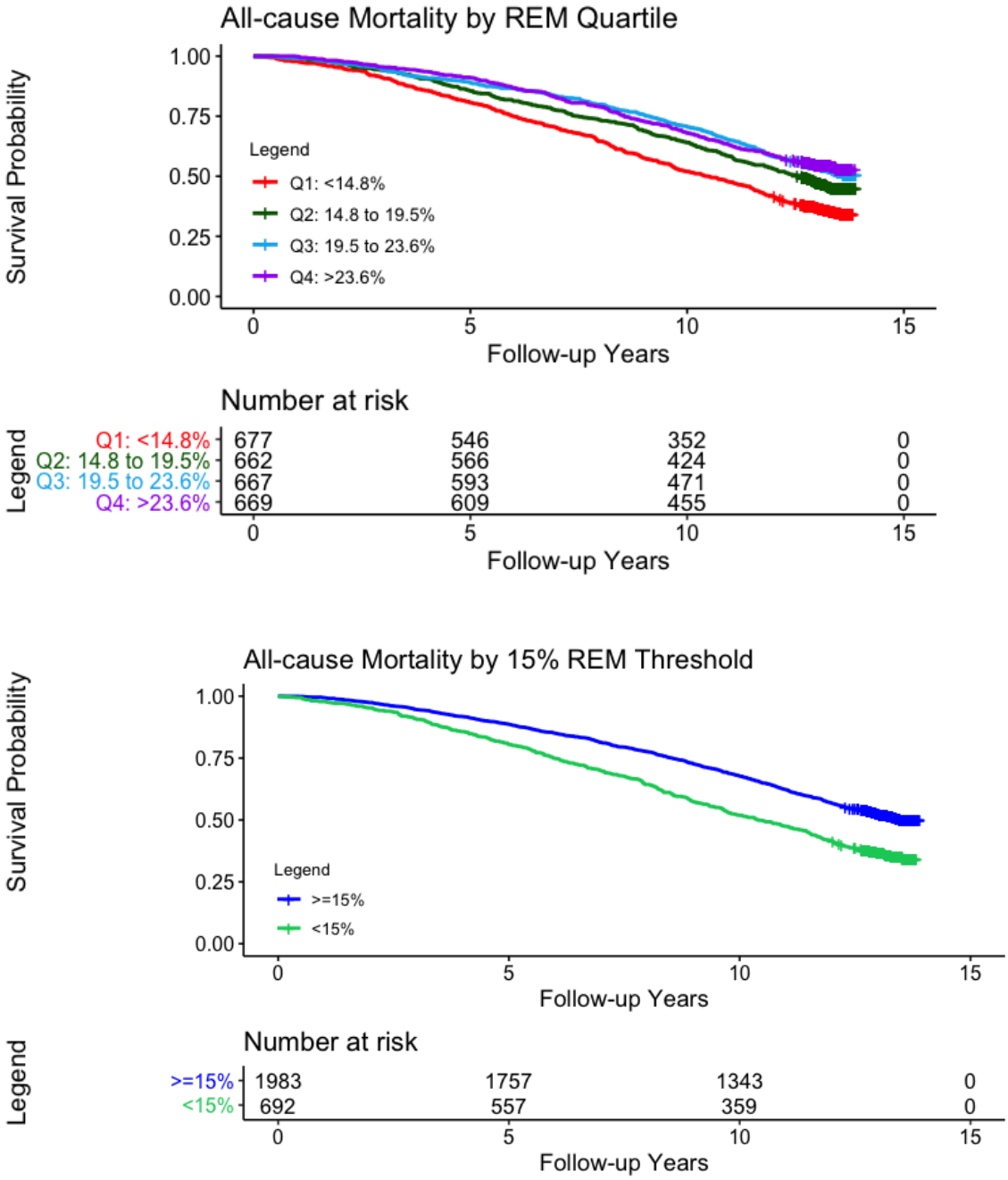


Figure 11: Unadjusted Cardiovascular Mortality Kaplan-Meier Plots by Percent REM Quartile (Q1 through Q4, Top) and 15% REM Threshold (Bottom) in the MrOS Cohort

The horizontal axis (x-axis) represents time in years, and the vertical axis (y-axis) shows the probability of surviving. The lines represent the different survival curve by group. A vertical drop in the curves indicates an event. A vertical tick mark means that a patient was censored at this time.

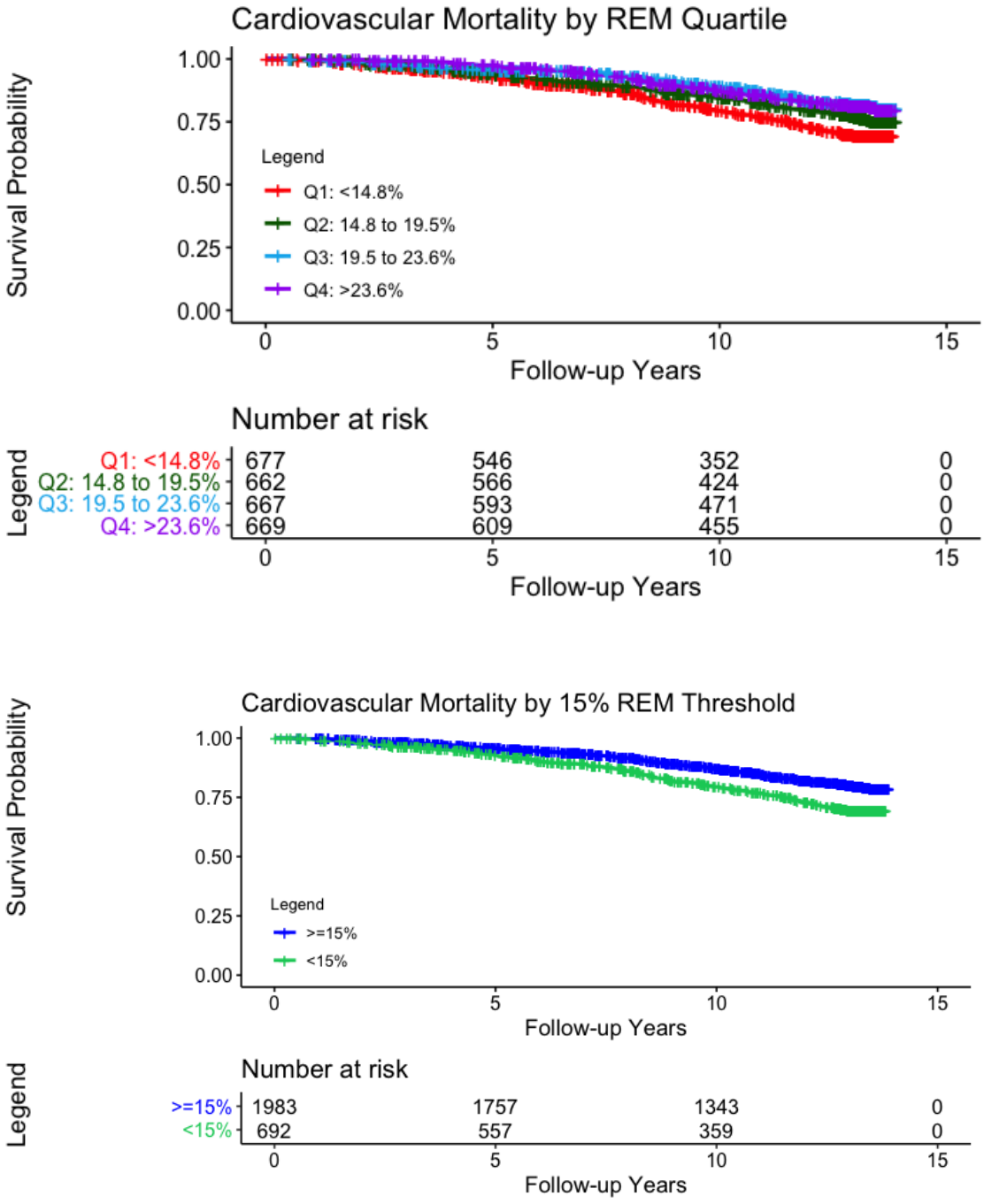


Figure 12: Unadjusted Cancer Mortality Kaplan-Meier Plots by Percent REM Quartile (Q1 through Q4, Top) and 15% REM Threshold (Bottom) in the MrOS Cohort

The horizontal axis (x-axis) represents time in years, and the vertical axis (y-axis) shows the probability of surviving. The lines represent the different survival curve by group. A vertical drop in the curves indicates an event. A vertical tick mark means that a patient was censored at this time.

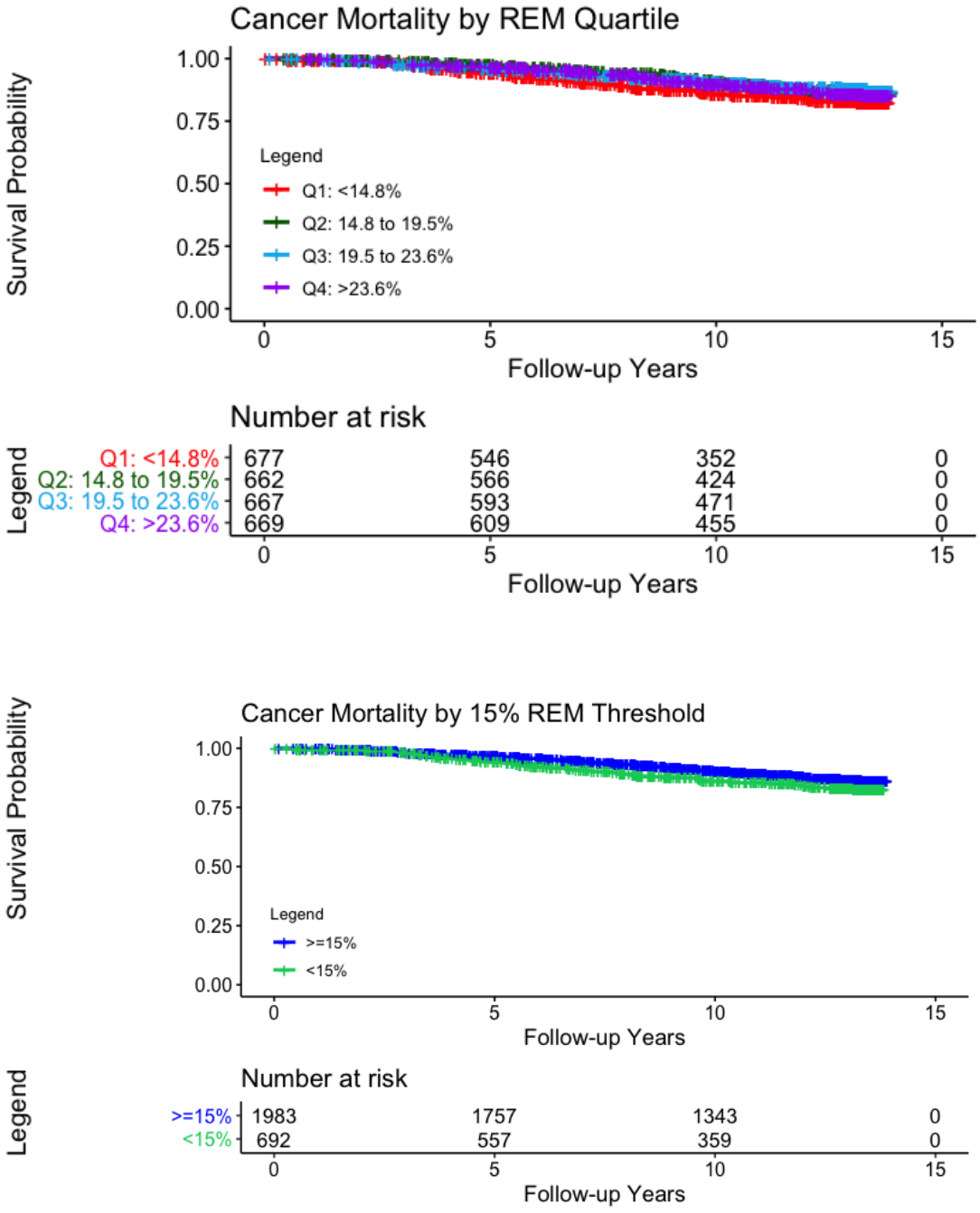


Figure 13: Unadjusted Other Mortality Kaplan-Meier Plots by Percent REM Quartile (Q1 through Q4, Top) and 15% REM Threshold (Bottom) in the MrOS Cohort

The horizontal axis (x-axis) represents time in years, and the vertical axis (y-axis) shows the probability of surviving. The lines represent the different survival curve by group. A vertical drop in the curves indicates an event. A vertical tick mark means that a patient was censored at this time.

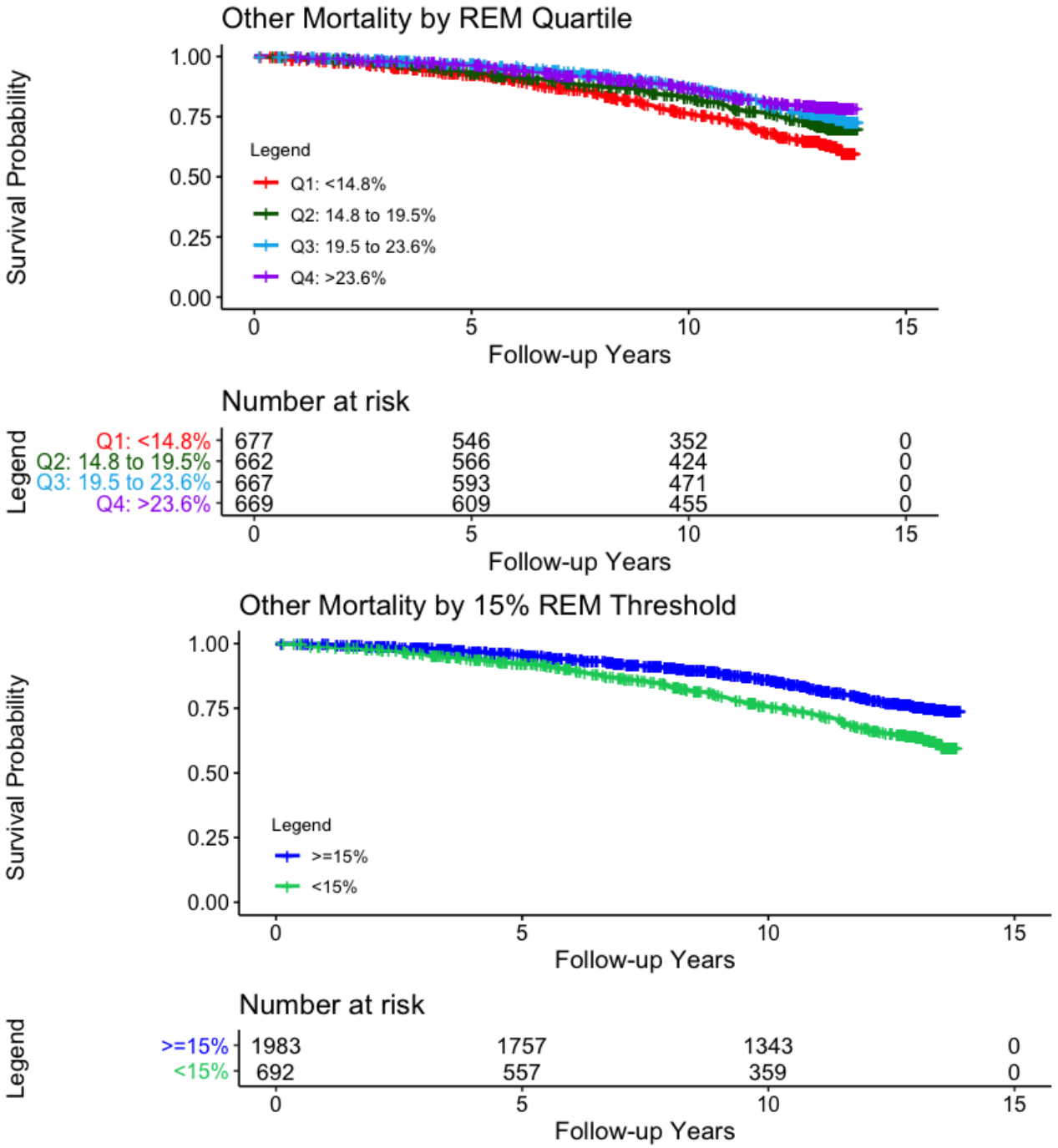


Figure 14: Variable Importance Based On Random Survival Forest Classifier Evaluating All Sleep Stages as Predictors of Survival in the MrOS Study

Percent REM was found to be the most important variable with a mean decrease in accuracy of 0.058. This technique ranks the variables from most to least predictive in context of one another.

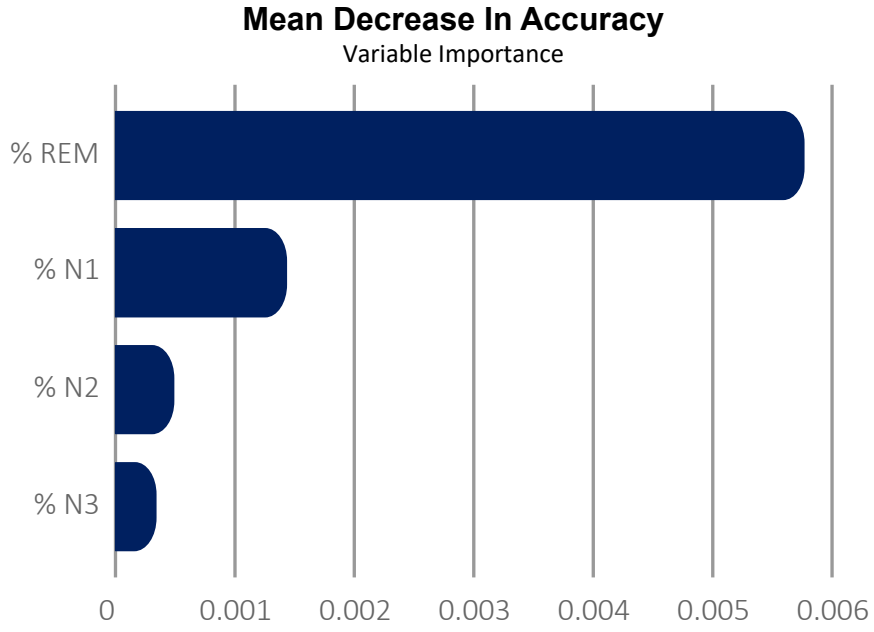


Figure 15: Conditional Inference Survival Tree Model evaluating all sleep stages as predictors of survival in the MrOS study

All sleep stages were included as potential predictors of all-cause mortality using data from the MrOS study. Percent REM was selected as the first and second nodes with cut-points of 15.4% and 10.9%. Percent non-REM stage 1 (%N1) was the 5th node with a cut-point of 13.6%.

.

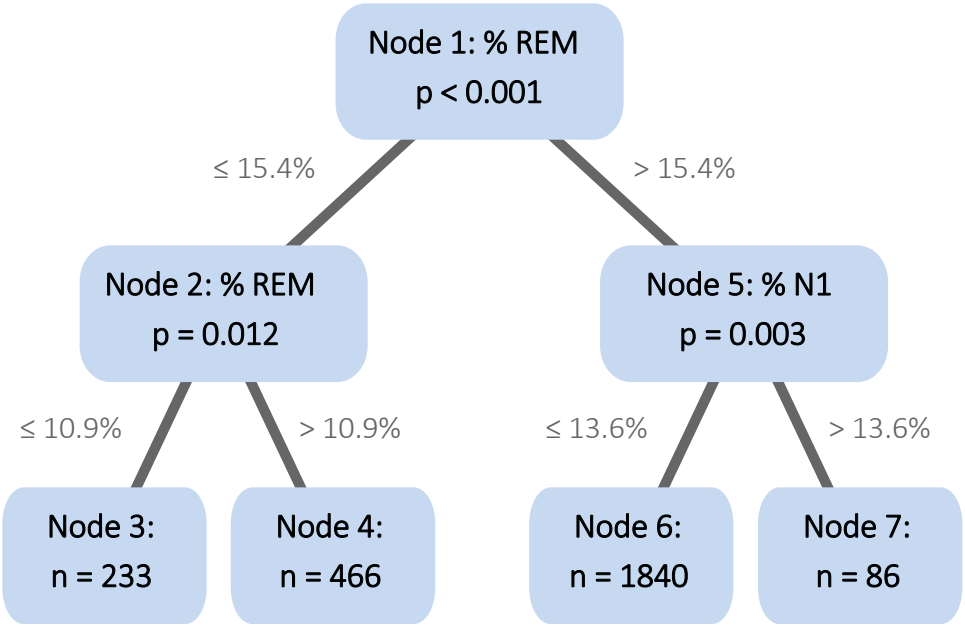
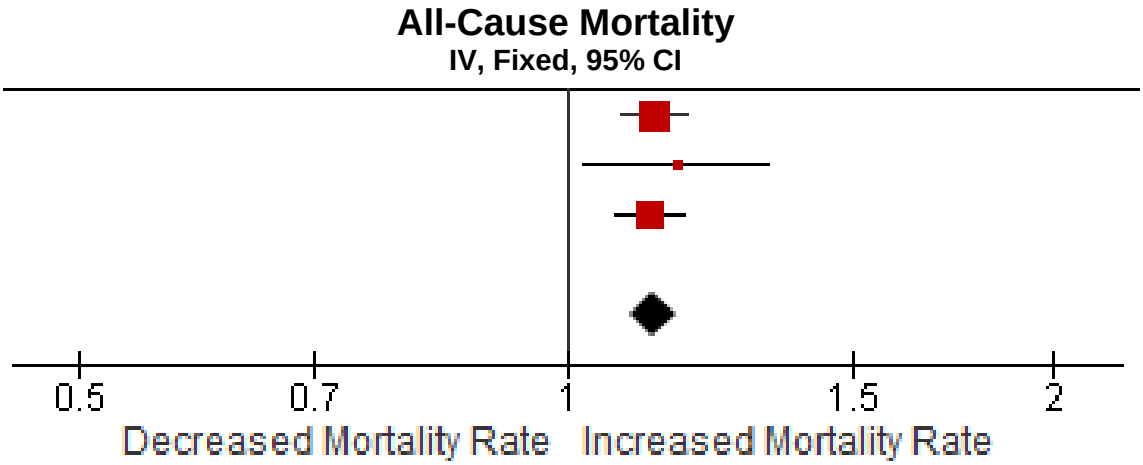


Figure 16: Meta-analysis Results and Forest Plot of All-Cause Mortality

The meta-analysis of the three cohorts yielded an overall hazard ratio of 1.13 (95% CI 1.10–1.17) for all-cause mortality with very little variation between trials ($Chi^2 = 0.31$, $I^2 = 0\%$). The results are graphically displayed below in the Forest Plot. Table 15 contains all data used for the meta-analysis.

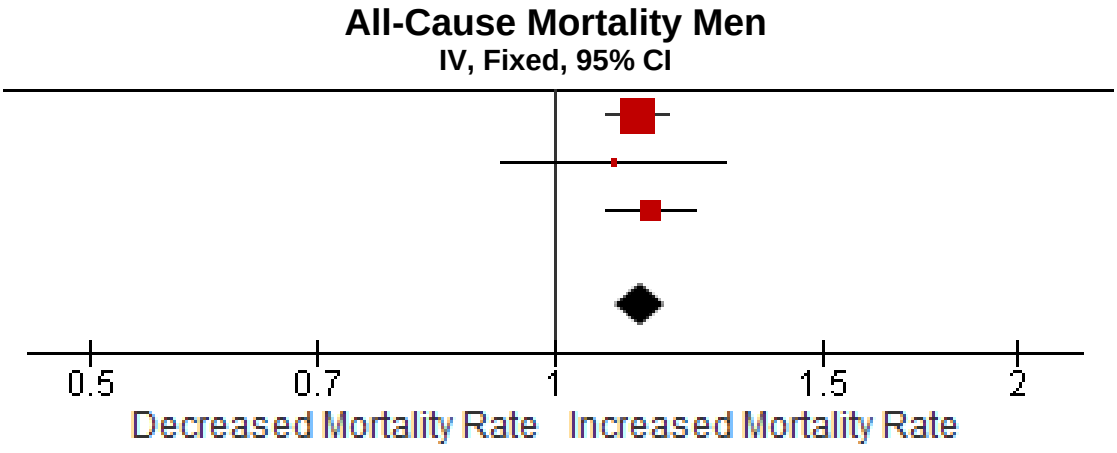


Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Fixed, 95% CI
MrOS	0.12459975	0.024	48.3%	1.13 [1.08, 1.19]
WSC	0.15791624	0.066962	6.2%	1.17 [1.03, 1.34]
SHHS	0.11813373	0.024721	45.5%	1.13 [1.07, 1.18]
Total (95% CI)			**100.0%**	**1.13 [1.10, 1.17]**

Heterogeneity: $Chi^2 = 0.31$, $df = 2$ ($P = 0.86$); $I^2 = 0\%$
Test for overall effect: $Z = 7.42$ ($P < 0.00001$)

Figure 17: Meta-analysis Results and Forest Plot of All-Cause Mortality in Men

The meta-analysis of the three cohorts yielded an overall hazard ratio of 1.14 (95% CI 1.10–1.18) for all-cause mortality in men with very little variation between trials ($Chi^2 = 0.48$, $I^2 = 0\%$). The results are graphically displayed below in the Forest Plot. Table 15 contains all data used for the meta-analysis.



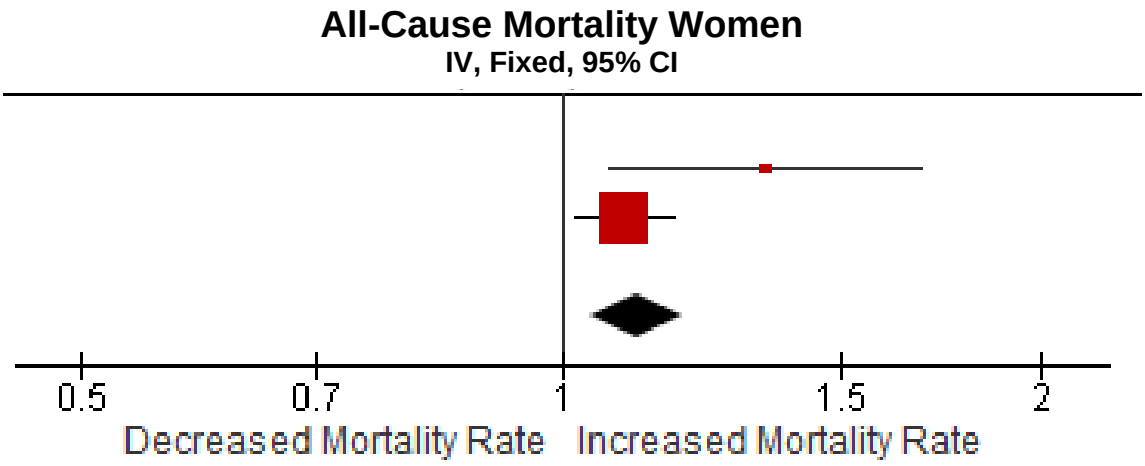
Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Fixed, 95% CI
MrOS	0.12459975	0.024	64.1%	1.13 [1.08, 1.19]
WSC	0.0902419	0.085866	5.0%	1.09 [0.92, 1.30]
SHHS	0.14618976	0.0346	30.9%	1.16 [1.08, 1.24]
Total (95% CI)			**100.0%**	**1.14 [1.10, 1.18]**

Heterogeneity: $Chi^2 = 0.48$, $df = 2$ ($P = 0.79$); $I^2 = 0\%$

Test for overall effect: $Z = 6.74$ ($P < 0.00001$)

Figure 18: Meta-analysis Results and Forest Plot of All-Cause Mortality in Women

The meta-analysis of the three cohorts yielded an overall hazard ratio of 1.11 (95% CI 1.04–1.19) for all-cause mortality in women with some variation between trials ($Chi^2 = 2.93$, $I^2 = 66\%$). The results are graphically displayed below in the Forest Plot. Table 15 contains all data used for the meta-analysis.

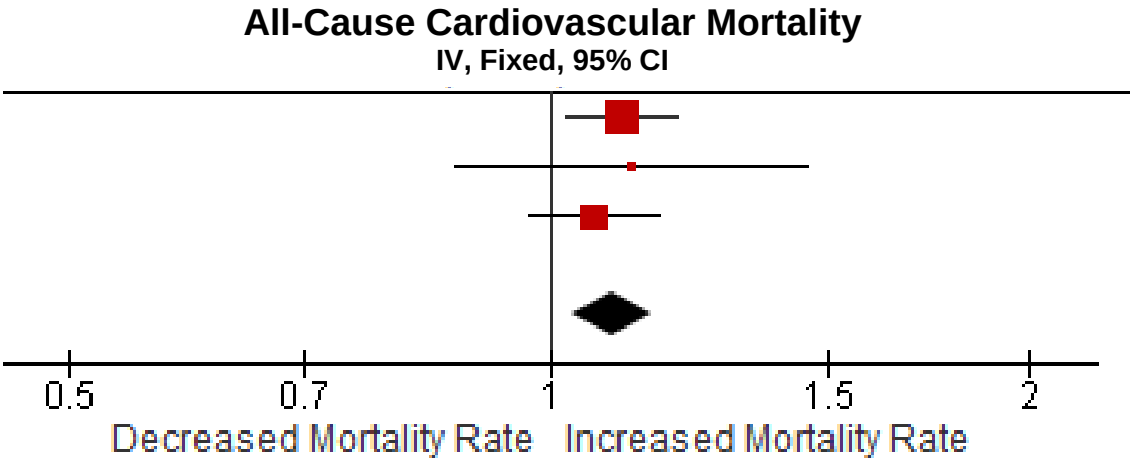


Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Fixed, 95% CI
MrOS	0	0	0.0%	1.13 [1.08, 1.19]
WSC	0.29523423	0.114315	9.1%	1.34 [1.07, 1.68]
SHHS	0.09014806	0.0360616	90.9%	1.09 [1.02, 1.17]
Total (95% CI)			**100.0%**	**1.11 [1.04, 1.19]**

Heterogeneity: $Chi^2 = 2.93$, df = 1 (P = 0.09); $I^2 = 66\%$
Test for overall effect: Z = 3.16 (P = 0.002)

Figure 19: Meta-analysis Results and Forest Plot of Cardiovascular Mortality

The meta-analysis of the three cohorts yielded an overall hazard ratio of 1.10 (95% CI 1.03–1.16) for cardiovascular related mortality with little variation between trials ($Chi^2 = 0.45$, $I^2 = 0\%$). The results are graphically displayed below in the Forest Plot. Table 15 contains all data used for the meta-analysis.



Study or Subgroup	log[Hazard Ratio]	SE	Weight	Hazard Ratio IV, Fixed, 95% CI
MrOS	0.1061602	0.0408842	55.0%	1.11 [1.03, 1.20]
WSC	0.11801856	0.129	5.5%	1.13 [0.87, 1.45]
SHHS	0.06593698	0.048215	39.5%	1.07 [0.97, 1.17]
Total (95% CI)			**100.0%**	**1.10 [1.03, 1.16]**

Heterogeneity: $Chi^2 = 0.45$, $df = 2$ ($P = 0.80$); $I^2 = 0\%$

Test for overall effect: $Z = 3.00$ ($P = 0.003$)