Daniel Vethe

Use of novel technology to advance the assessment and treatment of sleep and circadian rhythms disruption

Blue-depleted light environments, radar assessment, and automated digital therapy

Doctoral thesis

NTNU
Norwegian University of
Science and Technology

Daniel Vethe

Use of novel technology to advance the assessment and treatment of sleep and circadian rhythms disruption

Blue-depleted light environments, radar assessment, and automated digital therapy

Thesis for the Degree of Philosophiae Doctor

Trondheim, March 2023

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Mental Health

NTNU
Norwegian University of
Science and Technology

Populærvitenskapelig tittel: **Kan ny teknologi hjelpe oss å sove?**

Sammendrag

De siste tiårene har samfunnet vært preget av en rask teknologisk utvikling. Stadig økende bruk av kunstig belysning og skjermlys på kvelden byr på utfordringer for døgnrytmen og søvnen vår. Dette skjer mot et bakteppe hvor mange mennesker sliter med dårlig søvn fra før, men hvor god behandling for søvnproblemer er lite tilgjengelig i helsevesenet. Når mennesker legges inn på sykehus topper ofte søvnproblemene seg, men sykehusbelysning og rutiner er lite tilpasset søvn, og kan utgjøre et hinder for søvn samtidig som kroppen er i behov av søvn for tilfriskning. På tross av disse utfordringene, åpner teknologien likevel nye muligheter for måling av søvn og behandling.

Målet med denne doktorgraden har vært å teste tre slike teknologier i settinger hvor de potensielt kan være til stor nytte. Først undersøkte vi effekten av et blå-blokkert lys-miljø ved å la friske deltagere bo midlertidig i en nybygd sykehusavdeling med blå-blokkert lys. Grunnforskning på søvn og døgnrytmer har avdekket at det er kortbølgede lysfrekvenser som forårsaker negative effekter av lys på kvelden, og at utvikling av blå-blokkerte lyskilder kan redusere disse effektene. Det har imidlertid vært ukjent hvorvidt disse effektene kan gjenskapes når teknologien implementeres i store bygninger slik som et sykehus. Deretter undersøker vi hvorvidt man kan måle kontaktløst om et individ sover eller er våken ved hjelp av en svært nøyaktig radar. Dette vil kunne forenkle søvnregistrering i mange settinger, blant annet i sykehus hvor dette kan redusere oppvåkninger som følge av fysiske tilsyn gjennom natten. Radaren ble innfelt i taket over sengen, og ved å registrere mikro-bevegelser forsøker den å avgjøre om et individ er sovende eller våken, helt uten sensorer festet på kroppen. Dette ble gjort både for de friske deltagerne som overnattet i sykehuset og for pasienter som ble undersøkt for søvnplager. Til slutt undersøkte vi om en helautomatisk digital versjon av kognitiv atferdsterapi for insomni (CBT-I) kan anses å ikke være underlegen standard ansikt-til-ansikt CBT-I for pasienter henvist til en søvnpoliklinikk på sykehusnivå. Digital CBT-I har vært testet i mange studier, og den skalerbarheten som følger med digital behandling kan potensielt løse eksisterende kapasitets-utfordringer knyttet til å tilby CBT-I til pasienter med insomni. Det mangler imidlertid kunnskap om hvor god denne behandlingen er sammenlignet direkte med gull-standard ansikt-til-ansikt behandling.

Studiene vi har gjennomført viser at å oppholde seg i et blå-blokkert lysmiljø på kvelden gjør at friske deltagere blir mer A-mennesker, produserer mer melatonin på kvelden, sover litt lengre, har mer REM-søvn og samtidig bedre kvalitet på REM-søvnen sin. Radar måling av søvn fremstår som en teknologisk nyvinning med omtrent tilsvarende presisjon som en søvnmåler festet på håndleddet, men kontaktløshet er en tydelig fordel, samtidig som det er et potensial for økt presisjon med nye analysemetoder i fremtiden. Til slutt viste det seg at ansikt-til-ansikt CBT-I var bedre enn helautomatisk digital CBT-I i denne settingen. Vi kunne ikke påvise at helautomatisk digital CBT-I var ikke-underlegen ansikt-til-ansikt behandling.

Konklusjonen i denne doktorgraden er at blå-blokkerte lysmiljøer og radar-måling av søvn viser et stort potensial til å bedre søvnbehandling og søvnmåling. Spesielt gjelder dette i konteksten av et sykehus, hvor søvnvansker er utbredt og kartlegging av søvn er vanskelig å gjennomføre uten å forstyrre pasienten. Det oppfordres til videre studier av disse teknologiene i kliniske populasjoner. Helautomatisk digital søvnbehandling kan fortsatt være effektivt på andre nivåer i helsetjenesten, men akkurat for pasienter som henvises for insomni til spesialisthelsetjenesten, fremstår ansikt-til-ansikt CBT-I fortsatt som beste behandlingsalternativ.

Summary

Background and aims

Recent decades have witnessed an upsurge in interest in sleep problems across social, clinical and research settings. This interest has coincided with developments in technology that have not only contributed to the high prevalence of poor sleep (e.g., late night engagement with social media), but also underpinned new approaches to the management of sleep problems. For example, one cause of sleep and circadian rhythms disruption is evening and nighttime exposure to artificial light. However, sleep and circadian rhythms research has discovered that the negative effects of evening light exposure are primarily mediated by short-wavelengths, and that the use of blue-depleted light sources may mitigate against these effects. Further, new radar technology may offer novel ways to assess sleep patterns across a range of settings without inconveniencing an individual (as they do not need to wear any electrodes or other monitoring devices). Also, the introduction of fully automated digital sleep therapies can improve access to potentially efficacious interventions. This is especially important given the relative lack of availability of therapists who are trained in cognitive behavioral therapy for insomnia (CBT-I) or similar interventions.

Given the above advances in technology, this thesis offers a timely examination of three recent technological advances, namely: blue-depleted lighting, radar sleep assessment, and digital cognitive behavioral therapy for insomnia (dCBT-I). The four publications included in this thesis describe two randomized trials (one with healthy participants and one with individuals with insomnia), as well as cross-sectional analyses of data collected from a hospital outpatient neurophysiology sleep clinic. Papers I and II provide a detailed exploration of the effects on the sleep and circadian rhythms of young adults who take up short-term residence in a blue-depleted lighting environment. Paper III explores technological approaches to sleep monitoring and assesses the accuracy of the classification of sleep-wake states by a radar sensor system compared with polysomnography (PSG) and actigraphy. Finally, Paper IV examines interventions for sleep problems and describes a randomized trial of dCBT-I compared with face-to-face (FtF) CBT-I for patients referred to a secondary care sleep clinic.

Methods

The randomized cross-over trial of a blue-depleted light environment (BDLE) versus a standard light environment (SLE) (papers I and II) recruited 12 healthy young adults who resided for 5 days in a BDLE followed by 5 days in a SLE (or vice versa). Melatonin assessments were performed to measure circadian rhythms, PSG was used to assess sleep, neurocognitive testing was used to assess alertness, questionnaires assessed subjective sleepiness and side effects, and color discrimination ability was assessed with an objective test.

Paper III on radar assessment of sleep analyzed triple-registered sleep data from radar, actigraphy, and PSG. These analyses utilized data from *the BDLE trial* (n=12) and data collected from 28 individuals referred to an outpatient neurophysiological sleep clinic for ambulatory sleep assessments. *The randomized trial of dCBT-I versus FtF CBT-I* (paper IV) recruited 101 participants with insomnia from individuals referred to a secondary care sleep clinic that usually treats insomnia and circadian rhythms disorders. The trial used a non-inferiority design; the primary outcome was assessed using the Insomnia Severity Index (ISI), and the non-inferiority margin was defined as 2 points on the ISI.

Results

In paper I, we found reduced melatonin suppression in individuals residing in the BDLE (15%) compared with the SLE (45%; $p = 0.011$). A larger phase-advance of dim light melatonin onset (DLMO) was observed after residing in the BDLE (1:20 h) than after the SLE (0:46 h; $p = 0.008$). Further, the overnight sleep duration was 8.1 min longer ($p = 0.032$), rapid eye movement (REM) sleep was 13.9 min longer ($p < 0.001$), and neurocognitive arousal was lower ($p = 0.042$) in the BDLE. There were no significant differences in subjective sleepiness ($p = 0.16$) or side effects ($p = 0.09$), but there was lower color-discrimination ability in the BDLE ($p < 0.001$).

In paper II, we found that the fragmentation of REM sleep was lower after evening BDLE compared with SLE ($p = 0.0003$). Similarly, we found fewer REM sleep microarousals ($p = 0.0493$) after BDLE. While REM sleep fragmentation was related to

DLMO phase-shift, BDLE had a unique effect over and above this association. There were no changes in REM density or latency to the first REM sleep episode. Increased accumulation of REM sleep in BDLE was not at the expense of non-REM stage 3 sleep.

In paper III, we found that for real-time models, accuracies were above 92%, sensitivities above 95%, specificities above 83%, and all Cohen's kappa values were above 0.81 compared to polysomnography in the healthy participants. In the patient dataset, accuracies were above 81%, sensitivities about 89%, specificities above 53%, and Cohen's kappa values above 0.44. We found no evidence of significant inter-method bias for sleep variable estimates, but in the patient dataset, the limits of agreement were wide.

In paper IV, we found that both interventions significantly reduced insomnia severity ($p < 0.001$), although the magnitude of reduction in ISI scores was greater in FtF compared with dCBT-I. At the primary endpoint of week 33 the mean between-group difference in ISI was −2.8 (95% CI: −4.8 to −0.8; $p = 0.007$, Cohen's d = 0.7). At week 9 the mean between-group difference in ISI was −4.6 (95% CI −6.6 to −2.7; p < 0.001, Cohen's d = 1.2).

Conclusions

The work included in this thesis shows that BDLE is effective in terms of reducing the adverse effects of evening light exposure for healthy subjects while residing in a hospital ward. This serves as a proof-of-concept for further investigations in different clinical populations. Further, we show that radar assessment of sleep can be undertaken both in the hospital ward and outpatient setting, allowing contact-free, real-time monitoring of sleep-wake state. For outpatients referred for insomnia, we do not show that dCBT-I is non-inferior. Although dCBT-I offers other benefits such as facilitated dissemination of treatment, these findings indicate FtF CBT-I should remain the treatment of choice in the context of secondary care sleep clinics. This series of papers suggests that technology can and will play a role as efforts are made to improve the treatment and assessment of sleep and circadian rhythms disruption in both healthcare settings and beyond.

Acknowledgements

First and foremost, I would like to thank the participants for the time and effort they have dedicated to participating in these studies. Their contributions are the mainstay of the work presented in this thesis. Furthermore, I would like to extend my gratitude to my supervisors. Thank you, **Håvard Kallestad**, for your continuing mentoring and encouragement on the scientific work and your flexibility and support on the personal level many times through this process, especially over this last year. Thank you, **Jan Scott**, for your patience and persistent tutoring in both writing and the general scientific process. Thank you, **Gunnar Morken**, for your guidance on effectively writing this thesis. Thank you, **Knut Langsrud**, for mentoring my clinical work. I would also like to acknowledge the larger clinical psychiatry and research community at Østmarka. The trials planned and performed during these years are a joint team effort of research and clinical staff in all positions at the hospital department. I am grateful for being part of this process and for the opportunity I have been given by writing up this work.

I would also like to thank the research group, in particular the other PhD candidates and junior researchers, for the great times we have had along the way and for mutual encouragement. Thanks to all my friends and family for their support and inspiring words. Lastly, thank you Cecilie, for your patience, support and just being the most amazing wife and partner in all areas of life. Your light is so bright it would phase-shift any circadian rhythm.

List of papers

Paper I:

The evening light environment in hospitals can be designed to produce less disruptive effects on the circadian system and improve sleep

Daniel Vethe, Jan Scott, Morten Engstrøm, Øyvind Salvesen, Trond Sand, Alexander Olsen, Gunnar Morken, Hanne Siri A. Heglum, Kaia Kjørstad, Patrick M. Faaland, Cecilie L. Vestergaard, Knut Langsrud and Håvard Kallestad.

Sleep, 2021, Vol. 44, No. 3, zsaa194. https://doi.org/10.1093/sleep/zsaa194

Paper II:

Evening light environments can be designed to consolidate and increase the duration of REM-sleep

Daniel Vethe, Henning. J. Drews, Jan Scott, Morten Engstrøm, Hanne Siri A. Heglum, Janne Grønli, Jonathan P. Wisor, Trond Sand, Stian Lydersen, Kaia Kjørstad, Patrick. M. P. Faaland, Cecilie L. Vestergaard, Knut Langsrud and Håvard Kallestad.

Scientific Reports, 2022, *12*(1), 8719. https://doi.org/10.1038/s41598-022-12408-w

Paper III:

Distinguishing sleep from wake with a radar sensor: A contact-free real-time sleep monitor

Hanne Siri Amdahl Heglum, Håvard Kallestad, Daniel Vethe, Knut Langsrud, Trond Sand and Morten Engstrøm

Sleep, 2021, *44*(8), zsab060. https://doi.org/10.1093/sleep/zsab060

Paper IV:

Mode of delivery of Cognitive Behavioral Therapy for Insomnia: A randomized controlled non-inferiority trial of digital and face-to-face therapy

Håvard Kallestad, Jan Scott, Øystein Vedaa, Stian Lydersen, Daniel Vethe, Gunnar Morken, Tore Charles Stiles, Børge Sivertsen and Knut Langsrud

Abbreviations and acronyms

Melanopic DER	Melanopic Daylight Efficacy Ratio
Melanopic EDI	Melanopic Equivalent Daylight Illuminance
ASPD	Advanced Sleep Phase Disorder
AUC	Area Under the Curve
BDLE	Blue-depleted Light Environment
BZ	Benzodiazepine
BZRA	Benzodiazepine receptor agonist
C-CPT-3	Connors Continuous Performance Test-3
CBT-i	Cognitive Behavioral Therapy for Insomnia
CCT	Correlated Color Temperature
CFS	The Chalder Fatigue Scale
CI	Confidence Interval
CIE	International Commission on Illumination
CONSORT	Consolidated Standards of Reporting Trials
CST	Consumer Sleep Technologies
D65	Standard daylight.
DBAS	Dysfunctional Beliefs About Sleep scale
dCBT-i	Digital Cognitive Behavioral Therapy for Insomnia
DLMO	Dim Light Melatonin Onset
DMH	Dorsomedial Hypothalamus
DRN	Dorsal Raphe Nucleus
DS	Dataset
DSM	Diagnostic and Statistical Manual of Mental Disorders
DSPD	Delayed Sleep Phase Disorder
EDI	Equivalent Daylight Illuminance
EEG	Electroencephalography
EMA	Early Morning Awakenings
EMA	European Medicines Agency
ES	Effect Size
FM-100	Farnsworth–Munsell 100 Hue Color Vision Test
FtF	Face to Face
GP	General Practitioner
HADS	Hospital Anxiety and Depression Scale
ICD	International Classification of Disorders
ICU	Intensive Care Unit
ipRGC	Intrinsically Photosensitive Retinal Ganglion Cells
IR-UWB	Impulse-Radio Ultra-Wideband
ISI	Insomnia Severity Index
ITT	Intention-to-treat
KSS	Karolinska Sleepiness Scale
LAs	Limits of Agreement
LC	Locus coeruleus of pons in the brain stem
LE	Light environment
LED	Light Emitting Diode
LH	Lateral Hypothalamus
MID	Minimally Important Difference

NA	Noradrenaline
NREM	Non-rapid Eye Movement
NW	Number of Wakening
OPN	Olivary Pretectal Nucleus
OSA	Obstructive Sleep Apnea
PAG	Periaqueductal Gray
PAG	Periaqueductal Gray Area
PDSQ	Psychiatric Diagnostic Screening Questionnaire
PHb	Peri-habenular area of the dorsal thalamus
POA	Preoptic Hypothalamic Area
PP	Per Protocol
PSG	Polysomnography
RCT	Randomized Controlled Trial
REM	Rapid Eye Movement
RHT	The Retinohypothalamic Tract
RPM	Respirations Per Minute
SACR	Trondheim Sleep and Chronobiology Research group
SCN	Suprachiasmatic nucleus of the anterior hypothalamus
SD	Standard deviation
SE	Sleep Efficiency
SLE	Standard Light Enironment
SOL	Sleep Onset Latency
SRI	Sleep Regularity Index
SWA	Slow-wave Activity
SWS	Slow-wave Sleep
TST	Total Sleep Time
UKU	The Side Effect Rating Scale created by the Udvalg for Kliniske Undersøgelser, Scandinavian Society for Psychopharmacology
VLPO	Ventrolateral Preoptic Area
WASO	Wake After Sleep Onset

Table of contents

1 Introduction

Sleep is as natural to human beings as breathing. We sleep even before we are born. Daily we take part in the temporary cessation of our consciousness, mostly without questioning the process. We accept this and even enjoy the warm relaxation brought by a long night's sleep. However, the physical discomfort after a single night without sleep instills some degree of apprehensiveness in most. And when sleep remains troublesome over time, this may induce anxiousness and worry and eventually lead to a distraught search for solutions. Therefore, what is naturally a source of health, relaxation, and energy, often proves to be a source of frustration, fatigue, and discontent.

The rapid technological development over the last centuries is a source of pride in modern society. These advances offer both opportunities and challenges when it comes to sleep. After millions of years at the mercy of the sun's light-dark cycle, the advent of electric lighting has allowed us to reshape the light-dark cycle in the span of a few hundred years.

> *"Let there be light, and there was light. And he saw that the light was good, and He separated the light from the darkness. God called the light "day," and the darkness He called "night". And there was evening, and there was morning—the first day." – Genesis 1:3.*

With light and darkness under our control – by the flick of a switch or even a word spoken – we forgo the quiescent evenings spent dimly lit by a campfire, instead engaging in work, play, or consuming content, while artificial light from screens, devices, and lamps keep the darkness at bay.

> *"Technology is a useful servant but a dangerous master." – Christian Lous Lange*

However, light is also the most potent time-giver to the circadian system, and the body depends on the regular shift between light and darkness to tell time. As we continuously adapt the timing and intensity of light exposure to match plans, desires, or work schedules, the signal of light vs. darkness grows increasingly unreliable in terms of time-telling. With our circadian rhythms taking the toll, the likelihood of developing sleep-wake difficulties may increase.

While we often engage in the mental exercise of trying to separate modern man from nature, the human body undeniably displays the same circadian rhythms characteristics as other mammals on earth. Moreover, not only does artificial light challenge circadian rhythmicity, but technological devices and brightly lit rooms make our evening activities more engaging. This may pose a problem in the form of more mental activity and higher alertness in the late evening hours leading up to bedtime. When the use of evening light emitting devices is unrestricted, this has been found to delay self-selected bedtimes compared with limited use (Chinoy et al., 2018). Further, the organization of modern society has also reached a point where we depend on around-the-clock services, particularly within healthcare. Night-shift workers are exposed to brightly lit environments both night and day, which brings havoc to the circadian system. This has been found to have downstream negative health consequences for these individuals, so much so that the world health organization recently included night shift work as a probable carcinogen (Ward et al., 2019).

Another group at risk of being systematically exposed to ill-timed artificial lighting is patients in the healthcare system. Patients admitted to hospitals need acute treatment for their physical or mental illnesses. Some of these illnesses may lead to poor sleep on their own due to, e.g., pain or agitation, and it is usually the case that sleep is considered to be an essential part of recovery and healing. However, acute hospital wards may be some of the worst environments for inducing sleep (Boyko et al., 2017; Engwall et al., 2015, 2017), with daytime light exposure being too low, nighttime light exposure being too high, monitors beeping and blinking, and nurses entering and exiting patient-rooms during the night (Giménez et al., 2017; Pisani et al., 2015). This is a context in which technology is already being utilized deliberately to address the patient's specific illnesses but where its inadvertent effects on sleep may interfere with healing.

In addition to any direct adverse effects of new technology, there may also be more subtle changes that affect the general population. Sleep has long been a window of downtime from the commercial commodification of our time and attention (Jansson & Adams, 2021). However, over the last decades, a new industry, the consumer sleep-tech industry, has found ways to move into the sleep domain. Sleep trackers are embedded in watches, jewelry, or mattresses. Sleep masks have embedded electrodes and claim to measure brainwaves while trying to lull you into deeper, better sleep by playing rhythmical sounds. Advertisements

essentially deliver a universal message that more is to be gained from sleep, even if you thought your sleep was good. While it is probably true that some of these technologies may improve sleep, at least for certain individuals, many of these products will likely not be effective for most individuals with both normal or disturbed sleep. However, increased attention towards sleep may foster more rumination or worry about sleep. Even if it is false, receiving negative sleep feedback can worsen daytime function and perceived sleepiness (Gavriloff et al., 2018). The increasing concern with optimizing sleep has led researchers to coin the term "orthosomnia" (Baron et al., 2017), which describes how striving to attain perfect sleep may increase anxiousness and lead previously healthy sleepers to develop sleep problems and even insomnia.

Given that the technological advances in modern society may both have the potential to improve and impair sleep, we are faced with one obvious obstacle: Sorting the effective from the ineffective interventions. Moreover, we must be conscious of how and when we utilize technology to improve sleep. This means distinguishing for whom an intervention is beneficial and in what context, as an ineffective intervention may not just be neutral but add to negative sleep focus. In the current thesis, I describe some concrete issues related to sleep and how new technology may be used to target these issues specifically. This thesis contains three studies, split into four papers, that test the effectiveness of three different technologies. The first study concerns using new lighting technology to adapt light exposure to human circadian biology. The second concerns the use of new radar technology to make sleep monitoring more accurate and less invasive. The third is using the internet to distribute at-home, fully automated, digital treatment for insomnia.

1.1 Light and circadian rhythms

Circadian is derived from the Latin "circa" meaning "about" and "dia" meaning "day", thus circadian rhythms denote rhythms that oscillate approximately once every 24 hours. Many functions in the human body display a circadian rhythm, e.g., sleep-wake cycles, melatonin and cortisol secretion, core body temperature, and many others. A series of early seminal studies of circadian rhythms let participants reside in bunkers completely cut off from external time cues such as watches, incident light, or sound. In this environment, circadian rhythms continued to oscillate, suggesting humans - like other plants and animals - have an endogenous circadian oscillator or pacemaker that can operate without external light stimuli

(Aschoff, 1965; Aschoff et al., 1967). What was also evident was that the circadian rhythms in this "free-running" state followed a slightly longer than 24h period. Although the human circadian pacemaker can generate a circadian rhythm in the absence of light, in a typical environment, the circadian system receives light input from the retina and entrains to the time-giving (from the German "zeitgeber") signal of the light-dark cycle, synchronizing bodily functions to its 24h period.

The human visual system contains the photoreceptors rods that mediate scotopic vision under low light conditions and cones that mediate photopic high spatial acuity vision, including color vision, in normal light conditions. The rods contain the photosensitive pigment rod opsin that is maximally sensitive to light around 507nm. The cones are divided into three subtypes, of which S-cones contain the pigment cyanolabe that is sensitive to short wavelengths (λ_{max} 420nm), M-cones use the opsin chlorolabe (λ_{max} 535nm), and L-cones with the erythrolabe opsin (λ_{max} 565nm) (Lucas et al., 2014). Since early studies found that higher levels of ocular illuminance were needed for circadian rhythm responses than for vision, it was thought that the photopic system was primarily responsible for light's effects on circadian rhythms. However, several studies found that light had effects on circadian processes even in the absence of rods and cones (Freedman et al., 1999; Lucas et al., 1999) and that effects were stronger for shorter wavelengths than would be expected with the photopic system (Brainard, Hanifin, Rollag, et al., 2001).
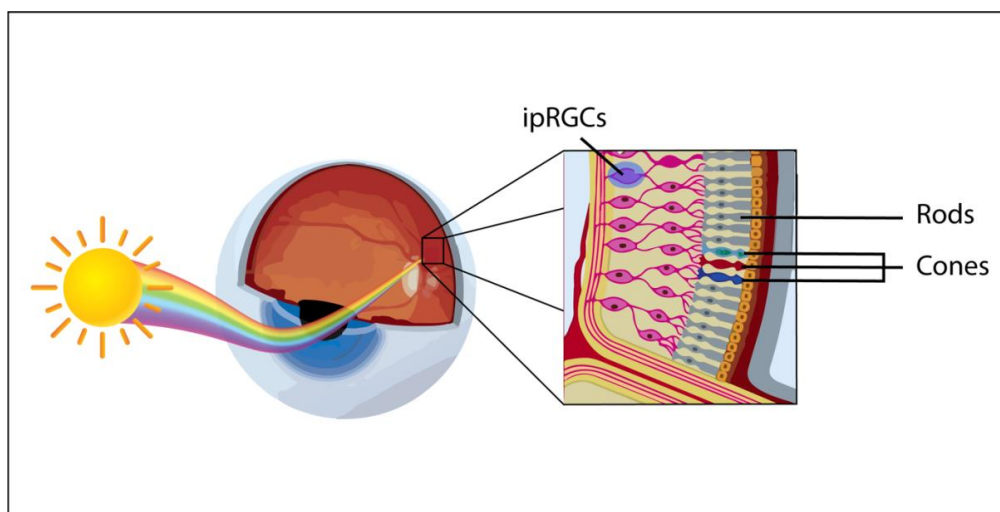
Figure 1. The constriction of the pupil controls the amount of light entering the eye.
Cone and rod photoreceptors are responsible for color and scotopic vision, respectively.
A subtype of retinal ganglion cells expresses the photosensitive pigment melanopsin,
which results in the third type of retinal photoreceptor: the intrinsically photosensitive
retinal ganglion cells (ipRGCs). Adapted from Blume et al. (Blume et al., 2019).

In the early 2000s, the discovery of a new photoreceptor, the intrinsically photosensitive
retinal ganglion cells (ipRGCs) (Brainard, Hanifin, Greeson, et al., 2001; Hattar et al., 2002;
Provencio et al., 2000; Thapan et al., 2001), paved the way for a new wave of circadian
rhythms research. The ipRGCs were found to project to the suprachiasmatic nucleus (SCN) of
the anterior hypothalamus, which had already been identified as the central circadian
oscillator in the brain (Ibuka & Kawamura, 1975; Mouret et al., 1978; Stephan & Nunez,
1977). It was known that melatonin secretion followed the circadian rhythms generated by the
SCN and thus also was entrained by light. Moreover, nighttime melatonin secretion could be
acutely suppressed by light (Lewy et al., 1980; Rollag & Niswender, 1976) and was already
used as an indicator of light input to the SCN. When exposed to light of differing
wavelengths, it was found that melatonin suppression was most potent in an action spectrum
that matched the spectral sensitivity of the ipRGCs with a peak around 480nm (see Figure 2)
(Brainard, Hanifin, Greeson, et al., 2001; Thapan et al., 2001). This led to the assumption that
ipRGCs mediated the observed effects of light on circadian rhythms and opened for new
study designs targeting the action spectrum of the ipRGCs. Seeing as it was a separate
photoreceptor system altogether, studies could now use light to target this specific

photoreceptor and measure effects on bodily functions that display circadian rhythm properties such as core temperature and melatonin production.
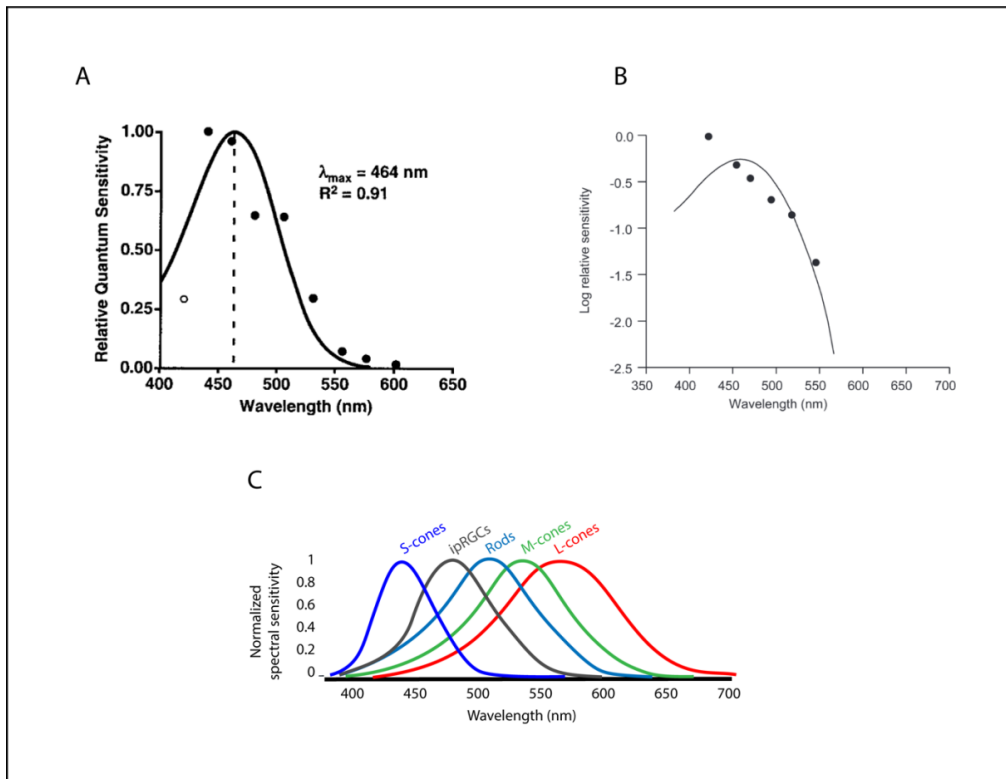
Figure 2. *The early action sprectra reported for melatonin suppression did not match well with spectral sensitivities for rods and cones, but was similar to that of the ipRGCs. A) Action spectrum for melatonin suppression found by Brainard and colleagues (Brainard, Hanifin, Greeson, et al., 2001). B) Similar action spectrum for melatonin suppression reported in Thapan and colleagues (Thapan et al., 2001). C) Spectral sensitivity of melanopsin versus the other photoreceptors of the human eye (adapted from Blume et al. (Blume et al., 2019)).*

The ipRGCs make up about 4-5% of the photoreceptors on the retina (Schmidt et al., 2011), and in contrast with the rods and cones, they express the photopigment melanopsin (Hattar et al., 2002). In addition to utilizing information from its photopigment, it also integrates and relays some information from rods and cones (Guler et al., 2008; Schmidt et al., 2011). Initially, it was thought that the ipRGCs' only function was to relay light information through

the retinohypothalamic tract (RHT) to the circadian pacemaker of the SCN, thus entraining the organism to the external light-dark cycle. However, it was later discovered that the ipRGCs innervate directly to a range of nuclei in the hypothalamus and limbic areas (Hannibal et al., 2015; Hattar et al., 2006). These include, e.g., the olivary pretectal nucleus (OPN) (central in the pupillary light reflex), the periaqueductal gray (PAG), and the ventrolateral preoptic area (VLPO). Interestingly the VLPO is central in switching between sleep and wake states (Saper et al., 2001), and there are also innervations from the SCN to the VLPO. Not surprisingly, light mediated by the ipRGCs has also been found to have numerous other non-visual or non-image-forming effects besides regulating circadian rhythms (Cajochen et al., 2005; Rahman et al., 2014). ipRGCs have even been found to aid visual perception under certain conditions (T. M. Brown et al., 2012; Zaidi et al., 2007).

1.1.1 Acute versus circadian rhythms effects of light

The many non-image-forming effects associated with ipRGC activation can broadly be divided into circadian rhythms-mediated/indirect effects and acute/direct effects. Via ipRGC pathways to the circadian pacemaker/oscillator in the SCN, light indirectly affects bodily functions through the entrainment and synchronization of the circadian clock. All human cells contain circadian oscillators entrained by the central pacemaker, ensuring proper timing and function across tissues. If circadian rhythms are disrupted, this negatively affects many functions heavily regulated by circadian rhythms, including sleep, core temperature, alertness, attention, etc.

Light exposure also has acute non-image-forming effects on humans in the domains of alertness, cognitive processes, sleep/wake cycle, pupillary light reflex, heart rate, body temperature, and melatonin secretion (Cajochen et al., 2005; Chang et al., 2013; Chellappa et al., 2011, 2017; Münch et al., 2006). Although circadian processes regulate sleep/wake cycles, light simultaneously directly modulates sleep (Zhang et al., 2021). Such acute effects may be independent of circadian processes arising immediately when exposed to light, or in the case of sleep, also prebedtime lighting may affect the following sleep episode.

It is generally agreed upon that the circadian rhythms effects of light are mediated through the SCN (LeGates et al., 2014; Rupp et al., 2019). However, the pathways underlying the acute non-visual effects of light are still being discovered. A recent study using animal models has

demonstrated that ipRGC innervations to the perihabenular area (PHb) may be responsible for light's acute effects on mood, whereas some acute effects on cognition may be mediated via the SCN but operate independently of circadian rhythms processes (Fernandez et al., 2018). Furthermore, distinct subtypes of ipRGC cells may be responsible for circadian rhythms and acute effects (Rupp et al., 2019).

1.1.2   The effects of light on sleep

Light's effect on sleep can also be divided into indirect circadian processes and direct acute effects. The circadian processes of sleep regulation have been well understood for some time and were incorporated into the influential two-process model for sleep-wake regulation some decades ago (Borbely, 1982). This model has laid the foundation for how we treat sleep disorders related to circadian rhythms and insomnia. As acute effects of light exposure have been further explored, it is evident that light also has considerable non-circadian effects on sleep.

1.1.2.1   The two-process model of sleep-wake regulation

The two-process model consists of two main processes, the circadian process C and the homeostatic sleep-dependent process S. At the time when this model was postulated, the exploration of the biological mechanisms sustaining and regulating circadian rhythms was still in its early beginnings. However, observations of rhythmicity in sleep propensity during sleep deprivation suggested that the circadian oscillator was central in sleep-wake regulation (Akerstedt & Froberg, 1977). Although sleep deprivation did not increase subsequent sleep duration proportionally, it was found that the amount of slow-wave activity (SWA) during deep sleep early in the night was dependent on the duration of prior wakefulness (Berger & Oswald, 1962) and that the amount of SWA declined with each sleep cycle through the night (Borbely, 1982; Borbély et al., 1981). It was therefore hypothesized that this represented a marker for sleep intensity and that the drive for sleep increases with wakefulness and declines with sleep. It follows from this model that process C and process S interact and that both factors must be considered to explain, e.g., sleep latency and the duration of sleep episodes (Borbely, 1982; Borbély et al., 2016).

A circadian drive for wakefulness, generated by the circadian pacemaker of the SCN, interacts with the homeostatic process S and translates into a 24h sleep-wake propensity rhythm. This circadian drive for wakefulness results from projections from the SCN to wake-promoting regions of the brainstem and the hypothalamus, regulating the wake-promoting orexins (Scammell et al., 2017), and inhibiting the sleep-promoting system of the VLPO (Gaus et al., 2002). During the daytime, the circadian drive for wakefulness increases. This happens simultaneously as melatonin levels are low or undetectable, cortisol concentrations steadily decrease after peaking in the morning, and body temperature increases. This circadian drive for wakefulness peaks in the biological evening (typically between 19:00-22:00 in healthy individuals) and compensates for the increasing homeostatic sleep pressure (Meyer et al., 2022). Shortly after, the circadian drive for wakefulness dissipates while melatonin concentrations increase and core body temperature decrease (Czeisler, 1999; Dijk & Czeisler, 1995). The circadian drive for wakefulness is at its lowest around the core body temperature nadir, approximately two hours before habitual wake time. During the biological day, we can easily stay awake for 16 hours even with rising homeostatic sleep pressure and may even find napping difficult. However, as the circadian drive for wakefulness dissipates during the biological night, it becomes more difficult to fight off sleep. On the other hand, if the homeostatic sleep pressure is unusually high, e.g., after sleep deprivation, sleep can be initiated during the daytime. Still, the circadian drive for wakefulness may hamper the duration and quality of sleep. Thus, an alignment of the dissipation of circadian drive for wakefulness and high homeostatic sleep pressure is needed if the conditions for good sleep are to be optimal.

1.1.2.2 *Timing of light and its effect on circadian rhythms phase*

The endogenous circadian rhythms generated by the SCN can be advanced, delayed, or stabilized by light exposure. The timing of light exposure in relation to the individual's circadian phase determines the effect's direction and size. By exposing participants to bright white light (10000 lux) at differing times, a phase-response curve can be generated (see Figure 3) (Khalsa et al., 2003; St Hilaire et al., 2012). The consequent phase-response curve is generated using the onset or midpoint of melatonin production to center values. Bright light exposure in the morning (after the core body temperature low point) has a phase-advancing effect. In contrast, evening and night light exposure has a phase-delaying effect (largest effect from Dim Light Melatonin Onset (DLMO) until the melatonin midpoint) (Khalsa et al., 2003;

Rüger et al., 2013; St Hilaire et al., 2012). Further investigations using monochromatic light with a wavelength of 480nm have found similar phase-response curves using a much lower intensity (~4% of the photon density) (Rüger et al., 2013), indicating that wavelengths around the λ_{max} of ipRGCs are more potent in eliciting circadian phase-shifting effects.
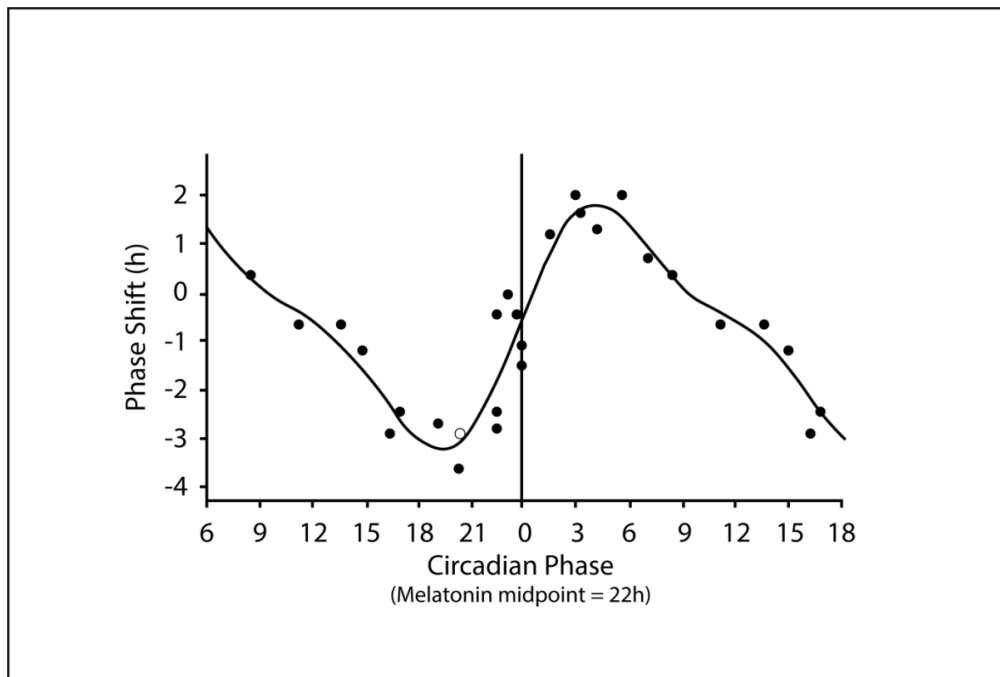
Figure 3. *The phase response curve for the exposure to 6.7 hours of bright light (10000 lux) (adapted from Khalsa et al. 2003). The melatonin midpoint from the pre-exposure melatonin assessment was used to center the data, with the melatonin midpoint being set to 22h. The core body temperature minimum was assumed to be 2h after the melatonin midpoint at 0h.*

1.1.2.3 Effects of daytime light on sleep

Effects of daytime light exposure are also reflected directly in sleep parameters. Compared with morning bright light exposure, evening light exposure increases sleep onset latency, whereas morning light exposure shortens it (Carrier & Dumont, 1995; Christian et al., 1992). Late exposure to daylight has been reported to increase the number of nocturnal awakenings and lead to less slow-wave sleep (SWS), whereas more daytime light exposure is associated with more SWS the following night (Wams et al., 2017). In a controlled laboratory study, 250

lux white or blue-enriched light was found to increase subsequent SWS and slow-wave activity (SWA) after 40 hours of extended wakefulness when compared to dim light exposure (8 lux) (Cajochen et al., 2019). This indicates that if daytime light intensities are too low, which likely was the case during the Covid-19 pandemic (Blume et al., 2020; Wright et al., 2020), this affects the subsequent sleep architecture. Interestingly, it has also been found that the duration of daylight, which in most parts of the world vary with seasons, has an effect where shorter days tend to prolong the melatonin production phase, creating a longer biological night and also leading to longer sleep times in a natural environment (Stothard et al., 2017). In pre-industrial societies, one study also indicates longer sleep in the winter months (Yetish et al., 2015).

1.1.2.4 *Effects of evening and nighttime light on sleep*

Modern evening light environments are increasingly permeated by artificial light in the form of indoor lighting and the backlit screens of numerous digital devices. This evening light exposure occurs as the human circadian system is most sensitive to circadian phase delays (Khalsa et al., 2003; Rüger et al., 2013; St Hilaire et al., 2012). Significant melatonin suppression has been found in response to a range of different evening artificial light sources (Cajochen et al., 2005; Chang et al., 2015; Santhi et al., 2012; Zeitzer et al., 2000). Moreover, artificial light exposure during the evening delays sleep onset (Chang et al., 2015; Y. Cho et al., 2015; Gordijn et al., 1999; Komada et al., 2000; Santhi et al., 2012), at least in part due to its phase-delaying properties. In addition, the acute effects of evening light exposure on alertness compound the issue and further complicate falling asleep (Chang et al., 2015; Rahman et al., 2014, 2017; Vandewalle et al., 2007, 2009).

The effects of evening light exposure also seem to persist into the following sleep episode. In a controversial study, Chang and colleagues (2015) compared reading a light-emitting e-book with an ordinary book in near darkness and found expected reductions in melatonin secretion and a later circadian phase. Further, subjective sleepiness before bedtime was reduced, and wake-electroencephalogram (EEG) measures indicated higher alertness (less delta-theta activity). When going to bed, participants had 10 min longer sleep latency, and overnight Rapid Eye Movement (REM)-sleep duration was shorter. A similar study by Grønli and colleagues (Grønli et al., 2016) found that reading on an iPad compared with an ordinary book in low room light levels delayed SWA dynamics about 30 minutes and reduced SWA after sleep onset. Reduced SWA in the first sleep cycle was also found for a blue-enriched

polychromatic light source (Chellappa et al., 2013). Using monochromatic light at different wavelengths, an early study by Munch et al. (2006) found that short-wavelength light (460nm) compared with longer wavelength light (555nm) exposure in the evening led to SWA being reduced in the first sleep cycle, and increased in the third, thereby shifting SWA dynamics later in the night. Moreover, they found reduced REM-sleep duration in the first and third sleep cycle. Thus, demonstrating that the effects of evening light exposure on sleep EEG also may be mediated by ipRGC activation.

It was previously assumed that disruptive effects on circadian rhythms and sleep were minimal at low intensities. Recently, it has been shown that humans are exquisitely sensitive to the effects of evening light exposure, even at low light intensities (Phillips et al., 2019; Prayag et al., 2019). Moreover, there is considerable interindividual variability in the size of this effect (Phillips et al., 2019), suggesting some individuals are more sensitive to the adverse effects of evening artificial light exposure. There is also some evidence of changes in non-image-forming effects with, e.g., age (Daneault et al., 2016), mood disorders (Berman et al., 2018; Hallam et al., 2009), SSRI usage (McGlashan et al., 2018), and DSPD (Aoki et al., 2001; Watson et al., 2018). These findings suggest that evening artificial light exposure may be harmful even at low intensities, especially for particularly vulnerable groups. A solution could be to exchange the modern indoor lifestyle with limited daylight and artificially lit evening environments and instead reside outdoors in nature with greater exposure to daylight and minimal exposure to light after sunset. Indeed, several studies have found that residing outdoors leads to reduced melatonin suppression and advanced circadian phase (Stothard et al., 2017; Wright et al., 2013). Although this may be beneficial for healthy circadian rhythms, evening light and screen use has become so intertwined with modern lifestyle that such transitions are not feasible for most.

1.1.3 *Health consequences of aberrant light exposure - sleep and other health issues*

Artificial evening lighting has dramatically increased the possibilities for work and leisure in the evening and nighttime, and presently it seems unthinkable that this use of evening light will change. Accompanied by less exposure to daylight, this has likely led to a weakening of the light-dark signal and thus also a widening of the chronotype spectrum leading to a higher proportion of late chronotypes (Roenneberg et al., 2019). There may also be behavioral components related to light, as the engaging nature of light-emitting screens may lead people

to delay bedtimes and not prioritize sleep (Hale & Guan, 2015), leading to shorter sleep and the buildup of homeostatic sleep pressure across week-days and subsequent attempts to "catch-up" on sleep debt over the weekends. This may be especially problematic for adolescents and young adults, that naturally have a later circadian phase but still have to get up early for school (Touitou et al., 2016). A 2015 review found that screen time was adversely related to sleep outcomes, primarily shorter sleep duration and later sleep timing (Hale & Guan, 2015). About one-third of the population report more than a 2-hour difference in sleep timing between weekdays and weekends (Roenneberg et al., 2012), indicating that they spend most of their time with some misalignment of their endogenous pacemaker and the timing of the sleep-wake period. Such circadian misalignment caused by later sleep timing on some days of the week is known as "social jetlag" (Wittmann et al., 2006). However, circadian misalignment can arise for a variety of reasons, e.g., shift work, trans-meridian travel, disrupted feeding rhythms, circadian rhythm disorders, or psychiatric illnesses (Baron & Reid, 2014), and there can potentially be circadian misalignment between various circadian rhythms (e.g., central and peripheral tissue and organ systems, hormone secretion, sleep-wake cycles, and feeding behavior). The inertia of these rhythms may also be different. Circadian misalignment, although with varying magnitude, is common in the general population.

The effects of evening artificial light exposure are challenging for human sleep but also extend to a range of other diseases related to circadian disruption or misalignment. Circadian disruption has been associated with a higher risk for depression, diabetes, metabolic abnormalities, obesity, immune impairment, poor cognitive performance, hypertension, and cancer (Chaput et al., 2022; Foster & Wulff, 2005; Hatori et al., 2017; Roenneberg et al., 2012). Notably, a recent 3-day simulated night-shift study found that although the traditional markers of the central circadian pacemaker, such as melatonin, cortisol, and PER3-expression, did not change, many other plasma metabolites showed nearly complete reversals (Skene et al., 2018), showing that circadian disruption can occur quickly and without complete circadian disruption, possibly explaining part of the elevated risk that night-shift workers have for metabolic disease. Furthermore, short sleep, defined as 4-7 hours, has been found to increase the risk of coronary heart disease, stroke, type 2 diabetes, obesity or weight gain, depression, workplace accidents, and mortality (Chaput et al., 2022; Kecklund & Axelsson, 2016).

1.1.4 Light, sleep disorders, and mental disorders

As we have seen, human sleep is sensitive to the misalignment of circadian rhythms and sleep timing. Increasing the use of artificial light in the evening and night constitutes a challenge for healthy sleep by potentially delaying DLMO, suppressing melatonin secretion, and increasing alertness. This may create problems falling asleep and waking up in the morning, which may be relevant even to healthy sleepers (Münch et al., 2020). However, it may pose a significant problem in circadian rhythms disorders such as delayed sleep phase disorder (DSPD) and advanced sleep phase disorder (ASPD). In these disorders, considerable misalignment between the individual's circadian phase and societal normative sleep timing often results in irregular sleep-wake patterns with intermittent short and long sleep durations (Baron & Reid, 2014). Recent evidence suggests that patients with DSPD have a higher sensitivity to the phase-shifting and melatonin-suppressing effects of light (Aoki et al., 2001; Watson et al., 2018). As such, DSPD may represent a group of individuals that biologically may struggle to maintain stable circadian rhythms in modern evening light environments.

The implication of sleep and circadian disruption in the pathophysiology of mental disorders is supported by extensive literature documenting the alterations of sleep physiology and circadian rhythmicity during mental illness (LeGates et al., 2014). The prevalence of sleep problems is elevated to such an extent in patients with depression that disordered sleep is included as one of the diagnostic criteria for mood disorders in both the ICD-10 (World Health Organization, 1992) and the DSM-5 (American Psychiatric Association, 2013). Previously, sleep disturbances were seen as secondary to the primary mental disorder. However, this view has shifted over the last decades, and a bidirectional relationship between sleep and mental health is now recognized. Sleep disturbance, usually disrupted sleep continuity, can occur before depression onset, making sleep disorders a potential prodromal symptom of depression (Perlis et al., 1997; Scott et al., 2021). In a similar vein, for patients with bipolar disorder, sleep time tends to shorten dramatically preceding the onset of a manic episode (Wehr et al., 1982). Patients with schizophrenia have a high prevalence of sleep onset and sleep maintenance insomnia, likely caused in part by the prevalent circadian disruption (Monti et al., 2013; Wulff et al., 2010). Moreover, the sleep architecture also tends to change with mental illness, with a shortening of REM sleep latency and increase in REM density and duration being reported for depression (Baglioni et al., 2016; Giles & Roffwarg, 1998; Lechinger et al., 2021; Palagini et al., 2013; Riemann et al., 2001, 2012, 2020; Steiger & Pawlowski, 2019; Wilson & Argyropoulos, 2005). This association may be particularly

interesting in the context of light, as REM-sleep propensity has been found to exhibit a circadian rhythmicity (Czeisler et al., 1980).

Associations between circadian disruption markers and mental disorders may partly explain sleep disruptions. Early studies indicated a phase-advancement of endogenous circadian rhythms in patients with depression (Wehr et al., 1979). However, more recent studies have repeatedly associated late chronotype with a greater risk for depression, and individuals suffering from depression report greater eveningness, which is also predictive of the severity of depression (Baron & Reid, 2014). In overweight adolescents, shorter duration of melatonin secretion and later circadian preference has been associated with worse mood (Simon et al., 2020). Patients with DSPD have also been found to have a high prevalence of clinically significant depressive symptoms, and later chronotype within DSPD is also associated with greater severity of depression symptoms (Abe et al., 2011). There have been some indications that rather than a specific phase-shift occurring in depression, a shorter phase-angle – the time between DLMO and sleep onset – is associated with more depressive symptoms among patients with major depression (Emens et al., 2009). In another study, the phase-angle between sleep-midpoint and core body temperature minimum predicted the severity of depression (Hasler et al., 2010). The depression group also had later DLMOs, later sleep onset, and greater eveningness chronotypes compared with healthy controls. In bipolar disorders, early studies found an association between the increased duration of daylight in the springtime in the northern hemisphere and a higher incidence of mania (P. A. Carney et al., 1988; Eagles, 1994), suggesting an increased sensitivity to light in this population.

1.1.5 *Blue-depleted light technology*
Circadian rhythms exist across species, and the role of light environments in regulating the circadian rhythms of physiological processes has long been recognized and utilized in the commercial growth of plants, poultry, and livestock, where photoperiods and spectral compositions are optimized to match the needs of the organism. Now there is increasing recognition that there may be a potential to optimize the pattern of light exposure to better complement the physiological and psychological needs of the human body. Specifically, there is a need to understand how artificial light can be used to mitigate circadian misalignment in humans (Münch et al., 2020).

The discovery of the adverse effects of evening light exposure has led to increased efforts to design and test light sources that can mitigate the problem. As the non-image forming effects are mediated mainly by the ipRGCs that have a peak spectral sensitivity around 480nm, the selective filtering of short-wavelength blue light in this range could be a way to create light sources that allow for vision while minimally stimulating the ipRGCs. In patients with DSPD, filtering short-wavelength blue light using blue-blocking glasses can phase-advance DLMO (Esaki et al., 2016). In healthy participants, the evening use of blue-blocking glasses has been found to reduce melatonin suppression and counteract the alerting effects of evening light exposure (Sasseville et al., 2006; van der Lely et al., 2015). Although blue-blocking glasses are physiologically effective countermeasures, their potential for widespread use is limited due to issues with adherence in some patient groups, and many perceive their use as inconvenient and less than stylish.

Light Emitting Diode (LED) technology, however, allows for programmable and tunable light spectra. A single multi-channel LED luminaire can contain multiple diodes. By coating diodes with layers transmitting different light spectra, these can be combined and adjusted within one luminaire to create virtually any light spectrum (Soler & Voss, 2021). Compared with a standard light source, evening light exposure to a blue-depleted LED light source has been found to reduce melatonin suppression and alertness as indicated by slower reaction times and increased power in the slow-to-medium wake-EEG range (delta-theta) (Rahman et al., 2017). Similarly, a study by Souman and colleagues with similar blue-depleted LED lighting found reduced melatonin suppression to the point that there were no differences in melatonin suppression between blue-depleted and dim light (Souman, et al., 2018). Other findings from home settings complement these results by showing that exposure to less light in the ipRGCs action spectra is associated with less melatonin suppression (Nowozin et al., 2017). These findings suggest that blue-depleted lighting may be a key in creating evening light environments that are more beneficial for human circadian rhythms and sleep.

1.1.6   *Application in hospital settings*

Evening blue-depleted light environments have potential in many contexts, especially where it is feasible to exert high control over ambient and incident light. A particularly promising application may be hospital units. In the patient population admitted to hospital inpatient units, sleep problems are highly prevalent (Cain & Phillips, 2021; Danielson et al., 2018; Doğan et al., 2005; Elliott et al., 2013; Schennach et al., 2019). There are several reasons for

this. First, patients may have difficulties sleeping as part of their acute illness, which is often the case for severe mental disorders (Wulff et al., 2010). Furthermore, light levels may often exceed those in residential home environments, with flashing lights from monitors making sleep arduous. Although closed eyelids reduce light intensity, they are still permeable to some influx of light. Artificial lighting turned on during sleep has been associated with less SWS (J. R. Cho et al., 2013). Circadian rhythms are also often disrupted by acute illnesses, a phenomenon associated with subsequent increases in morbidity and mortality (Craig & Mathieu, 2018).

In mental health care, there is a need for non-pharmacological interventions that can be used to target sleep-wake disruption and arousal. Early case series with bipolar disorder found it possible to stabilize mood by following regular bedtimes and wake times accompanied by an extended period of darkness in the evening (Wehr et al., 1998; Wirz-Justice et al., 1999). This was later tested in an inpatient setting, finding quicker reductions in symptoms of mania with acutely admitted patients (Barbini et al., 2005). In an attempt to get similar effects without extended darkness, Henriksen and colleagues compared blue-blocking glasses (glasses filtering all short-wavelength light) to clear lens glasses for inpatients acutely admitted with mania. They found that simulating extended "darkness" with blue-blocking glasses in the evening reduced manic symptoms within one week (Gottlieb et al., 2019; Henriksen et al., 2016). This line of research shows the potential benefits of utilizing such "chronotherapeutic" (from Greek "Khrónos" meaning "time") interventions in an acute psychiatric hospital setting. Potential benefits for other patient groups are currently unknown.

The effective dissemination of this chronotherapeutic intervention in acute psychiatry is challenging, as it requires acutely ill individuals to monitor their treatment protocols and the use of glasses. This may lead to low or intermittent adherence, which may impede treatment effect and even increase patient frustration. The blue-depleted LED lighting technology could potentially solve this issue, leaving the correct timing of light exposure to programmable light systems, thus automatizing the chronotherapy intervention (Gottlieb et al., 2019; Scott, Langsrud, et al., 2019). The implementation of such lighting systems has moved rapidly from the laboratory setting to hospitals and nursing homes. There are indications that cycled lighting (featuring both lower levels of short-frequency light in the evening and higher daytime light levels) is beneficial for circadian rhythmicity (West et al., 2019) and sleep duration (Giménez et al., 2017) in hospitals and that it improves subjective sleep and

depression symptoms in patients with Alzheimer's disease (Figueiro et al., 2019). Other studies indicate that cycled lighting does not affect critical care inpatients (Engwall et al., 2017) or inpatients with mood disorders (Okkels et al., 2019).

Although these trials find some indications of effectiveness with regard to cycled lighting, there has been a lack of control over incident light and cycled lighting has primarily been limited to patient rooms. Recent laboratory studies have shown that even low levels of melanopic EDI are sufficient to trigger melatonin suppression in the evening (Prayag et al., 2019). Further, it is difficult to differentiate between the effects of evening blue-depleted lighting and increased daytime light exposure by introducing cycled lighting. Moreover, these studies have been performed in clinical settings, where the invasiveness of polysomnography (PSG) and melatonin sampling limits assessments of physiological effects. Therefore, there is a knowledge gap regarding the physiological effects of evening Blue-Depleted Light Environments (BDLEs) on circadian rhythms and sleep outcomes and whether such physiological effects translate when BDLEs are incorporated into large, multi-room complexes.

1.2 Assessment of sleep using novel technology

1.2.1 *The challenge of sleep assessment*

Scientific knowledge about sleep depends on measuring sleep using various technologies. Both self-reported data and objective recordings can be used to capture different aspects of sleep. PSG is often called the gold standard of sleep assessment (Kales et al., 1968). However, before the development of the PSG assessment, sleep science was limited by the lack of suitable methods to observe the phenomenon of sleep. During the 1950s, Aserinsky, Kleitman, Dement, and Jouvet did pioneering work developing procedures for the use of EEG electrodes along with Electrooculography (EOG) and Electromyography (EMG) to distinguish sleep from wake and describing different sleep stages (Aserinsky & Kleitman, 1953; Dement & Kleitman, 1957; Jouvet et al., 1959). Rechtschaffen and Kales (Kales et al., 1968) did further work that standardized the sleep assessment procedure known as PSG, and sleep science picked up momentum now being able to assess the internal composition or architecture of sleep.

As PSG combines EEG, EOG, EMG, and other sensors, the affixing of PSG equipment is time-consuming and requires experienced sleep technicians. Moreover, many patients experience the equipment as cumbersome and obtrusive when trying to sleep, often giving rise to a "first-night" effect with less sleep and poorer quality sleep (Newell, 2012). After PSG has been completed, qualified personnel must subsequently score the recordings, which is time-consuming and resource-draining. The PSG assessment is essential in some contexts, e.g., with patients being assessed for certain sleep disorders or in sleep research investigating sleep at a microscopic level. However, the level of detail that the PSG offers is not needed in all situations, and its use is too expensive for extended periods in any case (De Zambotti et al., 2019). Thus, there is a need for less intrusive and time-consuming tools to be used for sleep assessment. Luckily, if accepting a lower level of detail, the arsenal of sleep-tracking devices is expanding rapidly with tech-industry developments.

Consumer sleep technologies (CSTs) are devices or programs aimed at improving or assessing the sleep of consumers (Ko et al., 2015). They include mobile apps (often using device functions like the camera or microphone), wearable devices, embedded devices (e.g., embedded in mattress or nightstand), accessory appliances, and conventional desktop/website resources. CST use is widespread, and the number of such devices available to the consumer market is increasing rapidly. Sleep-tracking devices can be considered a subgroup of CSTs, and include a number of wearables traditionally aimed at detecting physical movement. Newer devices also claim to use heart rate sensors and skin conductance to improve sleep-tracking capabilities (De Zambotti et al., 2019). The performance of CST sleep-trackers has been poor when independent researchers have tested their claims (Scott, Grierson, et al., 2019), although it may seem that some of the newer devices are improving in their ability to detect sleep (Chinoy et al., 2021). An issue with CST sleep trackers is that the sleep-detecting algorithms are proprietary and constantly evolving, making independent validation of the technology difficult. Moreover, the hardware is also evolving quickly, leading to a short life-cycle of such devices before new products are available. The lack of validation is worrying for consumer use but is particularly problematic for use in healthcare sleep assessment or sleep research (Haghayegh et al., 2019).

In sleep research, motion-based sleep detection has been the best low-cost, less invasive alternative to PSG since the first actigraphy procedures were validated in the 70s and 80s (Kripke et al., 1978; Mullaney et al., 1980; Webster et al., 1982). Actigraphy uses an on-body

accelerometer, commonly embedded in a wrist-worn actiwatch, to measure physical movement. Using relatively simple automatic classification models, activity-rest, and hence sleep-wake state is determined (Haghayegh et al., 2019; Van Someren, 2011). Importantly it is cost-effective, easily set up, demands less from the individual, and allows recording continuously for extended periods meaning one can assess the sleep-wake cycle across weeks at a time (Van Someren, 2011). Actigraphy data has typically been used to quantify sleep, but monitoring rest-activity cycles for extended periods can provide additional information about an individual's circadian rhythms (Cheng et al., 2021; Huang et al., 2021). Rest-activity cycles follow similar cosinor waveforms as endogenous circadian rhythms (Van Someren, 2011), reflecting their effect on our drive for activity and sleep.

When compared with PSG, actigraphy has shown a good ability to detect sleep (sensitivity) and high accuracy (ability to distinguish sleep from wake) in healthy subjects (Van Someren, 2011). However, it must be kept in mind that accuracy will tend to be higher in healthy participants as they have fewer wake epochs during the night, making it likely that a scoring method sensitive to sleep will yield a correct classification for most epochs. Actigraphy does tend to have a poorer ability to detect true wake epoch (specificity ranging from 26.9% to 77% in healthy subjects (De Zambotti et al., 2019), likely stemming from the fact that wake detection based solely on the physical movement of the wrist is limited (in the case of actiwatches). This is especially true when actigraphy is used in populations characterized by sleep difficulties (De Zambotti et al., 2019), such as patients with sleep disorders or mental disorders. Moreover, there are some settings where even the use of actigraphy may be considered too invasive and hence not feasible.

1.2.2   *Nightly monitoring of sleep and vital signs in hospital wards*

A thought-provoking editorial written by Cain and Phillips (Cain & Phillips, 2021) points to the puzzling fact that when admitted to hospitals, the body is in higher need of sleep to assist in recovery from illness, while the typical hospital environment tends to hinder sleep. As discussed in the section on light environments, aberrant lighting is an important factor. However, noise and being awoken by nurses carrying out observations are other obvious barriers (Danielson et al., 2018; Darbyshire & Young, 2013; Konkani & Oakley, 2012; MacKenzie & Galbrun, 2007). Nurses recognize the inherent conflict between the need for sleep and the need to monitor vital signs and sleep (Hope et al., 2018). More frequent

observation is associated with fever deaths and Intensive Care Unit (ICU)-admissions (Mitchell et al., 2010), but is at the expense of patients' sleep and potentially their healing. Therefore, there is a need to find ways to assess the patient's state without intermittently waking the patient.

This conflict between sleep and monitoring is particularly prominent in acute psychiatry. In the acute phase of severe mental illness, sleep problems are ubiquitous, and improving sleep is considered an essential part of the treatment (Strainge et al., 2019; Waters et al., 2012; Wulff et al., 2010). Consequently, monitoring sleep in this context is a way to assess improvement or deterioration in the patient's condition and thus may provide information relevant to planning further treatment. Two obstacles make actigraphy less suited for this application. First, wearable sensors attached to the body tend to be less accepted by patients in this population. Second, actigraphy assessment of sleep requires the actiwatch to be collected, and the data have to be downloaded and analyzed, making it both burdensome to staff and an offline or retrospective source of information. Instead, sleep monitoring usually consists of nurses physically entering patient rooms to assess whether patients are sleeping, thereby often interrupting sleep. This practice, which also serves to check patient safety, has even been suggested to be more harmful than beneficial for patients (Veale, 2019). In a hectic work environment, tailoring the frequency of monitoring to the individual patient's needs is difficult. Even if the frequency of observation was proper at admission, patients might soon be over-monitored if this is not continuously reviewed as they improve.

Another consequence of the difficulties of assessing sleep in acute psychiatry is that the literature surrounding inpatient sleep problems and their effect on outcomes is relatively scarce. This constitutes an ethical dilemma, given that vulnerable patient groups such as mental health-care inpatients are being understudied. A recent study using self-report questionnaires found that around 47% of inpatients in acute psychiatry still had sleep problems at discharge, but also noted the lack of comparable studies (Schennach et al., 2019). Another study note that patients typically are started on hypnotic or antipsychotic medication during inpatient stays, but that this does not seem to be enough to resolve the sleep problems (Waters et al., 2012). In outpatient mental health care, however, the literature regarding the effects of sleep on psychiatric and physical symptoms is vast. As the acute psychiatry setting differs in so many respects from outpatient care, it is problematic to apply the knowledge

gained from outpatient studies to this population (Strainge et al., 2019), still, conventional sleep assessment is not feasible with the most severely ill patient groups.

1.2.3   *Radar-based contact-free sleep monitoring*

A new application of radar sensors can potentially offer a new way to assess sleep in these contexts. Recent developments in radar technology have allowed for small and highly sensitive radar-sensor chips suited for embedding in the ceiling of patient rooms or used as a nightstand. The particular sensor of interest in the current thesis is an Impulse-Radio Ultra-Wideband (IR-UWB) radar. The sensor can detect movement from an individual, whether located in the bed or not, and the signal is not obstructed by bedding or clothing (Stone, 1997). The radar can detect movements of the limbs and, under favorable conditions, the spatial resolution is sufficient to detect respiration rate based on the movement of the chest when breathing, and the superimposed movements of the chest during heartbeats can under favorable conditions be used to assess heart rate (Lee et al., 2018; Wisland et al., 2016). This technology holds great potential as a contact-free, nearable (non-wearable) method of assessing sleep. Recordings can be made across multiple nights with no effort from staff. The radar recording has the potential of being scored in real-time, with results being fed with minimal delay to nurses or staff, informing whether patients are in bed or not and asleep or awake. If the accuracy of such a radar system can be validated for the detection of sleep-wake state, this offers significant advantages over actigraphy in the acute psychiatric hospital setting, addressing the specific issues outlined above. As the radar signal is much richer in data than wrist-based actigraphy, this technology offers more opportunities in the future as scoring algorithms improve, such as detecting certain vitals like heart rate and respiration rate. Moreover, real-time information on patient physical restlessness may assist in the evaluation of immediate patient safety.

1.3   Using the internet in the treatment of insomnia

1.3.1   *Insomnia symptoms and insomnia disorder diagnosis*
Insomnia is defined as sleep-continuity disturbance combined with complaints of daytime sleepiness, fatigue, cognitive impairment, mood disturbance, and dissatisfaction with or worry about sleep (Perlis et al., 2022). The sleep-disturbance refers to either disturbed sleep onset, increased number of awakenings, increased time awake after sleep onset, or early morning awakenings. Importantly, the sleep disturbance is present despite adequate sleep opportunity. The diagnosis of insomnia disorder is recognized in both the ICD-10 (World Health

Organization, 1992), the DSM-V (American Psychiatric Association, 2013), and the International Classification of Sleep Disorders-3 (American Academy of Sleep Medicine, 2014), however, the magnitude of sleep-disturbance is not specified. The diagnosis of insomnia is thus primarily based on subjective symptoms, both of sleep disturbance, level of distress, and daytime impairment, rather than objective measures of short sleep. Although brief symptoms of insomnia are widely prevalent in the general population, the symptoms should be present at least three days a week and persist for at least three months (American Psychiatric Association, 2013).

1.3.2 *Prevalence and consequences of insomnia*

At any given moment, between 10% and 15% of the population report symptoms that are in congruence with the criteria for insomnia disorder, with epidemiological studies from Norway pointing at an increasing trend over the last decade (Pallesen et al., 2001, 2014). Insomnia has been associated with an increased risk of a range of physical illnesses, such as cardiovascular disease, obesity, type 2 diabetes, neurodegenerative disease, and cognitive impairment (Riemann et al., 2017). Epidemiological studies using population data have also found that insomnia symptoms are a risk factor for fatal car-accidents (Laugsand et al., 2014), sick leave (Sivertsen, Øverland, Bjorvatn, et al., 2009), and work disability (Sivertsen, Øverland, Pallesen, et al., 2009). The effects of insomnia are especially prominent in mental disorders where there seems to be a bidirectional relationship between insomnia and mental illness, with the strongest indications for a causal mechanism of sleep problems on the development of mental disorders (D. Freeman et al., 2020). Patients with insomnia have an increased risk of developing depression (Baglioni et al., 2011), which may also be related to an early retirement due to disability (Paunio et al., 2015). On the other hand, treatment of insomnia may improve symptoms of comorbid mental disorders (D. Freeman et al., 2017; Manber et al., 2008).

Due to the high prevalence of insomnia symptoms and the associated daytime functional impairment, the associated economic disease burden is high. Attempts at quantifying the societal costs of insomnia suggest that insomnia ranks among the top ten most costly neuropsychiatric disorders (Gustavsson, 2010). The direct costs of treating insomnia are high, especially with psychotherapy that requires multiple sessions. However, the indirect costs of

reduced work productivity, sick leave, accidents, and early retirement or disability pension are likely also high, contributing to the economic disease burden (Riemann et al., 2017).

1.3.3 Theoretical models of insomnia

Several theoretical models describe the etiology and mechanisms of insomnia. Most of these build on Spielman's 3P-model of insomnia that distinguishes between predisposing, precipitating, and perpetuating factors (Spielman et al., 1987). Examples of predisposing factors are genetic risk (Palagini et al., 2014) and certain personality traits like neuroticism. Precipitating factors refer to events or stressors that trigger the initial insomnia symptoms. Examples may be major transitions in life, such as starting college, moving to a new city, or starting a new job, or interpersonal stress, such as divorce or conflict with a loved one. Chronic stress is also considered a precipitating factor of insomnia. Acute insomnia often resolves on its own once the stressor resolves (Ellis, 2012; Espie, 2001). The perpetuating factors usually manifest as insomnia develops into a chronic disorder. These include going to bed too early, staying in bed too long, napping to catch up on sleep, and an increased cognitive preoccupation with sleep (Espie et al., 2006; Spielman et al., 1987). These behaviors and attitudes are reasonable in the short term, but over time they reduce sleep pressure, increase worry and arousal, and contribute to the maintenance of insomnia.

1.3.3.1 The hyperarousal model of insomnia

Over the last decade, several findings have suggested that hyperarousal is common among patients with insomnia (Riemann et al., 2010). In terms of the 3P conceptualization of insomnia, chronically elevated arousal levels in both the cognitive, emotional, and physiological domains can be viewed as both predisposing and perpetuating factors of insomnia (Riemann et al., 2010). Evidence from EEG assessments of patients with insomnia lends support to this model, where increased power in the high spectral frequencies during non-REM sleep stages has been found in this group. Increased frequency of microarousals and interruptions during REM-sleep also suggests higher arousal and may contribute to the misperception of REM-sleep as waking (Feige et al., 2008; Riemann et al., 2012). In the hyperarousal model of insomnia, the instability of REM-sleep has been postulated as a mechanism by which overnight emotional dissolution of emotional distress is hampered, thereby perpetuating the hyperarousal and insomnia symptoms (Riemann et al., 2012; Van Someren, 2021). Some interesting recent findings indicate that reduced REM-sleep

consolidation impedes overnight amygdala adaptation (Wassing et al., 2019), and that slow dissolution of emotional distress contributes to hyperarousal (Wassing et al., 2016). In terms of the neurobiological regulation of sleep-wake, hyperarousal may result from increased activity in arousal-promoting regions relative to sleep-promoting systems (Saper et al., 2001, 2005).

1.3.4   Treatment of insomnia

Treatment of insomnia can be divided into pharmacological and non-pharmacological treatment options. Although many medications are used in the treatment of insomnia (antidepressants, antihistamines, low-dose antipsychotics, melatonin, etc.), pharmacological treatment most often consist of benzodiazepines (BZ) or benzodiazepine receptor agonists (BZRAs). In brief, these interventions are effective in the short-term (less than four weeks) (Riemann et al., 2017), with one Randomized Controlled Trial (RCT) reporting a 76,7 % response rate and a 47,7% remission rate (Pillai et al., 2017). Regarding the long-term effects of treatment with hypnotics, studies are more disparate with some indications of effect, but this may gradually decline (Riemann et al., 2017). Moreover, insomnia may rebound after withdrawal from pharmacological treatment. Due to the potential for drug tolerance and dependency, these medications are not recommended for long-term use (Riemann et al., 2017). Interestingly, a recent meta-analysis of 32 studies showed that around 60% percent of the effect of BZs and BZRAs could be attributed to placebo (Winkler & Rief, 2015), suggesting that cognitive factors may play a role in the efficacy of pharmacological treatment as well.

As a non-pharmacological intervention alternative, Cognitive Behavioral Therapy for Insomnia (CBT-I) is a multi-component therapy that typically consists of four components, including psychoeducation/sleep hygiene, stimulus control therapy, sleep restriction therapy, and cognitive therapy, with differences between manuals in how these components are implemented (Perlis et al., 2022). Across treatment studies, CBT-I is effective on the most common outcome measures of sleep such as sleep onset latency (SOL), total sleep time (TST), sleep efficiency (SE), and wake after sleep onset (WASO) (Trauer et al., 2015; Wu et al., 2015). There is a consensus that CBT-I is the treatment of choice for insomnia. Guidelines for the treatment of insomnia from both the U.S. (Brasure et al., 2016; Kathol & Arnedt, 2016; Qaseem et al., 2016) and in Europe (Riemann et al., 2017) conclude that cognitive

behavioral therapy for insomnia is the most effective treatment of insomnia, and should be the first line treatment for individuals seeking help for insomnia.

1.3.5   The challenge of disseminating treatment for insomnia disorder

The standard way of delivering CBT-I is through individual face-to-face consultation with a clinical therapist. However, very few clinicians receive appropriate education (Baglioni et al., 2020; Meaklim et al., 2020; Meltzer et al., 2009). Although there are attempts at making therapist training more available across the internet (Wilkerson et al., 2022), the number of therapists proficient in the use of CBT-I is still far too low compared with the demand for treatment (Thomas et al., 2016). The European Sleep Research Society has responded to this issue by creating a task force to find ways of disseminating therapist education (Baglioni et al., 2020). Nevertheless, the current situation is that many patients who seek help for insomnia will, instead of receiving the effective treatment of choice, most often receive other treatment options such as pharmacological treatment or sleep hygiene advice. A recent longitudinal study of adults with insomnia found that after five years, 40% of patients had not experienced remission (Morin et al., 2020), indicating few patients receive the most effective treatment. Although effective treatment exists for insomnia, the dissemination of treatment has turned out to be a substantial problem.

1.3.6   The internet to the rescue (for once)

One way to disseminate treatment more effectively is by utilizing digital versions of CBT-I (dCBT-I) through apps or websites. Many such adaptations have been made, with some intended to be used with supportive guidance from a clinician and some being completely automated and self-guided (Luik et al., 2017). Although dCBT-I versions vary in levels of digitalization and automation (Luik et al., 2019), they have overall been found effective in reducing insomnia severity in a meta-analysis including results from 11 randomized controlled trials (RCTs) (Zachariae et al., 2016). We have also previously found that a fully automated, self-guided version of dCBT-I reduced insomnia severity in a large community sample with a large effect size (Vedaa et al., 2020). The evidence regarding the effectiveness of dCBT-I has led some researchers to voice the opinion that dCBT-I should be made accessible to the community (Luik et al., 2019).

Most of the studies have tested dCBT-I against a placebo or a control group. As such, there is a knowledge gap concerning the effectiveness of dCBT-I when compared directly with

standard face-to-face (FtF) CBT-I. Although the evidence supporting the effectiveness of dCBT-I is strong, only three studies have compared it directly with FtF CBT-I with conflicting findings (van Straten & Lancee, 2020). One RCT compared therapist-guided dCBT-I with individual FtF CBT-I and found that both were effective in reducing insomnia severity but that FtF therapy was superior with a between-group Cohen's d effect size of 0.9 (Lancee et al., 2016). Second, a non-inferiority trial of therapist-guided dCBT-I against group CBT-I found that therapist-guided dCBT-I was non-inferior to the FtF group format (Blom et al., 2015). Finally, an RCT with military personnel compared self-guided (automatized) dCBT-I with FtF CBT-I and found both effective in reducing insomnia severity, with a trend favoring FtF CBT-I (Taylor et al., 2017).

It should be noted that the development of dCBT-I was not necessarily meant as a replacement for FtF CBT-I in the context of patients being referred for clinical treatment of insomnia. Moreover, studies of dCBT-I have focused primarily on convenience samples recruited via the media or the internet. Nevertheless, given the lack of providers of standard FtF CBT-I, plus the potential cost-effectiveness of digital versions – especially automated variants, there is a demand for the implementation of dCBT-I in the healthcare system (Luik et al., 2019). In light of this, there is a need for studies from the healthcare setting that compare dCBT-I with standard FtF CBT-I to investigate if fully automated dCBT-I is non-inferior to FtF CBT-I in these settings.

1.4   Aims of the current thesis

The rapid technological advances over the last decades have led to a discrepancy between the existing technological innovations aimed to improve sleep and knowledge of their effects in naturalistic settings. The overall aim of this thesis was to test three such technologies and evaluate their effects in real-world settings where they may hold high potential benefits if successful. The specific technologies investigated are an evening BDLE using tunable LED lighting technology installed in a hospital ward, contact-free sleep assessment using a novel radar sensor, and disseminating insomnia treatment through fully automated digital CBT-I. These efforts are divided into three studies, the results of which are reported in four papers.

The evening BDLE and the radar sensors are installed in a new-built acute psychiatric hospital unit at St Olavs University Hospital, The Hospital in Trondheim, Norway. The

implementation results from years of planning and cooperation between the Trondheim Sleep and Chronobiology Research group (SACR) and St.Olavs University Hospital Division of Mental Health Care. The papers on evening BDLE and radar assessment included in the current thesis represent efforts to validate and test the effects of these technologies with healthy participants as an initial step in the implementation into clinical practice. The paper on digital CBT-I was an effort to directly compare the effectiveness of dCBT-I to FtF CBT-I in the naturalistic setting of an outpatient sleep clinic at St. Olavs University Hospital, Division of Mental Health Care.

The specific aims of the four papers were:

1.4.1   Paper I

The primary aim (i) of the study was to test whether residing in an evening BDLE for five days influenced the timing of DLMO, melatonin suppression, and polysomnographic sleep variables compared with the effects of exposure to Standard Light Environment (SLE) in healthy participants. Secondary aims (ii) were to test whether there were effects on neurocognitive arousal and subjective sleepiness. Third (iii), we aimed to assess whether residing in the evening BDLE had detrimental effects on color perception and subjectively reported side effects.

1.4.2   Paper II

This paper builds on the findings from paper I and current evidence on the potential importance of the microstructure of REM sleep for affect regulation, mood, and sleep disorders. We first aimed (i) to test whether there were differential effects of residing in an evening BDLE compared with an SLE in terms of REM sleep fragmentation, REM sleep arousals, and REM density. Second (ii), we aimed to explore mechanisms of change by testing whether circadian phase shifts explain these changes in REM sleep parameters. Third (iii), we aimed to explore the temporal dynamics of the REM sleep overnight accumulation and compare this with the accumulation of non-REM stage 3 sleep (N3).

1.4.3   Paper III

The purpose of this paper was to test if the radar sensor could be used to detect body movements similarly to wrist-based actigraphy to distinguish sleep from wake. We first aimed

(i) to develop real-time sleep/wake classification models that can be used to analyze radar data. We then aimed (ii) to compare the results of these models with PSG and actigraphy and calculate estimates of sensitivity, accuracy, specificity, and Cohen's kappa. Third, we aimed (iii) to use Bland-Altman analyses to evaluate the agreement across radar, PSG, and actigraphy on the most common sleep outcome variables TST, SOL, WASO, SE, and number of awakenings (NW). Finally, we aimed (iv) to compare the performance of the different assessment modalities in two groups: one with healthy sleepers and one clinical sample with patients undergoing sleep assessment at a sleep clinic at St. Olavs University Hospital, Department of Neurology and Clinical Neurophysiology.

1.4.4 Paper IV

In this fourth paper, our primary aim (i) was to test if dCBT-I is non-inferior to FtF CBT-I on insomnia severity. The primary endpoint was set to be the six-month follow-up assessment, as longer-term outcomes were considered more indicative of effectiveness for both patients and the healthcare system. Secondary aims were to conduct between-group superiority analyses to (ii) test for potential differences in response or remission rates and (iii) test for differences in psychological distress, fatigue, and self-reported sleep-wake patterns.

2 Methods

2.1 Overview

The work contained in this thesis was performed in the context of St. Olav's University Hospital, the primary hospital for 325 000 inhabitants in the Trondheim area, while also being the university hospital for the region – about 730 000 inhabitants. The thesis comprises four papers where technologies designed to improve or assess sleep have been tested. The thesis can broadly be divided into two parts.

Part one, and the majority of the thesis, relates to a research project where an evening BDLE and radar technology has been installed in a new-build acute psychiatric hospital unit at the department of mental health. This has resulted in three of the papers included in the thesis. Papers I and II report the effects of a randomized cross-over trial with healthy participants residing in this hospital unit with the evening BDLE on circadian rhythms and sleep outcomes in healthy participants. From here on, this trial is referred to as *the BDLE trial*. Paper III focuses on the potential of using radar as a contact-free measure of sleep and includes the participants from *the BDLE trial* and additionally participants from a sleep clinic for organic sleep disorders at the department of Neurology and Clinical Neurophysiology.

Part two of the thesis stems from a clinical trial that compared the effects of digital versus FtF CBT-I for patients referred to a sleep clinic for insomnia and circadian rhythm disorders at the Division of Mental Health Care, St.Olav's University Hospital. From here on, this trial will be referred to as *the dCBT-I trial.* This has resulted in the fourth paper in this thesis, where primary and secondary outcomes are reported.

2.2 Study designs

2.2.1 Papers I and II

The BDLE trial utilized a randomized cross-over design. All 12 participants were exposed to both the experimental evening BDLE and the SLE for 5 days, with a randomized order of exposure and a 1-day wash-out period between conditions. Randomization was performed by the Unit of Applied Clinical Research at the Department of Medicine and Health Sciences, NTNU. Figure 4 shows an overview of the *BDLE trial* and includes study days, assessments, conditions, and the nomenclature used for the different phases of the trial.

2.2.2 Paper III

The design of this study was cross-sectional. The paper is based on two datasets with concurrent assessments of sleep with different sensor modalities. As part of the data collection in *the BDLE trial*, we also collected data from radar sensors to develop and test scoring algorithms. The first dataset, DS1, contains concurrent PSG, radar, and actigraphy recordings collected during *the BDLE trial*. The second dataset, DS2, was collected in the setting of an outpatient neurophysiological sleep clinic at St.Olav's University Hospital in Trondheim. DS2 comprises single-night sleep recordings with concurrent PSG, radar, and actigraphy recordings from 28 adult patients from this sleep clinic undergoing ambulatory sleep assessment.

2.2.3 Paper IV

The dCBT-I trial was a parallel group randomized non-inferiority trial where participants were allocated 1:1 to two modes of delivery of CBT-I, an automatized, self-guided digital version of CBT-I or CBT-I delivered FtF by an experienced clinical practitioner. The randomized non-inferiority design tests whether one of the treatments, in this case the dCBT-I, is not inferior by more than some predetermined meaningful margin than another treatment, usually considered a gold-standard treatment. In this case, this was FtF CBT-I delivered by a therapist. The Unit of Applied Clinical Research at the Department of Medicine and Health Sciences, NTNU, performed the randomization.

2.3 Study settings

2.3.1 Papers I, II, and III – the BDLE trial

In *the BDLE trial,* the sample consisted of healthy young adults without symptoms of psychiatric disorders or sleep disorders.

2.3.2 Paper III DS2 – the neurophysiology sleep clinic data

The sample was recruited from patients referred to the local Department of Neurology and Neurophysiology at St. Olav's University Hospital for overnight PSG sleep assessment.

2.3.3   Paper IV – The dCBT-I trial

The setting of *the dCBT-I trial* was an insomnia and circadian rhythm disorder sleep clinic under the Division of Mental Health Care at St.Olav's University Hospital. Prospective participants were patients referred to the sleep clinic for treatment of insomnia.

2.4   Recruitment and participants

2.4.1   Paper I, II, and III – Recruitment and eligibility criteria in the BDLE trial

Participants were recruited among students from the Norwegian University of Science and Technology, Trondheim, Norway, by way of adverts at the university campus. Participants were first screened in a semi-structured telephone interview before another interview was carried out with these individuals face to face. Participants were evaluated for inclusion if their usual sleep-wake patterns for weekdays consisted of bedtimes within 2230h-0000h and rise times within 0630h-0800h and that they did not have more than 2h intraindividual deviations between weekdays and weekends. Participants were further required to test negative for color blindness using the Ishihara plate test.

Exclusion criteria were comprised of any current medical or psychological disorders, ongoing use of prescription medication, a family history of severe mental illness, having worked night shifts in the last two years prior to the study, trans-meridian travel across more than one time-zone over the previous two months before the study, and/or ongoing use of nonprescription drugs or illicit substances (not including alcohol or nicotine).

2.4.2   Paper III DS2 – recruitment and eligibility for participants in the neurophysiological sleep clinic

Participants were recruited from patients referred for an overnight sleep examination at the Department of Neurology and Clinical Neurophysiology at St.Olavs University Hospital in Trondheim, Norway. This clinic receives referrals for a wide range of sleep disturbances, including sleep apnea conditions, hypersomnia, restless legs syndrome, and periodic limb movement disorder. Patients were consecutively included in the study, with the only inclusion criterion being informed consent. There were no exclusion criteria in this data-collection in order to test the radar assessment in a naturalistic population of patients in need of assessment and/or treatment for their sleep disorder.

2.4.3 Paper IV – recruitment and eligibility for participants in the dCBT-I trial

Prospective participants were patients over the age of 18 that had been referred to the local sleep clinic at the St. Olavs University Hospital in Trondheim, Norway, with insomnia symptoms. A semi-structured diagnostic interview based on the insomnia interview schedule (Morin, 1993) was used to assess whether prospective participants met the diagnostic criteria for insomnia disorder (DSM-5 (American Psychiatric Association, 2013)). The interviews were conducted by a licensed clinical psychologist or a trained psychiatrist with experience treating sleep disorders and insomnia in particular. Sleep diaries from the last 14 days before the interview were further used to confirm the diagnosis. Participants were required to confirm access to the internet and basic computer skills via self-report. Exclusion criteria were one or more of the following: the assessment interview finding evidence of either a circadian rhythms disorder or an organic sleep disorder (for sleep apnea specifically, both the assessment interview and an Oxygen Desaturation Index above 9 on an overnight oximetry recording was used); ongoing alcohol and/or substance use problem; working night shifts and not being able to stop working night shifts for the duration of the RCT; having previously tried CBT-I; having a medical condition where the administration of sleep restriction may potentially worsen the illness (e.g., an attack phase of epilepsy or multiple sclerosis); and/or not being fluent in Norwegian.

2.5 Assessments

2.5.1 Paper I, II, and III - Assessments in the BDLE Trial

For papers I and II, we tested the effects of BDLE on both objective measures and subjective assessments. Objective measures included the collection of saliva for melatonin assays, PSG, Connors Continuous Performance Test 3 (C-CPT-3), and the Farnsworth–Munsell 100 Hue Color Vision Test (FM-100). In addition, we used self-report questionnaires to assess the effects on subjective levels of sleepiness using the Karolinska Sleepiness Scale (KSS) and side effects of residing in the respective light environments using the Side Effect Rating Scale created by the Udvalg for Kliniske Undersøgelser, Scandinavian Society for Psychopharmacology (UKU side-effects scale).

For paper III, the radar assessment was added in addition to PSG and actigraphy to compare radar assessment with the gold standard of sleep assessment (PSG) and the most widely used device for sleep assessment over extended periods (actiwatch).

Study day	Planned Assessments		Environment		Study phase
-7 to -1	Actigraphy		Home environment (12 Participants)		Pre-randomization monitoring
1	Actigraphy, Melatonin		Dim light		Baseline assessment followed by Randomization
2 3 4 5 6	FM100 FM100 C-CPT-3 PSG PSG, Melatonin	All 5 days: Actigraphy, Radar, KSS	Blue-depleted light environment (6 Participants)	Standard light environment (6 Participants)	Period 1
7	Actigraphy, Melatonin, UKU		Dim light	Dim light	Intermission day assessment
8 9 10 11 12	FM100 FM100 C-CPT-3 PSG PSG, Melatonin	All 5 days: Actigraphy, Radar, KSS	Standard light environment	Blue-depleted light environment	Period 2
13	Actigraphy, Melatonin, UKU		Dim light	Dim light	Final assessment

Figure 4. *A detailed overview of the data collection during all phases of the evening BDLE-trial.*

2.5.1.1 Melatonin assessments

Saliva samples used to assess melatonin were collected on days 1,6,7,12, and 13 of the study. Saliva sampling was always performed hourly between 1900h and 2300h. Salivette Cortisol Code blue (Sarstedt AG & Co, Nünbrecht, Germany) sampling kits were used to collect the samples. Samples were centrifuged at 2200g for 10m immediately following sampling, and stored the first night at -20 degrees Celsius. The following day, the samples was moved to a -80 degrees Celsius freezer, and stored until analysis. Enzyme linked immunosorbent assay (Direct Saliva Melatonin, EK-DSM, Bühlmann, Schönenbuch, Switzerland) was used to analyze the samples.

2.5.1.1.1 Dim light melatonin assessments

On days 1, 7, and 13, melatonin was sampled in a dim light environment (<3 lux). The dim light environment was created by blacking out all windows. Participants entered the dim light environment at 1800h, thus spending an hour in dim light before the first saliva sampling. During dim light assessments, participants were sat sedentary, and eating and drinking were scheduled immediately following saliva sampling.

Melatonin assessments in BDLE and SLE

On days 6 and 12, melatonin was sampled in the respective light environments that participants were residing in, thus allowing for a naturalistic assessment of melatonin levels while exposed to differential light environments.

2.5.1.2 Polysomnography

Participants were mounted with the SOMNO HD (SOMNOmedicsGmbH, Randersacker, Germany) PSG equipment. Participants wore equipment for two consecutive nights the last two days in each light condition. EEG electrodes were fitted according to the 10-20 system (Klem et al., 1999). F3, F4, C3, C4, O1, and O2 electrodes were used, in addition to a mastoid reference electrode on the right and left side (M1 and M2). EOG electrodes were applied 1 cm lateral of and 2 cm under the eye cantus. EMG was recorded from the submental and bilateral anterior tibial muscles. All PSG equipment was fitted by trained technicians at the Department of Clinical Neurophysiology at St Olavs University Hospital, Trondheim. The scoring of PSG recordings was done according to the AASM scoring manual version 2.4 (Berry et al., 2017) and was performed by a clinical neurophysiologist with more than 10 years of scoring experience who was also blinded to participant allocation to conditions.

2.5.1.3 Neurocognitive arousal

The C-CPT-3 (Conners, 2014) was used as an objective assessment of participants' neurocognitive arousal. The cognitive test was completed once per light condition – corresponding to study days 4 and 10. The test was completed between 2100h and 2200h. The C-CPT-3 was conducted on a computer (blue-blocking screens (lowbluelights.com) were physically applied to the monitors in the BDLE condition). During the test, letters A–Z are presented in no particular order and with variable speed. The task consists of pressing a button when any letter – except for the letter X (20% of trials) – appears on-screen. Participants are instructed to react as quickly and accurately as they can. The test lasts for 14 minutes and totals 360 trials. An aggravation of response speed, consistency throughout testing, and accuracy may be expressions of reduced neurocognitive arousal. Thus we chose four common test variables that operationalize these outcomes: Reaction time, the standard deviation of the reaction times, omissions (failure to respond to targets), and commissions (response to nontargets) (Homack & Riccio, 2006; Wilhelmsen-Langeland et al., 2013).

2.5.1.4 Actigraphy

An actiwatch (Actiwatch Spectrum, Philips Respironics Inc., Murrysville, PA) was worn for 7 days before participants were randomized and throughout the rest of the study. Actigraphy data were used in several ways. Event-marker presses were used to record bedtimes, but in some cases, this was missing, and the bedtime recorded in the sleep-diary was used instead.

For paper I, sleep-wake data for each 30-s epoch through the 24h day was used to calculate the sleep-regularity index (SRI). The SRI is a percent-wise estimation of the regularity/irregularity of an individual's sleep-wake cycle across multiple days (Phillips et al., 2017). A high number indicates the individual typically sleeps at the same time each 24h day. In addition, actigraphy data were used to automatically estimate rise times using the actiwatch software (Actiware version 5.70.1, Philips Respironics Inc., Murrysville, PA).

For paper III, actigraphy data was used as a comparison to radar and PSG assessment of sleep. The alignment of the recording modalities was done by identifying the time shift with the highest cross-correlation between sequences of movement data. This way, the PSG signals could be synchronized in time with the actigraphy signals during post-processing. The standard actigraphy scoring algorithm provided by the manufacturer was not used. Instead, binned movement data was exported from the actiwatch and subjected to the same sleep-wake classification model development as the radar data. Thus, similar models were created for both data modalities, only with differing parameters. See Paper III Supplementary Table S1 in the appendix to this thesis for the full model specifications.

2.5.1.5 Radar

In *the BDLE trial*, a commercially available radar sensor (XeThru model X4M200, Novelda AS) was built into the ceiling of each hospital patient room that the participants slept in. In addition, another radar was mounted on a tripod at nightstand height by the bedside to validate the permanently mounted ceiling radar. The radar data were stored in baseband I/Q form. This opens for digital signal processing of the same data in new or improved ways in the future. For the current study, data were then processed using the pulse-Doppler method provided by the manufacturer (using the "Respiration_2" profile setting with a detection zone of 0.40-5.00m, and a respiration detector range of 8-30 respirations per minute (RPM)). This allows for the estimation of body movements and respiration rates at a 1 Hz resolution. For more details on the radar, see the appendix section with supplementary material for paper III.

2.5.1.6 Side effects

2.5.1.6.1 Farnsworth Munsell 100 Color Hue Test

The FM-100 was used to test whether BDLE affected the ability to discriminate colors. The test was administered once in each LE condition on study day 2 or 3 (in study period 1) and day 8 or 9 (in study period 2). The FM-100 is a color-hue sorting task, where 100 hues are to be sorted in a particular order depending on color. The errors made on the sorting task give a total score, where a higher score indicates more difficulties in color discrimination. An error score of less than 20 indicates a superior color discrimination ability, a score of 20–100 indicates average ability, whereas a score of more than 100 indicates low ability (Rigby et al., 1991).

2.5.1.6.2 Self-report questionnaires

The UKU side effect rating scale was chosen to assess self-reported side effects of the BDLE. The instrument was originally developed for use with newly developed psychotropic drugs. For this trial, it was chosen due to its broad scope of side-effects questions across the most important domains (i.e., psychiatric, neurological, autonomic, and other) (Lingjaerde et al., 1987). The questionnaire was completed on the last day in each LE, corresponding to study days 7 and 13. The scale has 40 items where participants indicate whether they have experienced the given symptom while exposed to the respective condition (scored on a scale from 0 - "not at all", to 3 - "much more than usual"). Some additional questions are also directed specifically to participants of either sex (3 for males and 5 for females), giving a possible total score range of 0-129 for men and 0-135 for women. Higher scores correspond to more experienced side effects. We considered potential increased sleepiness and longer/deeper sleep to be positive effects, not side effects, and we excluded these questionnaire items from analyses.

2.5.2 Paper II dataset 2 - Assessments in the neurophysiological sleep clinic data (DS2)

This data collection was performed in an outpatient neurophysiological sleep clinic. Ambulatory sleep assessments were undertaken using PSG, actigraphy, and radar. Patients were fitted with PSG equipment using the same procedure described above for *the BDLE trial*. PSG-recordings were scored by the same clinical neurophysiologist with more than 10

years of scoring experience. Actigraphy was also carried out using the same equipment and procedure described for *the BDLE trial*. In the case of the radar assessments, participants were equipped with a portable version of the radar sensor described in *the BDLE trial* and were instructed to place it on their nightstand or a provided tripod.

2.5.3   Paper IV - Assessments in the dCBT-I trial

Demographic information, including age, sex, socioeconomic status, and employment status, was collected prior to randomization.

2.5.3.1   Baseline comorbid disorders

Comorbid mental disorders at baseline were assessed using the Psychiatric Diagnostic Screening Questionnaire (PDSQ) (Zimmerman & Mattia, 2001). The PDSQ is a 111-item self-report questionnaire that targets symptoms of mental disorders prevalent in outpatient mental health care. The questionnaire is based on symptoms from the DMS-IV axis 1 mental disorders. Further, information on comorbid physical disorders, prior or ongoing treatment in mental health care, and sleep medication usage were gathered in the baseline clinical diagnostic interview and from medical case note recordings.

2.5.3.2   Insomnia Severity Index

The Insomnia Severity Index (ISI) was used to assess levels of insomnia at baseline, post-treatment (9 weeks post-randomization), and 6 months follow-up (33 weeks post-randomization). This is a 7-item self-report questionnaire, where each question is rated on a 5-point Likert scale from 0 (no problem) to 4 (very severe problem). The total score ranges from 0-28, where a higher score indicates more severe insomnia symptoms. The ISI has been found to have good psychometric properties and is recommended for use as a primary outcome in insomnia research (Buysse et al., 2006). The ISI is also widely used in interventional studies of insomnia treatments, facilitating the comparison of the current findings with those of other studies. We defined a decrease of 8 points or more from baseline as a response to treatment (Morin et al., 2011) and a decrease to an ISI-score of 7 or lower as a remission (American Psychiatric Association, 2013).

2.5.3.3   Sleep diaries

An online version of the Consensus Sleep Diary was used to assess daily sleep-wake parameters at baseline, post-treatment (9 weeks post-randomization), and 6-months follow-up (33 weeks post-randomization) (C. E. Carney et al., 2012). Participants were asked to track their SOL, WASO, NW, early morning awakenings (EMA), TST, and SE for at least 10 days over 14 consecutive days.

2.5.3.4   Dysfunctional Beliefs About Sleep

Common dysfunctional cognitions regarding sleep were assessed at baseline, post-treatment (9 weeks post-randomization), and 6-month follow-up (33 weeks post-randomization) using the Dysfunctional Beliefs About Sleep scale, 16-item brief version (DBAS-16) (Morin et al., 2007). Each item is scored from 1-10, where a higher score implies more dysfunctional beliefs about sleep. The mean of all 16 items was used as the outcome.

2.5.3.5   Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale (HADS) was used to assess psychological distress at baseline, post-treatment (9 weeks post-randomization), and 6-months follow-up (33 weeks post-randomization). The HADS is a frequently used self-report questionnaire with questions regarding anxiety and depression symptoms. This questionnaire has shown high reliability in rating psychological distress symptoms in hospital outpatient clinics (Zigmond & Snaith, 1983). The total score ranges between 0 and 39, with increasing scores indicative of more psychological distress.

2.5.3.6   The Chalder Fatigue Scale

The Chalder Fatigue Scale (CFS) was used to assess levels of chronic fatigue symptoms. This 13-item self-report questionnaire contains questions on both physical and mental fatigue. Each item is rated on a 4-point scale and is scored bimodally 0-0-1-1 (e.g., 0 = better than usual; 0 = no more than usual; 1= worse than usual; 1= much worse than usual). The scale has shown high validity and reliability (Chalder et al., 1993).

2.6 Statistical analyses

2.6.1 Statistical analyses Paper I

2.6.1.1 Primary outcomes

Melatonin

To estimate the effect of LE on circadian rhythms phase-shift, we used a linear mixed model to model the evening rise in saliva melatonin levels. A random intercept was used for the variable participant id, and a random slope was specified for the effect of LE by each participant ID. The combination of evening clock time, assessment day, and condition was used as a fixed factor. The melatonin concentration data were transformed to a logarithmic scale to meet the mixed model assumption of normality of the outcome variable. This model estimates melatonin concentration by LE for each hour of melatonin sampling on all days of melatonin assessments. A linear mixed model uses all available data to estimate the fixed effects, and thus uses existing data for participants with some missing data points. The model is robust under the missing at random assumption. The model-based values were then used to calculate the phase shift of DLMO and the melatonin suppression in each respective LE. Individual phase shifts and melatonin suppression levels were based on observed values to accurately reflect the measured data points. R statistical package (version 3.6.2., https://www.R-project.org/) was used to perform all statistical analyses, and an alpha-level of $p < 0.05$ was chosen for all analyses. Linear mixed-effects models were fitted using the R-package "lme4".

DLMO was the estimated clock time when melatonin concentrations in dim light surpassed a threshold of 4 pg/mL. The melatonin phase shift was calculated as the average DLMO in each LE minus the baseline DLMO. The area under the curve (AUC) of melatonin concentrations measured while participants were exposed to the respective LEs was calculated to estimate the melatonin suppression. This was then compared with the AUC of melatonin concentrations measured in dim light the following day. As such, melatonin suppression was calculated as the percent of melatonin AUC in each LE divided by the AUC in dim light.

Polysomnography

Similar mixed models were specified for the sleep outcome variables derived from the PSG assessments. A random intercept was specified for participant ID. A random slope was

included for the effect of LE by each participant ID. Effects were estimated for each LE, and as additional post-hoc analyses, the effect of LE by condition order was estimated.

2.6.1.2 Secondary outcomes

For the secondary outcomes of the KSS and the FM-100, the intraindividual differences between LEs were calculated. These intraindividual differences were tested for significance using a one-sample t-test. Due to some outcome variables from the C-CPT-3 and the UKU not being normally distributed, a Wilcoxon signed-rank test was used to test the effect of LE on these outcomes. For the UKU-side effects scale and the KSS, pairwise deletion was used in case of missing data. One participant had missing data on the KSS for the duration of the BDLE and was therefore removed from the analysis.

2.6.2 Statistical analyses Paper II

We used a linear mixed model with the following variables one at a time as dependent variables: REM sleep duration, REM-sleep fragmentation, and REM density. The models were fitted in two steps by first entering LE and phase-shift one at a time and then simultaneously as fixed effects. A random intercept was added for participant ID in all models. For modeling REM-sleep accumulation, the combination of LE and time as a percent of total sleep time rounded to the nearest integer was used as fixed effects. The difference between LEs at each percent of total sleep time was then estimated, and the resulting p-values were corrected for multiple comparisons using the Bonferroni correction.

Normality of residuals was checked by visual inspection of QQ-plots and with the Shapiro–Wilk test. In some analyses, there were some deviations from normality, in which case bootstrapping with 10000 resamples were performed, and bias-corrected and accelerated confidence intervals were used. Two-sided p-values <0.05 were considered significant. There were no missing data on any of the PSG outcomes. Statistical analyses were performed using the R statistical package (version 3.6.2., https:// www.R-project.org/). A statistical significance level of $p < 0.05$ was used for all analyses. Linear mixed-effects models were fitted (using the R-package "lme4").

2.6.3   Statistical analyses Paper III

2.6.3.1   Data preparation.

Data preparation involved matching the epoch lengths of the three data types involved in the analyses. We temporally aligned the recordings from the three data modalities with a method based on maximal correlation. As actigraphy was recorded in 15-s epochs, these were aggregated into 30-s epochs to match the PSG epochs. For the radar assessment, 1Hz radar data was similarly aggregated into 30-s bins to be able to directly compare with the other two data types. As the scored PSG recordings contain information on sleep stages, all sleep stages (REM, Non-REM Stage 1 (N1), Non-REM Stage 2 (N2), and Non-REM Stage 3 (N3)) were combined into sleep and given the score 0, whereas wake was given the score 1.

Dataset 1 (DS1) was split into two smaller datasets, DS1-train to be used for model parameter estimation and DS1- test to be used for testing. This was achieved by randomly assigning participants to either one. Dataset 2 was used as an independent validation dataset. For further details see paper III and paper III supplementary material.
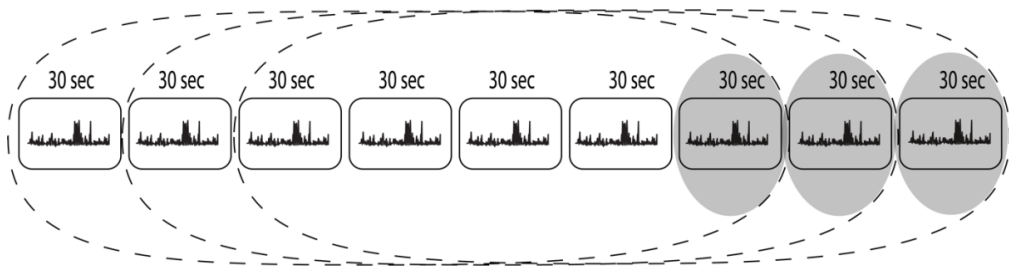
Figure 5. The figure represents the principle used in linear sliding sum models. A weighted sum activity value for each epoch is based on a time horizon of between 0-10 preceding and succeeding epochs. This value is subsequently compared with some threshold, which in our study was set to 0.5, to be determined as sleep or wake.

2.6.3.2 Model development.

Model development followed the same approach for all three data modalities, and was inspired by a type of model often seen for sleep-wake classification in actigraphy. These ´linear sliding sum models´ compute a value for each epoch as a weighted sum of its recorded activity and the activity recorded for a set number of preceding and succeeding epochs. They then use these computed values to make a classification decision for the epoch, typically by comparing it to some threshold and sometimes passing it through a set of heuristic rules (Haghayegh et al., 2019). We used PSG-derived sleep-wake state as the dependent variable and performed logistic regression to estimate the weights of the sliding sum model.

For actigraphy, the epoch activity count was used as the independent variable. In contrast, for the radar data both fast movement, slow movement, and RPM (the three data outputs from the radar sensor) were included separately as independent variables. All possible combinations of a time horizon of 0-10 epochs in each temporal direction were used giving a total of 121 models for each data modality. This resulted in a sleep/wake probability per epoch, and a 0.5 threshold was used to classify epochs as sleep or wake. The Cole-Kripke rescoring rules (Cole et al., 1992) was then applied, and in the current work, the rules were applied as follows: (1) the next 1 minute scored as sleep is rescored as wake if the preceding 4 minutes were scored as wake, (2) the next 3 minutes scored as sleep is rescored as wake if the preceding 10 minutes were scored as wake, and (3) the next 4 minutes scored as sleep is rescored as wake if the preceding 15 minutes were scored as wake. For details on model development and rescoring rules, please see the supplementary material of paper III.

2.6.3.3 Classification performance analysis

The classification models developed on DS1-train were applied to DS1-test and DS2. To assess classification performance, overall accuracy, sensitivity, and specificity were calculated by comparing classification results epoch-by-epoch with PSG-derived sleep-wake states. Cohen's kappa was calculated as a measure of agreement that takes into account the probability of correct classification by chance, given the high proportion of sleep epochs versus wake epochs. Cohen's kappa was regarded as; moderate between <0.41 and <0.60, substantial between <0.61 and <0.80, and near-perfect between <.081 and <0.99 (Ranganathan et al., 2017).

2.6.3.4   Sleep parameters

We calculated SOL, TST, WASO, SW, and NW based on the sleep-wake classification results. In the case of PSG, SOL was the time between bedtime and the first sleep-epoch, whereas, in the case of radar and actigraphy, it was the time between bedtime and the first 3-minute consecutive sleep period. For nights with PSG, participants indicated bedtime by pressing an event-marker button on the PSG equipment. For nights where PSG was not undertaken, the participants' scheduled bedtime at 2300h was used (de Souza et al., 2003). The first epoch after the last epoch of the night scored as sleep, was defined as the wake-time.

2.6.3.5   Statistical testing of sleep parameters

To test whether radar assessment or actigraphy was different from PSG in these sleep parameters, Student t-tests were performed. We used Cohen's *d* as a measure of effect size. To visually compare the agreement between data modalities, we plotted Bland-Altman plots with bias and 95% limits of agreement (LA). To identify proportional bias, we used the regression approach for nonuniform differences and then reported R^2 and added regression lines to the Bland-Altman plot for significant slopes. The absolute values of the difference between parameter estimates from PSG compared with parameter estimates of actigraphy and radar data were calculated and compared using forest plots (Bland & Altman, 1999). All analyses were carried out using MATLAB (versions R2018-R2020).

2.6.4   Paper IV

The non-inferiority analyses reported in this paper follow the Consolidated Standards of Reporting Trials (CONSORT) guidelines for non-inferiority trials (Piaggio et al., 2012).

2.6.4.1   Non-inferiority margin and power calculations

The primary outcome was the score on the ISI, and the primary endpoint was week 33 after randomization. A non-inferiority analysis is specifically designed to test whether an experimental treatment condition is not worse than an established treatment. It rests on a pre-specified inferiority margin (Δ), and if the experimental treatment performs poorer than the reference treatment by more than Δ, the experimental treatment is considered inferior. However, as long the difference between treatments is lower than Δ, non-inferiority is established. To test for non-inferiority, we used the confidence interval approach, which we interpreted as recommended in the CONSORT guidelines for the non-inferiority analysis (Piaggio et al., 2012). This tests whether the lower end of the 95 % confidence interval of the

difference between treatment groups was larger than the non-inferiority margin in the direction of dCBT-I. The between-group difference on ISI at endpoint week 9 was also analyzed using the non-inferiority approach.

Only one previous non-inferiority trial of dCBT-I was found, which compared dCBT-I to group CBT-I using a non-inferiority margin of 4 points on the ISI (Blom et al., 2015). As such, there is no established non-inferiority margin for trials of dCBT-I. The margin in the previous trial was decided based on specific ISI-scores that have previously been reported in terms of minimally important differences (6 points (Yang et al., 2009) and 8 points (Morin, 1993)) when reducing ISI-scores from before to after treatment for insomnia. For this trial, we instead used the effect sizes (ES) reported in the extensive literature on the efficacy of CBT-I. We determined the non-inferiority margin based on what would be considered a moderate effect size (Cohen's $d = 0.5$) in the treatment literature, corresponding to 2 points on the ISI given a standard deviation (SD) of 4.0. As ESs are typically in the magnitude of Cohen's $d = 1.0$ for RCTs of CBT-I (Blom et al., 2015; Zachariae et al., 2016), this margin was considered feasible to be able to distinguish a clinically meaningful effect from an effect with limited clinical relevance. Following these ES and SD assumptions, a sample size of 100 participants would be sufficient to give a statistical power of 80% to detect non-inferiority of dCBT-I with an alpha level of .05.

Secondary outcome were levels of psychiatric distress on the HADS, levels of fatigue on the CFS, sleep variables from the subjective sleep diaries, and responders and remitters defined by the size of improvement on the ISI. These outcomes were analyzed using a standard superiority methodology.

SPSS version 25 was used for the analysis of all outcomes. A statistician blinded to group allocation specified the models and performed the analyses. Linear mixed models were used with a random intercept for participant ID for all randomized participants (the ITT). Fixed effects were treatment group, time, and their interaction. Baseline values of the outcome variable were added as a covariate in the models to adjust for baseline insomnia severity levels. The interaction term was used to estimate the group effect at week 9 and week 33, a method recommended by Twist et al.(J et al., 2018). Similar linear mixed models were used for both the primary and the secondary outcomes. Pearson's chi-square test and the Newcombe Hybrid Score CI were used in the analyses of responders and remitters (Fagerland

et al., 2017). The same statistical analyses were performed for patients who had completed the treatment per protocol (PP). These patients had to have completed all sessions offered in FtF therapy or all modules in the dCBT-I treatment condition. Cohen's d was used as a measure of ES. The pre-post model estimated difference in mean scores divided by the baseline SD was used to calculate Cohen's d within each condition. To estimate the ES of treatment condition, the model estimated between-group difference was divided by the pooled SD at baseline (Grissom & Kim, 2012).

2.7   Ethics

All three data collections were conducted in accordance with the declaration of Helsinki and the revised declaration of Geneva (Parsa-Parsi, 2017). All participants gave informed written consent before participation. Regional Ethical Committees reviewed and approved the protocols before study start (for details, see below).

2.7.1   Papers I, II, and dataset 1 in paper III

The Regional Ethical Committee in Trondheim approved this protocol (Central Norway; REK: 2017/916). The trial was also registered on the ISRCTN website with reference number 12419665.

2.7.2   Paper III (dataset II)

The Regional Ethical Committee in Trondheim approved this study protocol (Central Norway; REK: 2017/309).

2.7.3   Paper IV

The study protocol was pre-approved by the Regional Ethical Committee of South-East Norway (Reference: 2013/1836). The RCT was registered on the Clinical Trials website with reference number NCT02044263.

3 Results

The published abstracts of papers I-IV are reported below:

3.1 Paper I

The evening light environment in hospitals can be designed to produce less disruptive effects on the circadian system and improve sleep

Daniel Vethe, Jan Scott, Morten Engstrøm, Øyvind Salvesen, Trond Sand, Alexander Olsen, Gunnar Morken, Hanne Siri A. Heglum, Kaia Kjørstad, Patrick M. Faaland, Cecilie L. Vestergaard, Knut Langsrud and Håvard Kallestad.

Study Objectives

Blue-depleted lighting reduces the disruptive effects of evening artificial light on the circadian system in laboratory experiments, but this has not yet been shown in naturalistic settings. The aim of the current study was to test the effects of residing in an evening blue-depleted light environment on melatonin levels, sleep, neurocognitive arousal, sleepiness, and potential side effects.

Methods

The study was undertaken in a new psychiatric hospital unit where dynamic light sources were installed. All light sources in all rooms were blue-depleted in one half of the unit between 06:30 pm and 07:00 am (melanopic lux range: 7–21, melanopic equivalent daylight illuminance [M-EDI] range: 6–19, photopic lux range: 55–124), whereas the other had standard lighting (melanopic lux range: 30–70, M-EDI range: 27–63, photopic lux range: 64–136), but was otherwise identical. A total of 12 healthy adults resided for 5 days in each light environment (LE) in a randomized cross-over trial.

Results

Melatonin levels were less suppressed in the blue-depleted LE (15%) compared with the normal LE (45%; $p = 0.011$). Dim light melatonin onset was phase-advanced more (1:20 h) after residing in the blue-depleted LE than after the normal LE (0:46 h; $p = 0.008$). Total sleep time was 8.1 min longer ($p = 0.032$), rapid eye movement sleep 13.9 min longer ($p < 0.001$),

and neurocognitive arousal was lower ($p = 0.042$) in the blue-depleted LE. There were no significant differences in subjective sleepiness ($p = 0.16$) or side effects ($p = 0.09$).

Conclusions

It is possible to create an evening LE that has an impact on the circadian system and sleep without serious side effects. This demonstrates the feasibility and potential benefits of designing buildings or hospital units according to chronobiological principles and provide a basis for studies in both nonclinical and clinical populations.

3.2 Paper II

Evening light environments can be designed to consolidate and increase the duration of REM-sleep

Daniel Vethe, Henning. J. Drews, Jan Scott, Morten Engstrøm, Hanne Siri A. Heglum, Janne Grønli, Jonathan P. Wisor, Trond Sand, Stian Lydersen, Kaia Kjørstad, Patrick. M. P. Faaland, Cecilie L. Vestergaard, Knut Langsrud and Håvard Kallestad.

Scientific Reports, 2022, *12*(1), 8719. https://doi.org/10.1038/s41598-022-12408-w

Evening exposure to short-wavelength light has disruptive effects on circadian rhythms and sleep. These effects can be mitigated by blocking short-wavelength (blue) frequencies, which has led to the development of evening blue-depleted light environments (BDLEs). We have previously reported that residing 5 days in an evening BDLE, compared with residing in a normal indoor light environment of similar photopic lux, advances circadian rhythms and increases the duration of rapid eye movement (REM) sleep in a randomized cross-over trial with twelve healthy participants. The current study extends these findings by testing whether residing in the evening BDLE affects the consolidation and microstructure of REM sleep in the same sample. Evening BDLE significantly reduces the fragmentation of REM sleep ($p = 0.0003$), and REM sleep microarousals in ($p = 0.0493$) without significantly changing REM density or the latency to first REM sleep episode. Moreover, the increased accumulation of REM sleep is not at the expense of NREM stage 3 sleep. BDLE further has a unique effect on REM sleep fragmentation ($p = 0.0479$) over and above that of circadian rhythms phase-shift, indicating a non-circadian effect of BDLE. If these effects can be replicated in clinical

populations, this may have a therapeutic potential in disorders characterized by fragmented REM sleep.

3.3   Paper III

Distinguishing sleep from wake with a radar sensor: A contact-free real-time sleep monitor

Hanne Siri Amdahl Heglum, Håvard Kallestad, Daniel Vethe, Knut Langsrud, Trond Sand and Morten Engstrøm

This work aimed to evaluate whether a radar sensor can distinguish sleep from wakefulness in real time. The sensor detects body movements without direct physical contact with the subject and can be embedded in the roof of a hospital room for completely unobtrusive monitoring. We conducted simultaneous recordings with polysomnography, actigraphy, and radar on two groups: healthy young adults (n = 12, four nights per participant) and patients referred to a sleep examination (n = 28, one night per participant). We developed models for sleep/wake classification based on principles commonly used by actigraphy, including real-time models, and tested them on both datasets. We estimated a set of commonly reported sleep parameters from these data, including total-sleep-time, sleep-onset-latency, sleep-efficiency, and wake-after-sleep-onset, and evaluated the inter-method reliability of these estimates. Classification results were on-par with, or exceeding, those often seen for actigraphy. For real-time models in healthy young adults, accuracies were above 92%, sensitivities above 95%, specificities above 83%, and all Cohen's kappa values were above 0.81 compared to polysomnography. For patients referred to a sleep examination, accuracies were above 81%, sensitivities about 89%, specificities above 53%, and Cohen's kappa values above 0.44. Sleep variable estimates showed no significant intermethod bias, but the limits of agreement were quite wide for the group of patients referred to a sleep examination. Our results indicate that the radar has the potential to offer the benefits of contact-free real-time monitoring of sleep, both for in-patients and for ambulatory home monitoring.

3.4 Paper IV

Mode of delivery of Cognitive Behavioral Therapy for Insomnia: A randomized controlled non-inferiority trial of digital and face-to-face therapy

Håvard Kallestad, Jan Scott, Øystein Vedaa, Stian Lydersen, Daniel Vethe, Gunnar Morken, Tore Charles Stiles, Børge Sivertsen and Knut Langsrud

Study Objectives: Digital Cognitive Behavioral Therapy for Insomnia (dCBT-I) has demonstrated efficacy in reducing insomnia severity in self-referred and community samples. It is unknown, however, how dCBT-I compares to individual face-to-face (FtF) CBT-I for individuals referred to clinical secondary services. We undertook a randomized controlled trial to test whether fully automated dCBT-I is non-inferior to individual FtF CBT-I in reducing insomnia severity. Methods: Eligible participants were adult patients with a diagnosis of insomnia disorder recruited from a sleep clinic provided via public mental health services in Norway. The Insomnia Severity Index (ISI) was the primary outcome measure. The non-inferiority margin was defined a priori as 2.0 points on the ISI at week 33. Results: Individuals were randomized to FtF CBT-I (n = 52) or dCBT-I (n = 49); mean baseline ISI scores were 18.4 (SD 3.7) and 19.4 (SD 4.1), respectively. At week 33, the mean scores were 8.9 (SD 6.0) and 12.3 (SD 6.9), respectively. There was a significant time effect for both interventions (p < 0.001); and the mean difference in ISI at week 33 was −2.8 (95% CI: −4.8 to −0.8; p = 0.007, Cohen's d = 0.7), and −4.6 at week 9 (95% CI −6.6 to −2.7; p < 0.001), Cohen's d = 1.2. Conclusions: At the primary endpoint at week 33, the 95% CI of the estimated treatment difference included the non-inferiority margin and was wholly to the left of zero. Thus, this result is inconclusive regarding the possible inferiority or non-inferiority of dCBT-I over FtF CBT-I, but dCBT-I performed significantly worse than FtF CBT-I. At week 9, dCBT-I was inferior to FtF CBT-I as the 95% CI was fully outside the non-inferiority margin. These findings highlight the need for more clinical research to clarify the optimal application, dissemination, and implementation of dCBT-I.

4 Discussion

The results presented in this thesis demonstrate the effectiveness of three different technologies in efforts to influence circadian rhythms and sleep, facilitate the assessment of sleep, and disseminate effective treatment of insomnia. These findings indicate that technology, while possibly a part of the widespread sleep problems in modern society, may simultaneously contribute to alleviating them if applied in the specific contexts where they demonstrate efficacy.

All the studies reported in this thesis are set in the context of a university hospital. *The BDLE trial* was performed with healthy participants in a new-built acute psychiatric ward. The radar assessment was both a part of *the BDLE trial* and collected data from ambulatory sleep assessments in an outpatient neurophysiological sleep clinic. *The dCBT-I trial* was set in a different hospital outpatient sleep clinic treating insomnia and circadian rhythm disorders. There is a need for ways to both assess and treat sleep disturbance in the hospital setting. Prevalence of sleep disturbance is high (Elliott et al., 2013; Engwall et al., 2015; Langsrud et al., 2016; Schennach et al., 2019; Wulff et al., 2010) with limited treatment options besides medication during admissions. Similarly, the options for monitoring patients' sleep-wake states during admissions are often limited to nurse observations that may contribute to disrupted sleep for inpatients (Veale, 2019). In outpatient sleep clinics, there is a shortage of practitioners trained in CBT-I, leading to long waitlists, and many patients do not receive what should be the first-line treatment for insomnia (Baglioni et al., 2020).

The scientific contributions of the work included in this thesis are different for these three technologies, given that they are at different stages of development. The BDLE technology is at an early stage, with laboratory studies demonstrating its effectiveness but a lack of studies investigating the physiological effects in naturalistic settings. Our *BDLE-trial* addresses this particular question, thus contributing an essential piece of evidence: BDLEs have measurable effects on circadian and sleep variables in healthy participants when installed in multi-room building complexes. Further work is needed to understand the considerable individual differences in effects and who will have clinically relevant effects of BDLE in terms of sleep and other health outcomes. Similarly, the radar assessment of sleep is also at an early stage of development, with studies like ours aiming to assess its accuracy against PSG sleep assessment. Our work demonstrates one approach to sleep-wake classification using radar sensors. At the current stage, this may hold potential as an auxiliary information source on

patients' sleep. However, given the low specificity, especially in the clinical sample, this information is not yet ready to be utilized in clinical decision-making. The dCBT-I technology has existed for nearly two decades and has progressed further in terms of studies establishing its efficacy. However, most studies have tested dCBT-I against wait-list or control conditions. We have attempted to discern its effectiveness compared with FtF CBT-I. This comparison adds information, particularly as it contradicts other studies by not establishing non-inferiority of dCBT-I, that may be useful as stakeholders plan implementations of dCBT-I versions into healthcare systems worldwide.

4.1 The effects of evening blue-depleted light environments

In the intersection between sleep and technology, evening and nighttime exposure to artificial lighting pose one of the primary challenges to human health by disrupting circadian rhythms and sleep and increasing alertness (T. M. Brown et al., 2022; Foster & Wulff, 2005; Münch et al., 2020). This has sparked efforts to use recent developments in lighting technology (programmable LEDs) to create evening light environments that reduce these adverse effects. In the first two papers of this thesis, we demonstrate that residing in a hospital ward with an integrated evening BDLE does lead to meaningful physiological effects on circadian rhythms, sleep and its microstructure, and waking alertness while simultaneously not affecting subjective sleepiness and having few side-effects outside of reduced color-discrimination ability.

4.1.1 Circadian rhythms effects of evening BDLE

When comparing the circadian phase of DLMO in both light environments, we found that, on average, DLMO was phase-advanced more after residing in the evening BDLE compared with after standard LE, indicating an apparent effect of BDLE on circadian phase-shift. This suggests the evening BDLE has properties that can be used to mitigate the delaying effects of evening artificial light exposure on the entrained phase of the circadian pacemaker (Burgess & Molina, 2014; Khalsa et al., 2003; Rüger et al., 2013; St Hilaire et al., 2012; Zeitzer et al., 2000). This is in line with other studies of evening blue-depleted lighting (Rahman et al., 2022; Santhi et al., 2012) or blue-blocking glasses (Esaki et al., 2016), that also find phase-advancement of DLMO. Moreover, it mimics the phase-advancing effects found when participants reside outdoors in natural evening darkness (Stothard et al., 2017; Wright et al.,

2013) or when at-home evening light levels are significantly reduced (Burgess & Molina, 2014).

We further found that melatonin suppression was reduced in the BDLE. In fact, we could not demonstrate significant suppression of melatonin in the BDLE compared with dim light. Although this does not necessarily mean the BDLE was similar to darkness, it suggests that residing in BDLE is close to darkness in terms of non-visual effects of light. Light exposure with properties that impedes the non-visual effects of light while still allowing for photopic vision has been eloquently termed 'virtual darkness' (Phelps, 2008). Our findings are in line with other studies of blue-depleted lighting (Nowozin et al., 2017; Papamichael et al., 2012; Rahman et al., 2017; Souman, et al., 2018; van de Werken et al., 2013), or blue-blocking glasses (Kayumov et al., 2005; Ostrin et al., 2017; Sasseville et al., 2006) performed in laboratory settings that also find reduced melatonin suppression.

Recent findings also indicate that the degree of melatonin suppression depends on the dosage of melanopic lux (Lucas et al., 2014; Nowozin et al., 2017). In our study, the melanopic EDI in the BDLE was less than 20, compared with the standard LE, where it was typically between 40-60 melanopic EDI. We did not completely remove blue light frequencies, as this would have created a light environment that was visually uncomfortable, but we based the degree of blue-depletion on previously published dose-response curves. However, recent studies have found that humans are exquisitely sensitive to the melatonin-suppressing effects of short-wavelength light, with light as low as 1.5 melanopic lux (Prayag et al., 2019). In that laboratory study, however, a pupil dilator was used to uncover the lowest possible light level that could lead to melatonin suppression. The melanopic illuminance threshold that will generate melatonin suppression is likely somewhat higher in naturalistic settings.

The emerging literature on blue-depleted lighting has typically been focused on demonstrating its efficiency in controlled laboratory settings. The experimental conditions in the lab (e.g., utilizing light domes or boxes and limiting the movement of participants) allow for strict control of intensity and spectral composition of the light exposure across time (see, e.g., (Nowozin et al., 2017; Rahman et al., 2017; Souman, et al., 2018)). The laboratory research has been crucial in demonstrating that the reduction of evening melanopic lux is not just a theoretical opportunity to reduce the harmful effects of artificial evening lighting but leads to physiologically measurable improvements in markers of circadian rhythms. However,

given the unnatural restrictions placed upon individuals in lab studies, it has remained unanswered whether these effects will translate when implemented in larger-scale housing or hospital settings, where individuals are free to move between rooms and where the intensity of light exposure will change with the direction of gaze. The novel contribution demonstrated in *the BDLE trial* is that the blue-depleted lighting technology can be successfully implemented across multiple rooms and corridors in a large-scale building complex while preserving the physiological benefits of reduced melatonin suppression and phase-advancement of circadian rhythms.

4.1.2   Reduced alertness during evening BDLE

It is known that short-wavelength light has an alerting effect on human psychophysiology. These effects manifest themselves in improved reaction times and cognitive functioning (Rahman et al., 2014; Rüger et al., 2006) but can also be found in wake-EEG recordings where the spectral density is shifted towards alertness-correlated wavelengths (Cajochen et al., 2000; Münch et al., 2011; Rahman et al., 2014). These effects are sometimes not reflected in subjective measures of sleepiness and arousal (Sasseville et al., 2015), suggesting objective effects may be minor. In paper I, we find some indications for increased alertness, as there was a significantly higher variability in participants' reaction times when assessed in the BDLE. Other studies on blue-depleted lighting (Rahman et al., 2017) or blue-blocking glasses (van der Lely et al., 2015) have also found reductions in attention and alertness as measured by reaction times. Another study using blue-depleted lighting did not find reduced alertness (Souman, et al., 2018), although the blue-depleted light solution in that study included a portion of the light spectrum below the action spectrum of ipRGCs. It should also be noted that while the evidence for an alerting effect of light is considerable, the effect of is not consistently replicated (Souman, et al., 2018), and thus seems less pronounced and more unstable than, e.g., the suppressing effects on melatonin.

The alerting effects of short-wavelength light may be mediated both directly through alertness-promoting regions of the brain (Rupp et al., 2019), and indirectly by suppressing melatonin and delaying the biological night. An early study of wake-EEG activity in the evening during bright or dim light exposure found that bright light was associated with reductions in theta-activity and slow eye movements, indicating increased cortical alertness (Cajochen et al., 2000). Moreover, several studies have found increases in alpha-activity

during evening light exposure (Münch et al., 2011; Rahman et al., 2014) – which is considered a marker of the circadian drive for alertness. Connecting these findings to melatonin as a marker for circadian rhythm, a study by Chellappa and colleagues found that short-wavelength shifted polychromatic light increased alertness and that this increase was strongly correlated with the degree of melatonin-suppression (Chellappa et al., 2011). There is ample evidence that bright and blue-shifted light exposure increases pre-bedtime objective alertness. Our findings in paper I add to the evidence that blue-depleted light interventions reduce objective forms of alertness.

The evidence for changes in subjective sleepiness, however, is less clear. In paper I, we do not find any changes in subjective sleepiness during the evening or the following morning. Other studies have found similar discrepancies between objective measures and subjective sleepiness (Chang et al., 2015; Rahman et al., 2017). Extrapolating from sleep deprivation studies, humans are notoriously poor at detecting lapses of attention and reduced cognitive function (Van Dongen et al., 2003). The absence of differences in subjective sleepiness, even in the presence of objective changes in alertness, melatonin, or sleep, may be explained by poor insight into objective cognitive functioning. The discrepancy may also support an absence of expectancy effects in other measures, as subjective sleepiness would likely be elevated in the case of expectancy effects.

4.1.3   Longer sleep duration after BDLE

We further demonstrate that residing in evening BDLE has an effect on sleep as assessed with PSG. Specifically, we find that TST increases with 8 minutes when residing in the evening BDLE. In the study by Rahman and colleagues, there was no difference in TST between blue-depleted and fluorescent lighting (Rahman et al., 2017). However, participants in that study received light exposure for only one evening in each condition, which might not have been enough to generate an overall increase in TST (Rahman et al., 2017). The authors report increased TST across both nights in the study if blue-depleted lighting was received first. They suggest that the disruptive effects of fluorescent lighting may be carried into subsequent nights, thus limiting the increase in TST for those that received blue-depleted as the second condition (Rahman et al., 2017). In a study by Münch and colleagues (Münch et al., 2006), there was no difference between being exposed to two hours of monochromatic blue or green light or darkness in the evening on TST. In that study, the light exposure was also limited to

one evening. In our *BDLE-trial*, participants resided for five days in each condition resulting in advance of DLMO, while bedtimes were kept approximately similar. This may have changed the phase-angle such that participants went to bed later in relation to their DLMO, which may have led to a reduced circadian drive for alertness and an increased homeostatic sleep pressure, potentially shortening sleep-onset latency and reducing awakenings. We do not, however, find any significant differences in sleep-onset latency or awakenings, but taken together, this may explain increased sleep time. It would be of interest to see whether the observed increase in TST would be larger had participants been given more opportunity to self-select bedtimes and rise times. In the current trial, however, the primary aim was to test the effects of BDLE exposure on markers of circadian rhythms. Due to the potential confounding effects of shifting the timing and duration of sleep window of opportunity on both homeostatic and circadian processes, we tried to keep sleep-wake schedules similar across conditions for this study.

Two other studies with blue-blocking glasses have reported outcomes for objectively assessed sleep. One study using blue-blocking glasses found that actigraphy-assessed sleep onset timing was advanced (Esaki et al., 2016). In contrast, the other did not find any difference in PSG-assessed sleep variables (van der Lely et al., 2015). Some evidence indicates that reading from a backlit e-book results in longer sleep-onset latency but no difference in TST compared with reading a book in dim light (Chang et al., 2015). Our finding of increased TST after BDLE has not previously been demonstrated but adds to the same general direction as studies demonstrating shorter sleep-onset latency. Future studies should test whether TST and sleep-onset latencies change if participants residing in BDLE are allowed to self-select bedtimes.

4.1.4 *More REM sleep after BDLE*

In addition to prolonged TST, we found a notable increase in REM sleep duration while residing in the evening BDLE. This increase was 13.9 minutes, thus more prominent than the increase in TST. REM sleep is known to display circadian rhythm characteristics, with an increased propensity to enter into REM sleep in the late biological night/early morning (Czeisler et al., 1980). This is reflected in findings that evening administration of exogenous melatonin phase advances DLMO, reduces REM sleep latency, and increases the duration of the first sleep cycle (Cajochen et al., 1997, 1998). Furthermore, other studies that have observed phase-delaying effects of DLMO after evening light exposure have also reported concomitant effects on REM sleep timing and duration (Chang et al., 2015; Münch et al.,

2006). Similarly, morning light exposure has been found to increase REM sleep duration in the early night, accompanying its known phase-advancing properties on DLMO (Gordijn et al., 1999; Sack et al., 1986). This mimics our findings of both phase-advancement of DLMO and increased REM-sleep duration in BDLE. We did not, however, find reduced REM-latency, but in the supplementary information of paper II, we report longer REM-sleep duration in the first sleep cycle in BDLE. In paper II, we found further support for this circadian rhythm of REM sleep propensity by showing that DLMO phase-shifts were associated directly with total overnight REM sleep duration. The association with phase-shift could not be separated from an association with BDLE, but we know from the primary outcomes in paper I that BDLE and phase-shift were related. Interestingly, the temporal accumulation of REM sleep stages (see paper II Figure 2a) exhibited a marked increase in REM sleep accumulation approximately 30 minutes earlier in BDLE, which approximates the 34-minute larger DLMO phase-advance. Given the association with phase-shifts, and this temporal advance of accumulation, the increased REM duration likely results from the phase-advance of circadian rhythms.

The neurological basis for the circadian rhythm regulation of REM sleep timing is currently unknown but may include at least two mechanisms or brain areas. First, orexin-neurons in the lateral hypothalamus (LH) show distinct wake-on, REM-off firing patterns, and seem to inhibit REM sleep by activating REM-off neurons in the locus coeruleus (LC) and the dorsal raphe nucleus (DRN) (R. E. Brown et al., 2012; Chemelli et al., 1999). These orexin neurons receive input both directly and indirectly from the SCN via the dorsomedial hypothalamus (DMH) (Aston-Jones et al., 2001; Chou et al., 2003). Patients with narcolepsy exhibit a loss of orexins and show a loss of diurnal control of REM sleep. Second, the preoptic hypothalamic area (POA) also receives direct and indirect input from the SCN (Chou et al., 2003; Deurveilher et al., 2002), and evidence from mouse models point to the level of activation of these areas being correlated with the amount of REM sleep (Lu et al., 2002). These areas could constitute regulating input into a REM-on, REM-off network in pontine areas with reciprocal inhibitory mechanisms responsible for the generation of the REM sleep state (Lu et al., 2006; Wang et al., 2021). Changes in the SCN input to these two areas may thus putatively enhance the REM-off or the REM-on side, thus temporally shifting the propensity to enter REM sleep in accordance with SCN circadian rhythms.

4.1.5 More consolidated REM sleep after BDLE

In paper II, we aimed to investigate the increased REM sleep in more detail by examining REM sleep consolidation and microstructure. We found that REM sleep was more consolidated in the BDLE condition with fewer interruptions of REM sleep episodes and fewer EEG-microarousals during REM sleep. These changes occurred without altering the REM-density. We also found that REM sleep fragmentation changes were associated with the phase-shift of DLMO. However, BDLE had an effect over and above the effect of phase-shift on REM sleep fragmentation. This was not the case for REM sleep duration, where both phase-shift and BDLE explained largely the same variance. Moreover, in paper I, we found that the effect of BDLE on REM sleep duration was strongest in period 2 of the study, which was also where the phase-shifting effects were strongest. However, the effect of BDLE on REM sleep consolidation was most prominent in period 1 of the study, which may temporally put the fragmentation effect of BDLE earlier than the phase-advance effect. Taken together, these findings may indicate that the consolidation of REM sleep as a result of BDLE may be driven by non-circadian mechanisms.

In the context of non-circadian effects on REM sleep, it is interesting to note that direct projections from ipRGCs innervate to the VLPO, a core region of sleep-wake regulation, and to the LH, which inhibits LC activity as discussed above (Daneault et al., 2016). The VLPO, among other sleep-regulating functions, projects to the periaqueductal gray area (PAG) (Lu et al., 2002), which has been proposed as a central modulator of a REM-on/REM-off flip-flop switch (Kaur et al., 2009; Lu et al., 2006). Regarding LC activity, there is an interesting line of research coupling LC hyperactivation to the instability of REM sleep (Aston-Jones et al., 2007; Poe et al., 2020). LC activity during wake and NREM sleep ensures noradrenaline (NA) release, which is crucial for the long-term potentiation of synapses (Poe et al., 2020) and, thus, learning and memory. The temporary cessation of LC firing just before the initiation of REM sleep and through REM sleep episodes briefly creates a low-NA environment that may allow for synaptic depotentiation of some memory circuits and facilitate differentiation between salient and less salient memories. REM sleep consolidation has received increasing attention in sleep research because it can be seen as an expression of increased LC activity (Wassing et al., 2019). Moreover, signs of increased LC activity can be found in a range of mental disorders (Aston-Jones et al., 2007), suggesting this may be an underlying common factor in psychopathology that is manifested in REM sleep alterations (Van Someren, 2021; Wassing et al., 2019).

An increase or decrease in REM sleep duration is often seen in conjunction with changes in the time course of slow wave activity (SWA – power density often between 0.5 – 4.0Hz) or SWS (sum of NREM stages 3 and 4). In particular, evening light exposure (both monochromatic blue and polychromatic white) tend to decrease the SWA in the early night and increase it later in the night (Chellappa et al., 2013; Christian et al., 1992; Grønli et al., 2016; Münch et al., 2006). This might indicate that evening light exposure has a delaying or suppressing effect on SWA in the first sleep cycle, which leads to a rebound of SWA later in the night. This could result from melatonin suppression or higher cortical alertness persisting into the early sleep episode. A potential underlying pathway could be the direct ipRGC projections to the VLPO (Daneault et al., 2016), which contains NREM-sleep promoting neurons (Gooley et al., 2003). A recent study demonstrated that such ipRGC stimulation of similar areas could explain the acute effects of light on NREM sleep in mice (Zhang et al., 2021). Suppose acute effects of light on sleep-latency and sleep in humans are partly based on similar mechanisms. In that case, increased ipRGC activation may lead to reduced NREM-promoting firing in the VLPO, thus delaying the ability to generate intensive NREM sleep in the early night. This may be supported by the effects of having the light stay on while sleeping, where one study found both reduced SWS and increased NREM 1 sleep (J. R. Cho et al., 2013).

We did not investigate the effects on SWA, as this was deemed outside the scope of paper II, but these mechanisms may be of interest in discussing the effects on REM sleep. Returning to the two-process theory of sleep (Borbely, 1982; Borbély et al., 2016), the ultradian NREM-REM cycle is thought to result from a reciprocal interaction of the sleep states. In the early night, a high SWA/NREM propensity due to high homeostatic sleep pressure inhibits long REM sleep episodes and forces a quick return to NREM stages. In the late night, a high circadian propensity for REM sleep paired with an already dissipated pressure for SWA/NREM sleep increases REM sleep episode durations and reduces NREM SWA. If excitation of ipRGCs leads to an early-night reduction of NREM intensity reflected in lower SWA, this explains the rebound of SWA often seen in the later sleep cycles (Cajochen et al., 1998; Münch et al., 2006). If reduced ipRGC activation facilitates the initiation and maintenance of NREM SWA, this would lower the propensity for NREM sleep faster, thus potentially explaining the more consolidated REM sleep in BDLE.

4.2 The potential of using radar in the assessment of sleep

An obvious area of sleep research that stands to benefit from the rapid technological development over the last decades is how we measure sleep. Sleep research, still a relatively young scientific discipline, is somewhat limited by the difficulties and costs associated with accurately measuring sleep using PSG. This limits the settings in which accurate sleep assessments may feasibly be undertaken. Moreover, actigraphy, considered the assessment modality of choice in such settings, has limited specificity and is not always well-tolerated by research subjects. This means that in some vulnerable groups, such as inpatients in psychiatry, we lack good ways to monitor sleep, and therefore, we are also limited in our ability to test the outcomes of treatments on sleep – which is considered a key intervention target for improvement in the comorbid mental disorder. Over the last decade, a range of different sensor types and assessment devices have been developed as alternatives to actigraphy or PSG. In paper III, we demonstrate that a UWB radar embedded in the ceiling or on a nightstand may be equally good or better than actigraphy while reducing the assessment's invasiveness by being completely contact-free.

The main findings in paper III were that radars, both embedded in the ceiling and on nightstands, had excellent or good agreement with PSG in terms of detecting sleep, and were thus similar to actigraphy in this regard. The real-time and non-real-time models did not differ much in performance. We could attain reliable estimates for the most common outcomes of TST, SOL, SE, NW, and WASO. We found an expected difference between the two datasets, where assessments performed on the healthy participants from *the BDLE trial* showed excellent agreement with PSG with small LAs. However, in the outpatient at-home sleep assessment, LAs were generally wider for both radar and actigraphy. Although comparable to actigraphy, LAs were slightly wider for radar in this setting. Finally, radar and actigraphy were inclined to overestimate SOL and underestimate awakenings.

Instead of the concrete outcome values, the novelty in paper III may lie more in the approach to implementing a new sensor technology. In this study, we have taken what may be described as a modular approach. A technological measuring device may be split into three modules: A basic hardware layer comprising the sensors, a pre-processing layer usually involving analog to digital conversion and some noise reduction filtering, and lastly, an

interpretation layer that includes software for scoring or classification of the pre-processed data. The development of new sleep measuring devices often involves a complete replacement of all three of these modules. This means the replacement of hardware, the development and tuning of pre-processing and filtering, and figuring out ways to meaningfully interpret the data in terms of sleep and waking. Instead, we wanted to see if it was possible to exchange only the hardware (and some signal processing) and utilize the already rigorously tested interpretation algorithms from actigraphy. The advantages to this approach are several. First, the algorithms used for sleep-wake classification in actigraphy are well known and relatively easy to comprehend to the researcher, leading to more transparency in how results are generated compared with, e.g., more complex artificial intelligence-based approaches. Second, this also facilitates understandable comparisons between new and existing hardware and increases the communicability of the results. Finally, as the pace of development of new hardware is high, the modular approach we describe in paper III means it can be reused as new hardware is developed. In sum, we believe this represents a useful way of approaching new technology that combines the aggregated knowledge from the field of actigraphy with the advantages of a new sensor modality.

4.2.1   Radar similar to actigraphy

In the healthy participant dataset, both radar and actigraphy demonstrated similar high sensitivities and accuracies, paired with strikingly high specificities and Cohen's kappa values. Although both radar and actigraphy showed good or excellent agreement with PSG, both were inclined to overestimate SOL and underestimate NW. This is a common problem with non-PSG, movement-based sleep monitoring, as the difference between resting wake and sleep is minuscule in terms of physical body movement. This results in the typical findings of high accuracy and sensitivity for sleep but low specificity. In healthy subjects, the odds of detecting true sleep (sensitivity) by chance are relatively high, as most epochs through the night consist of sleep. The algorithms are weighted to classify the absence of movement to indicate sleep. However, the identification of wakefulness (specificity) among the epochs with no or little movement is difficult and relies on finding suitable classification thresholds that can separate the two. With increasing levels of sleep problems, the specificity often declines, as patients with, e.g., insomnia tend to have long nightly awakenings with no or little movement as they try to fall asleep. Others that have investigated radar-based sleep assessment performance in healthy subjects have found similar results. O'Hare and colleagues

found that radar and actigraphy essentially displayed equivalent accuracy (0.81-0.82) and Cohens's kappa (0.51-0.52) and that both were inclined to overestimate TST and underestimate WASO (O'Hare et al., 2015). In contrast to our finding, they reported that both radar and actigraphy underestimated SOL. In summary, radar seems to have similar agreements with PSG as actigraphy and currently shares some of the same limitations.

In the dataset comprised of patients referred for ambulatory sleep assessment, radar and actigraphy both showed excellent similarity against PSG for WASO, TST, and SE. There was, however, a slight difference favoring actigraphy for specificity, Cohen's kappa, TST and WASO. In terms of the bias found for SOL and NW with healthy subjects, this was not identifiable in the patient dataset. However, the LAs were much wider around the estimates in this dataset, indicating more uncertainty. The task of classifying sleep in heterogenous patients with disordered sleep compared with healthy sleepers is more complex (Kushida et al., 2001), as mentioned above. As such, the findings of lower specificity in both actigraphy and radar were expected. Nevertheless, the numbers reported here are still within what may be considered a normal range of performance compared with previous studies comparing actigraphy with PSG (de Souza et al., 2003; Haghayegh et al., 2019; Kripke et al., 2010; Kushida et al., 2001; Paquet et al., 2007; Van De Water et al., 2011).

The slightly lower specificity in the patient dataset may be related to the ambulatory at-home setting in which these assessments were undertaken. Patients were instructed on setting up the nightstand radar (i.e., how close to the bed, not sleeping with a partner, etc.). Nevertheless, variations in these factors may influence the quality of the data, highlighting a need for clear instructions to patients. Further, in the patient dataset, the probability of Obstructive Sleep Apnea (OSA) events was likely higher. Given that the radar classification algorithm also utilized respiration frequency to indicate sleep or waking, this may have caused more confusion than clarity regarding sleep-wake. It may also be that the classification threshold (that, for the purposes of this study, was a simple 0.5 cutoff) could be tuned to increase specificity in this population. Although specificity was slightly higher in actigraphy for the patient dataset, the biases when comparing the modalities with each other produced relative biases similar to or smaller than what other studies have reported when comparing actigraphy algorithms with each other (Haghayegh et al., 2019; Paquet et al., 2007).
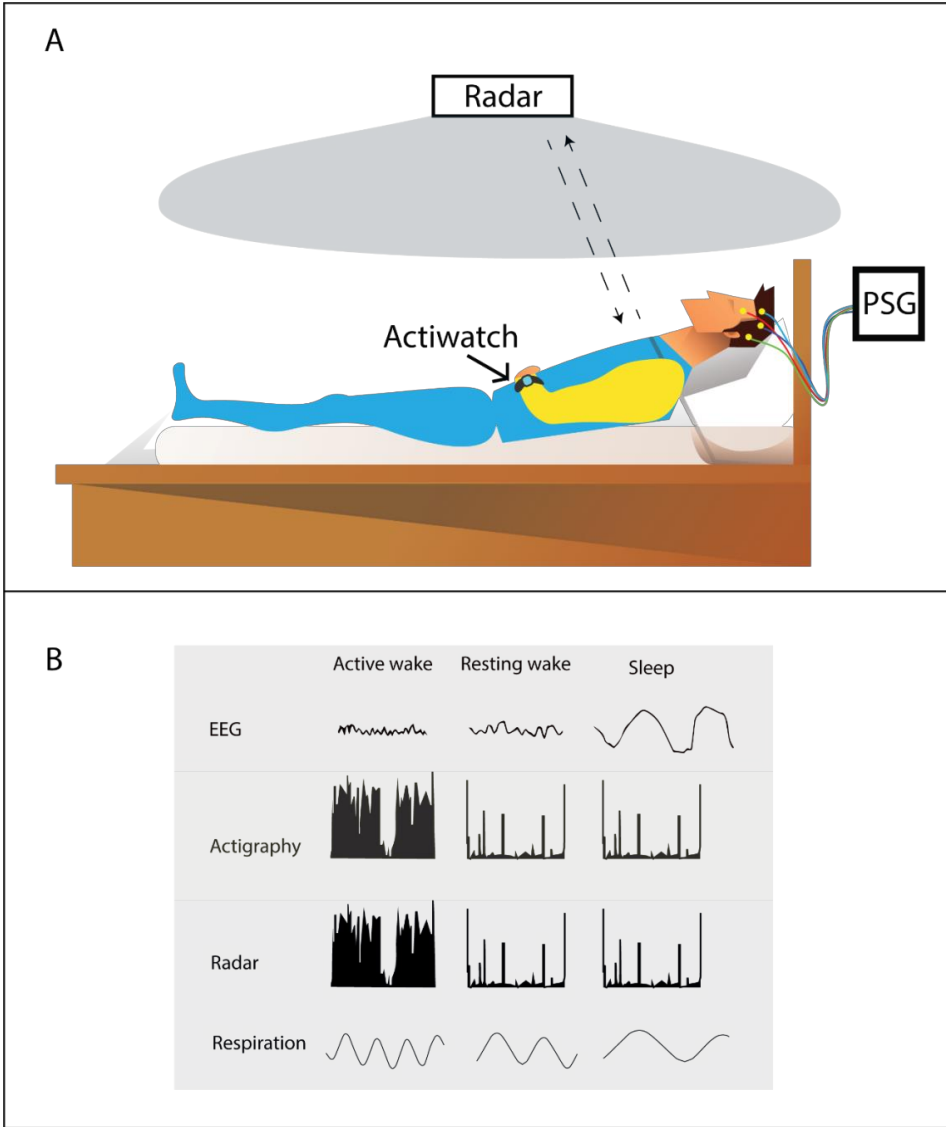
Figure 6. *A: Demonstration of the three sensor modalities as fitted in the BDLE trial of DS1. B: Representation of what the different sensors are able "see". The PSG is the gold-standard benchmark utilizing EEG, actigraphy detects differences in physical activity levels, whereas the radar data is richer compared with actigraphy and already contains additional information on respiration rate, with possibilities of discerning more information in the future as software improves further.*

4.3 Digital versus face-to-face CBT for insomnia

In papers I to III, we focused primarily on testing the efficacy of technological innovations in healthy participants. The technologies in these papers are at an early stage of testing, and the focus was on evaluating the effectiveness in healthy volunteers before the technology can be tested in clinical trials. In paper IV, however, we are focusing on a third technology that has already been thoroughly tested in people with sleep disturbances: digital versions of CBT for insomnia. Although many studies demonstrate the effectiveness of treating insomnia with dCBT-I, few directly compare the digital and FtF treatment modalities. The few studies that do perform direct comparisons primarily recruit participants through media and not in a naturalistic healthcare setting such as an outpatient sleep clinic. In paper IV, we sought to test if dCBT-I can be considered non-inferior to standard FtF CBT-I in such a setting, as this may have implications for the design of the services for patients that seek treatment for their insomnia in the mental health care system.

4.3.1.1 *Non-inferiority was not established for dCBT-I, and FtF was superior*

Previous studies have found that dCBT-I has similar effect sizes to FtF CBT-I (Gao et al., 2022; Luik et al., 2019; Ritterband et al., 2017; van Straten et al., 2018; Vedaa et al., 2020; Zachariae et al., 2016). In paper IV, our findings did not support the hypothesis that dCBT-I was non-inferior to FtF CBT-I. Instead, the data indicated inconclusiveness regarding non-inferiority at the primary endpoint of six-month follow-up. Moreover, post-treatment we found dCBT-I to be inferior to FtF CBT-I. These findings contrast with what has been found for digital versions of CBT in other fields, where effectiveness has been similar to FtF CBT (Andersson et al., 2014; Carlbring et al., 2018; Luo et al., 2020). This may be partly explained by differences in sampling, i.e., recruiting from a hospital outpatient clinic vs. convenience sampling may influence outcomes. Notably, there may also be significant differences between the fully-automated approach used in our study and therapist-guided digital therapy versions.

Our findings align with the findings reported in a superiority trial that compared therapist-guided dCBT-I to FtF CBT-I in a self-referred convenience sample, which found FtF treatment superior (Lancee et al., 2016). Our analyses also show that FtF CBT-I is superior to dCBT-I both post-treatment and six-month follow-up. Another study in active military personnel reported a clear trend toward FtF CBT-I being superior to dCBT-I (Taylor et al.,

2017). However, a non-inferiority trial of dCBT-I vs. group FtF CBT-I did find dCBT-I to be non-inferior to the group format (Blom et al., 2015). Interestingly, individual FtF CBT-I has also been found superior to FtF group CBT-I (Yamadera et al., 2013), which may indicate that in settings where group CBT-I is the available treatment, therapist-guided dCBT-I may be an equally effective alternative. Although group CBT-I may be more cost-effective compared with individual CBT-I, therapist-guided dCBT-I may outperform the group-format in this regard.

We did not find any differences in symptom severity for secondary outcomes except for variables apart from lower sleep medication use and less dysfunctional beliefs about sleep in FtF CBT-I at the six-months follow-up. This expands our previous findings of reduced sleep medication use after dCBT-I compared with patient education in a community sample (Vedaa et al., 2020). The automated dCBT-I does not address sleep medication use, whereas tapering medication is often a focus in FtF CBT-I and may thus explain this difference. As sleep medication use may serve to cope with sleep-related anxiety, addressing this may be part of lasting remission from insomnia, potentially explaining some of the higher efficacy found for the FtF modality. Similarly, FtF therapy may be able to address the individual's specific dysfunctional beliefs about sleep more effectively. Changes in dysfunctional beliefs have been found to mediate the effect of CBT-I in both the digital (Lancee et al., 2015) and FtF modality (Parsons et al., 2021), and may thus be one factor explaining the superiority of FtF in our trial.

4.3.1.2 *Adherence in both modalities*

Previous studies on dCBT-I have typically reported relatively high attrition rates (Espie et al., 2019; Ritterband et al., 2017; Vedaa et al., 2020). Fully automated dCBT-I may be especially vulnerable to attrition, given the lack of support and encouragement during treatment. In paper IV, we found that attrition rates, although lower than expected compared with the aforementioned studies, were higher in the digital compared with FtF treatment. Further, we found that 9% of the individuals invited to the trial were unwilling to participate, as they only accepted FtF treatment for their insomnia. Moreover, out of those starting treatment, only 63% completed all modules of the dCBT-I program, whereas 98% of participants in the FtF group attended all treatment sessions. Nevertheless, per-protocol analyses did reach the same conclusions as the ITT, suggesting the superiority of FtF therapy was not solely due to lower

attrition rates. Nevertheless, it does seem like a weakness of fully-automated dCBT-I is keeping motivation high enough to complete the treatment.

Although the efficacy of dCBT-I has been established (Luik et al., 2019), new digital variants of CBT-I are continuously being developed. These could potentially increase user engagement and reduce dropout compared with current programs. An interesting example demonstrated that a semi-automatic algorithm could be used to identify patients at risk for treatment failure early in the treatment and that the outcome of these patients was better if they received adaptations and more contact with a therapist (Forsell et al., 2019). In that study, participants at risk for poor outcomes had higher initial levels of insomnia symptoms. A recent study we performed in a community sample indicated that self-reported chronotype at baseline moderated the treatment outcome for dCBT-I (Faaland et al., 2022). More work is needed to identify factors that predict poor outcomes of dCBT-I. This information may be used to identify patients needing therapist guidance or being switched to FtF CBT-I altogether.

The identification of risk factors for poor outcomes of dCBT-I may be especially relevant in the case of patients seeking help with insomnia in the mental healthcare system, given our findings of the superiority of FtF. As mentioned, there may be systematic differences in patient characteristics if self-referred community samples are compared with patients referred to mental healthcare. We have previously performed a study using the same dCBT-I version as in paper IV, but in a community sample (Vedaa et al., 2020), allowing for some comparisons of the participant characteristics. Notably, in the clinical sample of paper IV, rates of sick leave or disability pension were higher, as well as baseline levels of fatigue, comorbidities, and sleep medication use. However, the levels of insomnia severity were similar across samples. Thus, it may not be insomnia severity but rather the complexity of symptoms and level of functioning that could inform the decision to provide patients with digital or face-to-face treatment options.

4.4   Methodological discussion

4.4.1   Study designs

4.4.1.1   The BDLE trial and the randomized cross-over design

For *the BDLE trial,* we utilized the randomized cross-over trial design of the AB/BA type. A strength of the randomized cross-over design is the within-subject approach to the research question (Li et al., 2015; Wellek & Blettner, 2012). In *the BDLE trial,* the within-participant variance was expected to be smaller than the between-participant variance for the main outcomes. Given that all participants undergo both the experimental and the control intervention, each participant can serve as their own control, which results in lower variance in the statistical modeling. This increases the statistical power to detect differences even with the small sample size of 12 individuals in this trial (see the section on sample size below). The within-subject design also limits the possibilities of confounding factors due to participants in the intervention and control group by chance differing in some regard (e.g., habitual bedtime). This is a strength when assessing the non-visual effects of light due to the observations of significant individual differences in both circadian rhythms and acute effects of light exposure (Phillips et al., 2019; Santhi et al., 2012). On the other hand, as participants were exposed to both light conditions, the obvious color difference unblinded participants to condition, which may increase the chances of placebo/expectancy effects. Even without blinding, we found no effects on subjective assessments of sleepiness, which may indicate that expectancy effects were not present to a large extent. Proper blinding of participants would also be troublesome in a parallel-group design, and the comparative strengths of the cross-over design still outweighed the difficulty of blinding.

There are also aspects of the cross-over trial that may complicate the interpretation of results, namely period effects and carry-over effects. First, period effects describe situations in which the outcome of interest changes as a function of time, thus influencing outcome estimates across intervention and control conditions. If period effects are present, they will typically not bias the intervention effect when equal numbers have been allocated to each AB/BA cross-over design sequence, such as in *the BDLE trial* (Dwan et al., 2019). Second, carry-over effects may arise if the effect of the intervention received in period 1 persists into period 2 where the control conditions is received. This leads to the observed effects of the control condition being influenced by the active intervention, either by improving or suppressing the outcome in the control period. Further, this results in the observed effect of the active

intervention to depend on the sequence (see Figure 7) in which intervention and control were received (Dwan et al., 2019; Fleiss et al., 1985; Li et al., 2015; Wellek & Blettner, 2012).
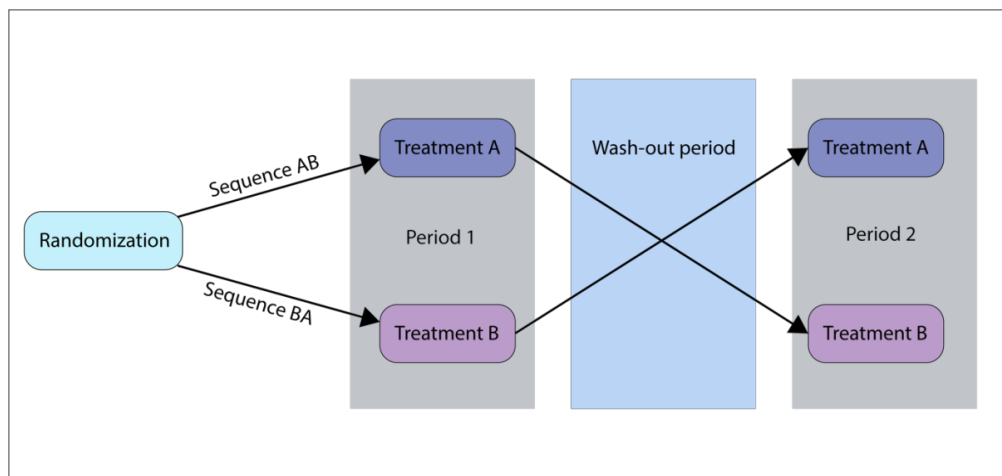
Figure 7. An overview of the AB/BA randomized cross-over design. Treatments A and B denote the experimental and control conditions. Period is the time period of the study in which a specific condition is received. Sequence is the sequence in which experimental and control condition is received.

Previously, it was often recommended to test for carry-over effects before testing for intervention effects, and if carry-over were present, to disregard any observations from period 2 and only test for between group effects in period 1, thus treating it as a parallel-group trial (Grizzle, 1965). However, Freeman (P. R. Freeman, 1989) showed that the carry-over effect and the intervention effect in period 1 are highly correlated, thus increasing the alpha-level for a test of only period 1. This argument has led two-stage testing to be largely abandoned. Instead, to decrease the probability of carry-over effects, an appropriate wash-out period is recommended between the two periods, in which the effect of the active treatment will be given enough time to subside (Dwan et al., 2019; Senn, 2002). In papers I and II, a limitation is that we only utilized a 1-day wash-out period, similar to most other studies of evening light exposure (Chang et al., 2015; Rahman et al., 2017; Stothard et al., 2017; Wright et al., 2013). The reasons for this were practical due to the limited time we had to perform *the BDLE trial* in the final construction phase of the hospital unit. In other circumstances, we would have considered a longer wash-out.

To address the limitation of short washout and facilitate the interpretation of the main effects, we perform some auxiliary tests to check for the presence of strong carry-over effects. In a 2x2 AB/BA design, carry-over effects can be modeled in either one of two ways: (1) by including a term for the sequence in which conditions were received, or (2) by including the period by condition interaction term (the sequence-term will be equal to the period by condition interaction term divided by two). In papers I and II, we chose to model this effect using the period-by-condition interaction and found a significant effect of order on DLMO. The period by condition interaction is not a direct reflection of carry-over and must be interpreted in the context of other outcomes in the trial. In period 1, participants in the intervention and control conditions had similar phase-shifts of DLMO, whereas there was a large difference in period 2. Although participants were asked to maintain stable sleep-wake schedules in the pre-randomization period, the transition to period 1 was accompanied by significant advancement of bedtimes and rise times (by 0:35h and 1:26h, respectively), and an increase in sleep regularity, but such changes were not observed between period 1 and 2 in the hospital unit. The phase-shift of DLMO during period 1 also correlated with the advancement in rise-times from the pre-randomization period to period 1, something that again was not observed from period 1 to period 2. Advancing the sleep period and morning light exposure is known to have potent phase-advancing effects on the timing of DLMO, and this may have led to a period effect on DLMO in both LEs in period 1. In light of this, we interpret these findings as a period effect masking potential effects of BDLE in the first period rather than a carry-over effect of BDLE on the circadian phase persisting into period 2. Nevertheless, it may also be a combination of a carry-over and a period effect, as the standard LE was not phase-delayed to baseline levels by the end of period 2, but remained advanced. Potential carryover effects should be addressed by the use of longer washout periods and alternative study designs, such as larger parallel group designs or stepped-wedge designs in future trials.

4.4.1.2 The DCBT-I trial – The non-inferiority design

Most randomized controlled trials (RCTs) are designed to test whether two conditions are different in terms of effectiveness with such a certainty that it justifies considering one superior to the other. Superior is the technical term for the result where the 95% CI of the estimated between-group difference lies wholly outside of 0, and trials designed to test the difference between two conditions are therefore called 'superiority trials'. In *the dCBT-I trial* the added benefit of digital therapy is not necessarily related to treatment efficacy but rather

to other factors such as cost, scalability, and allowing for wider treatment dissemination in an environment with too few therapists providing the gold-standard FtF CBT-I. Therefore, the question of interest is not symmetric (Piaggio et al., 2012). The non-inferiority design tests whether the new treatment is not worse than the reference treatment by some acceptable margin. The strength of this approach lies in its focus on, and assessment of, acceptable loss of efficacy weighed against other advantages.

For a non-inferiority margin in paper IV, we assumed a-priori that a difference in effect size between the treatments of 0.5 (Cohen's d) or more would be clinically meaningful. In trials of CBT-I this roughly corresponds to a 2-point difference on the ISI. We base this partly on evidence from a meta-analysis of studies that have computed minimally important differences (MIDs) in health-related quality-of-life instruments for patients with chronic diseases, which found that these differences most often converge to an effect size of approximately 0.5 (Cohen's d) (Norman et al., 2003). A challenge with non-inferiority trials, ours included, is the lack of consensus regarding the non-inferiority margins, with studies describing different justifications for the selection of margins. In the context of CBT-I, other non-inferiority trials have used both a 4-point (Blom et al., 2015; Garland et al., 2014), a 3-point (Gehrman et al., 2021), and a 1.67 point (Gehrman et al., 2020) non-inferiority margin. This leads to a seemingly arbitrary margin, with a wider margin increasing the chances of concluding with non-inferiority. If the 2-point ISI non-inferiority margin used in our trial were used for the three other non-inferiority trials, none of these would have concluded with non-inferiority (see Figure 8). At the moment, this issue requires the reader to be familiar with relevant outcome scales to interpret findings.
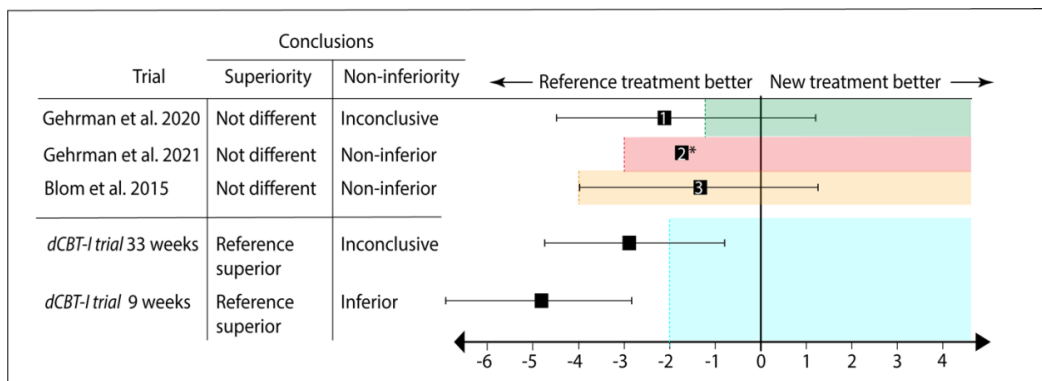
Figure 8: The figure shows the conclusions of previous non-inferiority trials comparing digital versions of CBT-I with FtF CBT-I. The colored boxes represent the non-inferiority margins on the ISI chosen for the trials. Point estimates represent estimated between-group differences on the ISI. Whiskers represent 95% CIs.

*Gehrman et al. 2021 did not report the CI, but described in text that did not include the non-inferiority margin and indicated non-inferiority.

A further source of confusion is that non-inferiority trials increase the number of possible conclusions regarding the new treatment: superior, non-inferior, inconclusive, or inferior (see Figure 9). Moreover, in paper IV, analysis at the primary end-point shows inconclusiveness regarding non-inferiority. At the same time, the significant between-group effect indicates superiority of the FtF reference. The superiority conclusions can be drawn from the same analysis already based on the ITT group. As such, the analysis corresponds to best-practice for superiority analyses, and there is no argument for statistical correction due to multiple comparisons (Committee for Proprietary Medicinal Products (CPMP), 2001). However, these results that combine inconclusiveness regarding non-inferiority with the superiority of the reference are a handful in terms of effectively communicating results. Although an extension of the CONSORT guidelines has been released for non-inferiority trials (Piaggio et al., 2012), not all researchers adhere to the guidelines, leading to further confusion for readers. Combined with the relatively lower familiarity of readers with the non-inferiority design, this may cause significant misunderstanding.
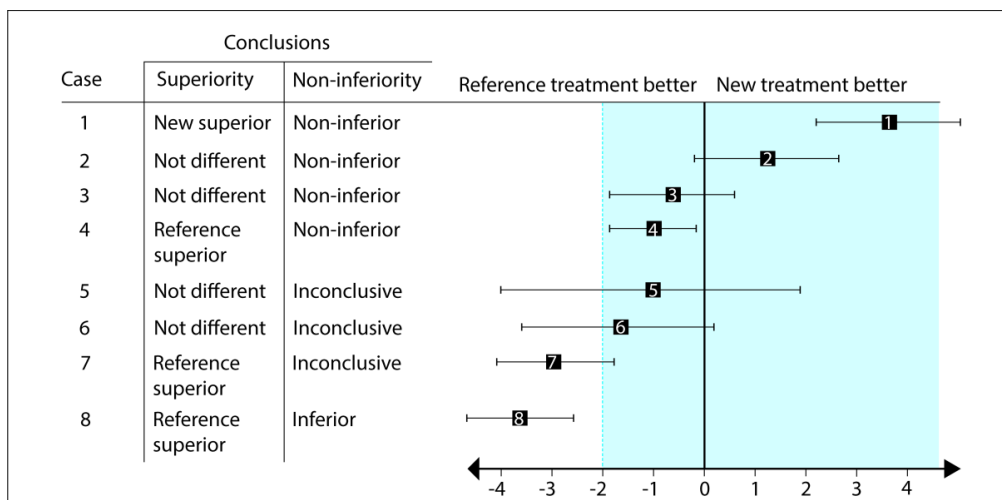
Figure 9: *The figure illustrates the numerous potential conclusions that can be derived from a non-inferiority trial. Cases 1-8 are fictive trials that represent possible between-group outcomes and their associated conclusions in the non-inferiority and superiority paradigms. The blue dotted line indicates the non-inferiority margin. Point estimates represent estimated between-group differences. Whiskers represent 95% CIs.*

In non-inferiority trials, the burden of proof is also shifted compared with superiority trials. This means the usually conservative ITT analysis may be anti-conservative, increasing the chances of declaring non-inferiority (Hills, 2017). In the ITT group, the true effect of the treatment may be diluted by other factors, thus reducing the size of the between-group difference. Per-protocol analyses, usually anti-conservative, will, in the case of non-inferiority trials, assist in making more conservative conclusions regarding non-inferiority (Hills, 2017). This does not necessarily mean substituting ITT with PP analyses, as the problems with disregarding attrition still apply, but should instead lead to ensuring the trials are appropriately designed and executed (Schumi & Wittes, 2011). This may be particularly important with dCBT-I, which, as mentioned, often suffers from high attrition rates (Espie et al., 2019; Ritterband et al., 2017; Vedaa et al., 2020). The European Medicines Agency (EMA) draws particular attention to this issue in their recommendations by pointing out that both the ITT and PP analysis have equal importance in non-inferiority trials, and they should reach the same conclusions if results are to be trusted (Committee for Proprietary Medicinal Products (CPMP), 2001). In paper IV, we used ITT for primary analyses, but PP analyses

were performed with similar conclusions. As with other aspects of trial quality, the shifted burden of proof requires that the FtF CBT-I is performed with high quality, as the probability of reaching a non-inferiority conclusion for dCBT-I increases with lower effectiveness in the comparator. Although the shortcomings of the non-inferiority design need to be remedied, the shift in the burden of proof remains a strength that ensures proper attention is given to the advantages of dCBT-I outside of treatment effect sizes and prompts a discussion on the acceptable loss of efficacy in different treatment settings.

4.4.2   Sample size and characteristics

4.4.2.1   Sample sizes

The sample size in *the BDLE trial* was small, with 12 participants. This was the result of several factors. First, we performed a-priori calculations on statistical power using previous findings from studies investigating the effects of evening light exposure on DLMO phase shifts (Sasseville et al., 2006; Solheim et al., 2018; Stothard et al., 2017; van der Lely et al., 2015). We estimated that eight participants would give us statistical power of 90% to detect a difference of 1.5 hours in DLMO. By recruiting 12 participants, we planned for a minimum of such power in the case of dropouts during the study. As the study was powered with DLMO phase-shift in mind, the 12 participants may not have been sufficient to detect differences in the other outcomes reported in papers I and II. This is particularly the case for the order analyses performed in both papers I and II, which halves the sample size. It should also be noted that paper II was not planned a priori. It was carried out due to the striking differences in REM sleep duration found in paper I. Paper II should therefore be regarded as post-hoc and exploratory by nature. Nevertheless, in paper II, the validity of the findings is strengthened by the fact that increased consolidation was detectable on two different outcome measures, both REM sleep episode interruptions and REM sleep arousals.

In *the dCBT-I trial* (Paper IV), similar a-priori power calculations were performed based on previous trials of dCBT-I (Blom et al., 2015; Garland et al., 2014; Lancee et al., 2016; Zachariae et al., 2016). We assumed a baseline SD of 4.0, which would give a moderate effect size (Cohen's d = 0.5) with a 2-point difference on the ISI. As discussed in relation to the non-inferiority design, we took such a moderate effect size to be a good estimate of a minimally important difference in a clinical setting, thus setting the non-inferiority margin at 2 points on the ISI. We then estimated that including 100 participants would give at least 80% statistical power to detect non-inferiority with such a non-inferiority margin at an alpha level

of 0.05. In *the dCBT-I trial* (paper IV) we did not adjust the target sample size to account for potential missing data (e.g., by adding 20% to the target sample size as in *the BDLE trial*). Missing data was minimal, however, with only 4 and 8 participant scores missing on the primary outcome at six-month follow-up. Nevertheless, had we included 120 participants, the increased power may have avoided "inconclusiveness" with regards to non-inferiority at 33 weeks follow-up. As a general side note, we base the above power calculations on t-tests. Whereas in the analyses of both *the BDLE trial* and *the dCBT-I trial*, linear mixed models were used, which have higher power than t-tests.

4.4.2.2 Sample characteristics

The participants in *the BDLE trial* were healthy subjects with no reported sleep difficulties. This limits the generalizability of the findings outside the healthy population. In particular, it is difficult to draw conclusions regarding potential benefits for patients admitted to the acute psychiatric ward based on this trial with healthy participants. See future directions for a further discussion on this.

In paper III, the radar was most precise for the healthy participant dataset (DS1) from *the BDLE trial*. However, the classification models were trained on participants from the same trial by randomly splitting into a training and test dataset, meaning both the participant characteristics and the physical assessment situation was identical between training and testing. This is highly beneficial for precision outcomes and may partly explain why the radar performs poorer in the ambulatory patient dataset. Thus, the results obtained for the radar in the healthy dataset might not be generalizable to other samples with more disordered sleep or other less controlled settings.

The context of *the dCBT-I trial* was different from most other trials where digital and FtF CBT-I have been compared as it was performed in the context of a secondary care sleep clinic. The patients receiving treatment at the clinic have been referred by their GP or from other secondary healthcare providers, such as the department of neurology, clinical neurophysiology, and other mental healthcare clinics. Although the severity of insomnia might not be substantially different (as discussed earlier), compared with participants recruited from the general population, patients recruited from a sleep clinic will typically have higher complexity in terms of symptoms, with a higher number of comorbid disorders and symptoms that may result in poorer daily functioning. Thus, a strength of *the dCBT-I trial* is

the use of a sample from a sleep clinic, which increases the validity and generalizability of the results for patients that seek insomnia treatment from the mental healthcare system. On the other hand, this setting also limits, to some extent, the generalizability to self-referred individuals with insomnia outside of the secondary mental healthcare setting, where it may be that dCBT-I is non-inferior to FtF. Finally, the dCBT-I version in our trial was fully automated, and it may be that a therapist-guided version with more therapist interaction would have performed better also in this sample.

4.4.3 Measurements

A limitation of *the BDLE trial* was that the DLMO assessments were not performed according to a constant routine procedure. It is known that posture, particularly standing, may increase the immediate concentration of melatonin, which is later reversed after 10 minutes of sitting (Benloucif et al., 2008; Deacon & Arendt, 1994). Participants were sitting most of the time during dim-light assessments but were allowed to stand, walk around the room, and go to the bathroom if necessary. This might thus constitute a source of noise in the data but is expected to influence assessments equally between conditions. Moreover, light exposure was never higher than 3lux, well below the acceptable thresholds (Benloucif et al., 2008), and thus accommodates recent discoveries that melatonin suppression occurs at lower levels than previously thought (Prayag et al., 2019).

Furthermore, in *the BDLE trial,* we did not have baseline measures with PSG. This means sleep variables could not be compared to the individual's sleep in the home setting. We did find that bedtimes and wake-up times changed from pre-randomization to period 1 in the study, which might have led to some changes in PSG in period 1. Moreover, PSG assessments tend to display a first-night effect, where many participants find it more difficult to fall asleep and display poorer sleep quality the first due to the cumbersome nature of the equipment (Newell, 2012). We did have two nights of PSG assessments in each condition. However, in planning the analyses, we decided that the increased power gained by utilizing both nights in each condition outweighed our concerns regarding the first-night effect.

Prior light history has also been found to moderate the effect of evening light exposure. Exposure to <200 lux during daytime (similar to an indoor light environment) for one week has been found to increase melatonin suppression when exposed to evening bright light compared with after a week of higher intensity daytime light exposure (Hébert et al., 2002).

Similarly, if exposed to dim light (<0.5 lux) during daytime for 3 days, a nighttime light exposure results in more melatonin suppression compared with after 3 days of daytime 200 lux light exposure (Smith et al., 2004). Similar moderating effects of prior light exposure have also been found for circadian phase shifts (Chang et al., 2011) and alerting effects (Chang et al., 2013) of evening or night light exposure. In *the BDLE trial*, we attempted to gather information on daytime light exposure, but this data was unavailable for analysis due to hardware failure on several of these sensors. Thus, we did not control for daytime light exposure in the analyses. If participants were exposed to different daytime light intensities, this could confound the results reported for circadian phase-shifts and melatonin suppression. The randomized design, however, reduces the chances of prior light history explaining the between-condition effects.

4.5   Future directions

4.5.1   *Upcoming challenges in blue-depleted lighting research*

The early studies on melatonin suppression spectral sensitivity found peaks around 460nm (Brainard, Hanifin, Rollag, et al., 2001; Thapan et al., 2001), indicating that there may be both a melanopsin and an S-cone contribution. However, a recent reanalysis of some of this data indicates melanopsin best explains the melatonin suppression (Prayag et al., 2019). Moreover, Spitschan and colleagues compared light spectra high or low in S-cone stimulation and found S-cone stimuli not to affect melatonin suppression (Spitschan et al., 2019). In support of this are findings from individuals with color-blindness (Ruberg et al., 1996) or complete blindness (Hull et al., 2018) showing near normal melatonin suppression in response to evening light. Although increasing evidence point to a dominance of melanopsin in the spectral sensitivity of melatonin suppression, there is a need for further studies disentangling the contributions of other photoreceptors (T. M. Brown et al., 2022; Vetter et al., 2022). Whether there is a contribution of rods will not significantly alter the spectral sensitivity of non-visual effects, given that rods and ipRGCs, to a large extent, have overlapping spectral sensitivity profiles, but further investigation using multi-photoreceptor models may discover slightly more accurate spectral sensitivities (T. M. Brown et al., 2022). For practical purposes, it seems like the use of the melanopsin action spectrum is a relatively precise way of predicting biological responses to light stimuli.

This increasing consensus in the scientific community (T. M. Brown, 2020; CIE Central Bureau, 2019; Lucas et al., 2014; Nowozin et al., 2017; Prayag et al., 2019; Spitschan et al.,

2019) has led the CIE (International Commission on Illumination) to release a standard to measure and quantify ipRCG-mediated responses to light (CIE Central Bureau, 2018). This internationally accepted standard, the melanopic Equivalent Daylight Illuminance (EDI), is an important step in comparing results across studies more effectively. The CIE also recently released a position statement that goes a long way to recommend the tuning of indoor light to achieve lower melanopic EDI during the evening and night (CIE Central Bureau, 2019). There is currently a high interest from academic scientists, industry, and healthcare stakeholders, in translational research on tunable indoor light environments. We are aware of several hospitals and nursing homes that have implemented versions of tunable lighting (Engwall et al., 2017, 2017; Figueiro et al., 2019; Giménez et al., 2017; Okkels et al., 2019; Schledermann et al., 2021; Vásquez-Ruiz et al., 2014), which most frequently takes the form of "cycled lighting" that combine some level of blue-depletion in the evening with higher melanopic EDI during daytime. With implementation moving at a rapid pace, it is crucial that rigorous trials evaluate the effectiveness of these interventions in influencing non-visual effects of light as discussed in relation to papers I and II, but also investigate any issues regarding implementation for both patients and staff.

The literature on non-visual effects of light exposure has demonstrated large interindividual differences. A recent study found substantial variations in individuals' sensitivity to the melatonin suppression effects by evening light exposure (Phillips et al., 2019). Furthermore, the non-image forming effects of light exposure are also known to be influenced by prior light history (Chang et al., 2011; Hébert et al., 2002), age (Daneault et al., 2016), differences in light sensitivity, circadian rhythm disorders (Aoki et al., 2001; Watson et al., 2018), medical illnesses, mental disorders (Crasson et al., 2004; Hasler et al., 2010; McGlashan et al., 2019), and medication status (McGlashan et al., 2018). The circadian system also receives input from other factors, such as sleep-wake and feeding-fasting cycles (Skene et al., 2018). There are also many acute and indirect non-image-forming effects, with new effects still being uncovered, and there is much work to be done to understand the interplay between these functions. All these factors should be addressed in more detail to understand for whom and given what circumstances evening BDLEs may be beneficial.

A challenge regarding the implementation of evening blue-depleted lighting is making sure that light levels allow the residents or workers in such spaces visually comfortable and safe lighting while at the same time reducing melanopic EDI (Zandi et al., 2021). The depletion of

blue-light frequencies is typically done by removing most or all short-wavelength frequencies, resulting in orange-colored lighting with low Correlated Color Temperature (CCT) and low melanopic EDI. As the human visual system is adapted to use information from the full range of the photopic system, the removal of a large portion of the spectrum is reflected in the lower color discrimination ability that we found in paper I, which has also been reported by others (Rahman et al., 2017). This change in color perception may be an obstacle to widespread implementation, as some will experience reduced visual comfort or enjoyment of media and screens in the evenings. More importantly, this may be of more consequence in some working environments, such as hospitals, where the reduced visual ability of staff may lead to severe consequences for inpatients. We recently performed a small study with nurse staff at the new acute psychiatric unit. We did not find any significant changes in nurses' sleep or functioning except a slight increase in sleepiness while working in the BDLE (Kjørstad et al., 2022). One qualitative study, with the staff in an elderly care-home two years after the implementation of "circadian lighting" reported high staff satisfaction with the new lighting, but that the ability to adjust the lighting to maintain visibility in all situations was perceived as important (Schledermann et al., 2021). These are small studies, however, and further investigations in larger samples need to be undertaken to uncover any negative or positive effects BDLE may have on staff.

Recent advances in lighting technology may alleviate the reduced color-discrimination ability. Several groups have now published findings with new LED light sources that blend several light spectra to combine wavelengths at the long end with wavelengths at the short end of the melanopsin action spectrum. The result is a metameric light spectrum that falls mostly outside the action spectrum of melanopsin-based light reception while still activating the S-cones to some extent, in addition to the M and L-cones, thus being perceived as white light. A recent study by Souman and colleagues demonstrated a 50% reduction in melatonin suppression comparing a metameric light with a standard light source (Souman, Borra, et al., 2018), which is comparable in size to our findings in BDLE. Allen and colleagues created metameric visual display units and found less melatonin suppression and higher subjective sleepiness when participants viewed movies on such displays compared with standard display units (Allen et al., 2018). Recent work by Zandi and colleagues also demonstrates that by using 6, 8, or 11-channel LED luminaires, metameric spectra could be generated in specific chromaticity regions that theoretically could be highly efficacious in reducing melatonin suppression (Zandi et al., 2021). They also show that metameric lighting can be generated with high color

fidelity, meaning less distortion of the color appearance of objects. The potential of metameric lighting should be addressed in further laboratory studies to evaluate its effects on other non-image-forming effects, such as circadian phase-shifting, alertness, and mood. If effectiveness is confirmed, metamers may solve some caveats related to the implementation of blue-depleted lighting.

4.5.2 *Potential application of BDLEs and radars in hospitals*

In *the BDLE trial*, the beneficial effects of evening BDLE on sleep and circadian rhythms are demonstrated in a group of young, healthy adults characterized by stable circadian rhythms profiles and normal sleep. The benefits of BDLE may be even more pronounced in populations where sleep is more disturbed or in settings where light levels impede good sleep. These two factors converge in the case of hospitals, where patients tend to have a high prevalence of sleep disturbances (Elliott et al., 2013; Engwall et al., 2015; Langsrud et al., 2016; Schennach et al., 2019; Wulff et al., 2010), and the where the light levels tend to be high (Bani Younis et al., 2021). In psychiatry, sleep disturbance is almost a trans-diagnostic symptom of mental illness, and the generation of sleep and stabilization of sleep-wake patterns is considered to be an important part of the treatment. These problems are currently treated with a combination of sleep medication and stabilization of bedtimes and rise times (Rehman et al., 2017). The line of research demonstrating the efficacy of "dark therapy", or "virtual dark therapy" using blue-blocking glasses, on both sleep and symptoms of mania in acutely admitted patients with bipolar disorder (Barbini et al., 2005; Gottlieb et al., 2019; Henriksen et al., 2016, 2020) suggests reducing evening light exposure may be a way to improve treatment of inpatients in psychiatry. It is currently unknown whether the mechanisms for such improvements are mediated via circadian rhythms, sleep, reduced alertness, or other factors. In these data from healthy participants, we find that a "virtual darkness" can be replicated in a hospital light environment by generating a BDLE. This has the advantage of remotely controlling the timing of exposure to light, which removes the burden for patients of having to comply with treatment schedules.

There is widespread interest in the implementation of programmable lighting systems in hospitals. Such chronotherapeutic interventions depend upon the ability to control all light exposure, both ambient and incident, and a hospital environment may be well suited to exert such rigorous control. However, implementation of tunable lighting has moved at a pace

where the clinical trials testing efficacy are lagging behind. Some studies using cycled lighting have found indications of increased rhythmicity (West et al., 2019) and actigraphy-assessed sleep duration (Giménez et al., 2017). Other studies indicate that cycled lighting has no effect on critical care inpatients (Engwall et al., 2017) or inpatients with mood disorders (Okkels et al., 2019). One recent study found cycled lighting increased TST, SE, and WASO paired with phase-advancement of rest-activity rhythms as assessed with actigraphy in patients admitted with depression (Canazei et al., 2022). In general, the clinical trials to date have been small or have suffered from limited control of incident light and other light sources. With plans of implementation continuing in many hospitals, there continues to be a need for larger trials in diverse clinical populations (Drews et al., 2020; Scott, Langsrud, et al., 2019). However, particular care needs to be taken in the design of such studies in hospital settings, as numerous factors influence a patient's recovery from illness. The added benefit of BDLE on these healing processes might be small and challenging to tease apart from other processes. There may be large variations in effects on health for individuals depending on their age and medical status (Boyce, 2022). Trials should be performed across diverse populations to investigate which patients may or may not benefit. It may also be that effects of BDLE are more pronounced over time, and we would encourage investigations on the clinical usefulness of BDLE in settings where patients are admitted over more extended periods.

In paper III, we test real-time assessment of sleep using the radar. Although real-time models performed slightly worse than non-real-time models, the advantages of real-time models may outweigh such performance differences in the hospital setting. Similarly, although the nightstand radar assessment performed slightly better, the possibility of embedding radar technology in ceilings likely outweighs this slight performance benefit. Currently, due to the invasiveness and resource-demanding nature of PSG, and that actigraphy does not allow for real-time sleep monitoring, effective use of sleep-wake monitoring is not economically or practically feasible in a hospital inpatient setting. This results in the utilization of real-time sleep-wake information being severely limited. In psychiatry, the most common form of sleep-wake monitoring is nurses physically entering patient rooms to intermittently check on patients (e.g., every other hour through the night). In this setting, intermittent observations aim to ensure patients do not engage in self-harming or suicidal behavior. However, it has been argued that such observations do not reduce such risk, as few suicide attempts occur during nighttime – and overwhelmingly occur during observations. Instead, intermittent observations lead to sleep deprivation for most inpatients (Veale, 2019). There is still much

work to be done before radar assessment of sleep is reliable enough to support clinical decision-making in hospitals. E.g., studies comparing radar to nurse observations should be performed. With time, however, the use of real-time monitoring of sleep-wake state could be a replacement for intermittent observations.

Another application of real-time models is the identification of patients with OSA symptoms. A recent review found a higher prevalence of OSA among, e.g., patients with depression or PTSD (Gupta & Simpson, 2015). Moreover, some medications commonly used in inpatient psychiatric care may exacerbate symptoms for patients with underlying OSA (Zolezzi & Heck, 2015). Both Crinion and colleagues (Crinion et al., 2020) and Zaffaroni and colleagues (Zaffaroni et al., 2013) have tested radar as a screening tool for OSA and found it especially useful in confirming more severe cases. This use of radar may provide a way to identify patients with OSA before administering medications that worsen such symptoms and identify patients with harmful side effects in real-time. Further studies are needed to establish the efficacy of radar sleep monitoring in sleep-disordered breathing, but it has potential in terms of improving assessment and, thus, treatment in a psychiatric hospital setting where this traditionally has been challenging.

4.5.3   *Potential of BDLE in the treatment of depression and insomnia*

A long line of research has associated depression with dysregulated REM-sleep (Riemann et al., 2020), which includes longer REM-sleep duration, shorter latency to the first REM-episode, and increased REM-density (Kupfer & Foster, 1972; Riemann et al., 2012, 2020). Increased REM-density has also been found to predict the onset of depression (Rush et al., 1986). Further, antidepressant medications suppress REM sleep (Wilson & Argyropoulos, 2005), and response to antidepressant treatment is associated with the degree of REM-suppression (Riemann et al., 2001), and with higher REM-density before treatment (Lechinger et al., 2021). After treatment, patients with persisting REM abnormalities may also be at a higher risk for relapse (Giles & Roffwarg, 1998; Steiger & Pawlowski, 2019). This was thought to indicate an increased pressure for REM sleep, to the extent that it is maladaptive, and early hopes were for these changes to represent a stable biomarker for depression. However, later findings demonstrated that altered pressure for REM sleep is exhibited in many psychiatric illnesses and is associated with increased comorbidity (Baglioni et al., 2016).

The proposed role of REM sleep microstates in emotion regulation and the frequent REM sleep alterations in mental disorders and insomnia may be part of the mood and cognitive symptoms in these disorders (Simor et al., 2020). Restless REM sleep, defined as the instability of REM sleep and high REM-density, has been found to reduce the overnight resolution of emotional distress (Riemann et al., 2012; Wassing et al., 2016, 2019). Being prevalent in insomnia, restless REM sleep has been proposed to lead to an accumulation of distress inducing a state of hyperarousal that may prompt sleep-onset and maintenance issues in insomnia and further exacerbate restless REM sleep – thus both being a risk factor for and a maintaining factor in insomnia (Van Someren, 2021; Wassing et al., 2016). As previously mentioned, it has further been speculated that this same process constitutes a trans-diagnostic risk factor for both mood and anxiety disorders, as insomnia strongly predicts the onset of both mood and anxiety disorders (Hertenstein et al., 2019), and that the treatment of insomnia may improve levels of depression (Manber et al., 2008). The studies using blue-blocking glasses for depression have so far been inconsistent (Hester et al., 2021), and the trial of cycled lighting for depression found no effect on depressive symptoms (Okkels et al., 2019). Future studies should investigate if increased REM sleep consolidation could mediate any potential effect of blue-blocking lighting in the evening.

In individuals with insomnia, the use of blue-blocking glasses has been reported to improve both objective and subjective sleep outcomes in symptoms (Shechter et al., 2018).
In paper I, we find a slight increase of about 8 minutes in TST with BDLE. This is similar to the objectively measured increase in TST found in a recent meta-analysis of CBT-I (Trauer et al., 2015). Importantly, this is with healthy participants, but it may indicate that adding BDLE to CBT-I could be tested to boost treatment, especially considering potential effects on REM sleep consolidation and arousal. Interestingly, the sleep-restriction component of CBT-I has been found to reduce measures of both subjective pre-sleep arousal and objective cortical arousal during sleep (Maurer et al., 2022), suggesting that reduced cortical arousal during sleep may underlie the effectiveness of sleep restriction. Recent findings also indicate that a subgroup of insomnia patients may have an underlying circadian disruption (Faaland et al., 2022; Flynn-Evans et al., 2017), suggesting BDLE might help these patients specifically. Indeed, one small study has reported some added benefits of adding blue-blocking glasses to CBT-I, particularly on subjective levels of anxiety and hyperarousal (Janků et al., 2020). Future studies should test the effects of BDLE on insomnia in larger samples. It may also be that such effects only exist for subgroups of patients, e.g., patients with circadian

misalignment or patients with sleep-wake state misperception, which may indicate increased arousal and REM sleep fragmentation.

4.5.4 *Implementation of BDLEs in the general population*

A seminal work by Geoffrey Rose greatly impacted the discourse regarding implementation strategies in the field of epidemiology (Rose, 1992). Rose argued that efforts to prevent illness in a societal context primarily take two forms. First, the high-risk strategy intends to identify individuals with symptoms or illness early and aim effective treatments at these individuals, thus reducing the risk for further illness. An example could be identifying patients that have or are developing circadian rhythm disorders with screenings at the primary care physician's office and prescribing tools to create at-home BDLEs (e.g., guidance on lighting, physical blue-blocking filters on screens, etc.). The disadvantages of this strategy are primarily the difficulties of identifying high-risk patients early and that interventions tend to be palliative and temporary (Hunt & Emslie, 2001). Second, the population strategy aims to implement interventions population-wide to achieve an incremental reduction in the risk of illness, e.g., encouraging the general population to adopt the principles of blue-depleted evening lighting at home. This approach acknowledges that illness always develops in a context (McLaren et al., 2010) and attempts to control the determinants of incidence by encouraging radical change to the norms of behavior in society (Hunt & Emslie, 2001; Rose, 1992). The discoveries of the non-visual effects over the last two decades have run in parallel with the increasing use of light-emitting screens (Hale et al., 2018) and a shift towards LED luminaires higher in short-wavelengths (Cain et al., 2020), creating a significant discrepancy between the behavioral norms for evening light consumption and the accrued scientific knowledge. This is a typical situation where there may be a high gain in terms of overall population health and societal disease burden by encouraging the population-wide interventions, such as circadian-friendly lighting technology in the home.

It is known that the general economic burden related to sleep loss is high across many developed countries (Hafner et al., 2016). Recent data indicate that higher light levels in the evenings are associated with an increased risk of a range of mental disorders (Burns et al., 2022). The effect of lighting on circadian rhythms and sleep likely contributes to a large extent to the disease burden in society, including reduced work productivity (Koritala & Çakmaklı, 2018). To further assess the benefits of using a population strategy, there is a need

for large-scale longitudinal studies that investigate the effects of evening light exposure on health outcomes, sick days, productivity, and incidents of accidents (T. M. Brown et al., 2022). For certain groups, such as individuals at high risk for circadian rhythm disorders (e.g., late chronotypes with more than 2 hours of social jetlag) (Meyer et al., 2022), the evidence may already be sufficient to prescribe at-home BDLEs as a preventive intervention to limit circadian rhythm misalignment. The cost of providing such equipment may be recovered by the potential benefits on sleep, health, and productivity (Koritala & Çakmaklı, 2018). However, although there may be both an economic and health gain on a population level, the benefit for the individual with healthy sleep may be minuscule and irrelevant to the individual. This has been referred to as the prevention paradox (Hunt & Emslie, 2001; Rose, 1992) and necessitates careful consideration of any potential risk, as small risks may easily outweigh these small benefits on the individual level.

The risks of adjusting evening lighting levels are considered to be relatively small. In our *BDLE trial*, we did not find indications for unwanted side effects, except for the discussed reductions in color-discrimination ability. However, it may be that advocating for population-wide adoption of evening BDLE may increase focus and attention towards optimizing sleep in many individuals, given that the visibly different lighting provides a constant potential reminder of sleep optimization. Such a development has been seen as a result of the increased use of sleep trackers in the general population, where many anxious individuals seek help for self-diagnosed sleep problems as a consequence of the increased perfectionist focus on sleep that has been termed "orthosomnia" (Baron et al., 2017; Gavriloff et al., 2018). Typically, the literature on tunable lighting seems to lean more towards techno-optimism (Danaher, 2022), focusing on the novelty, potential, and benefits of the new lighting technology and less on harmful effects. Industry is already pushing heavily for the adoption of new lighting products at a population-wide level. Thus, it is my view that the scientific community must take the role of voicing any potential concerns. As there is a risk of increasing negative focus toward sleep, this may outweigh the benefits of implementing BDLE on a population-wide level.

Rather than recommending widespread at-home adoption of BDLEs, it may be more feasible to recommend simpler changes in at-home evening light environments. The reduction of evening light intensity at home, although to very low levels (< 3 lux), has been found to phase-advance DLMO by approximately one hour (Burgess & Molina, 2014). Brown and colleagues recently published expert consensus recommendations for at-home light levels and

recommend that in the last 3 hours prior to bedtime, light levels should stay below a maximum of 10 melanopic EDI (T. M. Brown et al., 2022). This can be achieved without making costly and comprehensive changes to lighting fixtures. Most homes already have warm-white, often dimmable, light. Modern LED-luminaires often have melanopic Daylight Efficacy Ratio (DER) low enough to allow 30 photopic lux while keeping melanopic EDI below 10 lux (Schlangen & Price, 2021). Moreover, investigations using wearable light sensors have found that in 50% of homes, the evening light levels were already below recommended threshold (Cain et al., 2020). Public information campaigns may increase awareness of how simple changes in existing lighting may be combined with reduced screen time in the evening. This may provide healthier evening light environments and be an alternative to the population-wide implementation of BDLE. This may strike a balance between concerns of stimulating more anxiety regarding sleep and nudging the general population towards sleep and circadian-friendly evening light environments.

The more extensive alterations required (exchanging all fixtures and filters to block incident short-wavelength light) to create the ultra-low melanopic DER in a BDLE may be reserved for public settings in which higher light levels are needed (typically >100 lux) to support normal function, e.g., hospitals, nursing homes, etc. (T. M. Brown et al., 2022). Such applications might also mitigate the adverse effects of shift work on staff in public facilities. Attempts have been made to create models that take into account individual variation among staff in circadian rhythms timing and chronotype, to plan and use blue-filtering tools to reduce the negative health effects of shift work (Guarana et al., 2021). One study by Albala and colleagues demonstrated that an automatic blue-depleted task light could be used when nurses performed tasks in patient rooms during nighttime, and this increased caregiver satisfaction and patient anxiety (Albala et al., 2019). As already mentioned, we have not found negative effects of working in BDLE on nurses' sleep or work function (Kjørstad et al., 2022), but the study was limited by low participation rates and hence low power. More work is needed to determine the effects of implementing these new light systems on staff performance and health outcomes.

There is a rapid development of smart integrative lighting based on the metameric principles discussed earlier. More studies are needed to ascertain this technology's physiologic effectiveness in naturalistic settings. However, this technology may reduce the potential risks discussed counter to the population-wide adoption strategy of BDLEs. Smart integrative

lighting may allow the user to select from a menu that metameric BDLEs should be automatically generated in the evening. As this would lead to minimal changes in perceived color, the risks of increasing attention toward sleep and sleep optimization might be lower. With time, such solutions will likely find their way into home environments, which, if anything, should motivate the sleep research community to be at the forefront of testing the efficacy of such solutions.

4.5.5   Implementation of radar assessment of sleep in the general population

The radar assessment of sleep, almost more than the other technologies, could very easily be used as a consumer sleep technology (CST). In fact, the same radar sensor used in paper III is already commercially available as a CST. There may, however, be important caveats regarding who needs continuous sleep tracking and if there are unwanted effects of such tracking in healthy sleepers, e.g., similar negative attentional processes following sleep optimization discussed for BDLE. Some groups, like individuals with mood disorders, may benefit from long-term sleep monitoring as changes in sleep can precede worsening of psychiatric symptoms, allowing earlier identification of illness episodes and intervention on sleep. A range of other episodic illnesses (e.g., autoimmune diseases) are also aggravated by poor sleep and may have similar benefits from early identification and intervention of developing sleep disturbances. Moreover, there may be a potential benefit of using radar assessment as a part of CBT-I (Baron et al., 2017) to provide more objective measures of sleep, particularly in the case of patients with the paradoxical-insomnia subtype (Perlis et al., 2022).

4.5.6   Implementation of dCBT-I

CBT-I is an excellent example of an intervention targeted toward a specific group of individuals. dCBT-I, although its scalability allows for population-wide dissemination, can be characterized as a high-risk strategy of prevention (Hunt & Emslie, 2001; Rose, 1992). For the treatment of insomnia, we would advocate for a stepped-care model to disseminate CBT-I across the population effectively, similar to what Espie and colleagues (Espie, 2009; Espie et al., 2013), and later Baglioni and colleagues (Baglioni et al., 2020) have proposed. This would entail a hierarchy of treatment levels with increasing therapist assistance. The first step could be GPs prescribing access to a fully automated version of dCBT-I, such as the one tested in paper IV. A second step could be group-based FtF CBT-I, or therapist-assisted dCBT-I, based on patient preference or traits. Therapist-guided versions provide a middle-

ground that may help reduce the often high drop-out rates during completely automated therapy (Leerssen et al., 2022). Third could be individual FtF therapy with trained health care professionals that can tailor treatment further to the individual patient's symptoms and needs. If treatment is unsuccessful at a lower step, patients could be "stepped up" to a higher level of care. However, a potential problem for stepped care approaches where CBT-I is offered at all steps is that patients may not be motivated to go through additional rounds of the same intervention that they already experienced as ineffective, with the only difference being by whom the intervention is provided. Additional strategies could perhaps mitigate this. To curtail the number of patients having to go through unsuccessful treatments, screening at the primary care level would be necessary to sort people into steps based on, e.g., severity of insomnia, treatment modality preferences, the complexity of clinical symptoms outside of insomnia, and general functioning. In addition, measures should be taken to detect signs of unsuccessful treatment early and increase their level of assistance (as discussed in the context of paper IV) to minimize resignation or treatment fatigue. Although the model must be tailored to what resources are available locally in terms of trained CBT-I therapists, it demonstrates a way in which quality-controlled CBT-I can be made available to the public while reserving the scarce and expensive therapist resources to those clinical cases where this is most required.

4.6   Main conclusions

- Evening BDLEs phase-advanced DLMO phase, reduced melatonin suppression, and increased TST and REM sleep duration compared with SLE.
- There was some indication of reduced neurocognitive arousal in BDLE compared with SLE, but no effects on subjective sleepiness.
- The color-discrimination ability was reduced in BDLE compared with SLE, but no substantial subjective side effects were reported.
- Residing in the evening BDLE reduced the fragmentation of REM sleep and microarousals during REM sleep. There were no differences in REM sleep density.
- Both REM sleep fragmentation and REM sleep duration were associated with DLMO phase-shifts, but BDLE has an effect over and above that of phase-shift. This was not the case for REM sleep duration.
- REM sleep accumulated faster throughout the night in BDLE but did not seem to be at the expense of SWS (N3).

- Real-time sleep-wake classification models were developed to analyze radar data.
- Compared with PSG, real-time models performed on par with those often seen for actigraphy.
- Agreement between radar and PSG was generally excellent in healthy subjects, but the limits of agreement were wider for patients undergoing ambulatory sleep assessments.
- In healthy subjects, both radar and actigraphy showed high classification results. In the patient data, actigraphy showed higher specificities. Both radar and actigraphy tended to overestimate SOL and underestimate NW.
- dCBT-I was inferior to FtF CBT-I at 9 weeks, whereas results were inconclusive regarding non-inferiority at 33 weeks follow-up.
- FtF was superior to dCBT-I at both 9- and 33-weeks follow-up.
- Uptake and completion rates of treatment were lower for dCBT-I.
- There were no differences in other secondary outcomes of *the dCBT-I trial*.

4.7   Concluding remarks

The collection of papers included in this thesis collectively demonstrate that new technology may be used to improve the assessment and treatment of sleep and circadian rhythms disruption. The context for all the trials/data collections included in this thesis is a university hospital. These papers shed light on how technology may be specifically used to improve sleep management in a hospital setting. For the emerging technologies of BDLE and radar, these findings represent early steps and lay the foundations for larger studies with clinical samples to assess further potential. Our findings of measurable physiological effects of BDLE in healthy subjects could serve as a proof-of-concept and may spur further investigations in different clinical samples where the measurements that were undertaken here may not be feasible. Indeed, we are aware of several studies in the pipeline that test variants of BDLE with inpatients, and it is a matter of time before we know more about its efficacy for various clinical populations. Our findings regarding REM sleep consolidation must be considered exploratory but open new questions for future work on insomnia and mental disorders. The radar assessment of sleep offers advantages over PSG or actigraphy by offering contact-free sleep assessment. Moreover, the radar can provide real-time information on sleep-wake state, which, with further validation, may be used to monitor inpatient sleep-wake state. This may indirectly improve sleep by reducing the need for intermittent nurse observations that wake the patient. In the case of dCBT-I we do not establish non-inferiority to FtF treatment. dCBT-

I does offer other advantages, such as facilitated dissemination of evidence-based insomnia treatment that may be especially helpful at lower levels of care. However, these findings indicate FtF CBT-I should remain the treatment of choice in the setting of secondary care sleep clinics. The upcoming questions for dCBT-I concern the implementation into national health care systems. Future studies should investigate methods to identify patients likely to have good outcomes with digital therapy and systems to detect unsuccessful treatments. Moreover, the cost-benefit of different versions of stepped-care models should eventually be investigated to ensure the best possible implementation strategies. Taken together, this thesis suggests that technology has a role in efforts to improve the treatment and assessment of sleep and circadian rhythms disturbance.

5   References

Abe, T., Inoue, Y., Komada, Y., Nakamura, M., Asaoka, S., Kanno, M., Shibui, K., Hayashida, K., Usui, A., & Takahashi, K. (2011). Relation between morningness–eveningness score and depressive symptoms among patients with delayed sleep phase syndrome. *Sleep Medicine*, *12*(7), 680–684. https://doi.org/10.1016/j.sleep.2010.12.017

Akerstedt, T., & Froberg, J. (1977). Psychophysiological circadian rhythms in women during 72 h of sleep deprivation. *Waking Sleeping*, *1*, 387–394.

Albala, L., Bober, T., Hale, G., Warfield, B., Collins, M. L., Merritt, Z., Steimetz, E., Nadler, S., Lev, Y., & Hanifin, J. (2019). Effect on nurse and patient experience: Overnight use of blue-depleted illumination. *BMJ Open Quality*, *8*(3), e000692. https://doi.org/10.1136/bmjoq-2019-000692

Allen, A. E., Hazelhoff, E. M., Martial, F. P., Cajochen, C., & Lucas, R. J. (2018). Exploiting metamerism to regulate the impact of a visual display on alertness and melatonin suppression independent of visual appearance. *Sleep*, *41*(8). https://doi.org/10.1093/sleep/zsy100

American Academy of Sleep Medicine. (2014). *International classification of sleep disorders*.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5^{TM}, 5th ed.* (pp. xliv, 947). American Psychiatric Publishing, Inc. https://doi.org/10.1176/appi.books.9780890425596

Andersson, G., Cuijpers, P., Carlbring, P., Riper, H., & Hedman, E. (2014). Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: A systematic review and meta-analysis. *World Psychiatry*, *13*(3), 288–295. https://doi.org/10.1002/wps.20151

Aoki, H., Ozeki, Y., & Yamada, N. (2001). Hypersensitivity of melatonin suppression in response to light in patients with delayed sleep phase syndrome. *Chronobiology International*, *18*(2), 263–271. https://doi.org/10.1081/CBI-100103190

Aschoff, J. (1965). Circadian Rhythms in Man. *Science*, *148*(3676), 1427–1432. https://doi.org/10.1126/science.148.3676.1427

Aschoff, J., GERECKE, U., & Wever, R. (1967). Desynchronization of human circadian rhythms. *The Japanese Journal of Physiology*, *17*(4), 450–457.

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

Aston-Jones, G., Chen, S., Zhu, Y., & Oshinsky, M. L. (2001). A neural circuit for circadian regulation of arousal. *Nature Neuroscience*, *4*(7), 732–738. https://doi.org/10.1038/89522

Aston-Jones, G., Gonzalez, M., & Doran, S. (2007). Role of the locus coeruleus-norepinephrine system in arousal and circadian regulation of the sleep–wake cycle. In G. A. Ordway, M. A. Schwartz, & A. Frazer (Eds.), *Brain Norepinephrine* (1st ed., pp. 157–195). Cambridge University Press. https://doi.org/10.1017/CBO9780511544156.007

Baglioni, C., Altena, E., Bjorvatn, B., Blom, K., Bothelius, K., Devoto, A., Espie, C. A., Frase, L., Gavriloff, D., Tuuliki, H., Hoflehner, A., Högl, B., Holzinger, B., Järnefelt, H., Jernelöv, S., Johann, A. F., Lombardo, C., Nissen, C., Palagini, L., … Riemann, D. (2020). The European Academy for Cognitive Behavioural Therapy for Insomnia: An initiative of the European Insomnia Network to promote implementation and dissemination of treatment. *Journal of Sleep Research*, *29*(2). https://doi.org/10.1111/jsr.12967

Baglioni, C., Battagliese, G., Feige, B., Spiegelhalder, K., Nissen, C., Voderholzer, U., Lombardo, C., & Riemann, D. (2011). Insomnia as a predictor of depression: A meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders*, *135*(1–3), 10–19. https://doi.org/10.1016/j.jad.2011.01.011

Baglioni, C., Nanovska, S., Regen, W., Spiegelhalder, K., Feige, B., Nissen, C., Reynolds, C. F., & Riemann, D. (2016). Sleep and mental disorders: A meta-analysis of polysomnographic research. *Psychological Bulletin*, *142*(9), 969–990. https://doi.org/10.1037/bul0000053

Bani Younis, M., Hayajneh, F., & Alshraideh, J. A. (2021). Effect of noise and light levels on sleep of intensive care unit patients. *Nursing in Critical Care*, *26*(2), 73–78. https://doi.org/10.1111/nicc.12490

Barbini, B., Benedetti, F., Colombo, C., Dotoli, D., Bernasconi, A., Cigala-Fulgosi, M., Florita, M., & Smeraldi, E. (2005). Dark therapy for mania: A pilot study. *Bipolar Disorders*, *7*(1), 98–101. https://doi.org/10.1111/j.1399-5618.2004.00166.x

Baron, K. G., Abbott, S., Jao, N., Manalo, N., & Mullen, R. (2017). Orthosomnia: Are Some Patients Taking the Quantified Self Too Far? *Journal of Clinical Sleep Medicine*, *13*(02), 351–354. https://doi.org/10.5664/jcsm.6472

Baron, K. G., & Reid, K. J. (2014). Circadian misalignment and health. *International Review of Psychiatry*, *26*(2), 139–154. https://doi.org/10.3109/09540261.2014.911149

Benloucif, S., Burgess, H. J., Klerman, E. B., Lewy, A. J., Middleton, B., Murphy, P. J., Parry, B. L., & Revell, V. L. (2008). Measuring Melatonin in Humans. *Journal of Clinical Sleep Medicine*, *4*(1), 66–69. https://doi.org/10.5664/jcsm.27083

Berger, R. J., & Oswald, I. (1962). Effects of Sleep Deprivation on Behaviour, Subsequent Sleep, and Dreaming. *Journal of Mental Science*, *108*(455), 457–465. Cambridge Core. https://doi.org/10.1192/bjp.108.455.457

Berman, G., Muttuvelu, D., Berman, D., Larsen, J. I., Licht, R. W., Ledolter, J., & Kardon, R. H. (2018). Decreased retinal sensitivity in depressive disorder: A controlled study. *Acta Psychiatrica Scandinavica*, *137*(3), 231–240. https://doi.org/10.1111/acps.12851

Berry, R. B., Brooks, R., Gamaldo, C., Harding, S. M., Lloyd, R. M., Quan, S. F., Troester, M. T., & Vaughn, B. V. (2017). AASM Scoring Manual Updates for 2017 (Version 2.4). *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine*, *13*(5), 665–666. PubMed. https://doi.org/10.5664/jcsm.6576

Bland, J. M., & Altman, D. G. (1999). *Measuring agreement in method comparison studies*. 26.

Blom, K., Tarkian Tillgren, H., Wiklund, T., Danlycke, E., Forssén, M., Söderström, A., Johansson, R., Hesser, H., Jernelöv, S., Lindefors, N., Andersson, G., & Kaldo, V. (2015). Internet-vs. group-delivered cognitive behavior therapy for insomnia: A randomized controlled non-inferiority trial. *Behaviour Research and Therapy*, *70*, 47–55. https://doi.org/10.1016/j.brat.2015.05.002

Blume, C., Schmidt, M. H., & Cajochen, C. (2020). Effects of the COVID-19 lockdown on human sleep and rest-activity rhythms. *Current Biology*, *30*(14), R795–R797. https://doi.org/10.1016/j.cub.2020.06.021

Borbely, A. A. (1982). *A Two Process Model of Sleep Regulation*. *1*, 195–204.

Borbély, A. A., Baumann, F., Brandeis, D., Strauch, I., & Lehmann, D. (1981). Sleep deprivation: Effect on sleep stages and EEG power density in man. *Electroencephalography and Clinical Neurophysiology*, *51*(5), 483–493. https://doi.org/10.1016/0013-4694(81)90225-X

Borbély, A. A., Daan, S., Wirz-Justice, A., & Deboer, T. (2016). The two-process model of sleep regulation: A reappraisal. *Journal of Sleep Research*, *25*(2), 131–143. https://doi.org/10.1111/jsr.12371

Boyce, P. (2022). Light, lighting and human health. *Lighting Research & Technology*, *54*(2), 101–144. https://doi.org/10.1177/14771535211010267

Boyko, Y., Jennum, P., Nikolic, M., Holst, R., Oerding, H., & Toft, P. (2017). Sleep in intensive care unit: The role of environment. *Journal of Critical Care*, *37*, 99–105. https://doi.org/10.1016/j.jcrc.2016.09.005

Brainard, G. C., Hanifin, J. P., Greeson, J. M., Byrne, B., Glickman, G., Gerner, E., & Rollag, M. D. (2001). Action Spectrum for Melatonin Regulation in Humans: Evidence for a Novel Circadian Photoreceptor. *The Journal of Neuroscience*, *21*(16), 6405–6412. https://doi.org/10.1523/JNEUROSCI.21-16-06405.2001

Brainard, G. C., Hanifin, J. P., Rollag, M. D., Greeson, J., Byrne, B., Glickman, G., Gerner, E., & Sanford, B. (2001). Human Melatonin Regulation Is Not Mediated by the Three Cone Photopic Visual System. *The Journal of Clinical Endocrinology & Metabolism*, *86*(1), 433–436. https://doi.org/10.1210/jcem.86.1.7277

Brasure, M., Fuchs, E., MacDonald, R., Nelson, V. A., Koffel, E., Olson, C. M., Khawaja, I. S., Diem, S., Carlyle, M., Wilt, T. J., Ouellette, J., Butler, M., & Kane, R. L. (2016).

Psychological and Behavioral Interventions for Managing Insomnia Disorder: An Evidence Report for a Clinical Practice Guideline by the American College of Physicians. *Annals of Internal Medicine*, *165*(2), 113. https://doi.org/10.7326/M15-1782

Brown, R. E., Basheer, R., McKenna, J. T., Strecker, R. E., & McCarley, R. W. (2012). Control of Sleep and Wakefulness. *Physiological Reviews*, *92*(3), 1087–1187. https://doi.org/10.1152/physrev.00032.2011

Brown, T. M. (2020). Melanopic illuminance defines the magnitude of human circadian light responses under a wide range of conditions. *Journal of Pineal Research*, *69*(1). https://doi.org/10.1111/jpi.12655

Brown, T. M., Brainard, G. C., Cajochen, C., Czeisler, C. A., Hanifin, J. P., Lockley, S. W., Lucas, R. J., Münch, M., O'Hagan, J. B., Peirson, S. N., Price, L. L. A., Roenneberg, T., Schlangen, L. J. M., Skene, D. J., Spitschan, M., Vetter, C., Zee, P. C., & Wright, K. P. (2022). Recommendations for daytime, evening, and nighttime indoor light exposure to best support physiology, sleep, and wakefulness in healthy adults. *PLOS Biology*, *20*(3), e3001571. https://doi.org/10.1371/journal.pbio.3001571

Brown, T. M., Tsujimura, S., Allen, A. E., Wynne, J., Bedford, R., Vickery, G., Vugler, A., & Lucas, R. J. (2012). Melanopsin-Based Brightness Discrimination in Mice and Humans. *Current Biology*, *22*(12), 1134–1141. https://doi.org/10.1016/j.cub.2012.04.039

Burgess, H. J., & Molina, T. A. (2014). Home Lighting Before Usual Bedtime Impacts Circadian Timing: A Field Study. *Photochemistry and Photobiology*, *90*(3), 723–726. https://doi.org/10.1111/php.12241

Burns, A. C., Windred, D. P., Rutter, M. K., Olivier, P., Vetter, C., Saxena, R., Lane, J. M., Phillips, A. J. K., & Cain, S. W. (2022). *Low daytime light and bright night-time light are associated with psychiatric disorders: An objective light study in >85,000 UK Biobank participants* [Preprint]. Psychiatry and Clinical Psychology. https://doi.org/10.1101/2022.10.16.22280934

Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Lichstein, K. L., & Morin, C. M. (2006). Recommendations for a Standard Research Assessment of Insomnia. *Sleep*, *29*(9), 1155–1173. https://doi.org/10.1093/sleep/29.9.1155

Cain, S. W., McGlashan, E. M., Vidafar, P., Mustafovska, J., Curran, S. P. N., Wang, X., Mohamed, A., Kalavally, V., & Phillips, A. J. K. (2020). Evening home lighting adversely impacts the circadian system and sleep. *Scientific Reports*, *10*(1), 19110. https://doi.org/10.1038/s41598-020-75622-4

Cain, S. W., & Phillips, A. J. K. (2021). Do no harm: The beginning of the age of healthy hospital lighting. *Sleep*, *44*(3), zsab016. https://doi.org/10.1093/sleep/zsab016

Cajochen, C., Kräuchi, K., Danilenko, K. V., & Wirz-Justice, A. (1998). Evening administration of melatonin and bright light: Interactions on the EEG during sleep and

wakefulness. *Journal of Sleep Research*, *7*(3), 145–157. https://doi.org/10.1046/j.1365-2869.1998.00106.x

Cajochen, C., Kräuchi, K., Möri, D., Graw, P., & Wirz-Justice, A. (1997). Melatonin and S-20098 increase REM sleep and wake-up propensity without modifying NREM sleep homeostasis. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *272*(4), R1189–R1196. https://doi.org/10.1152/ajpregu.1997.272.4.R1189

Cajochen, C., Münch, M., Kobialka, S., Kräuchi, K., Steiner, R., Oelhafen, P., Orgül, S., & Wirz-Justice, A. (2005). High Sensitivity of Human Melatonin, Alertness, Thermoregulation, and Heart Rate to Short Wavelength Light. *The Journal of Clinical Endocrinology & Metabolism*, *90*(3), 1311–1316. https://doi.org/10.1210/jc.2004-0957

Cajochen, C., Reichert, C., Maire, M., Schlangen, L. J. M., Schmidt, C., Viola, A. U., & Gabel, V. (2019). Evidence That Homeostatic Sleep Regulation Depends on Ambient Lighting Conditions during Wakefulness. *Clocks & Sleep*, *1*(4), 517–531. https://doi.org/10.3390/clockssleep1040040

Cajochen, C., Zeitzer, J. M., Czeisler, C. A., & Dijk, D.-J. (2000). Dose-response relationship for light intensity and ocular and electroencephalographic correlates of human alertness. *Behavioural Brain Research*, *115*(1), 75–83. https://doi.org/10.1016/S0166-4328(00)00236-9

Canazei, M., Weninger, J., Pohl, W., Marksteiner, J., & Weiss, E. M. (2022). Effects of dynamic bedroom lighting on measures of sleep and circadian rest-activity rhythm in inpatients with major depressive disorder. *Scientific Reports*, *12*(1), 6137. https://doi.org/10.1038/s41598-022-10161-8

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. *Cognitive Behaviour Therapy*, *47*(1), 1–18. https://doi.org/10.1080/16506073.2017.1401115

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The Consensus Sleep Diary: Standardizing Prospective Sleep Self-Monitoring. *Sleep*, *35*(2), 287–302. https://doi.org/10.5665/sleep.1642

Carney, P. A., Fitzgerald, C. T., & Monaghan, C. E. (1988). Influence of Climate on the Prevalence of Mania. *British Journal of Psychiatry*, *152*(6), 820–823. https://doi.org/10.1192/bjp.152.6.820

Carrier, J., & Dumont, M. (1995). Sleep propensity and sleep architecture after bright light exposure at three different times of day. *Journal of Sleep Research*, *4*(4), 202–211. https://doi.org/10.1111/j.1365-2869.1995.tb00171.x

Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D., & Wallace, E. P. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research*, *37*(2), 147–153. https://doi.org/10.1016/0022-3999(93)90081-P

Chang, A.-M., Aeschbach, D., Duffy, J. F., & Czeisler, C. A. (2015). Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proceedings of the National Academy of Sciences*, *112*(4), 1232–1237. https://doi.org/10.1073/pnas.1418490112

Chang, A.-M., Scheer, F. A. J. L., & Czeisler, C. A. (2011). The human circadian system adapts to prior photic history: The human circadian system adapts to prior light history. *The Journal of Physiology*, *589*(5), 1095–1102. https://doi.org/10.1113/jphysiol.2010.201194

Chang, A.-M., Scheer, F. A. J. L., Czeisler, C. A., & Aeschbach, D. (2013). Direct Effects of Light on Alertness, Vigilance, and the Waking Electroencephalogram in Humans Depend on Prior Light History. *Sleep*, *36*(8), 1239–1246. https://doi.org/10.5665/sleep.2894

Chaput, J.-P., McHill, A. W., Cox, R. C., Broussard, J. L., Dutil, C., da Costa, B. G. G., Sampasa-Kanyinga, H., & Wright, K. P. (2022). The role of insufficient sleep and circadian misalignment in obesity. *Nature Reviews Endocrinology*. https://doi.org/10.1038/s41574-022-00747-7

Chellappa, S. L., Lasauskaite, R., & Cajochen, C. (2017). In a Heartbeat: Light and Cardiovascular Physiology. *Frontiers in Neurology*, *8*, 541. https://doi.org/10.3389/fneur.2017.00541

Chellappa, S. L., Steiner, R., Blattner, P., Oelhafen, P., Götz, T., & Cajochen, C. (2011). Non-Visual Effects of Light on Melatonin, Alertness and Cognitive Performance: Can Blue-Enriched Light Keep Us Alert? *PLoS ONE*, *6*(1), e16429. https://doi.org/10.1371/journal.pone.0016429

Chellappa, S. L., Steiner, R., Oelhafen, P., Lang, D., Götz, T., Krebs, J., & Cajochen, C. (2013). Acute exposure to evening blue-enriched light impacts on human sleep. *Journal of Sleep Research*, *22*(5), 573–580. https://doi.org/10.1111/jsr.12050

Chemelli, R. M., Willie, J. T., Sinton, C. M., Elmquist, J. K., Scammell, T., Lee, C., Richardson, J. A., Williams, S. C., Xiong, Y., Kisanuki, Y., Fitch, T. E., Nakazato, M., Hammer, R. E., Saper, C. B., & Yanagisawa, M. (1999). *Narcolepsy in orexin Knockout Mice: Molecular Genetics of Sleep Regulation*. 15.

Cheng, P., Walch, O., Huang, Y., Mayer, C., Sagong, C., Cuamatzi Castelan, A., Burgess, H. J., Roth, T., Forger, D. B., & Drake, C. L. (2021). Predicting circadian misalignment with wearable technology: Validation of wrist-worn actigraphy and photometry in night shift workers. *Sleep*, *44*(2), zsaa180. https://doi.org/10.1093/sleep/zsaa180

Chinoy, E. D., Cuellar, J. A., Huwa, K. E., Jameson, J. T., Watson, C. H., Bessman, S. C., Hirsch, D. A., Cooper, A. D., Drummond, S. P. A., & Markwald, R. R. (2021). Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*, *44*(5), zsaa291. https://doi.org/10.1093/sleep/zsaa291

Chinoy, E. D., Duffy, J. F., & Czeisler, C. A. (2018). Unrestricted evening use of light-emitting tablet computers delays self-selected bedtime and disrupts circadian timing and alertness. *Physiological Reports*, *6*(10), e13692. https://doi.org/10.14814/phy2.13692

Cho, J. R., Joo, E. Y., Koo, D. L., & Hong, S. B. (2013). Let there be no light: The effect of bedside light on sleep quality and background electroencephalographic rhythms. *Sleep Medicine*, *14*(12), 1422–1425. https://doi.org/10.1016/j.sleep.2013.09.007

Cho, Y., Ryu, S. H., Lee, B. R., Kim, K. H., Lee, E., & Choi, J. (2015). Effects of artificial light at night on human health: A literature review of observational and experimental studies applied to exposure assessment. *Chronobiology International*, *32*(9), 1294–1310. https://doi.org/10.3109/07420528.2015.1073158

Chou, T. C., Scammell, T. E., Gooley, J. J., Gaus, S. E., Saper, C. B., & Lu, J. (2003). Critical Role of Dorsomedial Hypothalamic Nucleus in a Wide Range of Behavioral Circadian Rhythms. *The Journal of Neuroscience*, *23*(33), 10691–10702. https://doi.org/10.1523/JNEUROSCI.23-33-10691.2003

Christian, C., Derk-Jan, D., & Alexander, B. (1992). Dynamics of EEG Slow-Wave Activity and Core Body Temperature in Human Sleep After Exposure to Bright Light. *Sleep*. https://doi.org/10.1093/sleep/15.4.337

CIE Central Bureau. (2018). *CIE System for Metrology of Optical Radiation for ipRGC-influenced Responses to Light*. CIE.

CIE Central Bureau. (2019). *CIE Position Statement on Non-Visual Effects of Light: Recommending Proper Light at the Proper Time*.

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic Sleep/Wake Identification From Wrist Activity. *Sleep*, *15*(5), 461–469. https://doi.org/10.1093/sleep/15.5.461

Committee for Proprietary Medicinal Products (CPMP). (2001). Points to consider on switching between superiority and non-inferiority: *Points to consider on switching between superiority and non-inferiority*. *British Journal of Clinical Pharmacology*, *52*(3), 223–228. https://doi.org/10.1046/j.1365-2125.2001.01397-3.x

Conners, C. K. (2014). *Conners Continuous Performance Test-3rd edition manual.* Multi-Health Systems.

Craig, T., & Mathieu, S. (2018). CANDLE: The critical analysis of the nocturnal distribution of light exposure – A prospective pilot study quantifying the nocturnal light intensity on a critical care unit. *Journal of the Intensive Care Society*, *19*(3), 196–200. https://doi.org/10.1177/1751143717748095

Crasson, M., Kjiri, S., Colin, A., Kjiri, K., L'Hermite-Baleriaux, M., Ansseau, M., & Legros, J. J. (2004). Serum melatonin and urinary 6-sulfatoxymelatonin in major depression. *Psychoneuroendocrinology*, *29*(1), 1–12. https://doi.org/10.1016/S0306-4530(02)00123-3

Crinion, S. J., Tiron, R., Lyon, G., Zaffaroni, A., Kilroy, H., Doheny, E., O'Hare, E., Boyle, P., Russell, A., Traynor, M., Kent, B. D., Ryan, S., & McNicholas, W. T. (2020). Ambulatory detection of sleep apnea using a non-contact biomotion sensor. *Journal of Sleep Research*, *29*(1). https://doi.org/10.1111/jsr.12889

Czeisler, C. A. (1999). Stability, Precision, and Near-24-Hour Period of the Human Circadian Pacemaker. *Science*, *284*(5423), 2177–2181. https://doi.org/10.1126/science.284.5423.2177

Czeisler, C. A., Zimmerman, J. C., Ronda, J. M., Moore-Ede, M. C., & Weitzman, E. D. (1980). Timing of REM Sleep is Coupled to the Circadian Rhythm of Body Temperature in Man. *Sleep*, *2*(3), 329–346. https://doi.org/10.1093/sleep/2.3.329

Danaher, J. (2022). Techno-optimism: An Analysis, an Evaluation and a Modest Defence. *Philosophy & Technology*, *35*(2), 54. https://doi.org/10.1007/s13347-022-00550-2

Daneault, V., Dumont, M., Massé, É., Vandewalle, G., & Carrier, J. (2016). Light-sensitive brain pathways and aging. *Journal of Physiological Anthropology*, *35*(1), 9. https://doi.org/10.1186/s40101-016-0091-9

Danielson, S. J., Rappaport, C. A., Loher, M. K., & Gehlbach, B. K. (2018). Looking for light in the din: An examination of the circadian-disrupting properties of a medical intensive care unit. *Intensive and Critical Care Nursing*, *46*, 57–63. https://doi.org/10.1016/j.iccn.2017.12.006

Darbyshire, J. L., & Young, J. D. (2013). An investigation of sound levels on intensive care units with reference to the WHO guidelines. *Critical Care*, *17*(5), R187. https://doi.org/10.1186/cc12870

de Souza, L., Benedito-Silva, A. A., Pires, M. L. N., Poyares, D., Tufik, S., & Calil, H. M. (2003). Further Validation of Actigraphy for Sleep Studies. *Sleep*, *26*(1), 81–85. https://doi.org/10.1093/sleep/26.1.81

De Zambotti, M., Cellini, N., Goldstone, A., Colrain, I. M., & Baker, F. C. (2019). Wearable Sleep Technology in Clinical and Research Settings. *Medicine & Science in Sports & Exercise*, *51*(7), 1538–1557. https://doi.org/10.1249/MSS.0000000000001947

Deacon, S., & Arendt, J. (1994). Posture influences melatonin concentrations in plasma and saliva in humans. *Neuroscience Letters*, *167*(1–2), 191–194. https://doi.org/10.1016/0304-3940(94)91059-6

Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, *9*(4), 673–690.

Deurveilher, S., Burns, J., & Semba, K. (2002). Indirect projections from the suprachiasmatic nucleus to the ventrolateral preoptic nucleus: A dual tract-tracing study in rat: Indirect SCN-VLPO pathways. *European Journal of Neuroscience*, *16*(7), 1195–1213. https://doi.org/10.1046/j.1460-9568.2002.02196.x

Dijk, D., & Czeisler, C. (1995). Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. *The Journal of Neuroscience*, *15*(5), 3526–3538. https://doi.org/10.1523/JNEUROSCI.15-05-03526.1995

Doğan, O., Ertekin, Ş., & Doğan, S. (2005). Sleep quality in hospitalized patients. *Journal of Clinical Nursing*, *14*(1), 107–113. https://doi.org/10.1111/j.1365-2702.2004.01011.x

Drews, H. J., Scott, J., Langsrud, K., Vethe, D., & Kallestad, H. (2020). Chronobiologically informed inpatient milieu in psychiatric institutions. *The Lancet Psychiatry*, *7*(12), 1013–1014. https://doi.org/10.1016/S2215-0366(20)30459-4

Dwan, K., Li, T., Altman, D. G., & Elbourne, D. (2019). CONSORT 2010 statement: Extension to randomised crossover trials. *BMJ*, *366*. https://doi.org/10.1136/bmj.l4378

Eagles, J. M. (1994). The relationship between mood and daily hours of sunlight in rapid cycling bipolar illness. *Biological Psychiatry*, *36*(6), 422–424. https://doi.org/10.1016/0006-3223(94)91216-5

Elliott, R., McKinley, S., Cistulli, P., & Fien, M. (2013). Characterisation of sleep in intensive care using 24-hour polysomnography: An observational study. *Critical Care*, *17*(2), R46. https://doi.org/10.1186/cc12565

Ellis, J. G. (2012). Acute insomnia: Current conceptualizations and future directions. *Sleep Medicine Reviews*, 10.

Emens, J., Lewy, A., Kinzie, J. M., Arntz, D., & Rough, J. (2009). Circadian misalignment in major depressive disorder. *Psychiatry Research*, *168*(3), 259–261. https://doi.org/10.1016/j.psychres.2009.04.009

Engwall, M., Fridh, I., Johansson, L., Bergbom, I., & Lindahl, B. (2015). Lighting, sleep and circadian rhythm: An intervention study in the intensive care unit. *Intensive and Critical Care Nursing*, *31*(6), 325–335. https://doi.org/10.1016/j.iccn.2015.07.001

Engwall, M., Fridh, I., Jutengren, G., Bergbom, I., Sterner, A., & Lindahl, B. (2017). The effect of cycled lighting in the intensive care unit on sleep, activity and physiological parameters: A pilot study. *Intensive and Critical Care Nursing*, *41*, 26–32. https://doi.org/10.1016/j.iccn.2017.01.009

Esaki, Y., Kitajima, T., Ito, Y., Koike, S., Nakao, Y., Tsuchiya, A., Hirose, M., & Iwata, N. (2016). Wearing blue light-blocking glasses in the evening advances circadian rhythms in the patients with delayed sleep phase disorder: An open-label trial. *Chronobiology International*, *33*(8), 1037–1044. https://doi.org/10.1080/07420528.2016.1194289

Espie, C. A. (2001). *Insomnia: Conceptual Issues in the Development, Persistence, and Treatment of Sleep Disorder in Adults*. 32.

Espie, C. A. (2009). "Stepped Care": A Health Technology Solution for Delivering Cognitive Behavioral Therapy as a First Line Insomnia Treatment. *Sleep*, *32*(12), 1549–1558. https://doi.org/10.1093/sleep/32.12.1549

Espie, C. A., Broomfield, N. M., MacMahon, K. M. A., Macphee, L. M., & Taylor, L. M. (2006). The attention–intention–effort pathway in the development of psychophysiologic insomnia: A theoretical review. *Sleep Medicine Reviews*, *10*(4), 215–245. https://doi.org/10.1016/j.smrv.2006.03.002

Espie, C. A., Emsley, R., Kyle, S. D., Gordon, C., Drake, C. L., Siriwardena, A. N., Cape, J., Ong, J. C., Sheaves, B., Foster, R., Freeman, D., Costa-Font, J., Marsden, A., & Luik, A. I. (2019). Effect of Digital Cognitive Behavioral Therapy for Insomnia on Health, Psychological Well-being, and Sleep-Related Quality of Life: A Randomized Clinical Trial. *JAMA Psychiatry*, *76*(1), 21. https://doi.org/10.1001/jamapsychiatry.2018.2745

Espie, C. A., Hames, P., & McKinstry, B. (2013). Use of the Internet and Mobile Media for Delivery of Cognitive Behavioral Insomnia Therapy. *Insomnia*, *8*(3), 407–419. https://doi.org/10.1016/j.jsmc.2013.06.001

Faaland, P., Vedaa, Ø., Langsrud, K., Sivertsen, B., Lydersen, S., Vestergaard, C. L., Kjørstad, K., Vethe, D., Ritterband, L. M., Harvey, A. G., Stiles, T. C., Scott, J., & Kallestad, H. (2022). Digital cognitive behaviour therapy for insomnia (dCBT-I): Chronotype moderation on intervention outcomes. *Journal of Sleep Research*, *31*(5). https://doi.org/10.1111/jsr.13572

Fagerland, M. W., Lydersen, S., & Laake, P. (2017). *Statistical analysis of contingency tables*. Chapman and Hall/CRC.

Feige, B., Al-Shajlawi, A., Nissen, C., Voderholzer, U., Hornyak, M., Spiegelhalder, K., Kloepfer, C., Perlis, M., & Riemann, D. (2008). Does REM sleep contribute to subjective wake time in primary insomnia? A comparison of polysomnographic and subjective sleep in 100 patients. *Journal of Sleep Research*, *17*(2), 180–190. https://doi.org/10.1111/j.1365-2869.2008.00651.x

Fernandez, D. C., Fogerson, P. M., Lazzerini Ospri, L., Thomsen, M. B., Layne, R. M., Severin, D., Zhan, J., Singer, J. H., Kirkwood, A., Zhao, H., Berson, D. M., & Hattar, S. (2018). Light Affects Mood and Learning through Distinct Retina-Brain Pathways. *Cell*, *175*(1), 71-84.e18. https://doi.org/10.1016/j.cell.2018.08.004

Figueiro, M. G., Plitnick, B., Roohan, C., Sahin, L., Kalsher, M., & Rea, M. S. (2019). Effects of a Tailored Lighting Intervention on Sleep Quality, Rest–Activity, Mood, and Behavior in Older Adults With Alzheimer Disease and Related Dementias: A Randomized Clinical Trial. *Journal of Clinical Sleep Medicine*, *15*(12), 1757–1767. https://doi.org/10.5664/jcsm.8078

Fleiss, J. L., Wallenstein, S., & Rosenfeld, R. (1985). Adjusting for baseline measurements in the two-period crossover study: A cautionary note. *Controlled Clinical Trials*, *6*(3), 192–197. https://doi.org/10.1016/0197-2456(85)90002-9

Flynn-Evans, E. E., Shekleton, J. A., Miller, B., Epstein, L. J., Kirsch, D., Brogna, L. A., Burke, L. M., Bremer, E., Murray, J. M., Gehrman, P., Rajaratnam, S. M. W., & Lockley, S. W. (2017). Circadian Phase and Phase Angle Disorders in Primary Insomnia. *Sleep*, *40*(12). https://doi.org/10.1093/sleep/zsx163

Forsell, E., Jernelöv, S., Blom, K., Kraepelien, M., Svanborg, C., Andersson, G., Lindefors, N., & Kaldo, V. (2019). Proof of Concept for an Adaptive Treatment Strategy to Prevent Failures in Internet-Delivered CBT: A Single-Blind Randomized Clinical Trial With Insomnia Patients. *American Journal of Psychiatry*, *176*(4), 315–323. https://doi.org/10.1176/appi.ajp.2018.18060699

Foster, R. G., & Wulff, K. (2005). The rhythm of rest and excess. *Nature Reviews Neuroscience*, *6*(5), 407–414. https://doi.org/10.1038/nrn1670

Freedman, M. S., Lucas, R. J., Soni, B., von Schantz, M., Muñoz, M., David-Gray, Z., & Foster, R. (1999). Regulation of Mammalian Circadian Behavior by Non-rod, Non-cone, Ocular Photoreceptors. *Science*, *284*(5413), 502–504. https://doi.org/10.1126/science.284.5413.502

Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L.-M., Nickless, A., Harrison, P. J., Emsley, R., Luik, A. I., Foster, R. G., Wadekar, V., Hinds, C., Gumley, A., Jones, R., Lightman, S., Jones, S., Bentall, R., Kinderman, P., Rowse, G., Brugha, T., … Espie, C. A. (2017). The effects of improving sleep on mental health (OASIS): A randomised controlled trial with mediation analysis. *The Lancet Psychiatry*, *4*(10), 749–758. https://doi.org/10.1016/S2215-0366(17)30328-0

Freeman, D., Sheaves, B., Waite, F., Harvey, A. G., & Harrison, P. J. (2020). Sleep disturbance and psychiatric disorders. *The Lancet Psychiatry*, *7*(7), 628–637. https://doi.org/10.1016/S2215-0366(20)30136-X

Freeman, P. R. (1989). The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Statistics in Medicine*, *8*(12), 1421–1432. https://doi.org/10.1002/sim.4780081202

Gao, Y., Ge, L., Liu, M., Niu, M., Chen, Y., Sun, Y., Chen, J., Yao, L., Wang, Q., Li, Z., Xu, J., Li, M., Hou, L., Shi, J., Yang, K., Cai, Y., Li, L., Zhang, J., & Tian, J. (2022). Comparative efficacy and acceptability of cognitive behavioral therapy delivery formats for insomnia in adults: A systematic review and network meta-analysis. *Sleep Medicine Reviews*, *64*, 101648. https://doi.org/10.1016/j.smrv.2022.101648

Garland, S. N., Carlson, L. E., Stephens, A. J., Antle, M. C., Samuels, C., & Campbell, T. S. (2014). Mindfulness-Based Stress Reduction Compared With Cognitive Behavioral Therapy for the Treatment of Insomnia Comorbid With Cancer: A Randomized, Partially Blinded, Noninferiority Trial. *Journal of Clinical Oncology*, *32*(5), 449–457. https://doi.org/10.1200/JCO.2012.47.7265

Gaus, S. E., Strecker, R. E., Tate, B. A., Parker, R. A., & Saper, C. B. (2002). Ventrolateral preoptic nucleus contains sleep-active, galaninergic neurons in multiple mammalian species. *Neuroscience*, *115*(1), 285–294. https://doi.org/10.1016/S0306-4522(02)00308-1

Gavriloff, D., Sheaves, B., Juss, A., Espie, C. A., Miller, C. B., & Kyle, S. D. (2018). Sham sleep feedback delivered via actigraphy biases daytime symptom reports in people with insomnia: Implications for insomnia disorder and wearable devices. *Journal of Sleep Research*, *27*(6), e12726. https://doi.org/10.1111/jsr.12726

Gehrman, P., Barilla, H., Medvedeva, E., Bellamy, S., O'Brien, E., & Kuna, S. T. (2020). Randomized trial of telehealth delivery of cognitive-behavioral treatment for insomnia vs. In-person treatment in veterans with PTSD. *Journal of Affective Disorders Reports*, *1*, 100018. https://doi.org/10.1016/j.jadr.2020.100018

Gehrman, P., Gunter, P., Findley, J., Frasso, R., Weljie, A. M., Kuna, S. T., & Kayser, M. S. (2021). Randomized Noninferiority Trial of Telehealth Delivery of Cognitive Behavioral Treatment of Insomnia Compared to In-Person Care. *The Journal of Clinical Psychiatry*, *82*(5). https://doi.org/10.4088/JCP.20m13723

Giles, D. E., & Roffwarg, H. P. (1998). Controlled Comparison of Electrophysiological Sleep in Families of Probands With Unipolar Depression. *Am J Psychiatry*, 8.

Giménez, M. C., Geerdinck, L. M., Versteylen, M., Leffers, P., Meekes, G. J. B. M., Herremans, H., de Ruyter, B., Bikker, J. W., Kuijpers, P. M. J. C., & Schlangen, L. J. M. (2017). Patient room lighting influences on sleep, appraisal and mood in hospitalized people. *Journal of Sleep Research*, *26*(2), 236–246. https://doi.org/10.1111/jsr.12470

Gooley, J. J., Lu, J., Fischer, D., & Saper, C. B. (2003). A Broad Role for Melanopsin in Nonvisual Photoreception. *The Journal of Neuroscience*, *23*(18), 7093–7106. https://doi.org/10.1523/JNEUROSCI.23-18-07093.2003

Gordijn, M. C. M., Beersma, D. G. M., Korte, H. J., & Hoofdakker, R. H. (1999). Effects of light exposure and sleep displacement on dim light melatonin onset. *Journal of Sleep Research*, *8*(3), 163–174. https://doi.org/10.1046/j.1365-2869.1999.00156.x

Gottlieb, J. F., Benedetti, F., Geoffroy, P. A., Henriksen, T. E. G., Lam, R. W., Murray, G., Phelps, J., Sit, D., Swartz, H. A., Crowe, M., Etain, B., Frank, E., Goel, N., Haarman, B. C. M., Inder, M., Kallestad, H., Jae Kim, S., Martiny, K., Meesters, Y., … Chen, S. (2019). The chronotherapeutic treatment of bipolar disorders: A systematic review and practice recommendations from the ISBD task force on chronotherapy and chronobiology. *Bipolar Disorders*, *21*(8), 741–773. https://doi.org/10.1111/bdi.12847

Grissom, R. J., & Kim, J. J. (2012). *Effect sizes for research: Univariate and multivariate applications*. Routledge.

Grizzle, J. E. (1965). The Two-Period Change-Over Design and Its Use in Clinical Trials. *Biometrics*, *21*(2), 467. https://doi.org/10.2307/2528104

Grønli, J., Byrkjedal, I. K., Bjorvatn, B., Nødtvedt, Ø., Hamre, B., & Pallesen, S. (2016). Reading from an iPad or from a book in bed: The impact on human sleep. A randomized controlled crossover trial. *Sleep Medicine*, *21*, 86–92. https://doi.org/10.1016/j.sleep.2016.02.006

Guarana, C. L., Barnes, C. M., & Ong, W. J. (2021). The effects of blue-light filtration on sleep and work outcomes. *Journal of Applied Psychology*, *106*(5), 784–796. https://doi.org/10.1037/apl0000806

Guler, A. D., Ecker, J. L., Lall, G. S., Haq, S., Altimus, C. M., Liao, H.-W., Barnard, A. R., Cahill, H., Badea, T. C., Zhao, H., Hankins, M. W., Berson, D. M., Lucas, R. J., Yau, K.-W., & Hattar, S. (2008). *Melanopsin cells are the principal conduits for rod–cone input to non-image-forming vision*. *453*, 5.

Gupta, M. A., & Simpson, F. C. (2015). Obstructive Sleep Apnea and Psychiatric Disorders: A Systematic Review. *Journal of Clinical Sleep Medicine*, *11*(02), 165–175. https://doi.org/10.5664/jcsm.4466

Gustavsson, A. (2010). *Cost of disorders of the brain in Europe 2010*. 62.

Hafner, M., Stepanek, M., Taylor, J., Troxel, W., & Stolk, C. (2016). *Why sleep matters -- the economic costs of insufficient sleep: A cross-country comparative analysis*. RAND Corporation. https://doi.org/10.7249/RR1791

Haghayegh, S., Khoshnevis, S., Smolensky, M. H., Diller, K. R., & Castriotta, R. J. (2019). Performance comparison of different interpretative algorithms utilized to derive sleep parameters from wrist actigraphy data. *Chronobiology International*, *36*(12), 1752–1760. https://doi.org/10.1080/07420528.2019.1679826

Hale, L., & Guan, S. (2015). Screen time and sleep among school-aged children and adolescents: A systematic literature review. *Sleep Medicine Reviews*, *21*, 50–58. https://doi.org/10.1016/j.smrv.2014.07.007

Hale, L., Kirschen, G. W., LeBourgeois, M. K., Gradisar, M., Garrison, M. M., Montgomery-Downs, H., Kirschen, H., McHale, S. M., Chang, A.-M., & Buxton, O. M. (2018). Youth Screen Media Habits and Sleep. *Child and Adolescent Psychiatric Clinics of North America*, *27*(2), 229–245. https://doi.org/10.1016/j.chc.2017.11.014

Hallam, K. T., Begg, D. P., Olver, J. S., & Norman, T. R. (2009). Abnormal dose-response melatonin suppression by light in bipolar type I patients compared with healthy adult subjects. *Acta Neuropsychiatrica*, *21*(5), 246–255. Cambridge Core. https://doi.org/10.1111/j.1601-5215.2009.00416.x

Hannibal, J., Kankipati, L., Strang, C. E., Peterson, B. B., Dacey, D., & Gamlin, P. D. (2015). *Central projections of intrinsically photosensitive retinal ganglion cells in the macaque monkey*. 25.

Hasler, B. P., Buysse, D. J., Kupfer, D. J., & Germain, A. (2010). Phase relationships between core body temperature, melatonin, and sleep are associated with depression severity: Further evidence for circadian misalignment in non-seasonal depression. *Psychiatry Research*, *178*(1), 205–207. https://doi.org/10.1016/j.psychres.2010.04.027

Hatori, M., Gronfier, C., Van Gelder, R. N., Bernstein, P. S., Carreras, J., Panda, S., Marks, F., Sliney, D., Hunt, C. E., Hirota, T., Furukawa, T., & Tsubota, K. (2017). Global rise

of potential health hazards caused by blue light-induced circadian disruption in modern aging societies. *Npj Aging and Mechanisms of Disease*, *3*(1), 9. https://doi.org/10.1038/s41514-017-0010-2

Hattar, S., Kumar, M., Park, A., Tong, P., Tung, J., Yau, K.-W., & Berson, D. M. (2006). *Central projections of melanopsin‐expressing retinal ganglion cells in the mouse*. 24.

Hattar, S., Liao, H.-W., Takao, M., Berson, D. M., & Yau, K.-W. (2002). Melanopsin-Containing Retinal Ganglion Cells: Architecture, Projections, and Intrinsic Photosensitivity. *Science*, *295*(5557), 1065–1070. https://doi.org/10.1126/science.1069609

Hébert, M., Martin, S. K., Lee, C., & Eastman, C. I. (2002). The effects of prior light history on the suppression of melatonin by light in humans. *Journal of Pineal Research*, *33*(4), 198–203. https://doi.org/10.1034/j.1600-079X.2002.01885.x

Henriksen, T., Grønli, J., Assmus, J., Fasmer, O. B., Schoeyen, H., Leskauskaite, I., Bjorke-Bertheussen, J., Ytrehus, K., & Lund, A. (2020). Blue-blocking glasses as additive treatment for mania: Effects on actigraphy-derived sleep parameters. *Journal of Sleep Research*. https://doi.org/10.1111/jsr.12984

Henriksen, T., Skrede, S., Fasmer, O. B., Schoeyen, H., Leskauskaite, I., Bjørke-Bertheussen, J., Assmus, J., Hamre, B., Grønli, J., & Lund, A. (2016). Blue-blocking glasses as additive treatment for mania: A randomized placebo-controlled trial. *Bipolar Disorders*, *18*(3), 221–232. https://doi.org/10.1111/bdi.12390

Hertenstein, E., Feige, B., Gmeiner, T., Kienzler, C., Spiegelhalder, K., Johann, A., Jansson-Fröjmark, M., Palagini, L., Rücker, G., Riemann, D., & Baglioni, C. (2019). Insomnia as a predictor of mental disorders: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *43*, 96–105. https://doi.org/10.1016/j.smrv.2018.10.006

Hester, L., Dang, D., Barker, C. J., Heath, M., Mesiya, S., Tienabeso, T., & Watson, K. (2021). Evening wear of blue-blocking glasses for sleep and mood disorders: A systematic review. *Chronobiology International*, *38*(10), 1375–1383. https://doi.org/10.1080/07420528.2021.1930029

Hills, R. K. (2017). Non-inferiority trials: No better? No worse? No change? No pain? *British Journal of Haematology*, *176*(6), 883–887. https://doi.org/10.1111/bjh.14504

Homack, S., & Riccio, C. A. (2006). Conners' Continuous Performance Test (2nd ed.; CCPT-II). *Journal of Attention Disorders*, *9*(3), 556–558. https://doi.org/10.1177/1087054705283578

Hope, J., Recio-Saucedo, A., Fogg, C., Griffiths, P., Smith, G. B., Westwood, G., & Schmidt, P. E. (2018). A fundamental conflict of care: Nurses' accounts of balancing patients' sleep with taking vital sign observations at night. *Journal of Clinical Nursing*, *27*(9–10), 1860–1871. https://doi.org/10.1111/jocn.14234

Huang, Y., Mayer, C., Cheng, P., Siddula, A., Burgess, H. J., Drake, C., Goldstein, C., Walch, O., & Forger, D. B. (2021). Predicting circadian phase across populations: A

comparison of mathematical models and wearable devices. *Sleep*, *44*(10), zsab126. https://doi.org/10.1093/sleep/zsab126

Hull, J. T., Czeisler, C. A., & Lockley, S. W. (2018). Suppression of Melatonin Secretion in Totally Visually Blind People by Ocular Exposure to White Light. *Ophthalmology*, *125*(8), 1160–1171. https://doi.org/10.1016/j.ophtha.2018.01.036

Hunt, K., & Emslie, C. (2001). Commentary: The prevention paradox in lay epidemiology— Rose revisited. *International Journal of Epidemiology*, *30*(3), 442–446. https://doi.org/10.1093/ije/30.3.442

Ibuka, N., & Kawamura, H. (1975). Loss of circadian rhythm in sleep-wakefulness cycle in the rat by suprachiasmatic nucleus lesions. *Brain Research*, *96*(1), 76–81. https://doi.org/10.1016/0006-8993(75)90574-0

J, T., L, B., T, H., J, R., M, W., & M, H. (2018). Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trials Communications*, *10*, 80–85. https://doi.org/10.1016/j.conctc.2018.03.008

Janků, K., Šmotek, M., Fárková, E., & Kopřivová, J. (2020). Block the light and sleep well: Evening blue light filtration as a part of cognitive behavioral therapy for insomnia. *Chronobiology International*, *37*(2), 248–259. https://doi.org/10.1080/07420528.2019.1692859

Jansson, A., & Adams, P. C. (2021). *Disentangling: The Geographies of Digital Disconnection*. Oxford University Press. https://books.google.no/books?id=H7o0EAAAQBAJ

Jouvet, M., Michel, F., & Courjon, J. (1959). Sur un stade d'activité électrique cérébrale rapide au cours du sommeil physiologique. *CR Soc Biol*, *153*, 1024–1028.

Kales, A., Rechtschaffen, A., University of California, L. Angeles. B. I. S., & NINDB Neurological Information Network (U.S.). (1968). *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects: Allan Rechtschaffen and Anthony Kales, Editors*. U. S. National Institute of Neurological Diseases and Blindness, Neurological Information Network. https://books.google.no/books?id=Z41IvQEACAAJ

Kathol, R. G., & Arnedt, J. T. (2016). Cognitive Behavioral Therapy for Chronic Insomnia: Confronting the Challenges to Implementation. *Annals of Internal Medicine*, *165*(2), 149. https://doi.org/10.7326/M16-0359

Kaur, S., Thankachan, S., Begum, S., Liu, M., Blanco-Centurion, C., & Shiromani, P. J. (2009). Hypocretin-2 Saporin Lesions of the Ventrolateral Periaquaductal Gray (vl PAG) Increase REM Sleep in Hypocretin Knockout Mice. *PLoS ONE*, *4*(7), 19.

Kayumov, L., Casper, R. F., Hawa, R. J., Perelman, B., Chung, S. A., Sokalsky, S., & Shapiro, C. M. (2005). Blocking Low-Wavelength Light Prevents Nocturnal Melatonin Suppression with No Adverse Effect on Performance during Simulated

Shift Work. *The Journal of Clinical Endocrinology & Metabolism*, *90*(5), 2755–2761. https://doi.org/10.1210/jc.2004-2062

Kecklund, G., & Axelsson, J. (2016). Health consequences of shift work and insufficient sleep. *BMJ*, i5210. https://doi.org/10.1136/bmj.i5210

Khalsa, S. B. S., Jewett, M. E., Cajochen, C., & Czeisler, C. A. (2003). A Phase Response Curve to Single Bright Light Pulses in Human Subjects. *The Journal of Physiology*, *549*(3), 945–952. https://doi.org/10.1113/jphysiol.2003.040477

Kjørstad, K., Faaland, P. M., Sivertsen, B., Kallestad, H., Langsrud, K., Vethe, D., Vestergaard, C. L., Harris, A., Pallesen, S., Scott, J., & Vedaa, Ø. (2022). Sleep and work functioning in nurses undertaking inpatient shifts in a blue-depleted light environment. *BMC Nursing*, *21*(1), 187. https://doi.org/10.1186/s12912-022-00973-4

Klem, G., Lüders, H., Jasper, H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr Clin Neurophysiol Suppl*, *52*, 3–6. PubMed.

Ko, P.-R. T., Kientz, J. A., Choe, E. K., Kay, M., Landis, C. A., & Watson, N. F. (2015). Consumer Sleep Technologies: A Review of the Landscape. *Journal of Clinical Sleep Medicine*, *11*(12), 7.

Komada, Y., Tanaka, H., Yamamoto, Y., Shirakawa, S., & Yamazaki, K. (2000). Effects of bright light pre-exposure on sleep onset process. *Psychiatry and Clinical Neurosciences*, *54*(3), 365–366. https://doi.org/10.1046/j.1440-1819.2000.00717.x

Konkani, A., & Oakley, B. (2012). Noise in hospital intensive care units—A critical review of a critical topic. *Journal of Critical Care*, *27*(5), 522.e1-522.e9. https://doi.org/10.1016/j.jcrc.2011.09.003

Koritala, B. S. C., & Çakmaklı, S. (2018). The human circadian clock from health to economics: The circadian clock: Health and economics. *PsyCh Journal*, *7*(4), 176–196. https://doi.org/10.1002/pchj.252

Kripke, D. F., Hahn, E. K., Grizas, A. P., Wadiak, K. H., Loving, R. T., & Poceta, J. S. (2010). Wrist actigraphic scoring for sleep laboratory patients: Algorithm development. *J. Sleep Res.*, 8.

Kripke, D. F., Mullaney, D. J., Messin, S., & Wyborney, V. G. (1978). Wrist actigraphic measures of sleep and rhythms. *Electroencephalography and Clinical Neurophysiology*, *44*(5), 674–676. https://doi.org/10.1016/0013-4694(78)90133-5

Kupfer, DavidJ., & Foster, F. G. (1972). INTERVAL BETWEEN ONSET OF SLEEP AND RAPID-EYE-MOVEMENT SLEEP AS AN INDICATOR OF DEPRESSION. *The Lancet*, *300*(7779), 684–686. https://doi.org/10.1016/S0140-6736(72)92090-9

Kushida, C. A., Chang, A., Gadkary, C., Guilleminault, C., Carrillo, O., & Dement, W. C. (2001). Comparison of actigraphic, polysomnographic, and subjective assessment of

sleep parameters in sleep-disordered patients. *Sleep Medicine*, *2*(5), 389–396. https://doi.org/10.1016/S1389-9457(00)00098-8

Lancee, J., Eisma, M. C., van Straten, A., & Kamphuis, J. H. (2015). Sleep-Related Safety Behaviors and Dysfunctional Beliefs Mediate the Efficacy of Online CBT for Insomnia: A Randomized Controlled Trial. *Cognitive Behaviour Therapy*, *44*(5), 406–422. https://doi.org/10.1080/16506073.2015.1026386

Lancee, J., van Straten, A., Morina, N., Kaldo, V., & Kamphuis, J. H. (2016). Guided Online or Face-to-Face Cognitive Behavioral Treatment for Insomnia: A Randomized Wait-List Controlled Trial. *Sleep*, *39*(1), 183–191. https://doi.org/10.5665/sleep.5344

Langsrud, K., Vaaler, A. E., Kallestad, H., & Morken, G. (2016). Sleep patterns as a predictor for length of stay in a psychiatric intensive care unit. *Psychiatry Research*, *237*, 252–256. https://doi.org/10.1016/j.psychres.2016.01.032

Laugsand, L. E., Strand, L. B., Vatten, L. J., Janszky, I., & Bjørngaard, J. H. (2014). Insomnia Symptoms and Risk for Unintentional Fatal Injuries—The HUNT Study. *Sleep*, *37*(11), 1777–1786. https://doi.org/10.5665/sleep.4170

Lechinger, J., Koch, J., Weinhold, S. L., Seeck-Hirschner, M., Stingele, K., Kropp-Näf, C., Braun, M., Drews, H. J., Aldenhoff, J., Huchzermeier, C., & Göder, R. (2021). REM density is associated with treatment response in major depression: Antidepressant pharmacotherapy vs. psychotherapy. *Journal of Psychiatric Research*, *133*, 67–72. https://doi.org/10.1016/j.jpsychires.2020.12.009

Lee, Y., Park, J.-Y., Choi, Y.-W., Park, H.-K., Cho, S.-H., Cho, S. H., & Lim, Y.-H. (2018). A Novel Non-contact Heart Rate Monitor Using Impulse-Radio Ultra-Wideband (IR-UWB) Radar Technology. *Scientific Reports*, *8*(1), 13053. https://doi.org/10.1038/s41598-018-31411-8

Leerssen, J., Lakbila-Kamal, O., Dekkers, L. M. S., Ikelaar, S. L. C., Albers, A. C. W., Blanken, T. F., Lancee, J., van der Lande, G. J. M., Maksimovic, T., Mastenbroek, S. E., Reesen, J. E., van de Ven, S., van der Zweerde, T., Foster-Dingley, J. C., & Van Someren, E. J. W. (2022). Treating Insomnia with High Risk of Depression Using Therapist-Guided Digital Cognitive, Behavioral, and Circadian Rhythm Support Interventions to Prevent Worsening of Depressive Symptoms: A Randomized Controlled Trial. *Psychotherapy and Psychosomatics*, *91*(3), 168–179. https://doi.org/10.1159/000520282

LeGates, T. A., Fernandez, D. C., & Hattar, S. (2014). Light as a central modulator of circadian rhythms, sleep and affect. *Nature Reviews Neuroscience*, *15*(7), 443–454. https://doi.org/10.1038/nrn3743

Lewy, A. J., Wehr, T. A., Goodwin, F. K., Newsome, D. A., & Markey, S. P. (1980). Light Suppresses Melatonin Secretion in Humans. *Science*, *210*(4475), 1267–1269. https://doi.org/10.1126/science.7434030

Li, T., Yu, T., Hawkins, B. S., & Dickersin, K. (2015). Design, Analysis, and Reporting of Crossover Trials for Inclusion in a Meta-Analysis. *PLOS ONE*, *10*(8), e0133023. https://doi.org/10.1371/journal.pone.0133023

Lingjaerde, O., Ahlfors, U. G., Bech, P., Dencker, S. J., & Elgen, K. (1987). The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatrica Scandinavica. Supplementum*, *334*, 1–100.

Lu, J., Bjorkum, A. A., Xu, M., Gaus, S. E., Shiromani, P. J., & Saper, C. B. (2002). Selective Activation of the Extended Ventrolateral Preoptic Nucleus during Rapid Eye Movement Sleep. *The Journal of Neuroscience*, *22*(11), 4568–4576. https://doi.org/10.1523/JNEUROSCI.22-11-04568.2002

Lu, J., Sherman, D., Devor, M., & Saper, C. B. (2006). *A putative flip–flop switch for control of REM sleep*. *441*, 6.

Lucas, R. J., Freedman, M. S., Munoz, M., Garcia-Fernandez, J.-M., & Foster, R. G. (1999). *Regulation of the Mammalian Pineal by Non-rod, Non-cone, Ocular Photoreceptors*. *284*, 4.

Lucas, R. J., Peirson, S. N., Berson, D. M., Brown, T. M., Cooper, H. M., Czeisler, C. A., Figueiro, M. G., Gamlin, P. D., Lockley, S. W., O'Hagan, J. B., Price, L. L. A., Provencio, I., Skene, D. J., & Brainard, G. C. (2014). Measuring and using light in the melanopsin age. *Trends in Neurosciences*, *37*(1), 1–9. https://doi.org/10.1016/j.tins.2013.10.004

Luik, A. I., Kyle, S. D., & Espie, C. A. (2017). Digital Cognitive Behavioral Therapy (dCBT) for Insomnia: A State-of-the-Science Review. *Current Sleep Medicine Reports*, *3*(2), 48–56. https://doi.org/10.1007/s40675-017-0065-4

Luik, A. I., van der Zweerde, T., van Straten, A., & Lancee, J. (2019). Digital Delivery of Cognitive Behavioral Therapy for Insomnia. *Current Psychiatry Reports*, *21*(7), 50. https://doi.org/10.1007/s11920-019-1041-0

Luo, C., Sanger, N., Singhal, N., Pattrick, K., Shams, I., Shahid, H., Hoang, P., Schmidt, J., Lee, J., Haber, S., Puckering, M., Buchanan, N., Lee, P., Ng, K., Sun, S., Kheyson, S., Chung, D. C.-Y., Sanger, S., Thabane, L., & Samaan, Z. (2020). A comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: A systematic review and meta-analysis. *EClinicalMedicine*, *24*, 100442. https://doi.org/10.1016/j.eclinm.2020.100442

MacKenzie, D. J., & Galbrun, L. (2007). Noise levels and noise sources in acute care hospital wards. *Building Services Engineering Research and Technology*, *28*(2), 117–131. https://doi.org/10.1177/0143624406074468

Manber, R., Edinger, J. D., Gress, J. L., Pedro-Salcedo, M. G. S., Kuo, T. F., & Kalista, T. (2008). Cognitive Behavioral Therapy for Insomnia Enhances Depression Outcome in Patients with Comorbid Major Depressive Disorder and Insomnia. *Sleep*, *31*(4), 489–495. https://doi.org/10.1093/sleep/31.4.489

Maurer, L. F., Espie, C. A., Omlin, X., Emsley, R., & Kyle, S. D. (2022). The effect of sleep restriction therapy for insomnia on sleep pressure and arousal: A randomized controlled mechanistic trial. *Sleep*, *45*(1), zsab223. https://doi.org/10.1093/sleep/zsab223

McGlashan, E. M., Coleman, M. Y., Vidafar, P., Phillips, A. J. K., & Cain, S. W. (2019). Decreased sensitivity of the circadian system to light in current, but not remitted depression. *Journal of Affective Disorders*, *256*, 386–392. https://doi.org/10.1016/j.jad.2019.05.076

McGlashan, E. M., Nandam, L. S., Vidafar, P., Mansfield, D. R., Rajaratnam, S. M. W., & Cain, S. W. (2018). The SSRI citalopram increases the sensitivity of the human circadian system to light in an acute dose. *Psychopharmacology*, *235*(11), 3201–3209. https://doi.org/10.1007/s00213-018-5019-0

McLaren, L., McIntyre, L., & Kirkpatrick, S. (2010). Rose's population strategy of prevention need not increase social inequalities in health. *International Journal of Epidemiology*, *39*(2), 372–377. https://doi.org/10.1093/ije/dyp315

Meaklim, H., Jackson, M. L., Bartlett, D., Saini, B., Falloon, K., Junge, M., Slater, J., Rehm, I. C., & Meltzer, L. J. (2020). Sleep education for healthcare providers: Addressing deficient sleep in Australia and New Zealand. *Sleep Health*, *6*(5), 636–650. https://doi.org/10.1016/j.sleh.2020.01.012

Meltzer, L. J., Phillips, C., & Mindell, J. A. (2009). Clinical psychology training in sleep and sleep disorders. *Journal of Clinical Psychology*, *65*(3), 305–318. https://doi.org/10.1002/jclp.20545

Meyer, N., Harvey, A. G., Lockley, S. W., & Dijk, D.-J. (2022). Circadian rhythms and disorders of the timing of sleep. *The Lancet*, *400*(10357), 1061–1078. https://doi.org/10.1016/S0140-6736(22)00877-7

Mitchell, I. A., McKay, H., Van Leuvan, C., Berry, R., McCutcheon, C., Avard, B., Slater, N., Neeman, T., & Lamberth, P. (2010). A prospective controlled trial of the effect of a multi-faceted intervention on early recognition and intervention in deteriorating hospital patients. *Resuscitation*, *81*(6), 658–666. https://doi.org/10.1016/j.resuscitation.2010.03.001

Monti, J. M., BaHammam, A. S., Pandi-Perumal, S. R., Bromundt, V., Spence, D. W., Cardinali, D. P., & Brown, G. M. (2013). Sleep and circadian rhythm dysregulation in schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *43*, 209–216. https://doi.org/10.1016/j.pnpbp.2012.12.021

Morin, C. M. (1993). *Insomnia: Psychological assessment and management.* (pp. xvii, 238). Guilford Press.

Morin, C. M., Belleville, G., Bélanger, L., & Ivers, H. (2011). The Insomnia Severity Index: Psychometric Indicators to Detect Insomnia Cases and Evaluate Treatment Response. *Sleep*, *34*(5), 601–608. https://doi.org/10.1093/sleep/34.5.601

Morin, C. M., Jarrin, D. C., Ivers, H., Mérette, C., LeBlanc, M., & Savard, J. (2020). Incidence, Persistence, and Remission Rates of Insomnia Over 5 Years. *JAMA Network Open*, *3*(11), e2018782. https://doi.org/10.1001/jamanetworkopen.2020.18782

Morin, C. M., Vallières, A., & Ivers, H. (2007). Dysfunctional Beliefs and Attitudes about Sleep (DBAS): Validation of a Brief Version (DBAS-16). *Sleep*, *30*(11), 1547–1554. https://doi.org/10.1093/sleep/30.11.1547

Mouret, J., Coindet, J., & Chouvet, G. (1978). *SUPRACHIASMATIC NUCLEI LESIONS IN THE RAT: ALTERATIONS IN SLEEP CIRCADIAN RHYTHMS*. 7.

Mullaney, D. J., Kripke, D. F., & Messin, S. (1980). Wrist-Actigraphic Estimation of Sleep Time. *Sleep*, *3*(1), 83–92. https://doi.org/10.1093/sleep/3.1.83

Münch, M., Kobialka, S., Steiner, R., Oelhafen, P., Wirz-Justice, A., & Cajochen, C. (2006). Wavelength-dependent effects of evening light exposure on sleep architecture and sleep EEG power density in men. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *290*(5), R1421–R1428. https://doi.org/10.1152/ajpregu.00478.2005

Münch, M., Scheuermaier, K. D., Zhang, R., Dunne, S. P., Guzik, A. M., Silva, E. J., Ronda, J. M., & Duffy, J. F. (2011). Effects on subjective and objective alertness and sleep in response to evening light exposure in older subjects. *Behavioural Brain Research*, *224*(2), 272–278. https://doi.org/10.1016/j.bbr.2011.05.029

Münch, M., Wirz-Justice, A., Brown, S. A., Kantermann, T., Martiny, K., Stefani, O., Vetter, C., Wright Jr, K. P., Wulff, K., & Skene, D. J. (2020). The Role of Daylight for Humans: Gaps in Current Knowledge. *Clocks & Sleep*, *2*(1), 61–85. https://doi.org/10.3390/clockssleep2010008

Newell, J. (2012). Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples. *Psychiatry Research*, 7.

Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of Changes in Health-related Quality of Life: The Remarkable Universality of Half a Standard Deviation. *Medical Care*, *41*(5), 582–592. https://doi.org/10.1097/01.MLR.0000062554.74615.4C

Nowozin, C., Wahnschaffe, A., Rodenbeck, A., Zeeuw, J. de, Hädel, S., Kozakov, R., Schöpp, H., Münch, M., & Kunz, D. (2017). Applying Melanopic Lux to Measure Biological Light Effects on Melatonin Suppression and Subjective Sleepiness. *Current Alzheimer Research*, *14*(10). https://doi.org/10.2174/1567205014666170523094526

O'Hare, E., Flanagan, D., Penzel, T., Garcia, C., Frohberg, D., & Heneghan, C. (2015). A comparison of radio-frequency biomotion sensors and actigraphy versus polysomnography for the assessment of sleep in normal subjects. *Sleep and Breathing*, *19*(1), 91–98. https://doi.org/10.1007/s11325-014-0967-z

Okkels, N., Jensen, L. G., Skovshoved, L. C., Arendt, R., Blicher, A. B., Vieta, E., & Straszek, S. (2019). Lighting as an aid for recovery in hospitalized psychiatric patients: A randomized controlled effectiveness trial. *Nordic Journal of Psychiatry*, *74*(2), 105–114. https://doi.org/10.1080/08039488.2019.1676465

Ostrin, L. A., Abbott, K. S., & Queener, H. M. (2017). Attenuation of short wavelengths alters sleep and the ip RGC pupil response. *Ophthalmic and Physiological Optics*, *37*(4), 440–450. https://doi.org/10.1111/opo.12385

Palagini, L., Baglioni, C., Ciapparelli, A., Gemignani, A., & Riemann, D. (2013). REM sleep dysregulation in depression: State of the art. *Sleep Medicine Reviews*, *17*(5), 377–390. https://doi.org/10.1016/j.smrv.2012.11.001

Palagini, L., Biber, K., & Riemann, D. (2014). The genetics of insomnia – Evidence for epigenetic mechanisms? *Sleep Medicine Reviews*, *18*(3), 225–235. https://doi.org/10.1016/j.smrv.2013.05.002

Pallesen, S., Nordhus, I. H., Nielsen, G. H., Havik, O. E., Kvale, G., Johnsen, B. H., & Skjøtskift, S. (2001). Prevalence of Insomnia in the Adult Norwegian Population. *Sleep*, *24*(7), 771–779. https://doi.org/10.1093/sleep/24.7.771

Pallesen, S., Sivertsen, B., Nordhus, I. H., & Bjorvatn, B. (2014). A 10-year trend of insomnia prevalence in the adult Norwegian population. *Sleep Medicine*, *15*(2), 173–179. https://doi.org/10.1016/j.sleep.2013.10.009

Papamichael, C., Skene, D. J., & Revell, V. L. (2012). Human Nonvisual Responses to Simultaneous Presentation of Blue and Red Monochromatic Light. *Journal of Biological Rhythms*, *27*(1), 70–78. https://doi.org/10.1177/0748730411431447

Paquet, J., Kawinska, A., & Carrier, J. (2007). Wake Detection Capacity of Actigraphy During Sleep. *Sleep*, *30*(10), 1362–1369. https://doi.org/10.1093/sleep/30.10.1362

Parsa-Parsi, R. W. (2017). The Revised Declaration of Geneva: A Modern-Day Physician's Pledge. *JAMA*, *318*(20), 1971. https://doi.org/10.1001/jama.2017.16230

Parsons, C. E., Zachariae, R., Landberger, C., & Young, K. S. (2021). How does cognitive behavioural therapy for insomnia work? A systematic review and meta-analysis of mediators of change. *Clinical Psychology Review*, *86*, 102027. https://doi.org/10.1016/j.cpr.2021.102027

Paunio, T., Korhonen, T., Hublin, C., Partinen, M., Koskenvuo, K., Koskenvuo, M., & Kaprio, J. (2015). Poor sleep predicts symptoms of depression and disability retirement due to depression. *Journal of Affective Disorders*, *172*, 381–389. https://doi.org/10.1016/j.jad.2014.10.002

Perlis, M. L., Giles, D. E., Buysse, D. J., Tu, X., & Kupfer, D. J. (1997). Self-reported sleep disturbance as a prodromal symptom in recurrent depression. *Journal of Affective Disorders*, *42*(2–3), 209–212. https://doi.org/10.1016/S0165-0327(96)01411-5

Perlis, M. L., Posner, D., Riemann, D., Bastien, C. H., Teel, J., & Thase, M. (2022). Insomnia. *The Lancet*, *400*(10357), 1047–1060. https://doi.org/10.1016/S0140-6736(22)00879-0

Phelps, J. (2008). Dark therapy for bipolar disorder using amber lenses for blue light blockade. *Medical Hypotheses*, *70*(2), 224–229. https://doi.org/10.1016/j.mehy.2007.05.026

Phillips, A. J. K., Clerx, W. M., O'Brien, C. S., Sano, A., Barger, L. K., Picard, R. W., Lockley, S. W., Klerman, E. B., & Czeisler, C. A. (2017). Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Scientific Reports*, *7*(1), 3216. https://doi.org/10.1038/s41598-017-03171-4

Phillips, A. J. K., Vidafar, P., Burns, A. C., McGlashan, E. M., Anderson, C., Rajaratnam, S. M. W., Lockley, S. W., & Cain, S. W. (2019). High sensitivity and interindividual variability in the response of the human circadian system to evening light. *Proceedings of the National Academy of Sciences*, *116*(24), 12019–12024. https://doi.org/10.1073/pnas.1901824116

Piaggio, G., Elbourne, D., Pocock, S. J., Evans, S. J. W., & Altman, D. G. (2012). *Reporting of Noninferiority and Equivalence Randomized Trials: Extension of the CONSORT 2010 Statement*. 11.

Pillai, V., Roth, T., Roehrs, T., Moss, K., Peterson, E. L., & Drake, C. L. (2017). Effectiveness of Benzodiazepine Receptor Agonists in the Treatment of Insomnia: An Examination of Response and Remission Rates. *Sleep*, *40*(2). https://doi.org/10.1093/sleep/zsw044

Pisani, M. A., Friese, R. S., Gehlbach, B. K., Schwab, R. J., Weinhouse, G. L., & Jones, S. F. (2015). Sleep in the Intensive Care Unit. *American Journal of Respiratory and Critical Care Medicine*, *191*(7), 731–738. https://doi.org/10.1164/rccm.201411-2099CI

Poe, G. R., Foote, S., Eschenko, O., Johansen, J. P., Bouret, S., Aston-Jones, G., Harley, C. W., Manahan-Vaughan, D., Weinshenker, D., Valentino, R., Berridge, C., Chandler, D. J., Waterhouse, B., & Sara, S. J. (2020). Locus coeruleus: A new look at the blue spot. *Nature Reviews Neuroscience*, *21*(11), 644–659. https://doi.org/10.1038/s41583-020-0360-9

Prayag, A. S., Najjar, R. P., & Gronfier, C. (2019). Melatonin suppression is exquisitely sensitive to light and primarily driven by melanopsin in humans. *Journal of Pineal Research*, *66*(4), e12562. https://doi.org/10.1111/jpi.12562

Provencio, I., Rodriguez, I. R., Jiang, G., Hayes, W. P., Moreira, E. F., & Rollag, M. D. (2000). A Novel Human Opsin in the Inner Retina. *The Journal of Neuroscience*, *20*(2), 600–605. https://doi.org/10.1523/JNEUROSCI.20-02-00600.2000

Qaseem, A., Kansagara, D., Forciea, M. A., Cooke, M., Denberg, T. D., & for the Clinical Guidelines Committee of the American College of Physicians. (2016). Management of

Chronic Insomnia Disorder in Adults: A Clinical Practice Guideline From the American College of Physicians. *Annals of Internal Medicine*, *165*(2), 125. https://doi.org/10.7326/M15-2175

Rahman, S. A., Flynn-Evans, E. E., Aeschbach, D., Brainard, G. C., Czeisler, C. A., & Lockley, S. W. (2014). Diurnal Spectral Sensitivity of the Acute Alerting Effects of Light. *Sleep*, *37*(2), 271–281. https://doi.org/10.5665/sleep.3396

Rahman, S. A., St. Hilaire, M. A., Grant, L. K., Barger, L. K., Brainard, G. C., Czeisler, C. A., Klerman, E. B., & Lockley, S. W. (2022). Dynamic lighting schedules to facilitate circadian adaptation to shifted timing of sleep and wake. *Journal of Pineal Research*, *73*(1). https://doi.org/10.1111/jpi.12805

Rahman, S. A., St. Hilaire, M. A., & Lockley, S. W. (2017). The effects of spectral tuning of evening ambient light on melatonin suppression, alertness and sleep. *Physiology & Behavior*, *177*, 221–229. https://doi.org/10.1016/j.physbeh.2017.05.002

Ranganathan, P., Pramesh, C., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in Clinical Research*, *8*(4), 187. https://doi.org/10.4103/picr.PICR_123_17

Rehman, A., Waite, F., Sheaves, B., Biello, S., Freeman, D., & Gumley, A. (2017). Clinician perceptions of sleep problems, and their treatment, in patients with non-affective psychosis. *Psychosis*, *9*(2), 129–139. https://doi.org/10.1080/17522439.2016.1206955

Riemann, D., Baglioni, C., Bassetti, C., Bjorvatn, B., Dolenc Groselj, L., Ellis, J. G., Espie, C. A., Garcia-Borreguero, D., Gjerstad, M., Gonçalves, M., Hertenstein, E., Jansson-Fröjmark, M., Jennum, P. J., Leger, D., Nissen, C., Parrino, L., Paunio, T., Pevernagie, D., Verbraecken, J., … Spiegelhalder, K. (2017). European guideline for the diagnosis and treatment of insomnia. *Journal of Sleep Research*, *26*(6), 675–700. https://doi.org/10.1111/jsr.12594

Riemann, D., Berger, M., & Voderholzer, U. (2001). Sleep and depression — results from psychobiological studies: An overview. *Biological Psychology*, *57*(1–3), 67–103. https://doi.org/10.1016/S0301-0511(01)00090-4

Riemann, D., Krone, L. B., Wulff, K., & Nissen, C. (2020). Sleep, insomnia, and depression. *Neuropsychopharmacology*, *45*(1), 74–89. https://doi.org/10.1038/s41386-019-0411-y

Riemann, D., Spiegelhalder, K., Feige, B., Voderholzer, U., Berger, M., Perlis, M., & Nissen, C. (2010). The hyperarousal model of insomnia: A review of the concept and its evidence. *Sleep Medicine Reviews*, *14*(1), 19–31. https://doi.org/10.1016/j.smrv.2009.04.002

Riemann, D., Spiegelhalder, K., Nissen, C., Hirscher, V., Baglioni, C., & Feige, B. (2012). REM Sleep Instability – A New Pathway for Insomnia? *Pharmacopsychiatry*, s-0031-1299721. https://doi.org/10.1055/s-0031-1299721

Rigby, H. S., Warren, B. F., Diamond, J., Carter, C., & Bradfield, J. W. (1991). Colour perception in pathologists: The Farnsworth-Munsell 100-hue test. *Journal of Clinical Pathology*, *44*(9), 745–748. https://doi.org/10.1136/jcp.44.9.745

Ritterband, L. M., Thorndike, F. P., Ingersoll, K. S., Lord, H. R., Gonder-Frederick, L., Frederick, C., Quigg, M. S., Cohn, W. F., & Morin, C. M. (2017). Effect of a Web-Based Cognitive Behavior Therapy for Insomnia Intervention With 1-Year Follow-up: A Randomized Clinical Trial. *JAMA Psychiatry*, *74*(1), 68. https://doi.org/10.1001/jamapsychiatry.2016.3249

Roenneberg, Pilz, Zerbini, & Winnebeck. (2019). Chronotype and Social Jetlag: A (Self-) Critical Review. *Biology*, *8*(3), 54. https://doi.org/10.3390/biology8030054

Roenneberg, T., Allebrandt, K. V., Merrow, M., & Vetter, C. (2012). Social Jetlag and Obesity. *Current Biology*, *22*(10), 939–943. https://doi.org/10.1016/j.cub.2012.03.038

Rollag, M. D., & Niswender, G. D. (1976). Radioimmunoassay of Serum Concentrations of Melatonin in Sheep Exposed to Different Lighting Regimens. *Endocrinology*, *98*(2), 482–489. https://doi.org/10.1210/endo-98-2-482

Rose, G. A. (1992). *The Strategy of Preventive Medicine*. Oxford University Press. https://books.google.no/books?id=eMZpQgAACAAJ

Ruberg, F. L., Skene, D. J., Hanifin, J. P., Rollag, M. D., English, J., Arendt, J., & Brainard, G. C. (1996). Melatonin regulation in humans with color vision deficiencies. *The Journal of Clinical Endocrinology & Metabolism*, *81*(8), 2980–2985. https://doi.org/10.1210/jcem.81.8.8768862

Rüger, M., Gordijn, M. C. M., Beersma, D. G. M., de Vries, B., & Daan, S. (2006). Time-of-day-dependent effects of bright light exposure on human psychophysiology: Comparison of daytime and nighttime exposure. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *290*(5), R1413–R1420. https://doi.org/10.1152/ajpregu.00121.2005

Rüger, M., St Hilaire, M. A., Brainard, G. C., Khalsa, S.-B. S., Kronauer, R. E., Czeisler, C. A., & Lockley, S. W. (2013). Human phase response curve to a single 6.5 h pulse of short-wavelength light: Blue light phase response curve. *The Journal of Physiology*, *591*(1), 353–363. https://doi.org/10.1113/jphysiol.2012.239046

Rupp, A. C., Ren, M., Altimus, C. M., Fernandez, D. C., Richardson, M., Turek, F., Hattar, S., & Schmidt, T. M. (2019). *Distinct ipRGC subpopulations mediate light's acute and circadian effects on body temperature and sleep*. 16.

Rush, A. J., Erman, M. K., Giles, D. E., Schlesser, M. A., Carpenter, G., Vasavada, N., & Roffwarg, H. P. (1986). Polysomnographic Findings in Recently Drug-Free and Clinically Remitted Depressed Patients. *Archives of General Psychiatry*, *43*(9), 878–884. https://doi.org/10.1001/archpsyc.1986.01800090068009

Sack, R. L., Lewy, A. J., Miller, L. S., & Singer, C. M. (1986). Effects of morning versus evening bright light exposure on rem latency. *Biological Psychiatry*, *21*(4), 410–413. https://doi.org/10.1016/0006-3223(86)90172-1

Santhi, N., Thorne, H. C., van der Veen, D. R., Johnsen, S., Mills, S. L., Hommes, V., Schlangen, L. J. M., Archer, S. N., & Dijk, D.-J. (2012). The spectral composition of evening light and individual differences in the suppression of melatonin and delay of sleep in humans: Artificial evening light suppresses melatonin and delays sleep. *Journal of Pineal Research*, *53*(1), 47–59. https://doi.org/10.1111/j.1600-079X.2011.00970.x

Saper, C. B., Chou, T. C., & Scammell, T. E. (2001). The sleep switch: Hypothalamic control of sleep and wakefulness. *Trends in Neurosciences*, *24*(12), 726–731. https://doi.org/10.1016/S0166-2236(00)02002-6

Saper, C. B., Scammell, T. E., & Lu, J. (2005). Hypothalamic regulation of sleep and circadian rhythms. *Nature*, *437*(7063), 1257–1263. https://doi.org/10.1038/nature04284

Sasseville, A., Martin, J. S., Houle, J., & Hébert, M. (2015). Investigating the contribution of short wavelengths in the alerting effect of bright light. *Physiology & Behavior*, *151*, 81–87. https://doi.org/10.1016/j.physbeh.2015.06.028

Sasseville, A., Paquet, N., Sévigny, J., & Hébert, M. (2006). Blue blocker glasses impede the capacity of bright light to suppress melatonin production. *Journal of Pineal Research*, *41*(1), 73–78. https://doi.org/10.1111/j.1600-079X.2006.00332.x

Scammell, T. E., Arrigoni, E., & Lipton, J. O. (2017). Neural Circuitry of Wakefulness and Sleep. *Neuron*, *93*(4), 747–765. https://doi.org/10.1016/j.neuron.2017.01.014

Schennach, R., Feige, B., Riemann, D., Heuser, J., & Voderholzer, U. (2019). Pre- to post-inpatient treatment of subjective sleep quality in 5,481 patients with mental disorders: A longitudinal analysis. *Journal of Sleep Research*, *28*(4), e12842. https://doi.org/10.1111/jsr.12842

Schlangen, L. J. M., & Price, L. L. A. (2021). The Lighting Environment, Its Metrology, and Non-visual Responses. *Frontiers in Neurology*, *12*, 624861. https://doi.org/10.3389/fneur.2021.624861

Schledermann, K. M., Bjørner, T., & Hansen, T. S. (2021). Danish Nursing Home Staff's Perceived Visual Comfort and Perceived Usefulness of a Circadian Lighting System. *Proceedings of the Conference on Information Technology for Social Good*, 91–96. https://doi.org/10.1145/3462203.3475881

Schmidt, T. M., Chen, S.-K., & Hattar, S. (2011). Intrinsically photosensitive retinal ganglion cells: Many subtypes, diverse functions. *Trends in Neurosciences*, *34*(11), 572–580. https://doi.org/10.1016/j.tins.2011.07.001

Schumi, J., & Wittes, J. T. (2011). Through the looking glass: Understanding non-inferiority. *Trials*, *12*(1), 106. https://doi.org/10.1186/1745-6215-12-106

Scott, J., Grierson, A., Gehue, L., Kallestad, H., MacMillan, I., & Hickie, I. (2019). Can consumer grade activity devices replace research grade actiwatches in youth mental health settings? *Sleep and Biological Rhythms*, *17*(2), 223–232. https://doi.org/10.1007/s41105-018-00204-x

Scott, J., Kallestad, H., Vedaa, O., Sivertsen, B., & Etain, B. (2021). Sleep disturbances and first onset of major mental disorders in adolescence and early adulthood: A systematic review and meta-analysis. *Sleep Medicine Reviews*, *57*, 101429. https://doi.org/10.1016/j.smrv.2021.101429

Scott, J., Langsrud, K., Vethe, D., Kjørstad, K., Vestergaard, C. L., Faaland, P., Lydersen, S., Vaaler, A., Morken, G., Torgersen, T., & Kallestad, H. (2019). A pragmatic effectiveness randomized controlled trial of the duration of psychiatric hospitalization in a trans-diagnostic sample of patients with acute mental illness admitted to a ward with either blue-depleted evening lighting or normal lighting conditions. *Trials*, *20*(1), 472. https://doi.org/10.1186/s13063-019-3582-2

Senn, S. (2002). *Cross-over trials in clinical research* (2nd ed). J. Wiley.

Shechter, A., Kim, E. W., St-Onge, M.-P., & Westwood, A. J. (2018). Blocking nocturnal blue light for insomnia: A randomized controlled trial. *Journal of Psychiatric Research*, *96*, 196–202. https://doi.org/10.1016/j.jpsychires.2017.10.015

Simon, S. L., Diniz Behn, C., Laikin, A., Kaar, J. L., Rahat, H., Cree-Green, M., Wright, K. P., & Nadeau, K. J. (2020). Sleep & Circadian Health are Associated with Mood & Behavior in Adolescents with Overweight/Obesity. *Behavioral Sleep Medicine*, *18*(4), 550–559. https://doi.org/10.1080/15402002.2019.1629444

Simor, P., van der Wijk, G., Nobili, L., & Peigneux, P. (2020). The microstructure of REM sleep: Why phasic and tonic? *Sleep Medicine Reviews*, *52*, 101305. https://doi.org/10.1016/j.smrv.2020.101305

Sivertsen, B., Øverland, S., Bjorvatn, B., Mæland, J. G., & Mykletun, A. (2009). Does insomnia predict sick leave? *Journal of Psychosomatic Research*, *66*(1), 67–74. https://doi.org/10.1016/j.jpsychores.2008.06.011

Sivertsen, B., Øverland, S., Pallesen, S., Bjorvatn, B., Nordhus, I. H., Maeland, J. G., & Mykletun, A. (2009). Insomnia and long sleep duration are risk factors for later work disability. The Hordaland Health Study. *Journal of Sleep Research*, *18*(1), 122–128. https://doi.org/10.1111/j.1365-2869.2008.00697.x

Skene, D. J., Skornyakov, E., Chowdhury, N. R., Gajula, R. P., Middleton, B., Satterfield, B. C., Porter, K. I., Van Dongen, H. P. A., & Gaddameedhi, S. (2018). Separation of circadian- and behavior-driven metabolite rhythms in humans provides a window on peripheral oscillators and metabolism. *Proceedings of the National Academy of Sciences*, *115*(30), 7825–7830. https://doi.org/10.1073/pnas.1801183115

Smith, K. A., Schoen, M. W., & Czeisler, C. A. (2004). Adaptation of Human Pineal Melatonin Suppression by Recent Photic History. *The Journal of Clinical*

Endocrinology & Metabolism, *89*(7), 3610–3614. https://doi.org/10.1210/jc.2003-032100

Soler, R., & Voss, E. (2021). Biologically Relevant Lighting: An Industry Perspective. *Frontiers in Neuroscience*, *15*, 637221. https://doi.org/10.3389/fnins.2021.637221

Solheim, B., Olsen, A., Kallestad, H., Langsrud, K., Bjorvatn, B., Gradisar, M., & Sand, T. (2018). Cognitive performance in DSWPD patients upon awakening from habitual sleep compared with forced conventional sleep. *Journal of Sleep Research*, e12730. https://doi.org/10.1111/jsr.12730

Souman, J. L., Borra, T., de Goijer, I., Schlangen, L. J. M., Vlaskamp, B. N. S., & Lucassen, M. P. (2018). Spectral Tuning of White Light Allows for Strong Reduction in Melatonin Suppression without Changing Illumination Level or Color Temperature. *Journal of Biological Rhythms*, *33*(4), 420–431. https://doi.org/10.1177/0748730418784041

Souman, J. L., Tinga, A. M., te Pas, S. F., van Ee, R., & Vlaskamp, B. N. S. (2018). Acute alerting effects of light: A systematic literature review. *Behavioural Brain Research*, *337*, 228–239. https://doi.org/10.1016/j.bbr.2017.09.016

Spielman, A. J., Caruso, L. S., & Glovinsky, P. B. (1987). A Behavioral Perspective on Insomnia Treatment. *Sleep Disorders*, *10*(4), 541–553. https://doi.org/10.1016/S0193-953X(18)30532-X

Spitschan, M., Lazar, R., Yetik, E., & Cajochen, C. (2019). No evidence for an S cone contribution to acute neuroendocrine and alerting responses to light. *Current Biology*, *29*(24), R1297–R1298. https://doi.org/10.1016/j.cub.2019.11.031

St Hilaire, M. A., Gooley, J. J., Khalsa, S. B. S., Kronauer, R. E., Czeisler, C. A., & Lockley, S. W. (2012). Human phase response curve to a 1 h pulse of bright white light: PRC to 1 h light pulses in humans. *The Journal of Physiology*, *590*(13), 3035–3045. https://doi.org/10.1113/jphysiol.2012.227892

Steiger, A., & Pawlowski, M. (2019). Depression and Sleep. *International Journal of Molecular Sciences*, *20*(3), 607. https://doi.org/10.3390/ijms20030607

Stephan, F. K., & Nunez, A. A. (1977). *Elimination of Circadian Rhythms in Drinking, Activity, Sleep, and Temperature by Isolation of the Suprachiasmatic Nuclei*. 16.

Stone, W. C. (1997). *NIST Construction Automation Program Report No. 3*. 200.

Stothard, E. R., McHill, A. W., Depner, C. M., Birks, B. R., Moehlman, T. M., Ritchie, H. K., Guzzetti, J. R., Chinoy, E. D., LeBourgeois, M. K., Axelsson, J., & Wright, Jr., K. P. (2017). Circadian Entrainment to the Natural Light-Dark Cycle across Seasons and the Weekend. *Current Biology*, *27*(4), 508–513. https://doi.org/10.1016/j.cub.2016.12.041

Strainge, L., Sullivan, M. C., Blackmon, J. E., Cruess, S. E., Wheeler, D., & Cruess, D. G. (2019). *PROMIS®-Assessed Sleep Problems and Physical Health Symptoms in Adult Psychiatric Inpatients*. 10.

Taylor, D. J., Peterson, A. L., Pruiksma, K. E., Young-McCaughan, S., Nicholson, K., & Mintz, J. (2017). Internet and In-Person Cognitive Behavioral Therapy for Insomnia in Military Personnel: A Randomized Clinical Trial. *Sleep*, *40*(6). https://doi.org/10.1093/sleep/zsx075

Thapan, K., Arendt, J., & Skene, D. J. (2001). An action spectrum for melatonin suppression: Evidence for a novel non-rod, non-cone photoreceptor system in humans. *The Journal of Physiology*, *535*(1), 261–267. https://doi.org/10.1111/j.1469-7793.2001.t01-1-00261.x

Thomas, A., Grandner, M., Nowakowski, S., Nesom, G., Corbitt, C., & Perlis, M. L. (2016). Where are the Behavioral Sleep Medicine Providers and Where are They Needed? A Geographic Assessment. *Behavioral Sleep Medicine*, *14*(6), 687–698. https://doi.org/10.1080/15402002.2016.1173551

Touitou, Y., Touitou, D., & Reinberg, A. (2016). Disruption of adolescents' circadian clock: The vicious circle of media use, exposure to light at night, sleep loss and risk behaviors. *Journal of Physiology-Paris*, *110*(4), 467–479. https://doi.org/10.1016/j.jphysparis.2017.05.001

Trauer, J. M., Qian, M. Y., Doyle, J. S., Rajaratnam, S. M. W., & Cunnington, D. (2015). Cognitive Behavioral Therapy for Chronic Insomnia: A Systematic Review and Meta-analysis. *Annals of Internal Medicine*, *163*(3), 191–204. https://doi.org/10.7326/M14-2841

Van De Water, A. T. M., Holmes, A., & Hurley, D. A. (2011). Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography - a systematic review: A systematic review of objective sleep measures. *Journal of Sleep Research*, *20*(1pt2), 183–200. https://doi.org/10.1111/j.1365-2869.2009.00814.x

van de Werken, M., Giménez, M. C., de Vries, B., Beersma, D. G. M., & Gordijn, M. C. M. (2013). Short-wavelength attenuated polychromatic white light during work at night: Limited melatonin suppression without substantial decline of alertness. *Chronobiology International*, *30*(7), 843–854. https://doi.org/10.3109/07420528.2013.773440

van der Lely, S., Frey, S., Garbazza, C., Wirz-Justice, A., Jenni, O. G., Steiner, R., Wolf, S., Cajochen, C., Bromundt, V., & Schmidt, C. (2015). Blue Blocker Glasses as a Countermeasure for Alerting Effects of Evening Light-Emitting Diode Screen Exposure in Male Teenagers. *Journal of Adolescent Health*, *56*(1), 113–119. https://doi.org/10.1016/j.jadohealth.2014.08.002

Van Dongen, H. P. A., Maislin, G., Mullington, J. M., & Dinges, D. F. (2003). The Cumulative Cost of Additional Wakefulness: Dose-Response Effects on Neurobehavioral Functions and Sleep Physiology From Chronic Sleep Restriction and Total Sleep Deprivation. *Sleep*, *26*(2), 117–126. https://doi.org/10.1093/sleep/26.2.117

Van Someren, E. J. W. (2011). Actigraphic monitoring of sleep and circadian rhythms. In *Handbook of Clinical Neurology* (Vol. 98, pp. 55–63). Elsevier. https://doi.org/10.1016/B978-0-444-52006-7.00004-6

Van Someren, E. J. W. (2021). Brain mechanisms of insomnia: New perspectives on causes and consequences. *Physiological Reviews*, *101*(3), 995–1046. https://doi.org/10.1152/physrev.00046.2019

van Straten, A., & Lancee, J. (2020). Digital cognitive behavioural therapy for insomnia: The answer to a major public health issue? *The Lancet Digital Health*, *2*(8), e381–e382. https://doi.org/10.1016/S2589-7500(20)30167-9

van Straten, A., van der Zweerde, T., Kleiboer, A., Cuijpers, P., Morin, C. M., & Lancee, J. (2018). Cognitive and behavioral therapies in the treatment of insomnia: A meta-analysis. *Sleep Medicine Reviews*, *38*, 3–16. https://doi.org/10.1016/j.smrv.2017.02.001

Vandewalle, G., Maquet, P., & Dijk, D.-J. (2009). Light as a modulator of cognitive brain function. *Trends in Cognitive Sciences*, *13*(10), 429–438. https://doi.org/10.1016/j.tics.2009.07.004

Vandewalle, G., Schmidt, C., Albouy, G., Sterpenich, V., Darsaud, A., Rauchs, G., Berken, P.-Y., Balteau, E., Degueldre, C., Luxen, A., Maquet, P., & Dijk, D.-J. (2007). Brain Responses to Violet, Blue, and Green Monochromatic Light Exposures in Humans: Prominent Role of Blue Light and the Brainstem. *PLoS ONE*, *2*(11), e1247. https://doi.org/10.1371/journal.pone.0001247

Vásquez-Ruiz, S., Maya-Barrios, J. A., Torres-Narváez, P., Vega-Martínez, B. R., Rojas-Granados, A., Escobar, C., & Ángeles-Castellanos, M. (2014). A light/dark cycle in the NICU accelerates body weight gain and shortens time to discharge in preterm infants. *Early Human Development*, *90*(9), 535–540. https://doi.org/10.1016/j.earlhumdev.2014.04.015

Veale, D. (2019). Against the stream: Intermittent nurse observations of in-patients at night serve no purpose and cause sleep deprivation. *BJPsych Bulletin*, *43*(4), 174–176. https://doi.org/10.1192/bjb.2018.116

Vedaa, Ø., Kallestad, H., Scott, J., Smith, O. R. F., Pallesen, S., Morken, G., Langsrud, K., Gehrman, P., Thorndike, F. P., Ritterband, L. M., Harvey, A. G., Stiles, T., & Sivertsen, B. (2020). Effects of digital cognitive behavioural therapy for insomnia on insomnia severity: A large-scale randomised controlled trial. *The Lancet Digital Health*, *2*(8), e397–e406. https://doi.org/10.1016/S2589-7500(20)30135-7

Vetter, C., Pattison, P. M., Houser, K., Herf, M., Phillips, A. J. K., Wright, K. P., Skene, D. J., Brainard, G. C., Boivin, D. B., & Glickman, G. (2022). A Review of Human Physiological Responses to Light: Implications for the Development of Integrative Lighting Solutions. *LEUKOS*, *18*(3), 387–414. https://doi.org/10.1080/15502724.2021.1872383

Wams, E. J., Woelders, T., Marring, I., van Rosmalen, L., Beersma, D. G. M., Gordijn, M. C. M., & Hut, R. A. (2017). Linking Light Exposure and Subsequent Sleep: A Field Polysomnography Study in Humans. *Sleep*, *40*(12). https://doi.org/10.1093/sleep/zsx165

Wang, Y.-Q., Liu, W.-Y., Li, L., Qu, W.-M., & Huang, Z.-L. (2021). Neural circuitry underlying REM sleep: A review of the literature and current concepts. *Progress in Neurobiology*, *204*, 102106. https://doi.org/10.1016/j.pneurobio.2021.102106

Ward, E. M., Germolec, D., Kogevinas, M., McCormick, D., Vermeulen, R., Anisimov, V. N., Aronson, K. J., Bhatti, P., Cocco, P., Costa, G., Dorman, D. C., Fu, L., Garde, A. H., Guénel, P., Hansen, J., Härmä, M. I., Kawai, K., Khizkhin, E. A., Knutsson, A., … Schubauer-Berigan, M. K. (2019). Carcinogenicity of night shift work. *The Lancet Oncology*, *20*(8), 1058–1059. https://doi.org/10.1016/S1470-2045(19)30455-3

Wassing, R., Benjamins, J. S., Dekker, K., Moens, S., Spiegelhalder, K., Feige, B., Riemann, D., van der Sluis, S., Van Der Werf, Y. D., Talamini, L. M., Walker, M. P., Schalkwijk, F., & Van Someren, E. J. W. (2016). Slow dissolving of emotional distress contributes to hyperarousal. *Proceedings of the National Academy of Sciences*, *113*(9), 2538–2543. https://doi.org/10.1073/pnas.1522520113

Wassing, R., Lakbila-Kamal, O., Ramautar, J. R., Stoffers, D., Schalkwijk, F., & Van Someren, E. J. W. (2019). Restless REM Sleep Impedes Overnight Amygdala Adaptation. *Current Biology*, *29*(14), 2351-2358.e4. https://doi.org/10.1016/j.cub.2019.06.034

Waters, F., Faulkner, D., Naik, N., & Rock, D. (2012). Effects of polypharmacy on sleep in psychiatric inpatients. *Schizophrenia Research*, 4.

Watson, L. A., Phillips, A. J. K., Hosken, I. T., McGlashan, E. M., Anderson, C., Lack, L. C., Lockley, S. W., Rajaratnam, S. M. W., & Cain, S. W. (2018). Increased sensitivity of the circadian system to light in delayed sleep–wake phase disorder. *The Journal of Physiology*, *596*(24), 6249–6261. https://doi.org/10.1113/JP275917

Webster, J. B., Kripke, D. F., Messin, S., Mullaney, D. J., & Wyborney, G. (1982). An Activity-Based Sleep Monitor System for Ambulatory Use. *Sleep*, *5*(4), 389–399. https://doi.org/10.1093/sleep/5.4.389

Wehr, T. A., Goodwin, F. K., Wirz-Justice, A., Breitmaier, J., & Craig, C. (1982). 48-Hour Sleep-Wake Cycles in Manic-Depressive Illness: Naturalistic Observations and Sleep Deprivation Experiments. *Archives of General Psychiatry*, *39*(5), 559–565. https://doi.org/10.1001/archpsyc.1982.04290050037008

Wehr, T. A., Turner, E. H., Shimada, J. M., Lowe, C. H., Barker, C., & Leibenluft, E. (1998). Treatment of a Rapidly Cycling Bipolar Patient by Using Extended Bed Rest and Darkness to Stabilize the Timing and Duration of Sleep. *Biological Psychiatry*, *43*(11), 822–828. https://doi.org/10.1016/S0006-3223(97)00542-8

Wehr, T. A., Wirz-Justice, A., Goodwin, F. K., Duncan, W., & Gillin, J. C. (1979). Phase Advance of the Circadian Sleep-Wake Cycle as an Antidepressant. *Science*, *206*(4419), 710–713. https://doi.org/10.1126/science.227056

Wellek, S., & Blettner, M. (2012). On the Proper Use of the Crossover Design in Clinical Trials. *Deutsches Ärzteblatt International*. https://doi.org/10.3238/arztebl.2012.0276

West, A. S., Sennels, H. P., Simonsen, S. A., Schønsted, M., Zielinski, A. H., Hansen, N. C., Jennum, P. J., Sander, B., Wolfram, F., & Iversen, H. K. (2019). The Effects of Naturalistic Light on Diurnal Plasma Melatonin and Serum Cortisol Levels in Stroke Patients during Admission for Rehabilitation: A Randomized Controlled Trial. *International Journal of Medical Sciences*, *16*(1), 125–134. https://doi.org/10.7150/ijms.28863

Wilhelmsen-Langeland, A., Saxvig, I. W., Pallesen, S., Nordhus, I.-H., Vedaa, Ø., Lundervold, A. J., & Bjorvatn, B. (2013). A Randomized Controlled Trial with Bright Light and Melatonin for the Treatment of Delayed Sleep Phase Disorder: Effects on Subjective and Objective Sleepiness and Cognitive Function. *Journal of Biological Rhythms*, *28*(5), 306–321. https://doi.org/10.1177/0748730413500126

Wilkerson, A. K., Wardle-Pinkston, S., Dietch, J. R., Pruiksma, K. E., Simmons, R. O., Bunnell, B. E., & Taylor, D. J. (2022). Web-based provider training of cognitive behavioral therapy of insomnia: Engagement rates, knowledge acquisition, and provider acceptability. *Cognitive Behaviour Therapy*, *51*(4), 343–352. https://doi.org/10.1080/16506073.2021.1996453

Wilson, S., & Argyropoulos, S. (2005). Antidepressants and Sleep: A Qualitative Review of the Literature. *Drugs*, *65*(7), 927–947. https://doi.org/10.2165/00003495-200565070-00003

Winkler, A., & Rief, W. (2015). Effect of Placebo Conditions on Polysomnographic Parameters in Primary Insomnia: A Meta-Analysis. *SLEEP*. https://doi.org/10.5665/sleep.4742

Wirz-Justice, A., Quinto, C., Cajochen, C., Werth, E., & Hock, C. (1999). A rapid-cycling bipolar patient treated with long nights, bedrest, and light. *Biological Psychiatry*, *45*(8), 1075–1077. https://doi.org/10.1016/S0006-3223(98)00289-3

Wisland, D. T., Granhaug, K., Pleym, J. R., Andersen, N., Støa, S., & Hjortland, H. A. (2016). Remote monitoring of vital signs using a CMOS UWB radar transceiver. *2016 14th IEEE International New Circuits and Systems Conference (NEWCAS)*, 1–4. https://doi.org/10.1109/NEWCAS.2016.7604841

Wittmann, M., Dinich, J., Merrow, M., & Roenneberg, T. (2006). Social Jetlag: Misalignment of Biological and Social Time. *Chronobiology International*, *23*(1–2), 497–509. https://doi.org/10.1080/07420520500545979

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. World Health Organization.

Wright, K. P., Linton, S. K., Withrow, D., Casiraghi, L., Lanza, S. M., Iglesia, H. de la, Vetter, C., & Depner, C. M. (2020). Sleep in university students prior to and during COVID-19 Stay-at-Home orders. *Current Biology*, *30*(14), R797–R798. https://doi.org/10.1016/j.cub.2020.06.022

Wright, K. P., McHill, A. W., Birks, B. R., Griffin, B. R., Rusterholz, T., & Chinoy, E. D. (2013). Entrainment of the Human Circadian Clock to the Natural Light-Dark Cycle. *Current Biology*, *23*(16), 1554–1558. https://doi.org/10.1016/j.cub.2013.06.039

Wu, J. Q., Appleman, E. R., Salazar, R. D., & Ong, J. C. (2015). Cognitive Behavioral Therapy for Insomnia Comorbid With Psychiatric and Medical Conditions: A Meta-analysis. *JAMA Internal Medicine*, *175*(9), 1461. https://doi.org/10.1001/jamainternmed.2015.3006

Wulff, K., Gatti, S., Wettstein, J. G., & Foster, R. G. (2010). Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nature Reviews Neuroscience*, *11*(8), 589–599. https://doi.org/10.1038/nrn2868

Yamadera, W., Sato, M., Harada, D., Iwashita, M., Aoki, R., Obuchi, K., Ozone, M., Itoh, H., & Nakayama, K. (2013). Comparisons of short-term efficacy between individual and group cognitive behavioral therapy for primary insomnia: Comparison of individual and Group CBT-I. *Sleep and Biological Rhythms*, *11*(3), 176–184. https://doi.org/10.1111/sbr.12019

Yang, M., Morin, C. M., Schaefer, K., & Wallenstein, G. V. (2009). Interpreting score differences in the Insomnia Severity Index: Using health-related outcomes to define the minimally important difference. *Current Medical Research and Opinion*, *25*(10), 2487–2494. https://doi.org/10.1185/03007990903167415

Yetish, G., Kaplan, H., Gurven, M., Wood, B., Pontzer, H., Manger, P. R., Wilson, C., McGregor, R., & Siegel, J. M. (2015). Natural Sleep and Its Seasonal Variations in Three Pre-industrial Societies. *Current Biology*, *25*(21), 2862–2868. https://doi.org/10.1016/j.cub.2015.09.046

Zachariae, R., Lyby, M. S., Ritterband, L. M., & O'Toole, M. S. (2016). Efficacy of internet-delivered cognitive-behavioral therapy for insomnia – A systematic review and meta-analysis of randomized controlled trials. *Sleep Medicine Reviews*, *30*, 1–10. https://doi.org/10.1016/j.smrv.2015.10.004

Zaffaroni, A., Kent, B., O'Hare, E., Heneghan, C., Boyle, P., O'Connell, G., Pallin, M., de Chazal, P., & McNicholas, W. T. (2013). Assessment of sleep-disordered breathing using a non-contact bio-motion sensor. *Journal of Sleep Research*, *22*(2), 231–236. https://doi.org/10.1111/j.1365-2869.2012.01056.x

Zaidi, F. H., Hull, J. T., Peirson, S. N., Wulff, K., Aeschbach, D., Gooley, J. J., Brainard, G. C., Gregory-Evans, K., Rizzo, J. F., Czeisler, C. A., Foster, R. G., Moseley, M. J., & Lockley, S. W. (2007). Short-Wavelength Light Sensitivity of Circadian, Pupillary, and Visual Awareness in Humans Lacking an Outer Retina. *Current Biology*, *17*(24), 2122–2128. https://doi.org/10.1016/j.cub.2007.11.034

Zandi, B., Stefani, O., Herzog, A., Schlangen, L. J. M., Trinh, Q. V., & Khanh, T. Q. (2021). OPEN Optimising metameric spectra for integrative lighting. *Scientific Reports*, 15.

Zeitzer, J. M., Dijk, D. J., Kronauer, R., Brown, E., & Czeisler, C. (2000). Sensitivity of the human circadian pacemaker to nocturnal light: Melatonin phase resetting and

suppression. *The Journal of Physiology*, *526*(Pt 3), 695–702. PubMed. https://doi.org/10.1111/j.1469-7793.2000.00695.x

Zhang, Z., Beier, C., Weil, T., & Hattar, S. (2021). The retinal ipRGC-preoptic circuit mediates the acute effect of light on sleep. *Nature Communications*, *12*(1), 5115. https://doi.org/10.1038/s41467-021-25378-w

Zigmond, A. S., & Snaith, R. P. (1983). The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

Zimmerman, M., & Mattia, J. I. (2001). A Self-Report Scale to Help Make Psychiatric Diagnoses: The Psychiatric Diagnostic Screening Questionnaire. *Archives of General Psychiatry*, *58*(8), 787. https://doi.org/10.1001/archpsyc.58.8.787

Zolezzi, M., & Heck, T. (2015). Obstructive sleep apnea: Management considerations in psychiatric patients. *Neuropsychiatric Disease and Treatment*, 2691. https://doi.org/10.2147/NDT.S90521

PAPER I

ORIGINAL ARTICLE

The evening light environment in hospitals can be designed to produce less disruptive effects on the circadian system and improve sleep

Daniel Vethe[1,2,], Jan Scott[1,3,], Morten Engstrøm[4,5], Øyvind Salvesen[6], Trond Sand[4,5,], Alexander Olsen[7,8,], Gunnar Morken[1,2], Hanne S. Heglum[4,9], Kaia Kjørstad[1,2], Patrick M. Faaland[1,2], Cecilie L. Vestergaard[1,2,], Knut Langsrud[1,2] and Håvard Kallestad[1,2,*]

[1]Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway, [2]Division of Mental Health Care, St. Olav's University Hospital, Trondheim, Norway, [3]Institute of Neuroscience, University of Newcastle, Newcastle, UK, [4]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway, [5]Department of Clinical Neurophysiology, St. Olav's University Hospital, Trondheim Norway, [6]Unit of Applied Clinical Research, Norwegian University of Science and Technology, Trondheim, Norway, [7]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway, [8]Department of Physical Medicine and Rehabilitation, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway, [9]Novelda AS, Trondheim, Norway

*Corresponding author. Håvard Kallestad, AFFU, St. Olavs hospital, PO Box 3250 Torgarden, NO-7006 Trondheim, Norway. Email: havard.kallestad@ntnu.no.

Abstract

Study Objectives: Blue-depleted lighting reduces the disruptive effects of evening artificial light on the circadian system in laboratory experiments, but this has not yet been shown in naturalistic settings. The aim of the current study was to test the effects of residing in an evening blue-depleted light environment on melatonin levels, sleep, neurocognitive arousal, sleepiness, and potential side effects.

Methods: The study was undertaken in a new psychiatric hospital unit where dynamic light sources were installed. All light sources in all rooms were blue-depleted in one half of the unit between 06:30 pm and 07:00 am (melanopic lux range: 7–21, melanopic equivalent daylight illuminance [M-EDI] range: 6–19, photopic lux range: 55–124), whereas the other had standard lighting (melanopic lux range: 30–70, M-EDI range: 27–63, photopic lux range: 64–136), but was otherwise identical. A total of 12 healthy adults resided for 5 days in each light environment (LE) in a randomized cross-over trial.

Results: Melatonin levels were less suppressed in the blue-depleted LE (15%) compared with the normal LE (45%; $p = 0.011$). Dim light melatonin onset was phase-advanced more (1:20 h) after residing in the blue-depleted LE than after the normal LE (0:46 h; $p = 0.008$). Total sleep time was 8.1 min longer ($p = 0.032$), rapid eye movement sleep 13.9 min longer ($p < 0.001$), and neurocognitive arousal was lower ($p = 0.042$) in the blue-depleted LE. There were no significant differences in subjective sleepiness ($p = 0.16$) or side effects ($p = 0.09$).

Conclusions: It is possible to create an evening LE that has an impact on the circadian system and sleep without serious side effects. This demonstrates the feasibility and potential benefits of designing buildings or hospital units according to chronobiological principles and provide a basis for studies in both nonclinical and clinical populations.

Statement of Significance

Evening and night exposure to blue light exert particularly disruptive effects on the circadian system, but tunable LED-systems allow for blue-depleted evening lighting more adapted to human circadian biology. We demonstrate that when a blue-depleted light environment is integrated into a large-scale building complex, this has quantifiable effects on the circadian system and sleep. Our study was performed in a new acute psychiatric hospital unit where healthy participants resided for 5 days, a similar duration as patient admissions. This shows that it is possible to use the evening light environment to target circadian disruption and sleep in such a setting and it may have additional applications in a range of settings where control over incident and ambient light is feasible.

Key words: circadian rhythms; lighting; sleep; arousal; hospitals

Introduction

Light is the most important time-giver for the human circadian system [1–3]. Throughout evolution, this system has continuously adapted to the signal created by the cyclic shifts in nature from daylight to nighttime darkness. The advent of electric lighting transformed visual perception in evening environments, increasing the number of hours available for both productivity and recreation. However, exposure to artificial light during normal dark periods exerts additional effects because of the nonvisual effects of light, such as delaying the circadian phase, suppressing melatonin production, delaying sleep onset, changing sleep architecture, and increasing alertness [4–6]. Furthermore, circadian disruption can have negative effects on both mental and somatic health and can contribute to depression, insomnia, metabolic abnormalities, obesity, immune impairment, poor cognitive performance, and a greater risk of cancer [7]. These substantial changes in timing, color, and intensity of evening light have occurred at a rapid pace with regard to the evolutionary perspective and likely have widespread and ongoing implications for health and sleep at a societal level [8]. In contrast, residing in natural darkness at night has been shown to phase-advance circadian rhythms and sleep timing [5, 9]. It remains unknown whether it is possible to create a scalable and usable indoor light environment (LE) that mitigates the negative effects of artificial evening light on circadian rhythms, sleep, and arousal.

The effects of light on the circadian system are primarily mediated by the intrinsically photosensitive retinal ganglion cells (ipRGCs) [10–12], which project to the circadian pacemaker in the suprachiasmatic nucleus of the hypothalamus. ipRGCs are primarily sensitive to short-wavelength, blue light (λ_{max} ~ 480 nm) [10], suggesting the potential to mitigate the negative effects of evening artificial light by modifying the spectral composition of the indoor LE [8]. It has previously been shown that selectively filtering out short-wavelength, blue light in the evening or night reduces melatonin suppression and alertness [13–18], indicative of impeded ipRGC signaling. However, these effects were established under laboratory conditions, whereas a naturalistic setting introduces large variability in irradiance and illuminance resulting from general movement, changes in direction of gaze, light emitting screens, and ambient light. Therefore, it remains to be determined whether meaningful physiological effects can be achieved when such lighting modifications are applied in general housing or institutional settings.

The application of evening blue-depleted lighting may extend to many situations where it is feasible to exert a high level of control over all light sources. It may, however, hold potential in hospitals in which sleep–wake disruptions are ubiquitous and light levels at times may exceed those in household settings. It is known that patients with critical illnesses often have disrupted circadian rhythms [19], a phenomenon that is associated with a subsequent increase in morbidity and mortality [20, 21]. Moreover, disrupted circadian rhythms and sleep–wake cycles are equally, if not more apparent, in severe mental disorders compared with physical illnesses [22], highlighting the need for non-pharmacological interventions that can be used to target sleep–wake disruption and arousal in clinical psychiatry. Notably, the application of extended darkness in the evening and night [23–25], or filtering out short-frequency, blue light using blue-blocking glasses [26] has been shown to reduce

manic symptoms in patients acutely admitted with mania [27]. Given the potential benefits to this population, we decided to examine the potential application of an evening blue-depleted LE as implemented in a large-scale, multiroom complex such as a hospital by integrating such a lighting system in a new psychiatric unit.

The new hospital unit contains two wards that have identical, mirror-image layouts, and similar levels of photopic lux, albeit different light spectrum compositions in the evening: evening blue-depleted or standard hospital LE. Prior to opening of the unit for patient admissions, we undertook a proof-of-concept study to evaluate the effects of the evening blue-depleted LE on healthy subjects. In a randomized cross-over trial, healthy, young adult volunteers resided for alternating periods of 5 consecutive days in each ward. Our main aims were to test whether residing in the evening blue-depleted LE influenced the timing of dim light melatonin onset (DLMO), melatonin suppression, and polysomnographic sleep variables as compared with the effects of exposure to standard LE. Secondary aims were to determine whether differences in neurocognitive arousal and subjective levels of sleepiness could be identified and whether any side effects might be associated with residing in the blue-depleted LE.

Methods

Study design

Participants completed a comprehensive screening procedure, 7-day pre-randomization monitoring, and the 13-day study protocol. Prospective participants were eligible for inclusion in the research if their habitual sleep–wake patterns were within normal parameters (defined for the purposes of the study as weekday bedtime between 10:30 pm and 12:00 am and weekday rise time between 06:30 am and 08:00 am), with small intraindividual variations (<2 h) between weekdays and weekends and they tested negative for color blindness on the Ishihara plate test.

Exclusion criteria were evidence of any current medical or psychological conditions, current use of prescription medication(s), family history of severe mental illness, current sleep disorders, night shift work in the preceding 2 years, trans-meridian travel in the preceding 2 months exceeding one time-zone, and/or current use of nonprescription drugs or illicit substances (not including alcohol or nicotine).

During the 7-day pre-randomization monitoring, all participants were asked to maintain a fixed sleep–wake schedule (bedtime: 11:00 pm–12:00 am, rise time: 07:00 am–08:00 am) and wear an actiwatch. Participants were asked to refrain from the ingestion of alcohol and/or caffeine after 12:00 pm for the duration of the project.

A randomized cross-over design (see Figure 1 and Supplementary Figure S1) was chosen for this study as within-participant variation was expected to be lower for our main outcomes compared with between-participant variation, thus allowing for a smaller number of participants. Participants resided for a total of 10 days (2 conditions of 5 days each) in late September 2017 in a new 40-bedded acute psychiatric unit at St. Olavs Hospital in Trondheim, Norway. Individuals were randomized to first reside 5 days in one of two wards followed by an intermission day and then to reside the next
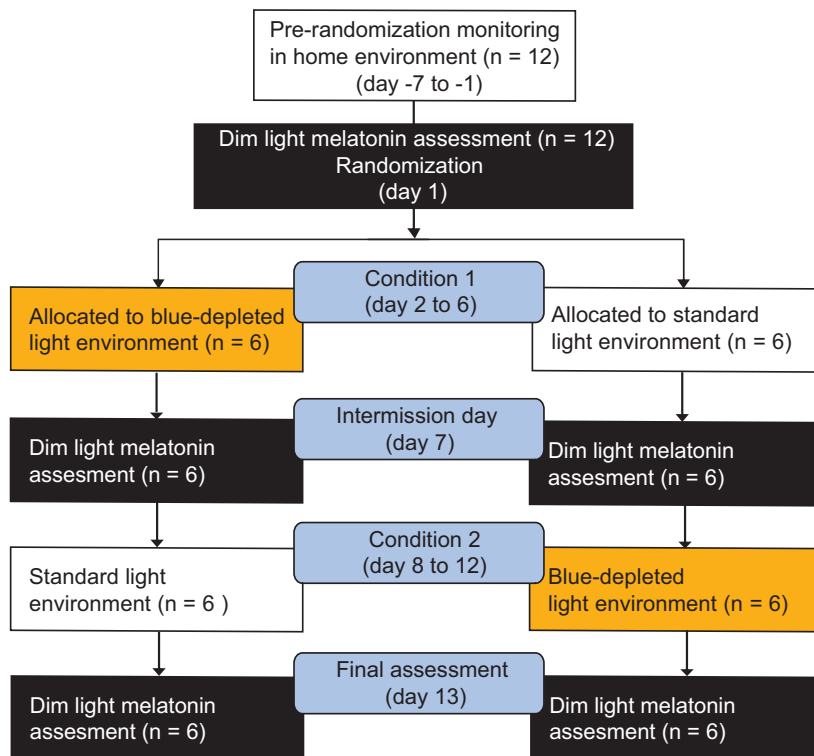
Figure 1. Overview of the study design. Flowchart of the randomized cross-over design describing the location of the participant at each key time point in the trial along with information regarding the nomenclature for the different phases in the study.

5 days in the other ward. The only difference between the two wards was the light spectrum to which residents were exposed during the evening and night (see Supplementary Figure S2 for an overview of the unit). One ward provided a blue-depleted LE in the bedrooms, bathrooms, hallways, and common areas from 06:30 pm until 06:50 am and standard hospital lighting throughout the day, whereas the other ward utilized standard LE at all times. In addition, participants in the blue-depleted LE were asked to use blue-blocking filters (lowbluelights.com) on their electronic media devices in the evenings. Each ward consisted of 20 bedrooms with common areas for socializing and dining.

Participants were given a timetable at the beginning of each 5-day condition that detailed the type and timing of assessments. Hospital staff members were present to maintain safety and assist in the day-to-day running of the wards (e.g. delivery of meals).

During residency at the hospital unit, all participants were awoken by 07:00 am and expected to leave the unit by 08:00 am and return by 05:00 pm. Between 05:00 pm and 06:00 pm, participants had dinner in a common dining room (with standard hospital lighting). At 06:00 pm, participants returned to the LE they were currently allocated to and were free to spend their time in their rooms or the common areas. Participants were requested to retire to their bedrooms for sleep by 11:00 pm and turned the lights off during the sleep opportunity.

Ethics

The protocol was approved by the Regional Ethical Committee in Trondheim (Central Norway; REK: 2017/916) and is registered on the ISRCTN website (Reference: 12419665). Written informed consent was obtained from all participants and the study was undertaken in accordance with the Revised Declaration of Geneva.

Overview of each hospital LE

The lighting fixtures and fittings in each ward were identical and included round downlights, recessed square lights, and built-in reading lights by the desk in the bedrooms (Glamox AS, Oslo, Norway).

Blue-depleted LE (experimental condition)
The blue-depleted LE was created using a LED lighting system that emits both colored and white light. The LED modules inside the light fittings contained a mix of red, green-white, and blue diodes that can be programmed individually to emit different light intensities at different times of the day. To create the evening blue-depleted LE, only the green-white and red diodes emitted light, whereas the blue diode was switched off. The green-white diode emitted a small amount of blue light as it is a blue-chip covered with yellow phosphorus. Blue-blocking

window filters were also deployed in the evening and all televisions had permanent blue-blocking filters (Supplementary Figure S3 shows an example of the bedrooms).

From 07:00 am to 06:00 pm, the light was comprised of standard hospital light (3,000 K). From 06:00 pm to 06:30 pm, there was a transition period from normal to blue-depleted lighting. All light sources were blue-depleted from 06:30 pm to 06:50 am. From 06:50 am to 07:00 am, the lighting underwent a further transition returning it to standard hospital lighting.

Normal LE (control condition)
In this ward, the light spectrum remained constant throughout the 24 h cycle (3,000 K).

Light measurements

Prior to commencing the trial, the light spectrum was assessed using a Mavospec Base light meter (Gossen Foto- Und Lichtmesstechnik GmbH, Nürnberg, Germany). The light measurements demonstrated that the LE in the two wards had similar levels of photopic lux but the levels of melanopic lux were lower in the blue-depleted LE than in the standard LE (see Table 1 for details) [28]. Light exposure will vary with the direction of gaze. However, for the purposes of this study, light measurements were performed horizontally at eye level at standardized locations and times in both units. These locations were patient rooms (1 m into the room, facing windows, standing, with measurement performed horizontally at eye level [160 cm]); bathrooms (standing in front of mirror, with measurement performed horizontally at eye level [160 cm]); common areas such as the TV room (seated in a sofa, facing TV-screen, with measurement performed horizontally at eye level [100 cm]); and hallway (standing beneath a hallway luminaire: the brightest lit area in the hallway).

Assessments

Melatonin
Saliva samples were collected hourly between 07:00 pm and 11:00 pm on study days 1, 6, 7, 12, and 13 using Salivette Cortisol Code blue (Sarstedt AG & Co, Nümbrecht, Germany). Immediately

following sample collection, the samples were centrifuged at 2,200*g* for 10 min and frozen at –18°C overnight, before they were moved into storage at –80°C the following day. Samples were analyzed using enzyme-linked immunosorbent assay (Direct Saliva Melatonin, EK-DSM, Bühlmann, Schönenbuch, Switzerland).

Melatonin assessment in dim light Participants were exposed to dim light (<3 lux) from 06:00 pm until 11:00 pm on three separate occasions (days 1, 7, and 13). The clock time at which melatonin levels were >4 pg/mL was defined as the DLMO.

Melatonin assessment in blue-depleted and standard LE Evening melatonin concentrations were assessed on study days 6 and 12 (i.e. when participants had been exposed to the different LEs for 5 consecutive days).

Sleep
Participants underwent polysomnography (PSG) on study days 5, 6, 11, and 12. Electrodes were applied to the scalp according to the 20/20 system for electroencephalography recording (F3, F4, C3, C4, O1, O2); electrooculogram, submental electromyogram, electrocardiogram, peripheral pulse oximetry, and electrodes on the legs were also measured. The PSG data were collected using SOMNO HD (SOMNOmedicsGmbH, Randersacker, Germany). Signals were sampled at 256 or 128 Hz, low-pass filtered, and stored at 128 Hz. Evaluation of sleep stages (according to the American Academy of Sleep Medicine criteria version 2.4) [29] was undertaken by a clinical neurophysiologist with >10 years of experience with PSG who was blinded to key participant details (such as current LE or order of LE exposure). Time spent in each sleep stage, sleep onset latency, time awake after sleep onset, rapid eye movement (REM) sleep onset latency, and sleep efficiency (percentage of time in bed spent asleep) were estimated.

Subjective sleepiness
Participants rated their subjective sleepiness on the Karolinska Sleepiness Scale (KSS), a 9-point Likert scale (from 1 = "extremely

Table 1. Light measurements in the two LEs

Light measurements	Photopic Lux	Cyanopic Lux	EDI	Melanopic Lux	EDI	Rhodopic Lux	EDI	Chloropic Lux	EDI	Erythropic Lux	EDI	Irradiance (µW/cm²)	photon flux	Log photon flux	peak irradiance (nm)
Standard LE															
Patient room	93	34	34	49	44	58	52	78	75	92	94	28.3	8.28E+13	13.9	605
Patient bathroom	64	17	17	30	27	37	33	53	51	64	65	18.9	5.58E+13	13.8	610
Common area, TV room	86	30	30	46	41	54	48	72	69	86	87	26.5	7.75E+13	13.9	610
Common area, hallway	136	48	48	70	63	83	74	113	109	135	138	42.3	1.24E+14	14.1	610
Blue-depleted LE															
Patient room	87	1	1	16	15	24	21	52	47	97	97	28.3	8,66E+13	13,9	625
Patient bathroom	49	2	2	7	6	11	10	29	26	54	55	14.8	4.49E+13	13.7	620
Common area, TV room	55	0	1	9	8	14	12	32	28	61	61	17.5	5.35E+13	13.7	620
Common area, hallway	124	1	2	21	19	33	28	74	66	139	139	39.7	1.21E+14	14.1	620

Light measurements taken inside the hospital in both the blue-depleted and the standard LE. α-opic illuminance (lux) levels for each of the five photopigments are given in concordance with Lucas et al [28], whereas the α-opic equivalent daylight illuminance (EDI) are reported according to the CIE S026:2018 standard [71].

alert" to 9 = "extremely sleepy—fighting sleep"), at 06:00 pm, 08:00 pm, and 10:00 pm on each night spent in the unit, and again at 07:00 am, the following morning [30].

Neurocognitive arousal

As a neurocognitive measure of arousal, participants completed the Connors Continuous Performance Test-3 (C-CPT-3) [31] between 09:00 pm and 10:00 pm on study days 4 and 10. In brief, the C-CPT-3 is a computerized test in which the letters A–Z are presented consecutively on a monitor (physical blue-blocking filters were used in front of monitors in both LEs). The test consists of 360 trials and lasts for 14 min. The participants are asked to press a response button each time a letter (targets, 80% of trials) was presented but to withhold their response when the letter was X (nontargets, 20% of trials). The participants were asked to respond as quickly and accurately as possible. Outcomes are reported on four commonly derived measures of response speed, consistency throughout testing, and accuracy; that is, reaction time, standard deviation of the reaction times, omissions (failure to respond to targets), and commissions (response to nontargets) [32, 33].

Actigraphy

Participants wore an actiwatch on their nondominant wrist (Actiwatch Spectrum, Philips Respironics Inc., Murrysville, PA), for both the 7 days of pre-randomization monitoring and throughout the 13-day study period. Sleep–wake data (divided into 30-s epochs) were used to estimate the sleep regularity index (SRI). The SRI indicates the percentage of epochs in each 24-h in which sleep–wake states are similar to the corresponding epoch in the previous 24 h [34]. Rise times were generated automatically from actigraphy recordings using an actigraphy software program (Actiware version 5.70.1, Philips Respironics Inc., Murrysville, PA). For bedtimes, participants indicated time they went to bed using an event-marker press on the actiwatch. If this information was missing, we used the reported bedtime recorded in the sleep diary.

Side effects

Subjective side effects Participants completed the Committee of Clinical Investigations (UKU) side effect rating scale on days 7 and 13. Although generally used in the evaluation of new psychotropic drugs, this scale was selected for the present trial as it assesses potential side effects across several important domains (i.e. psychiatric, neurological, autonomic, and other) [35]. The scale has 40 core items in addition to some sex-specific ratings (3 for males and 5 for females). The total score ranges from 0 to 129 for men and 0 to 135 for women, with higher scores indicating the presence of more side effects. However, as increased sleepiness and longer/deeper sleep duration represent desired benefits rather than side effects of exposure to blue-depleted LE, these items were excluded from our analysis.

Color perception Ability to discriminate colors was assessed using the Farnsworth–Munsell 100 Hue Color Vision Test (FM-100) on two occasions (day 2 or 3 and day 8 or 9). The FM-100 measures the amount of errors an individual makes in a color-hue sorting task. Superior color discrimination ability is defined as an error score <20, average ability as a score of 20–100, and low ability by a score >100 [36].

Sample size

A previous study reported that 3 days in a natural light–dark environment (without artificial lighting at night) advanced melatonin onset by 1.4 h and 6 nights advanced DLMO by 2.6 h (without artificial lighting at night) [5]. Other studies had reported that blue-blocking glasses at night had the same effect as darkness on melatonin secretion [15, 37]. As such, we assumed a priori that melatonin onset could be advanced by approximately 1.5 h following 5-day exposure to a blue-depleted LE with an estimated standard deviation of 45 min (a SD similar to that reported for melatonin onset in healthy controls [38] and the above-mentioned studies [5, 15, 37]. We estimated that a sample size of eight individuals would give a 90% chance of detecting these differences with a significance level of 0.05 (two-sided testing). As dropouts from the cross-over design were difficult to predict reliably, we determined we should recruit 12 participants (to allow for 30% attrition).

Randomization

The random allocation sequence was generated by the Unit of Applied Clinical Research (Department of Medicine and Health Sciences, NTNU). The members of the research group could not influence the process in any way.

Statistics

All statistical analyses were performed using R statistical package (version 3.5.2., R Core Team, Vienna, Austria, https://www.R-project.org/) and all figures were generated using GraphPad Prism (version 8.1.2, GraphPad Software, San Diego, California, www.graphpad.com/). A statistical significance level of $p < 0.05$ was chosen for all analyses. A within-subject approach was used in the main analyses to estimate the effects between the two LEs. The analyses were performed by a statistician who was blinded to participant allocation.

Melatonin

The effect of LE on concentrations of salivary melatonin was assessed using a linear mixed model. The combination of study day (day of melatonin assessment), LE, and hour was taken as the fixed effect. The combination of participant ID number and LE were taken as random effects. The assumption of normality was met using a logarithmic scale for the outcome variable. This model estimates the concentrations of melatonin for each of the 5 h included in the melatonin assessments undertaken in the different LEs and dim light. This method implicitly accounts for missing values. The modeled values were in turn used to calculate melatonin suppression and DLMO phase-shifts. As the interaction between LE and study day was specified in the model, any differences in the effects of LE for the order of exposure to each LE could be estimated. Individual levels of melatonin suppression and phase-shifts were calculated from the observed melatonin concentration values.

Melatonin suppression Melatonin suppression is reported as a percentage, which represents the level of melatonin in the two different LEs relative to the level of melatonin the following night in dim light. It was calculated using the area under the

curve (AUC) from 07:00 pm to 11:00 pm on days 6 and 12 divided by the AUC from 07:00 pm to 11:00 pm on days 7 and 13.

Phase shift of DLMO Phase shift was taken as the difference between timing of DLMO after residing in an LE for 5 days (days 7 or 13) and timing of DLMO at baseline (day 1). Linear interpolation was utilized to find the timing of DLMO.

Sleep
Similar linear mixed models were specified to test overall differences in PSG variables between the LEs and by condition order, and to test differences in bedtime, rise time, and SRI between pre-randomization monitoring, condition 1, and condition 2.

Subjective sleepiness, arousal, and side effects
For mean subjective sleepiness (KSS) in the evening and morning, intraindividual differences were calculated and tested for significance using one-sample Student's *t*-tests. The same approach was used for the color perception test (FM-100). As some measures from the C-CPT-3 and scores on the UKU side effects rating scale were not normally distributed, we used a Wilcoxson signed-rank test to examine intraindividual differences between the two LEs for these outcomes.

Missing values were pairwise deleted in the calculation of individual mean scores for the UKU side effects scale and the KSS. One individual was missing all KSS-scores in the blue-depleted LE and was thus excluded from those analyses. There were no missing values on the C-CPT-3 or the FM-100.

Results

Subjects and study design

As shown in Figure 1, 12 healthy young adults (mean age ± SD: 23.0 ± 3.1 years, 7 women) completed the eligibility screening and pre-randomization monitoring and participated in the 13-day trial protocol. Complete data were obtained for the main outcome assessments of melatonin and PSG. Prior to undertaking the analyses, we checked data distributions, etc.,

for outliers. This resulted in the exclusion of one melatonin concentration value associated with one participant to avoid over-estimating the difference in melatonin suppression between conditions (at 11:00 pm when residing in the blue-depleted LE this participant had an extreme concentration of melatonin which was considered most likely to be a measurement error). All but one of the secondary outcome assessments had complete data (one individual did not complete the KSS evaluations).

Melatonin levels and DLMO phase shift

Melatonin suppression was significantly lower when participants resided in the blue-depleted compared with the standard LE (mean difference = 27%, 95% confidence interval (CI): 4% to 51%, $p = 0.020$). Melatonin suppression was 18% (95% CI: −3% to 34%, $p = 0.09$) in the blue-depleted LE and 45% (95% CI: 29% to 57%, $p < 0.001$) in the standard LE. From observed values, 10 out of 11 individuals with complete suppression data exhibited lower levels of melatonin suppression in the blue-depleted LE (see Supplementary Figure S4 for details). DLMO occurred 0:34 h (95% CI: 0:10 to 0:54 h) earlier following residence in the blue-depleted compared with the standard LE ($p = 0.008$). Compared with DLMO at baseline, DLMO was phase-advanced by 1:20 h (95% CI: 1:00 to 1:38 h, $p < 0.001$) following residence in the blue-depleted LE, and by 0:46 h (95% CI: 0:25 to 1:09 h, $p < 0.001$) after residing in the standard LE (Figure 2). From observed values, 11 of 12 individuals presented greater phase advancement of DLMO after residing in the blue-depleted LE (Supplementary Figure S5).

Changes in total sleep time and REM sleep

Total sleep time (TST) was 8.1 min longer in the blue-depleted compared with the standard LE ($p = 0.032$). Furthermore, participants exhibited 13.9 min more REM sleep when residing in the blue-depleted compared with the standard LE ($p < 0.001$). As shown in Table 2, no significant differences were observed between LEs with regard to the variables sleep onset latency, REM sleep onset latency, wake after sleep onset, sleep efficiency, or time in non-REM sleep stages 1–3.
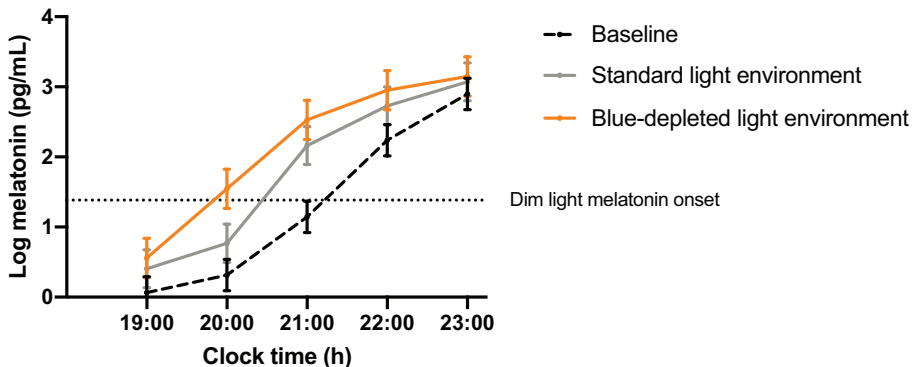
Figure 2. Melatonin concentration by hour and condition. Estimated mean dim light melatonin log concentrations hourly between 07:00 pm and 11:00 pm at baseline and after residing 5 nights in the blue-depleted and the standard LE. Error bars indicate the estimate ± standard error of the mean. The dotted line indicates the 4 pg/mL threshold for DLMO. Melatonin concentrations between 07:00 pm and 11:00 pm differed significantly when individuals resided in the blue-depleted LE compared with the standard LE.

Table 2. Sleep as measured by polysomnography

Sleep variables (min)	Blue-depleted LE		Standard LE		Difference		
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	P
Total sleep time	440.2	432.6 to 447.8	432.1	424.5 to 439.7	8.1	0.7 to 15.5	0.03
Time in REM	89.7	80.8 to 98.6	75.8	66.9 to 84.7	13.9	6.0 to 21.7	0.001
Time in N1	30.8	26.6 to 35.1	33.5	29.3 to 37.8	−2.7	−7.9 to 2.4	0.30
Time in N2	225.6	213.7 to 237.6	224.8	212.8 to 236.7	0.9	−8.3 to 10.1	0.85
Time in N3	94.8	84.3 to 105.4	98.7	88.1 to 109.3	−3.9	−10.9 to 3.0	0.27
Sleep onset latency	10.3	7.6 to 12.9	11.9	9.3 to 14.6	−1.7	−5.1 to 1.7	0.34
REM onset latency	128.0	110.7 to 145.3	125.2	107.9 to 142.5	2.8	−21.0 to 26.6	0.82
Wake after sleep onset	23.0	16.8 to 29.3	22.6	16.3 to 28.8	0.5	−5.8 to 6.7	0.89
Sleep efficiency*	93.0	91.6 to 94.5	92.6	91.2 to 94.1	0.4	−1.0 to 1.8	0.59

Estimates of means in PSG-measured sleep-variables during the last 2 nights residing in the evening blue-depleted LE or the standard LE, and the estimated mean differences between LEs. Estimates, 95% confidence intervals and p-values were calculated from a mixed model with n = 12 participants. REM, rapid eye movement; N1, non-REM sleep stage 1; N2, non-REM sleep stage 2; N3, non-REM sleep stage 3.
*Sleep efficiency is given in percent.

Table 3. Neurocognitive test outcomes

C-CPT3 test variables (ms)	Blue-depleted LE		Standard LE		Estimated difference	
	Mean	SD	Mean	SD	Z	P
Hit reaction time	356.7	27.1	369.9	60.2	−0.18	0.85
Hit reaction time SD	73.0	13.0	65.3	16.2	−2.03	0.04
Commissions*	32.4	17.1	27.3	16.0	−0.98	0.32
Omissions*	0.3	0.3	0.3	0.5	−0.09	0.93

Mean scores with standard deviations (SD) on Connor's Continuous Performance Test-3 variables in both LEs and the estimated differences between the LEs (Wilcoxon signed-rank test with n = 12 participants).
*Commissions and omissions are reported in percent of targets.

Subjective levels of sleepiness and neurocognitive arousal

No significant difference (mean difference = 0.18, 95% CI: −0.09 to 0.47, p = 0.16) was detected between mean evening subjective sleepiness scores for the 11 individuals residing in the blue-depleted (4.98 ± 2.05) compared with the standard LE (4.79 ± 1.85). The mean morning subjective sleepiness also did not significantly differ (mean difference = 0.02, 95% CI: −0.82 to 0.86, p = 0.96) between the 11 individuals with complete data for subjective sleepiness residing in the evening blue-depleted versus standard LE (5.91 ± 2.24 vs. 6.04 ± 2.24, respectively). Supplementary Figure S6 details the mean subjective sleepiness scores as measured at different times of the day. Participants exhibited higher variability in their response times (standard deviations of hit reaction times) throughout the C-CPT-3 computerized response test in the blue-depleted compared with the standard LE (p = 0.042). No statistically significant differences were detected between the two lighting conditions in mean hit reaction time or number of omission or commission errors (see Table 3 for details).

Side effects and color perception

Participants reported very few side effects on the UKU rating scale in either the blue-depleted (0.17 ± 0.42) or standard LE (0.12 ± 0.39), with no significant difference between the LEs (Z = −1.69, p = 0.091). Tiredness/fatigue represented the side effect most frequently reported in the blue-depleted LE (7 reports compared with 2 in the standard LE). Participants made 152 more errors (95% CI: 128 to 177, p < 0.001) on the FM-100 color-hue sorting task when residing in the blue-depleted (192 ± 33.4) compared with the standard LE (39.3 ± 23.1). Mean number of errors in the blue-depleted and normal LE were categorized as evincing low and average ability, respectively.

Post hoc analyses of melatonin-data by condition order

For participants who first resided in the blue-depleted LE and then resided in the standard LE, melatonin suppression was respectively, 18% (95% CI: −16% to 41%, p = 0.26) in the first condition and 62% (95% CI: 47% to 73%, p < 0.001) in the second condition (mean difference: 44%, 95% CI: 16% to 79%, p = 0.002).

For participants who first resided in the standard LE and then resided in the blue-depleted LE, melatonin suppression was respectively 34% (95% CI: 8% to 53%, p = 0.013) in the first condition and 18% (95% CI: −10% to 39%, p = 0.19) in the second condition (mean difference: 16%, 95% CI: −17% to 50%, p = 0.33). There was no statistically significant effect of order on the difference between conditions (mean difference: 28%, 95% CI: −19 to 73, p = 0.23).

For participants who first resided in the blue-depleted LE, DLMO was phased advanced by 0:55 h (95% CI: 0:30 to 1:20, p < 0.001) compared with baseline. After residing in the standard LE as the second condition, there was no significant change in DLMO (mean difference: −0:19 h [delay], 95% CI: −0:49 to 0:14, p = 0.19) compared with DLMO after residing in the blue-depleted LE.

For participants who first resided in the standard LE, DLMO was phase advanced by 0:55 h am (95% CI 0:30 to 1:33, *p* < 0.001) compared with baseline. After residing in the blue-depleted LE as the second condition, DLMO was further phase advanced by another 0:50 h (95% CI: 0:11 to 1:15, *p* = 0.01) compared with DLMO after residing in the standard LE. There was an effect of order in that the effect of the blue-depleted LE was larger in condition 2 (mean difference: 1:10 h, 95% CI: 0:20 to 2:01, *p* = 0.009). Findings for melatonin concentrations by the condition are shown in Supplementary Figure S7, D and E.

Post hoc analyses of PSG by condition order

For participants first residing in the blue-depleted LE, there were no statistically significant differences between LE in PSG variables. For participants first residing in the standard LE, REM sleep was 19.6 min longer (*p* < 0.001) and TST was 11.9 min longer (*p* = 0.02) in the blue-depleted LE compared with in the standard LE. There was a significant order effect for one PSG variable in that the blue-depleted LE had a larger effect on reducing sleep onset latency in condition 1 (mean difference: 8.5 min, 95% CI: −16.7 to −0.2, *p* = 0.04).

Post hoc analyses of sleep times across study phases

These post hoc exploratory analyses are reported as they provide insights regarding potential effects on individuals of residing in a regularized environment (e.g. inpatient unit with fixed rise times, meal times, or bedtimes) during the two study conditions representing time spent in the unit compared with their usual living environment.

Participants went to bed 0:35 h (95% CI: 17 to 0:52, *p* < 0.001) earlier in condition 1 compared with the pre-randomization period (when the participants resided at home). There were no significant differences in bedtimes between condition 1 and condition 2 (mean difference: 0:10 h, 95% CI −0:09 to 0:30, *p* = 0.31). Furthermore, participants rose 1:26 h (95% CI: 1:09 to 1:44 h) earlier in condition 1 than in the pre-randomization period (*p* < 0.001); no significant differences in rise times were found between condition 1 and condition 2 (*p* = 0.92). There was statistically significant correlation between phase-shifts in bedtime and phase-shifts in DLMO for the pre-randomization period to condition 1 (*r* = 0.66, 95% CI: 0.13 to 0.89, *p* = 0.02), but no statistically significant correlation for condition 1 to condition 2 (*r* = 0.26, 95% CI: −0.36 to 0.73, *p* = 0.40). The SRI increased significantly from 85% (95% CI: 83% to 88%) in the pre-randomization period to 95% (95% CI: 93% to 98%) in study condition 1 (*p* < 0.001). However, there was no statistically significant difference in the SRI from condition 1 to condition 2 (*p* = 0.14). Details are provided in Supplementary Figure S7, A–C.

Discussion

In this study, we demonstrated that it is possible to create an evening LE in a large, multiroom complex such as a hospital that had a meaningful effect on objective measures of circadian rhythms, sleep, and arousal, albeit little or no influence on subjective assessments of sleepiness or side effects. Specifically, we found that when healthy adults reside for 5 consecutive days in an evening blue-depleted LE, they exhibit substantially reduced suppression of melatonin production and phase-advancement of endogenous circadian rhythms compared with when residing for a similar period in standard LE conditions. Moreover, melatonin levels in the blue-depleted LE did not differ from those in a dim LE (<3 lux), suggesting that it is possible to design a well-tolerated LE that is similar to near-darkness; that is, "virtual darkness" [27], with regard to its effect on melatonin production. Furthermore, residence in the blue-depleted LE also increased TST and time in REM sleep. We suggest that not only may these effects be relevant for general housing and the healthy population, but the potential therapeutic effect of these adaptations may be even more pronounced in hospital settings.

In particular, sleep disturbances are virtually ubiquitous in critically ill inpatients admitted to hospital units [39–42]. In inpatient psychiatry, sleep disturbances receive particular attention as they constitute trans-diagnostic symptoms of most major mental disorders [22, 43, 44]. These have primarily been treated with medication; however, although chronotherapeutic treatments of these symptoms have been tested [27], their use to date has been limited owing to low feasibility in the clinic as they require acutely ill individuals to adhere to strict treatment regimens. In addition, whereas several hospital units and nursing homes have been built with variations of circadian lighting [8], most have focused on altering indoor daylight properties rather than exerting rigorous control over evening ambient and electric light. To the best of our knowledge, this is the first demonstration that creating such an evening blue-depleted LE is possible in a multiroom complex and the first evaluation of the effects of residing in such an environment using a randomized cross-over trial including objective markers of circadian rhythms, sleep, and arousal. Furthermore, the minimal individual input required by participants together with the failure to detect serious side effects suggests that this design could be applicable in numerous inpatient settings, allowing for effective dissemination of a non-pharmacological intervention targeting circadian rhythms and sleep in hospitals, and in psychiatry in particular. Thus, these findings with healthy adults in a hospital environment constitute an important step toward the implementation of chronotherapeutic interventions in the hospital setting.

In the current study, the two units had identical but mirrored layouts and similar levels of photopic lux and irradiance, whereas levels of melanopic, cyanopic, and rhodopic lux were lower in the blue-depleted LE [28]. A challenge in hospital settings is to create an LE that has meaningful physiological effects without major side effects but is also sufficiently bright to allow hospital staff to perform necessary tasks. Based on previously published work regarding the dose–response relationship between melanopic illuminance and melatonin suppression, we decided to maintain melanopic lux below approximately 20 in the blue-depleted LE [10, 17, 45, 46]. Consistent with this, our findings regarding melatonin suppression are in line with previous work, although one recent study found that suppression can occur at lower levels of melanopic lux under optimal laboratory conditions [47]. Recent research has also shown that large differences exist between individuals with regard to the response of the circadian system to light [48]. Individual differences were also observed in our data on individual participants; however, 11 out of 12 participants exhibited larger circadian effects in the blue-depleted than in the normal LE (see Supplementary Figures S1 and S2). It

has also been shown that certain patient groups, such as those with bipolar disorders [27], seasonal affective disorder [49], and circadian rhythm disorders [50, 51] display greater circadian responses to light than nonpatient groups. Additionally, some of the most widely used medications in psychiatric disorders, selective serotonin reuptake inhibitors, appear to increase the sensitivity of the circadian system to light [52, 53], and some evidence exists of decreased retinal sensitivity in individuals with depressive disorder [54]. Thus, it is likely that a blue-depleted LE may exert differential effects on melatonin suppression in particular patient groups or those taking certain medications compared with healthy control populations, highlighting the need to evaluate how different patient groups will respond to changes in the LE in clinical trials [55].

Notably, a longer TST was observed following residence in the blue-depleted LE. In particular, the 8 min difference was similar to that reported in a meta-analysis of PSG data from treatment trials of cognitive-behavioral therapy for insomnia, which is considered the gold standard of insomnia treatment [56]. This suggests the potential of more pronounced effects in psychiatric inpatient populations, among whom levels of disrupted sleep are higher [43, 44, 57], in turn implying that the blue-depleted LE may be sufficient to meaningfully improve sleep for inpatients. Furthermore, we also observed increased duration of REM sleep. REM sleep propensity has been shown to be influenced by circadian rhythms [58] and reductions in REM sleep have been observed in the first sleep cycle following blue-light exposure concomitant with phase-delay of circadian rhythms [6, 59]. These findings are complimented by REM sleep increases in the first sleep cycle following administration of melatonin [60]. REM sleep has also been implicated in emotional brain processing [61], which may be relevant in mental illness. However, findings from clinical samples are ambiguous and REM dysregulation has been observed in affective disorders [62, 63]. Thus, the effect of a blue-depleted LE on REM sleep may have clinical implications that need to be addressed in future trials.

Nevertheless, although we observed an effect on objective markers of circadian rhythms and sleep, no differences were detected in subjective sleepiness between conditions in the evening or morning. This is similar to prior reports indicating a lack of differences in subjective sleepiness following exposure to blue-depleted light in the evening [13]. This finding may be important for hospital staff working evening or night shifts in a blue-depleted LE. However, we did observe that participants in the blue-depleted LE exhibited higher variability in response-times during the continuous performance of a neurocognitive test, indicating that levels of neurocognitive arousal were lower in the blue-depleted LE. Interestingly, this discrepancy between subjective sleepiness and objective arousal was also observed in the above-cited study of blue-depleted lighting [13]. The decreased arousal may be explained in part by a lower circadian drive for alertness resulting from circadian phase-advancement in the blue-depleted LE [64] but also from a reduction in the direct alerting effects of short-wavelength light [65, 66]. Notably, decreased pre-sleep arousal may be beneficial for agitated inpatients in a psychiatric unit and may also facilitate sleep onset.

Participants reported very low numbers of side effects in both LEs in the current study. No overall difference was detected in side effect scores between the LEs, suggesting that the blue-depleted LE did not have an obvious adverse impact on participants. The most frequently reported side effect in the blue-depleted LE compared with the standard LE was increased fatigue/tiredness. Participants were also less adept at discrimination of color hues in the blue-depleted LE, scoring in the low-ability range compared with the normal-ability range in the standard LE. This finding may be of practical consequence when designing hospital units with changes in LE; for example, not using amber colors for signs or markers in medical charts.

As exploratory analyses, we also tested potential order effects in the trial. First, we did not find significant suppression of melatonin in the blue-depleted LE irrespective of the order of exposure. Second, we found an effect of the order on the difference in DLMO phase-shifts between conditions, with the largest effect of the blue-depleted LE when it was received as the second condition. However, there may be an additional effect of going from a home environment to a highly regularized inpatient environment with fixed bedtimes, rise times, and mealtimes, which may also have impacted the results. Participants were going to bed about 35 min earlier, rising on average 1.5 h earlier, and had higher sleep regularity in condition 1 compared with that during the pre-randomization monitoring. They therefore regularly woke and were exposed to light earlier in the morning. Advancing bedtime, rise time, and exposure to daylight in the morning have a strong phase-advancing effect [67]; thus, any additional effect of the evening blue-depleted LE may be masked by these factors. In contrast, rise times and sleep regularity did not differ between condition 1 and condition 2. In support of this, we also found that the change in bedtimes from the pre-randomization period to condition 1 correlated with the change in DLMO, but this was not the case between condition 1 and condition 2. This may indicate that when adjusted to the sleep–wake schedule, the effect of the evening blue-depleted compared with standard LE is more prominent.

Chronotherapeutic interventions that remove all light in the evening or block blue light with orange-tinted glasses [27] have demonstrated efficacy in reducing symptoms of mania in acutely admitted inpatients. These effects may be mediated by circadian or sleep pathways or may be the result of direct pathways influencing mood-centers in the brain, which have recently been identified in rodents [68]. Conversely, a recent exploratory study found no effects of blue-depleted light emitting diode (LED)-lighting on psychiatric symptoms in an inpatient psychiatric unit for affective disorders [69]. However, in this study incident or ambient light sources, such as TVs, mobile phones, or windows, were not controlled; moreover, the experimental light system was only installed in patient rooms, not corridors and common areas, and participants were thus free to enter and exit the experimental condition. This is problematic as the circadian system has recently been shown to be sensitive to very low levels of melanopic lux [47] or brief light exposures at night [17], and because increased sensitivity has been reported in several relevant patient groups [27, 49]. In contrast, the hospital unit in the current study is designed to allow a high level of control over both incident and ambient light. However, some extra effort may be needed from hospital staff to ensure adequate control over light in a psychiatric hospital unit, such as making blue-blocking filters available for mobile devices. In addition, these findings may be relevant across a variety of hospital units and might also possibly extend to other settings where control over ambient and incident light may be feasible such as trains, planes, and hotels.

Several limitations should be considered when interpreting the results from this study. First, an apparent effect was detected from residing in the hospital unit for the first week regardless of LE, suggesting that some degree of stabilization of sleep–wake rhythms could be achieved via a more regular routine. However, we did not design this study to investigate effects by condition order and these analyses, therefore, had limited statistical power. Second, researchers were given permission to access the hospital unit for only a limited time period (before it opened to acute admissions); thus, the study design had to consider such practicalities. Ultimately, as only a one-day washout phase was possible between the LEs, carryover effects cannot be excluded. However, this intermission is similar to that in other studies of the effects of light on evening melatonin and sleep [6, 13]. Third, participants resided in the unit from 05:00 pm until 08:00 am. During the day, they had to leave the unit owing to construction work. Participants in the current study thus may have been exposed to higher levels of light during daytime than the average patient housed in the hospital unit. Given that prior light history may exert protective effects against light exposure at night [70], the effects of a blue-depleted LE may be larger for an inpatient remaining indoors for most of the day. Fourth, 20 out of 48 PSG recordings were stopped at the designated wakeup time at 07:00 am. Therefore, when analyzing the data, the 07:00 am endpoint was applied to all recordings to minimize the influence of these events. Fifth, owing to the color of the light it was impossible to blind participants with regard to the specific LE in which they were residing. However, we did not observe a difference in subjective sleepiness scores, suggesting that expectancy effects were limited. Sixth, as we focused on a group of healthy young adults and the degree of some effects may differ in older adults with psychiatric disorders, some of our findings may not generalize across these populations. Seventh, we did not perform continuous measurements of light exposure at the eye level of the participants. Thus, we cannot control for the light exposure each participant was subjected to. Nevertheless, the current proof-of-concept study offered important insights regarding the effect of the modified hospital lighting system and provided a unique opportunity to apply physiological measurements such as melatonin assays and PSG, which are not always feasible for use with inpatients with severe mental disorders.

In conclusion, we have shown that the evening LE in a naturalistic setting can be modified according to chronobiological principles to have beneficial effects on the circadian system and sleep, without side effects. This offers translational relevance as it could readily be provided as a potential therapeutic intervention to large numbers of hospitalized patients with little increase in staff or patient burden.

Supplementary material

Supplementary material is available at *SLEEP* online.

Disclosure statements

Non-financial disclosure: None declared.
Financial disclosure: None declared.

Author contributions

H.K. and K.L. conceived the study idea. D.V., K.L., and H.K. designed the study, with support from T.S., M.E., and A.O. D.V. and H.S.H. performed the data collection. D.V., H.K., and Ø.S. planned and performed the statistical analyses, with support from A.O. M.E. scored the PSG recordings. D.V., J.S., and H.K. wrote the initial draft with critical revisions from K.L., C.L.V., K.K., P.M.F., H.S.H., Ø.S., A.O., G.M., T.S., and M.E.

References

1. Boivin DB, *et al*. Dose-response relationships for resetting of human circadian clock by light. *Nature.* 1996;**379**(6565):540–542.
2. Czeisler CA, *et al*. Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science.* 1999;**284**(5423):2177–2181.
3. Klerman EB, *et al*. Simulations of light effects on the human circadian pacemaker: implications for assessment of intrinsic period. *Am J Physiol.* 1996;**270**(1 Pt 2):R271–R282.
4. Cho Y, *et al*. Effects of artificial light at night on human health: A literature review of observational and experimental studies applied to exposure assessment. *Chronobiol Int.* 2015;**32**(9):1294–1310.
5. Stothard ER, *et al*. Circadian entrainment to the natural light-dark cycle across seasons and the weekend. *Curr Biol.* 2017;**27**(4):508–513.
6. Chang A-M, *et al*. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proc Natl Acad Sci USA.* 2015;**112**(4):1232–1237.
7. Foster RG, *et al*. The rhythm of rest and excess. *Nat Rev Neurosci.* 2005;**6**(5):407–414.
8. Münch M, *et al*. The role of daylight for humans: gaps in current knowledge. *Clocks Sleep.* 2020;**2**(1):61–85.
9. Wright KP Jr, *et al*. Entrainment of the human circadian clock to the natural light-dark cycle. *Curr Biol.* 2013;**23**(16):1554–1558.

10. Brainard GC, *et al.* Action spectrum for melatonin regulation in humans: evidence for a novel circadian photoreceptor. *J Neurosci.* 2001;**21**(16):6405–6412.

11. Provencio I, *et al.* A novel human opsin in the inner retina. *J Neurosci.* 2000;**20**(2):600–605.

12. Hattar S, *et al.* Melanopsin-containing retinal ganglion cells: architecture, projections, and intrinsic photosensitivity. *Science.* 2002;**295**(5557):1065–1070.

13. Rahman SA, *et al.* The effects of spectral tuning of evening ambient light on melatonin suppression, alertness and sleep. *Physiol Behav.* 2017;**177**:221–229.

14. van de Werken M, *et al.* Short-wavelength attenuated polychromatic white light during work at night: limited melatonin suppression without substantial decline of alertness. *Chronobiol Int.* 2013;**30**(7):843–854.

15. Sasseville A, *et al.* Blue blocker glasses impede the capacity of bright light to suppress melatonin production. *J Pineal Res.* 2006;**41**(1):73–78.

16. Kayumov L, *et al.* Blocking low-wavelength light prevents nocturnal melatonin suppression with no adverse effect on performance during simulated shift work. *J Clin Endocrinol Metab.* 2005;**90**(5):2755–2761.

17. Nowozin C, *et al.* Applying melanopic lux to measure biological light effects on melatonin suppression and subjective sleepiness. *Curr Alzheimer Res.* 2017;**14**(10):1042–1052.

18. Rahman SA, *et al.* Spectral modulation attenuates molecular, endocrine, and neurobehavioral disruption induced by nocturnal light exposure. *Am J Physiol Endocrinol Metab.* 2011;**300**(3):E518–E527.

19. Gazendam JAC, *et al.* Altered circadian rhythmicity in patients in the ICU. *Chest.* 2013;**144**(2):483–489.

20. Ely EW, *et al.* The impact of delirium in the intensive care unit on hospital length of stay. *Intensive Care Med.* 2001;**27**(12):1892–1900.

21. Carlson DE, *et al.* The absence of circadian cues during recovery from sepsis modifies pituitary-adrenocortical function and impairs survival. *Shock.* 2008;**29**(1):127–132.

22. Wulff K, *et al.* Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nat Rev Neurosci.* 2010;**11**(8):589–599.

23. Wirz-Justice A, *et al.* A rapid-cycling bipolar patient treated with long nights, bedrest, and light. *Biol Psychiatry.* 1999;**45**(8):1075–1077.

24. Wehr TA, *et al.* Treatment of rapidly cycling bipolar patient by using extended bed rest and darkness to stabilize the timing and duration of sleep. *Biol Psychiatry.* 1998;**43**(11):822–828.

25. Barbini B, *et al.* Dark therapy for mania: a pilot study. *Bipolar Disord.* 2005;**7**(1):98–101.

26. Henriksen TE, *et al.* Blue-blocking glasses as additive treatment for mania: a randomized placebo-controlled trial. *Bipolar Disord.* 2016;**18**(3):221–232.

27. Gottlieb JF, *et al.* The chronotherapeutic treatment of bipolar disorders: A systematic review and practice recommendations from the ISBD task force on chronotherapy and chronobiology. *Bipolar Disord.* 2019;**21**(8):741–773.

28. Lucas RJ, *et al.* Measuring and using light in the melanopsin age. *Trends Neurosci.* 2014;**37**(1):1–9.

29. Berry RB, *et al.* AASM scoring manual updates for 2017 (Version 2.4). *J Clin Sleep Med.* 2017;**13**(5):665–666.

30. Akerstedt T, *et al.* Subjective and objective sleepiness in the active individual. *Int J Neurosci.* 1990;**52**(1–2):29–37.

31. Conners CK. *Conners Continuous Performance Test—3rd Edition Manual.* Toronto: Multi-Health Systems; 2014.

32. Homack S, *et al.* Conners' continuous performance test (2nd ed.; CCPT-II). *J Atten Disord.* 2006;**9**(3):556–558.

33. Wilhelmsen-Langeland A, *et al.* A randomized controlled trial with bright light and melatonin for the treatment of delayed sleep phase disorder: effects on subjective and objective sleepiness and cognitive function. *J Biol Rhythms.* 2013;**28**(5):306–321.

34. Phillips AJK, *et al.* Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing. *Sci Rep.* 2017;**7**(1):3216.

35. Lingjaerde O, *et al.* The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatr Scand Suppl.* 1987;**334**:1–100.

36. Rigby HS, *et al.* Colour perception in pathologists: the Farnsworth-Munsell 100-hue test. *J Clin Pathol.* 1991;**44**(9):745–748.

37. van der Lely S, *et al.* Blue blocker glasses as a countermeasure for alerting effects of evening light-emitting diode screen exposure in male teenagers. *J Adolesc Health.* 2015;**56**(1):113–119.

38. Solheim B, *et al.* Cognitive performance in DSWPD patients upon awakening from habitual sleep compared with forced conventional sleep. *J Sleep Res.* 2018;**28**(5):e12730.

39. Elliott R, *et al.* Characterisation of sleep in intensive care using 24-hour polysomnography: an observational study. *Crit Care.* 2013;**17**(2):R46.

40. Foreman B, *et al.* Sleep in the neurological intensive care unit: feasibility of quantifying sleep after melatonin supplementation with environmental light and noise reduction. *J Clin Neurophysiol.* 2015;**32**(1):66–74.

41. Freeman D, *et al.* Sleep disturbance and psychiatric disorders. *Lancet Psychiatry.* 2020;**7**(7):628–637.

42. Andersen JH, *et al.* Sleep in the intensive care unit measured by polysomnography. *Minerva Anestesiol.* 2013;**79**(7):804–815.

43. Langsrud K, *et al.* Sleep patterns as a predictor for length of stay in a psychiatric intensive care unit. *Psychiatry Res.* 2016;**237**:252–256.

44. Schennach R, *et al.* Pre- to post-inpatient treatment of subjective sleep quality in 5,481 patients with mental disorders: A longitudinal analysis. *J Sleep Res.* 2019;**28**(4):e12842.

45. Brainard GC, *et al.* Sensitivity of the human circadian system to short-wavelength (420-nm) light. *J Biol Rhythms.* 2008;**23**(5):379–386.

46. Zeitzer JM, *et al.* Sensitivity of the human circadian pacemaker to nocturnal light: melatonin phase resetting and suppression. *J Physiol.* 2000;**526**(Pt 3):695–702.

47. Prayag AS, *et al.* Melatonin suppression is exquisitely sensitive to light and primarily driven by melanopsin in humans. *J Pineal Res.* 2019;**66**(4):e12562.

48. Phillips AJK, *et al.* High sensitivity and interindividual variability in the response of the human circadian system to evening light. *Proc Natl Acad Sci USA.* 2019;**116**(24):12019–12024.

49. Nathan PJ, *et al.* Melatonin sensitivity to dim white light in affective disorders. *Neuropsychopharmacology.* 1999;**21**(3):408–413.

50. Watson LA, *et al.* Increased sensitivity of the circadian system to light in delayed sleep–wake phase disorder. *J Physiol.* 2018;**596**(24):6249–6261.

51. Aoki H, *et al.* Hypersensitivity of melatonin suppression in response to light in patients with delayed sleep phase syndrome. *Chronobiol Int.* 2001;**18**(2):263–271.

52. McGlashan EM, *et al.* The SSRI citalopram increases the sensitivity of the human circadian system to light in an acute dose. *Psychopharmacology (Berl).* 2018;**235**(11):3201–3209.

53. McGlashan EM, *et al*. Decreased sensitivity of the circadian system to light in current, but not remitted depression. *J Affect Disord*. 2019;**256**:386–392.

54. Berman G, *et al*. Decreased retinal sensitivity in depressive disorder: a controlled study. *Acta Psychiatr Scand*. 2018;**137**(3):231–240.

55. Scott J, *et al*. A pragmatic effectiveness randomized controlled trial of the duration of psychiatric hospitalization in a trans-diagnostic sample of patients with acute mental illness admitted to a ward with either blue-depleted evening lighting or normal lighting conditions. *Trials*. 2019;**20**(1):472.

56. Trauer JM, *et al*. Cognitive behavioral therapy for chronic insomnia: a systematic review and meta-analysis. *Ann Intern Med*. 2015;**163**(3):191–204.

57. Haynes PL, *et al*. Examination of insomnia and insomnia treatment in psychiatric inpatients. *Int J Ment Health Nurs*. 2011;**20**(2):130–136.

58. Czeisler CA, *et al*. Timing of REM sleep is coupled to the circadian rhythm of body temperature in man. *Sleep*. 1980;**2**(3):329–346.

59. Münch M, *et al*. Wavelength-dependent effects of evening light exposure on sleep architecture and sleep EEG power density in men. *Am J Physiol Regul Integr Comp Physiol*. 2006;**290**(5):R1421–R1428.

60. Cajochen C, *et al*. Melatonin and S-20098 increase REM sleep and wake-up propensity without modifying NREM sleep homeostasis. *Am J Physiol*. 1997;**272**(4 Pt 2):R1189–R1196.

61. Goldstein AN, *et al*. The role of sleep in emotional brain function. *Annu Rev Clin Psychol*. 2014;**10**:679–708.

62. Benca RM, *et al*. Sleep and psychiatric disorders. A meta-analysis. *Arch Gen Psychiatry*. 1992;**49**(8):651–668; discussion 669.

63. Baglioni C, *et al*. Sleep and mental disorders: A meta-analysis of polysomnographic research. *Psychol Bull*. 2016;**142**(9):969–990.

64. Rahman SA, *et al*. Diurnal spectral sensitivity of the acute alerting effects of light. *Sleep*. 2014;**37**(2):271–281.

65. Vandewalle G, *et al*. Light as a modulator of cognitive brain function. *Trends Cogn Sci*. 2009;**13**(10):429–438.

66. Vandewalle G, *et al*. Brain responses to violet, blue, and green monochromatic light exposures in humans: prominent role of blue light and the brainstem. *PLoS One*. 2007;**2**(11):e1247.

67. Czeisler CA, *et al*. Bright light resets the human circadian pacemaker independent of the timing of the sleep-wake cycle. *Science*. 1986;**233**(4764):667–671.

68. Fernandez DC, *et al*. Light affects mood and learning through distinct retina-brain pathways. *Cell*. 2018;**175**(1):71–84.e18.

69. Okkels N, *et al*. Lighting as an aid for recovery in hospitalized psychiatric patients: a randomized controlled effectiveness trial. *Nord J Psychiatry*. 2020;**74**(2):105–114.

70. Hébert M, *et al*. The effects of prior light history on the suppression of melatonin by light in humans. *J Pineal Res*. 2002;**33**(4):198–203.

71. CIE Central Bureau. *CIE System for Metrology of Optical Radiation for ipRGC-Influenced Responses to Light*. Vienna: CIE; 2018.

SUPPLEMENTARY INFORMATION PAPER I

Supplemental Material

Title

The evening light environment in hospitals can be designed to produce less disruptive effects on the circadian system and improve sleep

Authors

Daniel Vethe[1,2], Jan Scott[1,3], Morten Engstrøm[4,5], Øyvind Salvesen[6], Trond Sand[4,5], Alexander Olsen[7,8], Gunnar Morken[1,2], Hanne S. Heglum[4,9], Kaia Kjørstad[1,2], Patrick M. Faaland[1,2], Cecilie L. Vestergaard[1,2], Knut Langsrud[1,2], and Håvard Kallestad[1,2]*.

Affiliations

[1] Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway.

[2] Division of Mental Health Care, St. Olav's University Hospital, Trondheim, Norway.

[3] Institute of Neuroscience, University of Newcastle, Newcastle, England.

[4] Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway.

[5] Department of Clinical Neurophysiology, St. Olav's University Hospital, Trondheim Norway.

[6] Unit of Applied Clinical Research, Norwegian University of Science and Technology, Trondheim, Norway.

[7] Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway.

[8] Department of Physical Medicine and Rehabilitation, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

[9] Novelda AS, Trondheim, Norway.

* Corresponding Author: Håvard Kallestad. Email: havard.kallestad@ntnu.no

Address: AFFU, St. Olavs hospital, PO Box 3250 Torgarden, NO-7006 Trondheim, Norway

Figure S1. Raster plot of the study design detailing the timing of light exposure throughout the study.

Figure S2. Overview of the hospital unit. Bedrooms are located on the outer walls, whereas common rooms are facing the atriums. The two halves of the unit are identical except for the light environments in the evenings. In the current study, only bedrooms facing north were used by participants. Participants ate breakfast and dinner in a cantina facing north between the two halves of the unit.

Figure S3. Blue-depleted light environment bedroom. Image of a bedroom in the blue-depleted light environment in the evening. Blue-blocking filters were employed between the two window-panels in the evening to block ambient blue light from entering the room.

Figure S4. Individual suppression of melatonin by condition. Percentage suppression of observed melatonin levels for each participant in each light environment compared with levels in dim light the following day. Error-bars indicate the mean ± the standard error of the mean of observed values from 11 participants. As one participant had twice the melatonin levels when residing in the blue-depleted light environment relative to dim light the following night (due to an extreme concentration value at 2300h), this was considered most likely to be an error, and this data-point was omitted.

Figure S5. Individual phase-shifts by condition. Observed dim light melatonin onset phase shift for each participant from baseline to after exposure to each light environment. Error-bars indicate the mean ± the standard error of the mean (SEM) in each light environment.

Figure S6. Subjective sleepiness by hour and condition. Mean scores on the Karolinska

Sleepiness Scale (KSS) throughout the evenings in the two light environments, and in the

mornings at 07:00 h. Error-bars indicate the mean ± standard error of the mean in each light

environment. One participant lost the KSS score sheet from one light environment, leaving 11

participants with scores from both conditions.

Figure S7. Bedtime, rise time, sleep regularity index by study phases and melatonin by condition order. Summary of post-hoc mixed model analyses showing differences between the study phases regarding sleep-wake stability (A, B, C) and changes in dim light melatonin onset (DLMO) phase shifts by condition order (D, E). All error-bars indicate the estimate ± the standard error of the mean. **(A, B, C)** Comparison of the mean bedtime (A), the mean rise time (B), and the Sleep Regularity Index (C) between the pre-randomization study phase, condition 1, and condition 2. **(D, E)** Log melatonin concentrations by time and condition for condition order blue-depleted LE first (D) and condition order standard LE first (E). The threshold for dim light

melatonin onset is indicated by the dotted grey line. The baseline estimates have been added as a

dashed black line in both (D) and (E) for reference.

Table S1. Sleep as measured by polysomnography according to condition order.

Sleep variables (min)	Order: Blue-depleted – Standard LE(n = 6)								Order: Standard – Blue-depleted LE (n = 6)							
	1: Blue-depleted LE		2: Standard LE		Difference				1: Standard LE		2: Blue-depleted LE		Difference		
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	p		Estimate	95% CI	Estimate	95% CI	Estimate	95% CI	p
Total sleep time	440.3	429.6 to 451.1	436.1	425.4 to 446.9	4.2	-6.2 to 14.7	0.43		428.2	417.4 to 438.9	440.1	429.3 to 450.8	11.9	1.47 to 22.4	0.02
Time in REM	82.5	69.9 to 95.1	74.4	61.8 to 87.0	8.1	-2.9 to 19.1	0.15		77.2	64.6 to 89.8	96.8	84.2 to 109.4	19.6	8.5 to 30.6	< 0.001
Time in N1	32.1	26.1 to 38.2	33.2	27.2 to 39.2	-1.1	-8.4 to 6.2	0.77		33.9	27.9 to 39.9	29.5	23.5 to 35.5	-4.4	-11.7 to 2.9	0.24
Time in N2	237.7	220.8 to 254.7	232.8	215.8 to 249.7	4.9	-8.1 to 18.0	0.46		216.7	199.8 to 233.7	213.6	196.6 to 230.5	-3.1	-16.2 to 9.9	0.64
Time in N3	88.7	73.7 to 103.7	96.3	81.3 to 111.3	-7.6	-17.5 to 2.2	0.13		101.1	86.1 to 116.1	100.9	85.9 to 115.9	-0.2	-10.0 to 9.6	0.97
Sleep onset latency	7.5	3.7 to 11.3	10.5	6.7 to 14.3	-3.0	-7.8 to 1.8	0.22		13.4	9.6 to 17.2	13.0	9.2 to 16.8	-0.33	-5.2 to 4.5	0.89
REM onset latency	124.5	100.1 to 149.0	122.2	97.7 to 146.7	2.4	-31.3 to 36.1	0.89		128.3	103.8 to 152.8	131.5	107.0 to 156.0	3.2	-30.5 to 36.9	0.85
Wake after sleep onset	17.8	8.9 to 26.6	20.0	11.2 to 28.8	-2.3	-11.1 to 6.6	0.62		25.1	16.3 to 34.0	28.3	19.5 to 37.1	3.2	-5.7 to 12.0	0.48
Sleep efficiency [a]	94.6	92.5 to 96.6	93.5	91.4 to 95.5	1.1	-0.8 to 3.0	0.26		91.8	89.7 to 93.8	91.4	89.3 to 93.5	-0.3	-2.3 to 1.6	0.72

Estimates of means in PSG-measured sleep-variables by condition order and the estimated mean differences between light environments (LEs) within the condition order from mixed model analyses. CI = confidence interval, REM = Rapid eye movement, N1 = Non-REM sleep stage 1, N2 = Non-REM sleep stage 2, N3 = Non-REM sleep stage. [a]Sleep efficiency is given in percent.

Table S2. Itemized list of side-effect scores.

Side effects	Blue-depleted		Standard	
	Mean	SE	Mean	SE
Concentration Difficulties	0.2	0.1	0.0	0.0
Astheniat/Lassitude/lncreased Fatigability	0.7	0.2	0.2	0.1
Failing Memory	0.3	0.1	0.1	0.1
Depression	0.1	0.1	0.0	0.0
Tension/lnner Unrest	0.1	0.1	0.1	0.1
Reduced Duration of Sleep	0.5	0.3	0.9	0.3
Increased Dream Activity	0.3	0.2	0.3	0.2
Dystonia	0.1	0.1	0.2	0.2
Rigidity	0.1	0.1	0.1	0.1
Hypokinesia/Akinesia	0.0	0.0	0.1	0.1
Hyperkinesia logic	0.3	0.1	0.0	0.0
Tremor	0.1	0.1	0.0	0.0
Akathisia	0.8	0.2	0.5	0.2
Epileptic Seizures	0.0	0.0	0.0	0.0
Paraesthesias	0.2	0.1	0.1	0.1
Headache	0.3	0.1	0.0	0.0
Accommodation Disturbances	0.1	0.1	0.1	0.1
Increased Salivation	0.0	0.0	0.0	0.0
Reduced Salivation	0.2	0.2	0.2	0.1
Nausea/Vomiting	0.2	0.1	0.1	0.1
Diarrhoea	0.3	0.2	0.2	0.1
Constipation	0.3	0.2	0.0	0.0
Micturition Disturbances	0.0	0.0	0.0	0.0
Polyuria/Polydipsia	0.1	0.1	0.2	0.2
Orthostatic Dizziness	0.2	0.1	0.2	0.1
Rash	0.1	0.1	0.2	0.1
Pruritus	0.3	0.1	0.3	0.2
Photosensitivity	0.0	0.0	0.0	0.0
Increased Pigmentation	0.0	0.0	0.0	0.0
Weight gain	0.5	0.2	0.8	0.2
Weight loss	0.1	0.1	0.0	0.0
Galactorrhoea	0.0	0.0	0.0	0.0
Gynaecomastia	0.0	0.0	0.0	0.0
Increased Sexual Desire	0.3	0.1	0.1	0.1
Menorrhagia	0.0	0.0	0.0	0.0
Intermenstrual Bleeding	0.1	0.1	0.0	0.0
Amenorrhoea	0.0	0.0	0.0	0.0
Orgastic Dysfunction	0.0	0.0	0.0	0.0
Dry Vagina	0.0	0.0	0.0	0.0
Erectile Dysfunction	0.0	0.0	0.0	0.0
Ejaculatory Dysfunction	0.0	0.0	0.0	0.0
Premature Ejaculation	0.0	0.0	0.0	0.0

Mean scores on each item in the Committee of Clinical Investigations (UKU) side effect rating scale in both light environments. SE = standard error of the mean.

List of captions for supplementary material

Figure S1. Raster plot of the study design detailing the timing of light exposure throughout the study.

Figure S2. Overview of the hospital unit. Bedrooms are located on the outer walls, whereas common rooms are facing the atriums. The two halves of the unit are identical except for the light environments in the evenings. In the current study, only bedrooms facing north were used by participants. Participants ate breakfast and dinner in a cantina facing north between the two halves of the unit.

Figure S3. Blue-depleted light environment bedroom. Image of a bedroom in the blue-depleted light environment in the evening. Blue-blocking filters were employed between the two window-panels in the evening to block ambient blue light from entering the room.

Figure S4. Individual suppression of melatonin by condition. Percentage suppression of observed melatonin levels for each participant in each light environment compared with levels in dim light the following day. Error-bars indicate the mean ± the standard error of the mean of all participants. As one participant had twice the melatonin levels when residing in the blue-depleted light environement relative to dim light the following night (due to an extreme concentration value at 2300h), this was considered most likely to be an error, and this data-point was omitted.

Figure S5. Individual phase-shifts by condition. Observed dim light melatonin onset phase shift for each participant from baseline to after exposure to each light environment. Error-bars indicate the mean ± the standard error of the mean (SEM) in each light environment.

Figure S6. Subjective sleepiness by hour and condition. Mean scores on the Karolinska Sleepiness Scale (KSS) throughout the evenings in the two light environments, and in the mornings at 07:00 h. Error-bars indicate the mean ± standard error of the mean in each light environment. One participant lost the KSS score sheet from one light environment, leaving 11 participants with scores from both conditions.

Figure S7. Bedtime, rise time, sleep regularity index by study phases and melatonin by condition order. Summary of post-hoc mixed model analyses showing differences between the study phases regarding sleep-wake stability (A, B, C) and changes in dim light melatonin onset (DLMO) phase shifts by condition order (D, E). All error-bars indicate the estimate ± the standard error of the mean. (A, B, C) Comparison of the mean bedtime (A), the mean rise time (B), and the Sleep Regularity Index (C) between the pre-randomization study phase, condition 1, and condition 2. (D, E) Log melatonin concentrations by time and condition for condition order blue-depleted LE first (D) and condition order standard LE first (E). The threshold for dim light melatonin onset is indicated by the dotted grey line. The baseline estimates have been added as a dashed black line in both (D) and (E) for reference.

Table S1. Sleep as measured by polysomnography according to condition order. Estimates of means in PSG-measured sleep-variables by condition order and the estimated mean differences between light environments (LEs) within the condition order from mixed model analyses. CI =

confidence interval, REM = Rapid eye movement, N1 = Non-REM sleep stage 1, N2 = Non-REM sleep stage 2, N3 = Non-REM sleep stage. aSleep efficiency is given in percent.

Table S2. Itemized list of side-effect scores. Mean scores on each item in the Committee of Clinical Investigations (UKU) side effect rating scale. SE = standard error of the mean.

PAPER II

scientific reports

Check for updates

OPEN Evening light environments can be designed to consolidate and increase the duration of REM-sleep

Daniel Vethe[1,2✉], H. J. Drews[1], J. Scott[1,3], M. Engstrøm[4,5], H. S. A. Heglum[4,6], J. Grønli[7], J. P. Wisor[8], T. Sand[4,5], S. Lydersen[9], K. Kjørstad[1,2], P. M. P. Faaland[1,2], C. L. Vestergaard[1,2], K. Langsrud[1,2] & H. Kallestad[1,2]

Evening exposure to short-wavelength light has disruptive effects on circadian rhythms and sleep. These effects can be mitigated by blocking short-wavelength (blue) frequencies, which has led to the development of evening blue-depleted light environments (BDLEs). We have previously reported that residing 5 days in an evening BDLE, compared with residing in a normal indoor light environment of similar photopic lux, advances circadian rhythms and increases the duration of rapid eye movement (REM) sleep in a randomized cross-over trial with twelve healthy participants. The current study extends these findings by testing whether residing in the evening BDLE affects the consolidation and microstructure of REM sleep in the same sample. Evening BDLE significantly reduces the fragmentation of REM sleep ($p = 0.0003$), and REM sleep microarousals in ($p = 0.0493$) without significantly changing REM density or the latency to first REM sleep episode. Moreover, the increased accumulation of REM sleep is not at the expense of NREM stage 3 sleep. BDLE further has a unique effect on REM sleep fragmentation ($p = 0.0479$) over and above that of circadian rhythms phase-shift, indicating a non-circadian effect of BDLE. If these effects can be replicated in clinical populations, this may have a therapeutic potential in disorders characterized by fragmented REM sleep.

Artificial light is an important aid to visual acuity in the evening and at nighttime. This light exposure also has extensive non-image-forming (NIF) effects on human sleep and circadian rhythms[1–5]. These effects are primarily driven by intrinsically photosensitive retinal ganglion cells (ipRGCs) that both entrain the central circadian pacemaker to the light environment and directly influence sleep[6]. The photosensitivity of ipRGCs is predominantly in the short-wavelength blue light range ($\lambda_{max} \approx 480$ nm). As such, there is an opportunity to design artificial evening light environments with less impact on these systems. This has led to the development of acute hospital facilities that utilize an evening blue-depleted light environment (BDLE) in order to improve sleep and circadian rhythmicity for inpatients[7]. In a previous publication[8], we reported findings from a pilot study undertaken with 12 healthy participants who resided in a new psychiatric unit prior to its official opening and who were exposed to both an evening BDLE and a standard light environment (standard LE) for 5 days each. The study demonstrated that, in addition to a phase-advancing effect on circadian rhythms and an increase in sleep duration, there was also a significant increase in the duration of rapid eye movement (REM) sleep. Noting that REM sleep is a heterogenous sleep stage[9], with alternations between active (phasic) and comparatively quiescent (tonic) substages, the downstream effects of increased duration might depend on REM sleep fragmentation and microstructure.

[1]Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway. [2]Division of Mental Health Care, St. Olav's University Hospital, Trondheim, Norway. [3]Institute of Neuroscience, University of Newcastle, Newcastle, UK. [4]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway. [5]Department of Clinical Neurophysiology, St. Olav's University Hospital, Trondheim, Norway. [6]Novelda AS, Trondheim, Norway. [7]Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway. [8]Sleep and Performance Research Center and Department of Translational Medicine and Physiology, Elson S. Floyd College of Medicine, Washington State University, Spokane, WA, USA. [9]Regional Centre for Child and Youth Mental Health and Child Welfare, Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway. ✉email: daniel.vethe@ntnu.no

1

	Mean	SD	Range
Age	23.0	3.1	20–28
Baseline DLMO	21:11	0:37	19:59–21:57
Baseline bedtime	23:54	0:26	23:09–00:36
Baseline rise-time	08:22	0:29	07:47–09:03
Baseline total sleep time	08:03	0:28	07:26–08:54

Table 1. Baseline sleep and circadian rhythm characteristics of the sample (N = 12). N = 12 participants for all variables and analyses. Sample means, standard deviations and ranges of baseline characteristics. Rise-times, bed-times and total sleep time values are reported as sample mean values, standard deviations and ranges of the underlying individual participants' mean, over the seven days prior to randomization (days − 7 to − 1). Baseline DLMO refers to the DLMO-assessment undertaken on day 1 of the study. *DLMO* Dim Light Melatonin Onset.

In adults, REM-sleep normally occupies 20–25% of the total sleep time and supports several functions such as brain temperature regulation, synaptic plasticity, and complex cognitive processes such as memory consolidation. Notably, research indicates that REM sleep, and particularly phasic REM sleep, plays a role in the overnight processing of emotionally salient information and thus emotion regulation[9–13]. Therefore, we can hypothesize that an increase in the duration of REM sleep after exposure to a BDLE might have many potential benefits for individuals. However, recent findings in insomnia suggest that the degree of REM sleep fragmentation modulates the favorable effect of REM sleep on amygdala reactivity[14–16]. This "restless" REM-sleep may then increase emotional distress and theoretically be a risk factor for insomnia and mental illness[15]. In the case of depression, increased REM sleep duration in combination with shorter latency to first REM sleep episode and increased REM-density is common, and is understood as a maladaptive disinhibition of REM-sleep that increases vulnerability and impacts treatment responses[10,11,17–23]. Although this research has mostly been undertaken in clinical populations, these characteristics of REM sleep might offer indications whether the increased REM sleep duration is adaptive or dysfunctional. Thus, they represent useful elements to explore in the context of evening light and its effects on REM sleep.

There are at least two conceivable mechanisms by which evening light exposure may have effects on REM sleep. First, REM sleep regulation is tied to circadian rhythms[24], with increased propensity to enter REM sleep in the late night/early biological morning. A phase-advance of circadian rhythms, as we and others have reported for evening BDLE[8,25], may result in more REM sleep earlier in the night, possibly at the expense of slow-wave sleep (SWS; N3) that has higher propensity in the first sleep cycles, and also is associated with numerous physiological health benefits[26]. Second, artificial evening light exposure results in higher arousal and alertness before bedtime[2,27,28], and may reduce slow wave activity (0.75–4 Hz) in the first sleep cycle[6]. This suggests a way by which arousal may be carried into the sleep period, possibly also altering REM sleep fragmentation or microstructure. Moreover, in the case of insomnia, pre-bedtime hyperarousal, has been linked to fragmented REM sleep. We and others have found that evening blue-depleted light diminishes pre-bedtime arousal[25]. This represents a non-circadian pathway for the effects of evening light exposure on REM sleep. Due to REM sleep being associated to both circadian and arousal processes, distinguishing the mechanisms by which evening light may modify REM sleep is a challenge.

In summary, we previously reported a significant increase in in REM sleep duration when healthy participants resided in an evening BDLE[8]. In this study, we explore this finding in more detail by first examining whether there are any differential effects of residing in an evening BDLE compared with a standard LE on REM sleep fragmentation and REM-density. Second, to explore potential mechanisms of change, we test if changes in REM sleep parameters are associated with phase-shift of circadian rhythms. Finally, we examine the temporal dynamic of the accumulated REM sleep, and contrast that with the accumulation of non-REM stage 3 sleep (N3).

Results

The sample of healthy participants consisted of 7 women and 5 men. See Table 1 for baseline descriptive statistics. Participants had a similar number of sleep cycles with 4.3 cycles (95% CI 3.9–4.7) in the BDLE and 4.3 cycles (95% CI 3.9–4.8) in the standard LE. The mean duration of sleep cycles was also similar in both conditions with 103.6 min (95% CI 94.4–113.2) in the BDLE and 103.6 min (95% CI 94.5–113.3) in the standard LE. For further descriptive information regarding REM sleep cycles, REM fragmentation, and REM microarousals see Supplementary Figs. S1, S2, S3, S4, S5, and S6.

REM-sleep. As reported previously[8], individuals demonstrated a significantly longer REM-sleep duration (see Table 2) when residing in the evening BDLE (89.7 min, 95% CI 80.8–98.6) compared with the standard LE condition (75.8, 95% CI 66.9–84.7)[8].

As shown in Table 2, further analyses demonstrate that individuals show less REM-sleep fragmentation in the evening BDLE (6.6%, 95% CI 3.8, 10.1) compared with in the standard LE condition (10.3%, 95% CI 7.7–14.2) (see Fig. 1). Moreover, there was also a reduction in cortical microarousals during REM sleep in the BDLE with 24.6 arousals (95% CI 15.6–35.4) compared with 37.2 arousals (95% CI 28.0–48.2) in the standard LE. There was no significant difference in REM-density between the conditions.

Step	REM sleep duration			REM sleep fragmentation			REM sleep arousals			REM-density		
	Estimate	95% CI	p	Estimate	95% CI	p	Estimate	95% CI	p	Estimate	95% CI	p
1												
BDLE	13.88	5.82 to 21.88	0.0018*	− 3.71	− 6.79 to − 1.50	0.0003*	− 12.58	− 25.38 to − 0.04	0.0493*	0.007	− 0.011 to 0.025	0.432
Phase shift	− 22.9	− 35.23 to − 10.55	0.0007*	4.56	1.00 to 9.86	0.0092*	15.11	− 2.22 to 32.96	0.089	0.005	− 0.025 to 0.034	0.757
2												
BDLE	6.65	− 4.95 to 18.24	0.27	− 3.68	− 7.87 to − 0.02	0.0479*	− 10.34	− 26.24 to 6.32	0.22	0.024	− 0.003 to 0.051	0.090
Phase shift	− 15.21	− 33.25 to 2.83	0.11	0.06	− 5.79 to 6.80	0.99	4.74	− 18.71 to 25.73	0.68	0.036	− 0.009 to 0.080	0.126

Table 2. Effects of phase shift and light environment on REM sleep duration, REM sleep fragmentation, REM sleep microarousals, and REM-density. Results from four linear mixed models with total REM sleep duration, REM sleep fragmentation, REM sleep microarousals, or REM-density as dependent variable and participant ID as random effect. The models are fitted in two steps. In step I, BDLE and Phase shift are included separately as covariates, and in Step II, they are included simultaneously. The estimates, 95% confidence intervals, and *p*-values were calculated from mixed models with N = 12 participants. BDLE = Blue-depleted light environment. * = significant effect of BDLE or phase shift, at *P* < 0.05.
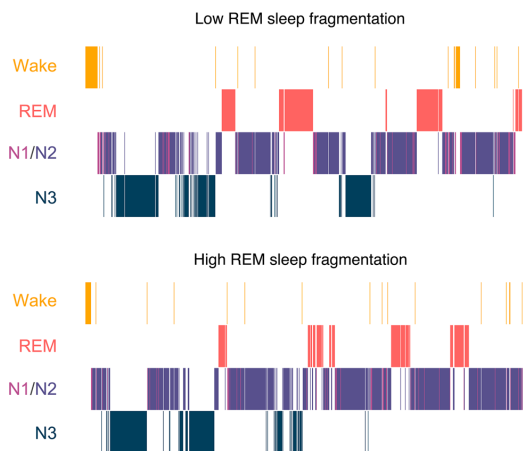
Figure 1. Example hypnograms of two nights with either low REM sleep fragmentation or high REM sleep fragmentation. Both examples are from the same individual with the low REM sleep fragmentation being in the BDLE condition and the high REM sleep fragmentation being in the standard LE condition.

Associations between changes in REM-variables and the phase-shift of dim light melatonin onset (DLMO). As shown in Table 2, phase advancement of the DLMO was associated with longer duration of REM-sleep and reduced REM-fragmentation (see Supplementary Fig. S7). When BDLE and phase-shift were simultaneously entered in the second step, there were no unique associations between either construct and REM-sleep duration. However, there was a unique effect of BDLE over and above that of phase-shift for REM-fragmentation (see Table 2).

Post hoc analyses by condition order. As reported in the previous publication[8], there was an effect of order of exposure to the two conditions on phase-shift of DLMO. Given the association of circadian rhythms and REM sleep it was considered relevant to also test for order-effects in the current REM-outcome variables. The effect of BDLE on REM sleep fragmentation and REM sleep arousal was only significant in Period 1. However, there was no significant order effect of condition on any of the REM sleep variables (see Table 3).

Accumulation of REM-sleep and N3. Time in REM-sleep accumulated more rapidly when individuals were residing in the evening BDLE compared with the standard LE. As shown in Fig. 2a, the difference between conditions reached statistical significance from about 40% of the total sleep time onwards, increasing in magnitude after about 75% of total sleep time until the end of the sleep period. There were no significant differences between LE conditions in the accumulation of N3 sleep (see Fig. 2b).

3

	BDLE in period 1			BDLE in period 2			Order effect		
	Estimate	95% CI	P	Estimate	95% CI	P	Estimate	95% CI	P
REM duration	5.3	− 13.5 to 24.2	0.56	22.4	3.5 to 41.2	0.024*	17.0	− 17.7 to 51.8	0.32
REM fragmentation	− 5.9	− 13.2 to − 0.4	0.036*	− 1.5	− 7.7 to 5.0	0.66	− 4.5	− 16.8 to 6.3	0.46
REM sleep micro-arousals	− 18.66	− 39.09 to 0.61	0.058	− 6.45	− 26.9 to 13.4	0.51	12.22	− 18.4 to 43.1	0.428
REM density	0.002	− 0.05 to 0.05	0.96	0.01	− 0.04 to 0.07	0.62	0.01	− 0.08 to 0.12	0.82

Table 3. Effects of BDLE on REM sleep outcomes depending on whether BDLE is given in period 1 or period 2, and the effect of order on the BDLE-effect. Results from three linear mixed models with total REM sleep duration, REM sleep fragmentation, REM sleep microarousals, or REM-density as dependent variable and participant ID as random effect. The order effects are reported as the interaction between period and condition. N = 12 participants. BDLE = Blue-depleted light environment. * = significant effect at $P < 0.05$.
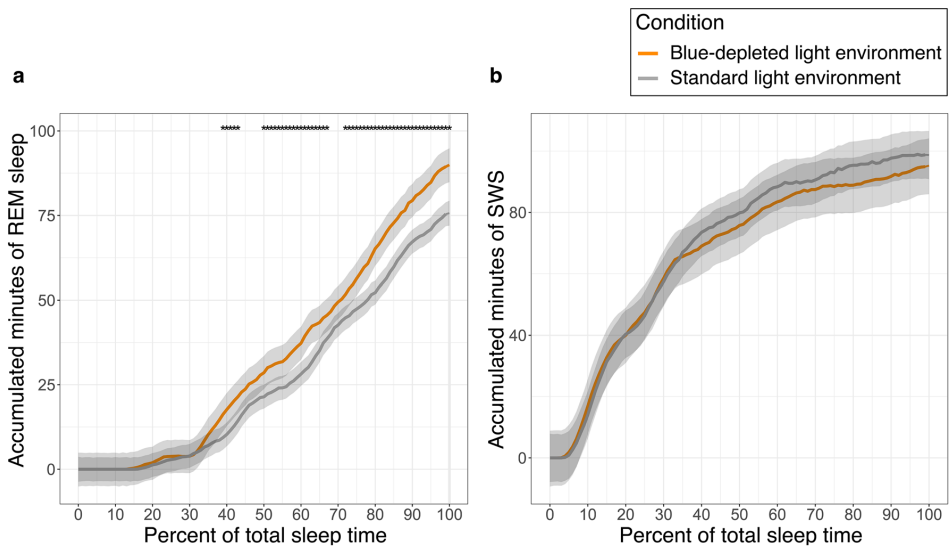
Figure 2. Average accumulated minutes in (**a**) REM sleep and (**b**) slow-wave sleep by percentage of the total sleep time. 95% confidence intervals, and p-values were calculated from a mixed model with N = 12 participants. Accumulated minutes as dependent variable, percent of total sleep time as a 101 level factor and light environment and their interaction as covariates, and participant ID as random effect. N = 12 participants. The p-values are Bonferroni-corrected for 101 comparisons.

Discussion

In our previous study we reported a phase-advance of circadian rhythms and an increase in REM sleep duration when residing in an evening BDLE[8]. The current study extends these findings by showing that residing in an evening BDLE also a reduces the fragmentation of REM sleep and the number of cortical microarousals during REM sleep in the same sample. However, REM-density did not change after exposure to an evening BDLE. This suggests that the duration of REM sleep was longer, more consolidated, and with fewer microarousals, but that its microstructure, in terms of the balance between phasic and tonic REM sleep, remained unchanged in these healthy individuals. To the best of our knowledge, this is the first demonstration of an intervention based on light exposure affecting the consolidation of REM sleep.

REM-sleep follows a circadian rhythm with higher propensity to enter REM sleep in the second half of the night[24]. Therefore, a further aim of this study was to explore if any changes in REM sleep variables were associated with the phase-shift of DLMO. Although we did find this for the duration of REM sleep, the effect of phase-shift could not be distinguished from that of BDLE. It is known that, evening administration of melatonin induces a phase-advance of circadian rhythms and increase REM sleep duration, most prominently in the first sleep-cycle[29]. Conversely, evening exposure to short-wavelength light phase-delays circadian rhythms and reduces REM-sleep duration[2], especially in the first REM sleep episode[6]. In the current study, it can be hypothesized that the effect of BDLE on REM sleep duration may be mediated through a phase-advance of circadian rhythms. Moreover, we have previously reported that the phase-advance was about 30 min larger in BDLE compared with standard

LE, which roughly corresponds to the increased accumulation of REM-sleep starting about 30 min earlier in BDLE. One previous study has found that increased SWS accumulation may be at the expense of REM sleep after different light exposures[30]. The faster accumulation of REM-sleep after BDLE in our study, does not seem to be at the expense of N3 sleep.

Similar to the duration of REM sleep, we also found an effect of phase-shift on REM sleep fragmentation. However, BDLE had an additional, albeit marginally significant effect over and above the effect of phase-shift, explaining nearly all the variance in the effect of phase-shift. This may suggest that there is a non-circadian effect of evening light exposure on REM sleep fragmentation. Moreover, in the previous publication[8] we found that the effect of BDLE on phase-shift was stronger in period 2 of the study. In the current study, we find in contrast that the effects of BDLE on REM sleep fragmentation were stronger in period 1, although there was no significant order effect. This further supports the notion that there may be additional non-circadian driven effects of BDLE. Interestingly, recent studies both with animal models and in humans have identified non-circadian pathways by which light can exert direct NIF-effects[31,32]. A non-circadian effect could potentially be mediated by lower arousal as an effect of the BDLE. Increased arousal has indeed been found after evening short-wavelength light exposure[2,27,33], and this effect seems to be carried over into the sleep period[2,6,33]. On the contrary, lower levels of wake EEG-derived alertness has been observed after BDLE[25], and we also found reduced neurocognitive arousal in our previous report on this sample[8]. In the current study we found a trend toward less microarousals from NREM sleep (see supplementary Table S1). Hence, the current finding indicates that BDLE increases the stability of REM sleep, and that this effect does not seem to be directly related to circadian processes, but may be the result of lower pre-bedtime arousal.

These findings could also be put in context of the two process theory of sleep–wake regulation[34,35]. This model explains how both homeostatic sleep pressure (process S) and the circadian pacemaker (process C) interact to regulate sleep. It is well known that EEG delta-activity in NREM sleep increases with longer pre-sleep wakefulness in an hourglass manner, and that it decays as a function of consolidated NREM sleep. The probability of entering REM sleep increases with duration of NREM sleep through each sleep cycle, and the duration of REM sleep episodes increases with each NREM-REM-cycle, indicating that stable REM sleep may require a degree of dissipation of process S. A hypothetical increase in NREM delta-activity in BDLE, potentially caused by lower arousal at sleep onset, may possibly contribute to consolidation of REM sleep. In this framework process S dissipation may interact with a phase-advance of process C, which also increases propensity for REM sleep[24]. However, an analysis of power frequency domain is outside the scope of this paper.

Our study focuses on healthy young adults, but if the findings are confirmed, we believe there are elements of this research that will translate to clinical practice in psychiatry. Increased REM sleep is not beneficial in all circumstances, as evidenced by the high prevalence of disinhibited REM-sleep in mood disorders[10,11,17–22]. Disinhibited REM sleep has been suggested to represent a maladaptive stress response that maintains depressive episodes[10,22,23] and impacts treatment responses[11]. However, we did not find an effect of evening BDLE on REM-density. Moreover, we did not find effects on latency to first REM-episode in our previous report on this sample[8]. These results suggest that the increased REM sleep duration in the BDLE is not a marker for disinhibited REM sleep. Rather it may be an effect of the phase-advance of circadian rhythms, potentially in combination with reduced fragmentation allowing for longer sustained episodes of REM-sleep. This may be a key finding given the potential therapeutic application of blue-depleted lighting systems in psychiatric care.

Mental disorders are also linked to fragmented REM sleep, which in combination with increased REM-density has been defined as restless REM sleep[14,16,17]. Restless REM sleep, has been found to impede overnight emotional processing, which may lead to accumulated distress, hyperarousal, and a sustaining of the restless REM sleep[15]. This mechanism has been proposed to be a transdiagnostic risk factor for mental illness, in particular insomnia, but also mood disorders. This is a different explanation of the REM sleep alterations in major depression, in that fragmented REM sleep over time leads to a shortage of time spent in REM sleep eventually resulting in a disinhibition and rebound of REM sleep in major depression[17]. The restless REM sleep hypothesis of mental illness offers another mechanism by which these findings may have beneficial effects if translated into clinical practice.

Future directions.   In the context of the restless REM sleep model of insomnia[15], the current findings on BDLE-consolidation of REM sleep may also suggest a potential therapeutic role of evening BDLE for patients with insomnia. This may particularly be the case for the subgroup of insomnia patients reporting sleep–wake misperception, which may indicate high arousal and fragmentation in REM sleep[15]. Indeed, one small study found additional benefits of adding blue-blocking glasses to cognitive behavioral therapy for insomnia on subjective levels of anxiety and hyperarousal[36], and another small study found effects of blue-blocking glasses on both subjective and objective sleep measures compared with clear glasses[37]. Moreover, given the rationale for restless REM sleep to exacerbate symptoms in patients with major depression and insomnia, it would also be of interest to test potential effects of BDLE on REM sleep parameters and clinical outcomes. Furthermore, other NIF-effects of evening light exposure has been found to have high interindividual variability[38]. This is likely the case for the current effects on REM sleep fragmentation (see Supplementary Fig. 8), suggesting future studies are needed to uncover the characteristics of individuals that have large effects of BDLE. Moreover, it is also conceivable that the effects of consolidating REM sleep may first be consequential if the outcome is not short-term improvement, but rather the alleviation of clinical symptoms over time or the prevention of future illness episodes. Previous case-series using extended darkness[39] or blue-blocking glasses with patients with rapid cycling bipolar disorder found that when used consistently over years, they seemed to stabilize mood and prevent new illness episodes.

Strengths and limitations.   A strength of the current study is the use of a randomized cross-over trial design which increases power by allowing for individuals to be compared to themselves in the two LEs. Moreover, it enables inference on the causality between conditions and the outcomes. Furthermore, we had complete data on all measures. However, several limitations must be considered in the interpretation of the study findings. First, this was an opportunistic study testing hypotheses that were developed after the primary analyses were completed. The sample size was small, and the power-calculations for outcome measures in the original study does not apply to the outcome measures in these secondary analyses, increasing the probability of type II errors. Second, the participants were healthy young adult individuals that were screened for any physical illness, mental disorders, and sleep disturbance. This limits the generalizability of findings to any clinical population, and to the aged population in which photoreception may be compromised. Finally, there was only a 1-day wash-out between study periods, which increases chances for carry-over effects.

Conclusion
This is the first study demonstrating that residing in an evening BDLE reduces the fragmentation of REM sleep and microarousals in REM sleep while REM density and N3 sleep are unaltered. Given the widespread interest in the spectral tuning of evening lighting from both hospitals, nursing homes, as well as private consumers, particular care should be taken to test these effects in larger samples. Effects should also be tested in clinical samples with clinical outcome-measures, to further understand the health consequences of the potential beneficial effects of BDLE on REM sleep fragmentation.

Methods
The current study reports secondary analyses of data collected during a short-term (13 days) randomized cross-over trial of 12 healthy young adults. The original study protocol was approved by the Regional Ethical Committee in Trondheim (Central Norway; REK: 2017/916). The protocol was designed and the study performed in accordance with the declaration of Helsinki. The study rationale, protocol, procedures are described on the ISRCTN website (Reference 12419665) and primary findings are published elsewhere[8]. Here, we briefly summarize key information relevant to the secondary analyses. It should be noted that this is an opportunistic study (arising because of recent findings reported by other research groups), and was not planned a priori, so the analyses must be regarded as exploratory.

Summary of the cross-over trial.   Overview: The trial was undertaken between September and October 2017 and tested the effects of residing in an evening BDLE on melatonin levels, sleep, neurocognitive arousal, and subjective sleepiness. The study was undertaken in a new-build acute psychiatric unit (prior to its opening to the public) at St. Olavs Hospital in Trondheim, Norway, where dynamic light sources were installed in one of two wards. After providing written informed consent, participants underwent a 7-day pre-randomization monitoring phase. Then participants completed a 13-day protocol where DLMO was assessed on day 1, 7, and 13. On days 2–6 and 8–12, they resided for 5 + 5 days in the LEs with the order of exposure being the result of randomization (see also Fig. 3). The randomization was performed by the Unit of Applied Clinical Research (Department of Medicine and Health Sciences, NTNU).

Participants.   Healthy participants were recruited to take part in the study. Participants met inclusion criteria if their habitual sleep–wake patterns were within normal parameters (here defined as weekday bedtime between 22:30 h and 24:00 h, weekday rise time between 06:30 h and 08:00 h, and small intraindividual variations (< 2 h) between weekdays and weekends) and they tested negative on the Ishihara plate test for color blindness. Exclusion criteria were current medical or psychological conditions, current use of prescription medication(s), family history of severe mental illness, current sleep disorders, night shift work in the preceding 2 years, trans-meridian travel in the preceding 2 months exceeding one time-zone, and/or current use of nonprescription drugs or illicit substances (not including alcohol or nicotine). Moreover, for the last 7 days preceding randomization, participants were requested to maintain a fixed sleep wake-schedule (bedtimes: 23:00 h–00:00 h, risetimes: 07:00 h–08:00 h). They were further asked to refrain from ingesting alcohol and/or caffeine after 12:00 h for the duration of the study.

Daily routine: Participants were awoken at 07:00 h each morning, left the unit by 08:00 h and returned by 17:00 h for a shared meal. At 18:00 h participants entered their assigned LE and were free to spend their time in their private rooms or the common areas. Participants retired to their bedrooms for sleep by 23:00 h and turned the lights off. Hospital staff were present at all times to ensure the safety of the participants and assist in the running of the unit including meals.

The evening light environments.   *Evening BDLE:* The light environment was created using a Light Emitting Diode (LED) lighting system containing red, blue and green-white diodes that could be individually programmed to emit different colored light. From 18:30 h until 06:50 h, this system generated blue-depleted lighting in bedrooms, bathrooms, hallways and common areas using a combination of the green-white and red diodes. Blue blocking filters automatically descended to cover windows in the evening and were retracted in the morning, whereas televisions had permanent blue-blocking filters. In addition, participants were asked to use physical blue-blocking filters (lowbluelights.com) on their electronic media devices in the evenings. During daytime, all diodes were used to generate a standard white hospital light.

Standard LE: The light environment had standard white hospital light at all times. TV screens and other electronic devices were used as normal, without blue-blocking filters.
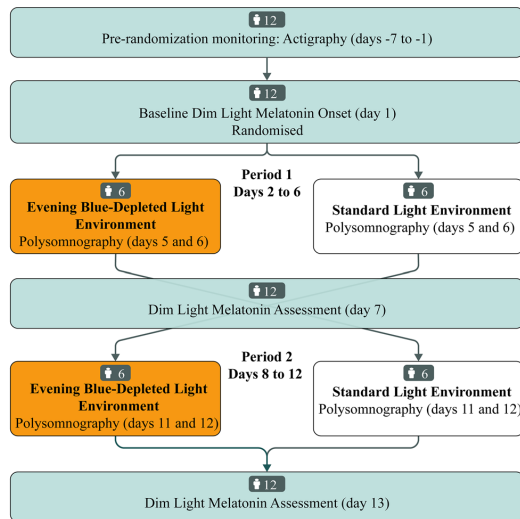
Figure 3. Overview of the study design, flow of participants in the study periods, and timing of relevant assessments. Adapted from the CONSORT guidelines extended to cross-over trials[43].

During the blue-depleted light period light levels (photopic lux) were similar in both wards, whereas levels of melanopic lux were lower in the evening BDLE (for detailed light measurements, see Vethe et al.[8]).

Assessments. See also Fig. 3 for an overview of the timing of exposure to the experimental conditions and the timing of assessments.

Polysomnographic recordings. Polysomnographic (PSG) recordings were used to assess sleep on the last two nights (4th and 5th) in each LE. The 10–20 system for electroencephalography recording was used in the mounting of the PSG equipment and included the F3, F4, C3, C4, O1 and O2 electrodes. Electrooculogram, submental electromyogram, electrocardiogram, peripheral pulse oximetry and electrodes on the legs were also used. The SOMNO HD (SOMNOmedicsGmbH, Randersacker, Germany) PSG-equipment were used to collect the data. Signals were sampled at 256 or 128 Hz, low-pass filtered, and stored at 128 Hz. Sleep stage scoring was performed according to the American Academy of Sleep Medicine rules[40] by a clinical neurophysiologist with > 10 years of experience with PSG-scoring who was blinded to participant details (individual characteristics, LE condition, etc.). After sleep stage scoring REM sleep epochs were visually inspected once more and all rapid eye movements (REMs) were marked by 3-s mini-epochs (using similar procedures as e.g. Lechinger et al.[11] and Feige et al.[41]). REM sleep microarousals was scored using an automatic scoring algorithm embedded in the PSG-scoring software DOMINO. Duration-criterion for cortical arousal was set to 3 s, and arousals were required to be accompanied by an EMG increase.

Dim light melatonin onset assessments (DLMO). We performed melatonin assessments to estimate the phase shift of DLMO from baseline until the first evening after having resided in the different evening LEs for 5 days. This translates into: a baseline assessment before entering the first light condition (day 1); an intermediary evening after 5 days in the first LE (day 7); and after 5 days in the second LE (day 13). Participants stayed in a dim light environment (< 3 lx) from 18:00 h until 23:00 h where saliva samples were collected every 60 min following a standard protocol[8]. DLMO was defined as the clock time when melatonin levels surpassed a 4 pg/mL threshold.

Pre-randomization sleep-monitoring. During the pre-randomization monitoring phase participants registered their sleep in sleep diaries upon awakening every morning, and wore an actiwatch (Actiwatch Spectrum, Philips Respironics Inc., Murrysville, PA). Participants were instructed to press an event-marker button on the actiwatch to indicate bedtime. For days where an event-marker press was missing, bedtimes indicated in the sleep diaries were used instead. Actigraphy data (sampled using 30 s epochs) was analyzed using an automatic scoring program (Actiware version 5.70.1, Philips Respironics Inc., Murrysville, PA) to calculate the rise-times and total sleep times during the pre-randomization monitoring phase.

Statistics. We used a linear mixed model with the following variables one at a time as dependent variables: REM sleep duration, REM sleep fragmentation, REM sleep microarousals, and REM-density. The models were fitted in two steps by first entering LE and phase-shift one at a time, and then simultaneously as fixed effects. A random intercept was added for participant ID in all models. This model utilizes the cross-over design such

that the effects is estimated intra-individually, resulting in higher precision than could have been achieved with a parallel group design. For the modeling of REM and N3-accumulation the combination of LE and time as percent of total sleep time rounded to the nearest integer was used as fixed effects. The difference between LEs at each percent of total sleep time was then estimated, and the resulting p-values were corrected for multiple comparisons using the Bonferroni correction.

Normality of residuals was checked by visual inspection of QQ-plots and with the Shapiro–Wilk test. In some analyses, there were some deviations from normality, in which case bootstrapping with 10,000 resamples were performed and bias-corrected and accelerated confidence intervals were used. Two-sided p-values < 0.05 were considered significant.

Statistical analyses were performed using R statistical package (version 3.6.2., https://www.R-project.org/). Linear mixed effects models were fitted (using the R-package "lme4").

Post hoc analyses of order effects. To investigate if there was an effect condition order we used a linear mixed model with condition, study period number, and their interaction as fixed effects. Dependent variables were the respective REM main-outcome variables. Participant ID was entered as a random intercept.

Rem sleep fragmentation, REM sleep microarousals, and REM-density. REM sleep fragmentation was the total overnight number of interruptions of REM sleep by non-REM sleep or wake bouts that was occurring within REM sleep episodes[42] divided by the total duration of REM sleep. A REM sleep episode was defined as starting on the first REM sleep epoch after a minimum of 15 min of continuous non-REM sleep and ended at the last epoch of REM sleep preceding at least 15 min of continuous non-REM sleep.

REM-density was calculated by dividing the sum duration of the 3-s mini-epochs marking REMs within REM-epochs by the total duration of REM sleep over the night[11].

Phase-shift of DLMO. Phase shift of DLMO was calculated by estimating the mean change in DLMO from baseline to the DLMO-assessment after each LE. Fitted values from a linear mixed model described in the previous publication[8] was used to calculate individual phase-shifts.

Accumulation of REM sleep and N3. For the accumulation of N3 and REM sleep, the epochs of the respective stages were cumulatively summed for each percent (rounded to the nearest integer) of the sleep period.

References

1. Wright, K. P. *et al.* Entrainment of the human circadian clock to the natural light-dark cycle. *Curr Biol.* **23**(16), 1554–1558. https://doi.org/10.1016/j.cub.2013.06.039 (2013).
2. Chang, A. M., Aeschbach, D., Duffy, J. F. & Czeisler, C. A. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proc. Natl. Acad. Sci. USA* **112**(4), 1232–1237. https://doi.org/10.1073/pnas.1418490112 (2015).
3. Gordijn, M. C. M., Beersma, D. G. M., Korte, H. J. & Hoofdakker, R. H. Effects of light exposure and sleep displacement on dim light melatonin onset. *J. Sleep Res.* **8**(3), 163–174. https://doi.org/10.1046/j.1365-2869.1999.00156.x (1999).
4. Cajochen, C., Kräuchi, K., Danilenko, K. V. & Wirz-Justice, A. Evening administration of melatonin and bright light: Interactions on the EEG during sleep and wakefulness. *J. Sleep Res.* **7**(3), 145–157. https://doi.org/10.1046/j.1365-2869.1998.00106.x (1998).
5. Cajochen, C. *et al.* Evidence that homeostatic sleep regulation depends on ambient lighting conditions during wakefulness. *Clocks Sleep* **1**(4), 517–531. https://doi.org/10.3390/clockssleep1040040 (2019).
6. Münch, M. *et al.* Wavelength-dependent effects of evening light exposure on sleep architecture and sleep EEG power density in men. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **290**(5), R1421–R1428. https://doi.org/10.1152/ajpregu.00478.2005 (2006).
7. Drews, H. J., Scott, J., Langsrud, K., Vethe, D. & Kallestad, H. Chronobiologically informed inpatient milieu in psychiatric institutions. *Lancet Psychiatry* **7**(12), 1013–1014. https://doi.org/10.1016/S2215-0366(20)30459-4 (2020).
8. Vethe, D. *et al.* The evening light environment in hospitals can be designed to produce less disruptive effects on the circadian system and improve sleep. *Sleep* https://doi.org/10.1093/sleep/zsaa194 (2020).
9. Simor, P., van der Wijk, G., Nobili, L. & Peigneux, P. The microstructure of REM sleep: Why phasic and tonic?. *Sleep Med. Rev.* **52**, 101305. https://doi.org/10.1016/j.smrv.2020.101305 (2020).
10. Riemann, D., Krone, L. B., Wulff, K. & Nissen, C. Sleep, insomnia, and depression. *Neuropsychopharmacol.* **45**(1), 74–89. https://doi.org/10.1038/s41386-019-0411-y (2020).
11. Lechinger, J. *et al.* REM density is associated with treatment response in major depression: Antidepressant pharmacotherapy vs. psychotherapy. *J. Psychiatr. Res.* **133**, 67–72. https://doi.org/10.1016/j.jpsychires.2020.12.009 (2021).
12. Tempesta, D., Socci, V., De Gennaro, L. & Ferrara, M. Sleep and emotional processing. *Sleep Med. Rev.* **40**, 183–195. https://doi.org/10.1016/j.smrv.2017.12.005 (2018).
13. Goldstein, A. N. & Walker, M. P. The role of sleep in emotional brain function. *Annu. Rev. Clin. Psychol.* **10**(1), 679–708. https://doi.org/10.1146/annurev-clinpsy-032813-153716 (2014).
14. Wassing, R. *et al.* Restless REM sleep impedes overnight amygdala adaptation. *Curr. Biol.* **29**(14), 2351-2358.e4. https://doi.org/10.1016/j.cub.2019.06.034 (2019).
15. Van Someren, E. J. W. Brain mechanisms of insomnia: New perspectives on causes and consequences. *Physiol. Rev.* **101**(3), 995–1046. https://doi.org/10.1152/physrev.00046.2019 (2021).
16. Wassing, R. *et al.* Slow dissolving of emotional distress contributes to hyperarousal. *Proc. Natl. Acad. Sci. USA* **113**(9), 2538–2543. https://doi.org/10.1073/pnas.1522520113 (2016).
17. Riemann, D. *et al.* REM sleep instability—A new pathway for insomnia?. *Pharmacopsychiatry* **45**(05), 167–176. https://doi.org/10.1055/s-0031-1299721 (2012).
18. Wilson, S. & Argyropoulos, S. Antidepressants and sleep: A qualitative review of the literature. *Drugs* **65**(7), 927–947. https://doi.org/10.2165/00003495-200565070-00003 (2005).

19. Riemann, D., Berger, M. & Voderholzer, U. Sleep and depression—Results from psychobiological studies: An overview. *Biol. Psychol.* **57**(1–3), 67–103. https://doi.org/10.1016/S0301-0511(01)00090-4 (2001).
20. Giles, D. E. & Roffwarg, H. P. Controlled comparison of electrophysiological sleep in families of probands with unipolar depression. *Am. J. Psychiatry.* **155**(2), 192–199 (1998).
21. Steiger, A. & Pawlowski, M. Depression and sleep. *IJMS* **20**(3), 607. https://doi.org/10.3390/ijms20030607 (2019).
22. Palagini, L., Baglioni, C., Ciapparelli, A., Gemignani, A. & Riemann, D. REM sleep dysregulation in depression: State of the art. *Sleep Med. Rev.* **17**(5), 377–390. https://doi.org/10.1016/j.smrv.2012.11.001 (2013).
23. Baglioni, C. *et al.* Sleep and mental disorders: A meta-analysis of polysomnographic research. *Psychol. Bull.* **142**(9), 969–990. https://doi.org/10.1037/bul0000053 (2016).
24. Czeisler, C. A., Zimmerman, J. C., Ronda, J. M., Moore-Ede, M. C. & Weitzman, E. D. Timing of REM sleep is coupled to the circadian rhythm of body temperature in man. *Sleep* **2**(3), 329–346. https://doi.org/10.1093/sleep/2.3.329 (1980).
25. Rahman, S. A., St. Hilaire, M. A. & Lockley, S. W. The effects of spectral tuning of evening ambient light on melatonin suppression, alertness and sleep. *Physiol. Behav.* **177**, 221–229. https://doi.org/10.1016/j.physbeh.2017.05.002 (2017).
26. Léger, D. *et al.* Slow-wave sleep: From the cell to the clinic. *Sleep Med. Rev.* **41**, 113–132. https://doi.org/10.1016/j.smrv.2018.01.008 (2018).
27. Lockley, S. W. *et al.* Short-wavelength sensitivity for the direct effects of light on alertness, vigilance, and the waking electroencephalogram in humans. *Sleep* **29**(2), 161–168. https://doi.org/10.1093/sleep/29.2.161 (2006).
28. Cho, Y. *et al.* Effects of artificial light at night on human health: A literature review of observational and experimental studies applied to exposure assessment. *Chronobiol. Int.* **32**(9), 1294–1310. https://doi.org/10.3109/07420528.2015.1073158 (2015).
29. Cajochen, C., Kräuchi, K., Möri, D., Graw, P. & Wirz-Justice, A. Melatonin and S-20098 increase REM sleep and wake-up propensity without modifying NREM sleep homeostasis. *Am. J. Physiol.* **272**(4), R1189–R1196. https://doi.org/10.1152/ajpregu.1997.272.4.R1189 (1997).
30. Wams, E. J. *et al.* Linking light exposure and subsequent sleep: A field polysomnography study in humans. *Sleep* **40**(12), zsx165. https://doi.org/10.1093/sleep/zsx165 (2017).
31. LeGates, T. A., Fernandez, D. C. & Hattar, S. Light as a central modulator of circadian rhythms, sleep and affect. *Nat. Rev. Neurosci.* **15**(7), 443–454. https://doi.org/10.1038/nrn3743 (2014).
32. Fernandez, D. C. *et al.* Light affects mood and learning through distinct retina-brain pathways. *Cell* **175**(1), 71-84.e18. https://doi.org/10.1016/j.cell.2018.08.004 (2018).
33. Münch, M. *et al.* Effects on subjective and objective alertness and sleep in response to evening light exposure in older subjects. *Behav. Brain Res.* **224**(2), 272–278. https://doi.org/10.1016/j.bbr.2011.05.029 (2011).
34. Borbely, A.A. *A Two Process Model of Sleep Regulation.* 11.
35. Borbély, A. A., Daan, S., Wirz-Justice, A. & Deboer, T. The two-process model of sleep regulation: A reappraisal. *J. Sleep Res.* **25**(2), 131–143. https://doi.org/10.1111/jsr.12371 (2016).
36. Janků, K., Šmotek, M., Fárková, E. & Kopřivová, J. Block the light and sleep well: Evening blue light filtration as a part of cognitive behavioral therapy for insomnia. *Chronobiol. Int.* **37**(2), 248–259. https://doi.org/10.1080/07420528.2019.1692859 (2020).
37. Shechter, A., Kim, E. W., St-Onge, M. P. & Westwood, A. J. Blocking nocturnal blue light for insomnia: A randomized controlled trial. *J. Psychiatr. Res.* **96**, 196–202. https://doi.org/10.1016/j.jpsychires.2017.10.015 (2018).
38. Phillips, A. J. K. *et al.* High sensitivity and interindividual variability in the response of the human circadian system to evening light. *Proc. Natl. Acad. Sci. USA* **116**(24), 12019–12024. https://doi.org/10.1073/pnas.1901824116 (2019).
39. Wehr, T. A. *et al.* Treatment of a rapidly cycling bipolar patient by using extended bed rest and darkness to stabilize the timing and duration of sleep. *Biol. Psychiatry* **43**(11), 822–828. https://doi.org/10.1016/S0006-3223(97)00542-8 (1998).
40. Berry, R. B. *et al.* AASM scoring manual updates for 2017 (Version 2.4). *J. Clin. Sleep Med.* **13**(5), 665–666. https://doi.org/10.5664/jcsm.6576 (2017).
41. Feige, B. *et al.* Does REM sleep contribute to subjective wake time in primary insomnia? A comparison of polysomnographic and subjective sleep in 100 patients. *J. Sleep Res.* **17**(2), 180–190. https://doi.org/10.1111/j.1365-2869.2008.00651.x (2008).
42. Drews, H. J. *et al.* Bed-sharing in couples is associated with increased and stabilized REM sleep and sleep-stage synchronization. *Front Psychiatry* **11**, 583. https://doi.org/10.3389/fpsyt.2020.00583 (2020).
43. Dwan, K., Li, T., Altman, D. G. & Elbourne, D. CONSORT 2010 statement: Extension to randomised crossover trials. *BMJ* **366**, l4378. https://doi.org/10.1136/bmj.l4378 (2019).

Author contributions

H.K. and K.L. originated the idea for the original data collection leading to this study. The data collection was planned by D.V., K.L., and H.K., with assistance from T.S., and M.E. D.V. and H.S.A.H. collected the data. D.V., H.J.D., H.K., J.G., and J.P.W. conceived the idea for the current study. Planning and performing of the statistical analyses was done by D.V., H.J.D., and S.L. PSG recordings were scored by M.E. The initial draft was written by D.V., H.J.D, J.S., and H.K. with critical revisions from K.L., C.L.V., K.K., P.M.P.F., H.S.A.H., S.L., T.S., J.G., J.P.W., and M.E.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-12408-w.

Correspondence and requests for materials should be addressed to D.V.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SUPPLEMENTARY INFORMATION PAPER II

Supplements
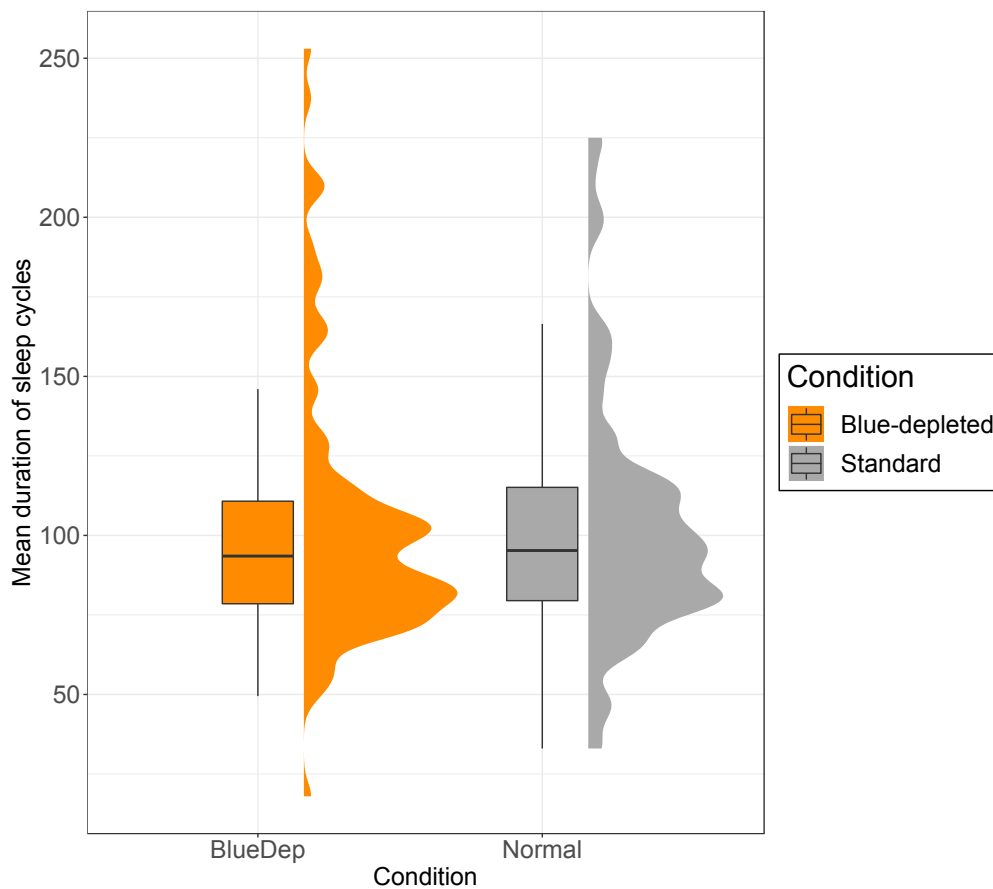
Supplementary figure S1. Boxplot indicating median duration of completed sleep cycles.

Hinges indicate the first and third quartile. Whiskers extend to 1.5*the inter quartile range

(distance between 1st and 3rd quartile).
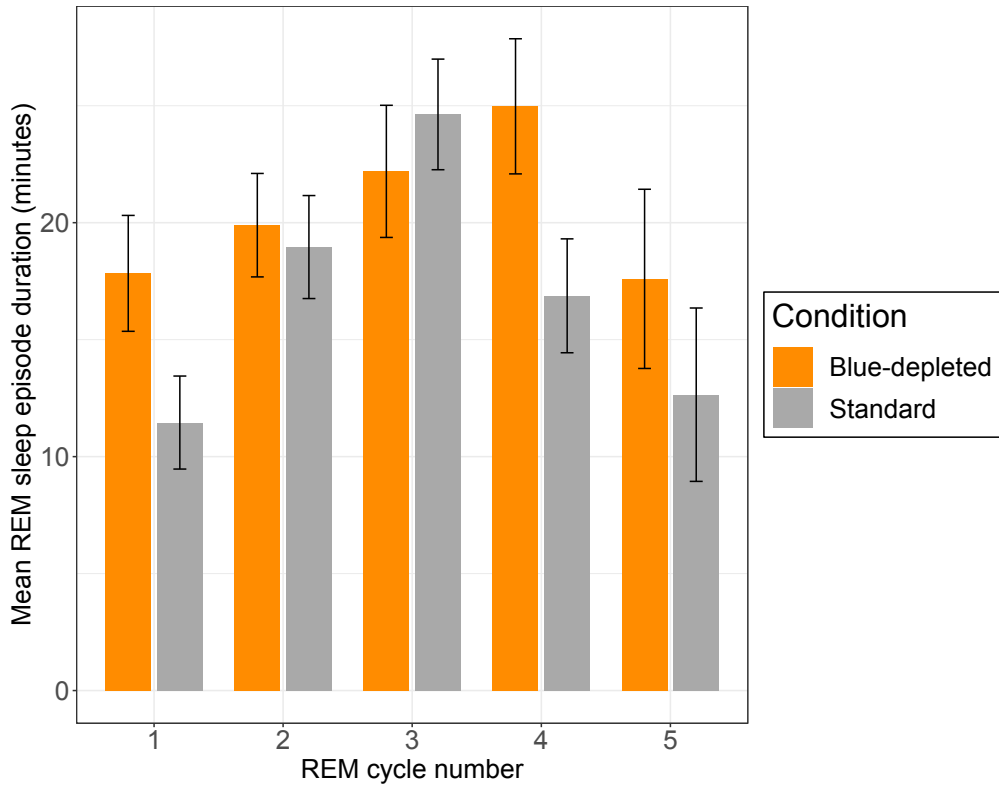
Supplementary figure S2. Mean duration of rem episode in each sleep cycle by condition.

Error bars indicate means ± SEM. Due to low N (5 nights) in cycle 6, these data-points were

not included in the figure.
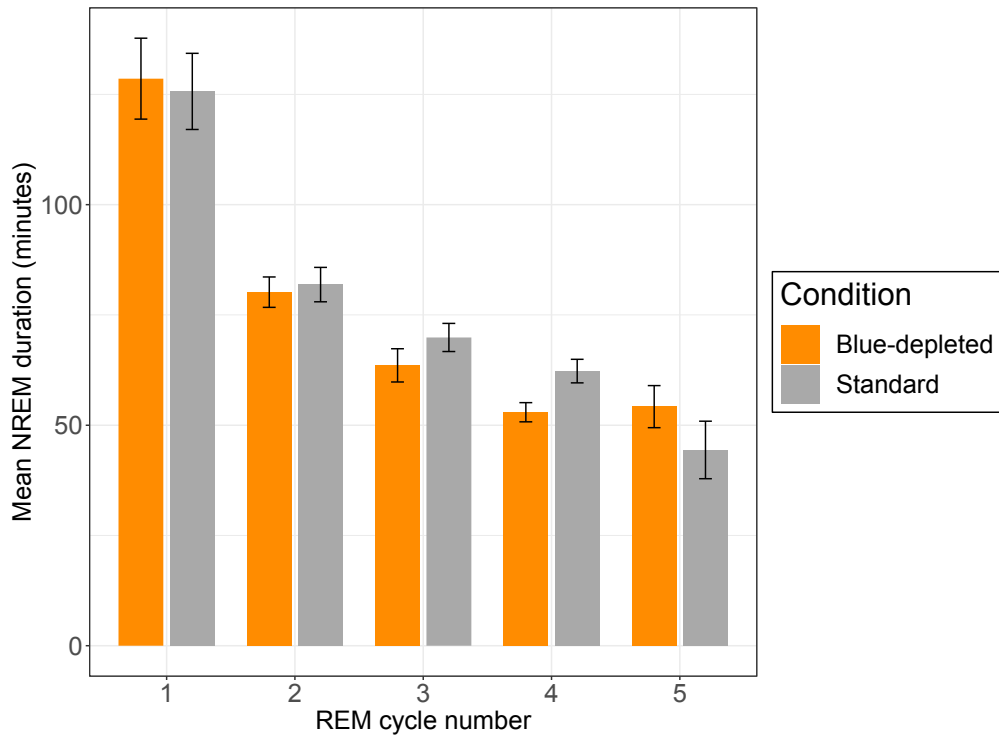
Supplementary figure S3. Mean duration of NREM sleep during each sleep cycle by condition. Error bars indicate means ± SEM. Due to low N (5 nights) in cycle 6, these data-points were not included in the figure.
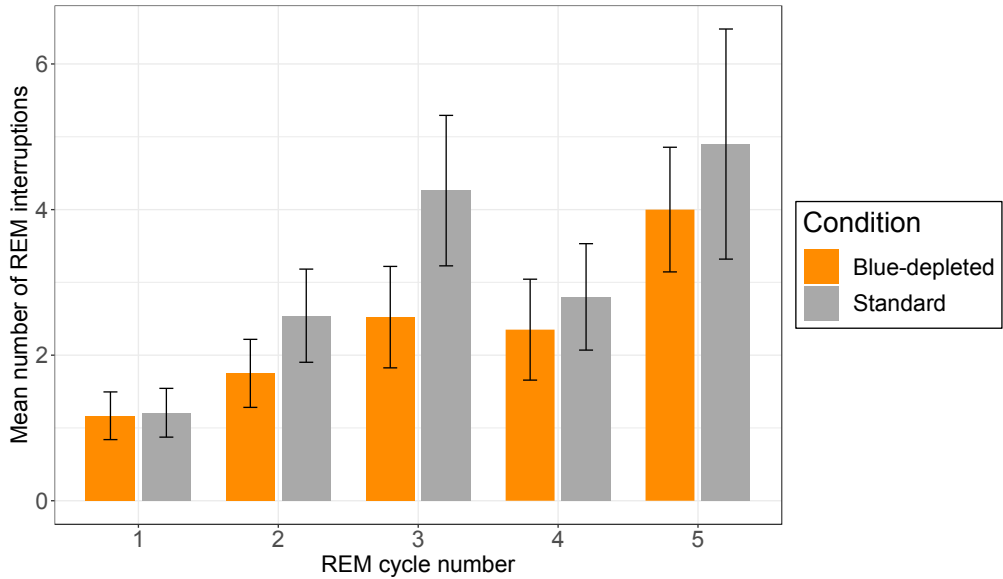
Supplementary figure S4. Mean number of REM sleep interruptions in each REM sleep cycle. Error bars indicate means ± SEM. Due to low N (5 nights) in cycle 6, these data-points were not included in the figure.
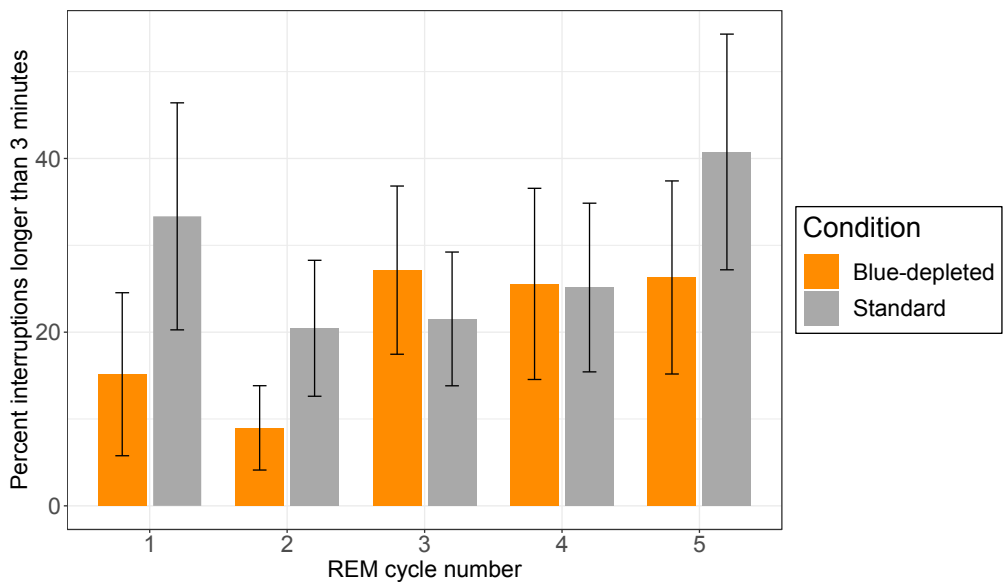
Supplementary figure S5. Percent of REM sleep interruptions lasting longer than 3 minutes by cycle number. Error bars indicate means ± SEM. Due to low N (5 nights) in cycle 6, these data-points were not included in the figure.
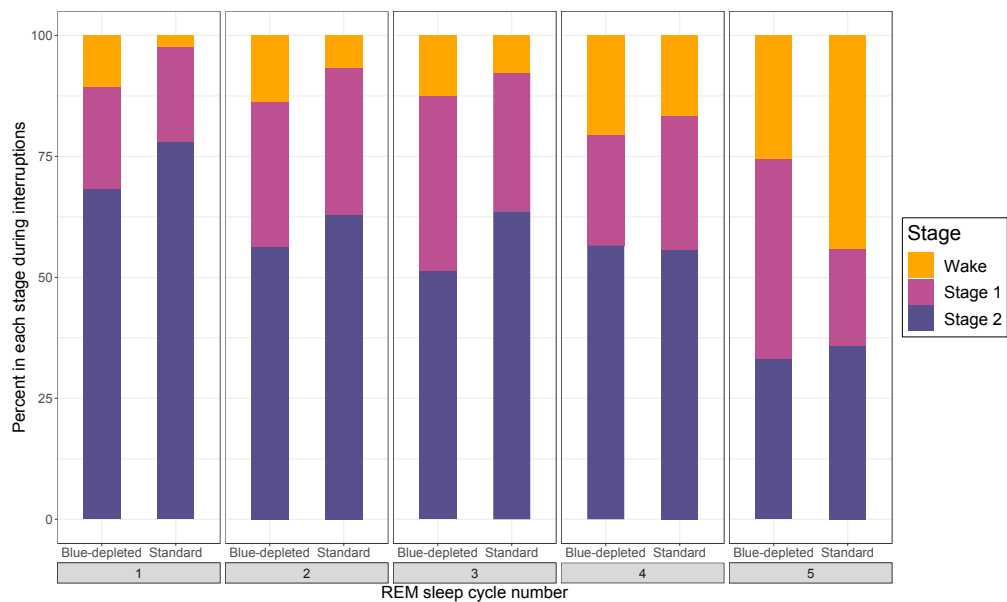
Supplementary figure S6. Distribution of sleep stages during the interruptions from REM sleep. Due to low N (5 nights) in cycle 6, these data-points were not included in the figure.
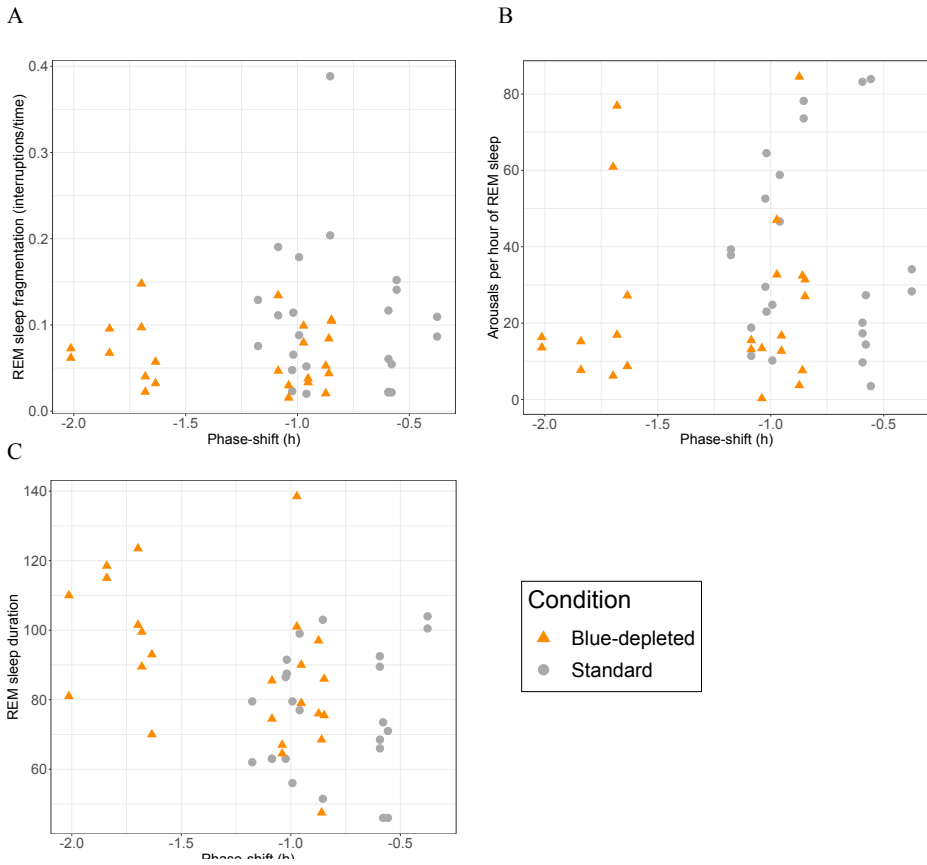
Supplementary figure S7. Observed values of (A) REM sleep fragmentation, (B) REM sleep arousals, and (C) REM sleep duration plotted against phase shift and separated by color/shape to indicate condition.
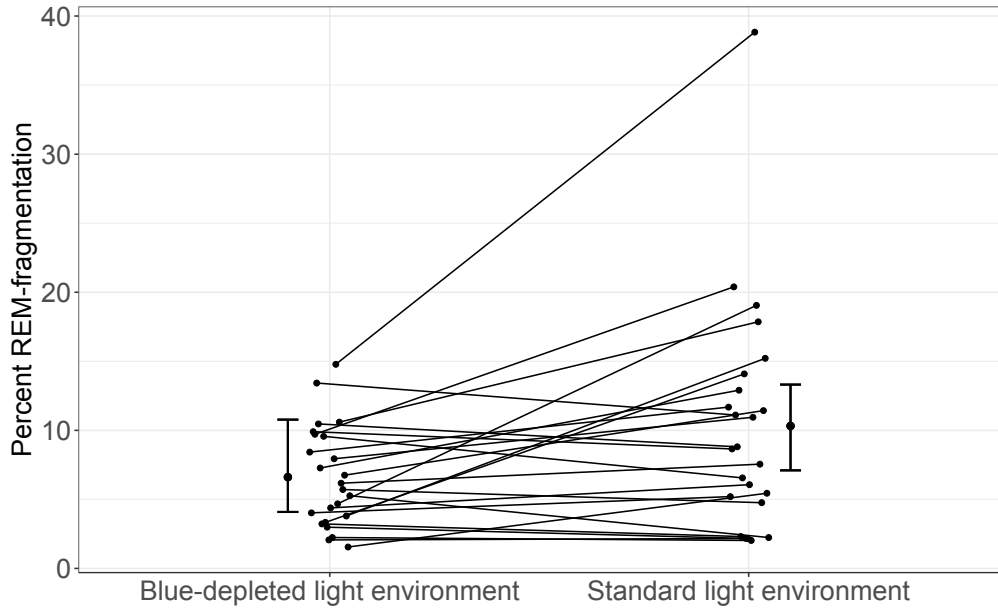
Supplementary figure S8. Individual REM sleep fragmentation in each condition. Percent REM sleep fragmentation is the number of interruptions during REM sleep divided by total duration of REM sleep in minutes times 100. Error bars indicate estimated means ± 95% confidence intervals from the linear mixed model (N = 12). The omission of the outlier value in standard LE did not meaningfully change the results of the analysis.

			NREM microarousals per hour	
Step		Estimate	95% CI	*p*
1	BDLE	-9.50	-19.48 to 0.49	0.063
	Phase shift	10.71	-1.21 to 23.50	0.079
2	BDLE	-6.48	-4.95 to 18.24	0.27
	Phase shift	-6.36	-33.25 to 2.83	0.38

Supplementary Table S1. Results from a similar mixed model to the main analyses fitted with NREM microarousals per hour of NREM sleep as the dependent variable. Independent variables are BDLE and phase shift. In the first step variables are tested separately, and in the second step variables are entered together. N=12 participants. BDLE=Blue-depleted light environment

PAPER III

Original Article

Distinguishing sleep from wake with a radar sensor: a contact-free real-time sleep monitor

Hanne Siri Amdahl Heglum[1,2,*,•], Håvard Kallestad[4,5,•], Daniel Vethe[4,5,•],
Knut Langsrud[4,5], Trond Sand[1,3,•] and Morten Engstrøm[1,3]

[1]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim,
Norway, [2]Novelda AS, Trondheim, Norway, [3]Department of Neurology and Clinical Neurophysiology, St. Olavs University
Hospital, Trondheim, Norway, [4]Department of Mental Health, Norwegian University of Science and Technology,
Trondheim, Norway and [5]Division of Mental Health Care, St. Olavs University Hospital, Trondheim, Norway

*Corresponding author. Hanne Siri Amdahl Heglum, Department of Neuromedicine and Movement Science, NTNU, Edvard Griegs Gate 8, Trondheim,
N-7491 Trondheim, Norway. Email: hanne.s.a.heglum@ntnu.no.

Abstract

This work aimed to evaluate whether a radar sensor can distinguish sleep from wakefulness in real time. The sensor detects body movements without direct physical contact with the subject and can be embedded in the roof of a hospital room for completely unobtrusive monitoring. We conducted simultaneous recordings with polysomnography, actigraphy, and radar on two groups: healthy young adults ($n = 12$, four nights per participant) and patients referred to a sleep examination ($n = 28$, one night per participant). We developed models for sleep/wake classification based on principles commonly used by actigraphy, including real-time models, and tested them on both datasets. We estimated a set of commonly reported sleep parameters from these data, including total-sleep-time, sleep-onset-latency, sleep-efficiency, and wake-after-sleep-onset, and evaluated the inter-method reliability of these estimates. Classification results were on-par with, or exceeding, those often seen for actigraphy. For real-time models in healthy young adults, accuracies were above 92%, sensitivities above 95%, specificities above 83%, and all Cohen's kappa values were above 0.81 compared to polysomnography. For patients referred to a sleep examination, accuracies were above 81%, sensitivities about 89%, specificities above 53%, and Cohen's kappa values above 0.44. Sleep variable estimates showed no significant intermethod bias, but the limits of agreement were quite wide for the group of patients referred to a sleep examination. Our results indicate that the radar has the potential to offer the benefits of contact-free real-time monitoring of sleep, both for in-patients and for ambulatory home monitoring.

Statement of Significance

This work shows that a contact-free radar sensor with our real-time actigraphy-inspired algorithm can detect body movements and provide valid estimates of sleep, wakefulness, and related parameters. The performance was best for healthy volunteers in a psychiatric hospital environment, with the radar placed either on the nightstand or permanently mounted in the ceiling. The radar generally performed on par with actigraphy for sleep/wake classification. Contact-free recording is an advantage for patients with less tolerance for wearable devices, e.g. in psychiatric hospitals. Future studies should attempt to improve the performance for home use and further validate this tool for sleep/wake classification in a wider population, particularly in a real in-hospital setting. Sleep-stage classification from radar data should also be explored.

Key words: sleep; radar; actigraphy; polysomnography; sleep monitoring; ambulatory home monitoring

---

Introduction

The polysomnography (PSG) test [1] is considered the gold standard of sleep studies, but it is not always an ideal option [2, 3]. The multiple on-body sensors can be experienced as invasive and uncomfortable, and the time and expense involved in laborious manual interpretation make it a poor choice for long-term monitoring or population studies. Fortunately, electrophysiological signals are not the only way to get a measure of sleep. Body movement modalities have been used for this purpose since the first actigraphs were validated in the late 1970s–early 1980s [4–6]. These devices typically record activity with an on-body accelerometer and determine activity–rest from those signals, commonly through simple automatic classification models [7, 8]. Compared with PSG, they are cost-effective, noninvasive, reasonably reliable for estimating periods of sleep (although also recognized as having unfortunately low specificity for periods of wakefulness during the night), and particularly well suited for monitoring activity–rest cycles over time. These properties have granted them a position of ubiquity in sleep medicine and research as an alternative when PSG is not feasible or desirable [9–11].

However, sometimes even a wrist actigraph can be too demanding. For patients in a psychiatric hospital setting, with, for example, disorganized behavior, psychosis, or suicidal intents, the use of on-body sensor equipment of any kind can be challenging, and potentially dangerous. And although actigraphy data are collected prospectively, they are analyzed retrospectively, so they can generally not be used to monitor sleep in real time. Thus, they are less useful for hospital night staff who needs to keep track of the sleep/wake state of patients at any given time. So, even though sleep problems are virtually ubiquitous among inpatients with mental disorders [12], and although assessment of sleep can be crucial in diagnosis and decision making, psychiatric units typically remain limited to intermittent staff observations. This solution is far from ideal; the reliability is limited, and it has been argued that the downsides in terms of sleep disruption and privacy infringement for the patients might outweigh the benefits of this type of monitoring by a fairly wide margin [13].

This work seeks to evaluate if a radar sensor can be used to provide an unobtrusive contact-free objective measure of sleep. The sensor in question, an Impulse-Radio Ultra-Wideband (IR-UWB) radar, can detect a range of movements from a person in a room without requiring them to wear equipment on their body; from the big movements of limbs to the very small motions induced by respiration. Under favorable conditions, even the miniscule chest movements caused by a beating heart can be recognized and recorded [14, 15]. Clothes or beddings do not impede these signals [16], and since the registered data cannot be used to directly identify individuals, a degree of privacy is preserved. The radar is also capable of communicating its signals in real time.

The general aim of the present study was to evaluate whether the body movement-derived signals recorded with the radar sensor can be used to distinguish sleep from wake at least as well as the activity signals from wrist actigraphy. In addition, we aimed to evaluate whether contact-free sleep monitoring can be done in real time. Our main objectives were to develop sleep/wake classification models for the radar data, including real-time models, and to compare the classification performance of the radar with PSG (the gold standard) and actigraphy. Models were first applied to a homogenous dataset from young healthy adults recorded in a hospital environment, and then to a heterogenous dataset from ambulatory sleep clinic patients. The specific objectives for the evaluation were: (1) Estimate classification performance for sleep-wakefulness detection by accuracy, sensitivity, specificity, and Cohen's kappa statistics using both real-time and future-dependent models, and (2) estimate agreement between methods for the main sleep outcome variable total sleep time (TST) using Bland–Altman analysis, and then similarly for four secondary sleep variables; sleep onset latency (SOL), wake after sleep onset (WASO), sleep efficiency (SE%), and number of awakenings (NW).

Methods

Data collection

Two sets of data were collected for this study from different populations. The study protocol for the randomized cross-over trial from which Dataset 1 (DS1) was acquired, and the protocol was approved by the Regional Ethical Committee in Trondheim, (Central Norway; REK: 2017/916) and is registered on the ISRCTN website (reference number 12419665). The study was undertaken in accordance with the Revised Declaration of Geneva [17] and written informed consent was obtained from all participants. The protocol for gathering Dataset 2 (DS2) was approved by the Regional Ethical Committee in Trondheim, (Central Norway; REK: 2017/309). Written informed consent was obtained from all participants.

Both datasets consist of simultaneously collected data from PSG, actigraphy, and IR-UWB radar.

PSG

PSG recordings were recorded using Somnomedics Somno HD equipment (Somnomedics GmbH, Randersacker, Germany). Six EEG electrodes were placed according to the International (10–20) system [18]; F3, F4, C3, C4, O1, O2, plus a mastoid reference left side (M1) for the electrodes on the right side and a mastoid reference right side for the left side electrodes (M2); two electrooculografic electrodes (EOG) applied 1 cm laterally and, respectively, 2 cm above and below the right and left lateral eye cantus. EOG-reference electrodes were applied to the left (M1) and the right (M2) mastoids. Surface electromyography (EMG) was registered from the submental and bilateral anterior tibial muscles.

Actigraphy

Phillips Actiwatch (Actiwatch Spectrum, Philips Respironics Inc., Murrysville) placed on the nondominant wrist were used for both datasets. Automatic scoring algorithms from the manufacturer were not used in this work. Instead, binned movement data were exported from the actigraphs using Actiware (version 5.70.1; Philips Respironics Inc., Murrysville, PA), and then treated in the same way as the simultaneously recorded radar data. The actigraphy and radar data were subjected to identical sleep/wake classification model development, resulting in models of similar form but with different parameters. These models and their parameters are reported in their entirety in Supplementary Table S1.

IR-UWB *radar*

The radar used was the XeThru model X4M200, a commercially available radar sensor developed by Novelda AS. All radar data for this work were stored in baseband I/Q [19] form, to enable different or improved digital signal processing (DSP) at a later date, and then subjected to the pulse-Doppler signal processing provided by the manufacturer. Their Respiration_2 profile was used, to obtain body movement and respiration rate estimates at a rate of 1 Hz. This profile has a detection zone of 0.40–5.00 m, and a respiration detector range of 8–30 respirations-per-minute (RPM). A more detailed description of the radar can be found in Supplementary Material, or in the manufacturer's datasheet [20].

Dataset 1

Dataset 1 (DS1) consists of data from twelve healthy participants of 20–30 years (5 male). These data were collected as part of a randomized cross-over trial meant to evaluate the effect of the light conditions in a state-of-the-art acute psychiatric hospital unit at St. Olavs Hospital, Trondheim, Norway [21]. All 40 patient rooms in this building have radar sensors installed in the ceiling. This study was conducted in the last phase of the building construction period before the unit was opened for patient admissions. Prospective participants were eligible for inclusion if their habitual sleep/wake cycle was normal; i.e. weekday bedtime between 22:30 h and 00:00 h and weekday rise time between 06:30 h and 08:00 h, with small intraindividual variations (<2 h) between weekdays and weekends, and no colour blindness. Exclusion criteria were evidence of any current medical or psychological condition, current use of prescription medication(s), family history of severe mental illness, current sleep problems, night shift work in the preceding 2 years, trans-meridian travel in the preceding 2 months, and/or current use of non-prescription drugs or illicit substances. From September 23, 2017, to October 5, 2017, these participants resided for a total of 10 days in the hospital; five consecutive days in each light environment, with randomized order of exposure and one day of washout in between. Between 08:00 h and 17:00 h participants had to leave the unit to follow their normal daily life of work or studies. They spent the remaining time in the hospital, and from 18:00 h to 07:00 h they were confined to their assigned light environment, with a set bedtime at 23:00 h every night. Each participant wore an Actiwatch set to 15-s epochs every day for the duration of the study, also during the daytime hours. Each room had one radar sensor mounted in the ceiling, and one placed on a nightstand next to the bed. These recorded continuously at 17 frames-per-second (FPS) for the duration of the study. Each participant underwent a total of four nights of PSG; two consecutive nights in each light condition.

Dataset 2

Dataset 2 was collected to observe a broad cross-section of heterogenous troubled sleepers. From 2017 to 2020, patients referred to the Department of Clinical Neurophysiology at St. Olavs Hospital in Trondheim, Norway, for overnight sleep examination for sleep problems of any kind, could be asked to participate. The only inclusion criterion was their willingness to participate and their informed consent, and there were no

exclusion criteria. Participants were outfitted with portable PSG equipment and sent home for ambulatory sleep monitoring, as per standard practice at this department. (Some participants were originally referred to as respiratory polygraphy, a simpler examination than PSG. These patients were "upgraded" to full PSG upon consenting to participate in this study.) In addition, they were given a Phillips Actiwatch actigraph set to 30-s epoch length, and a portable radar sensor configured to record baseband data at 300 FPS—this higher framerate was chosen to enable a more detailed analysis of these recordings at a later date. For the present work, the radar data was downsampled to 17 FPS and processed in the same way as the radar data from DS1. The participants were instructed to place the radar sensor on their nightstand (or on a provided camera stand, if they did not have a nightstand), and to be alone in their bed on the night of the recording.

Data preparation

Each PSG recording was manually scored by a specialist in clinical neurophysiology according to the AASM Manual for the Scoring of Sleep and Associated Events, version 2.4 [22], and then exported to ASCII-format from the Domino software. The raw radar data was processed using software provided by Novelda AS to output activity and respiration estimates at 1 Hz. Differences in internal clocks for radar, actigraphy, and PSG were corrected with an alignment method based on maximum correlation. To match the length of the PSG epochs, 15-s actigraphy activity counts from DS1 and 1 Hz radar data were aggregated into 30-s bins by taking the mean. The three data types output by the radar digital signal processing [fast movement, slow movement, and respirations per minute (RPM)] were scaled to unify their order of magnitude. Further details about data types and preparations can be found in Supplementary Material.

Sleep/wake classification model development

The inspiration for the sleep/wake classification approach taken in this work comes from actigraphy. Of the four most common methods for processing wrist actigraphic data from adults, three are based on linear-sliding sum models over a time-horizon [7]. Activity data is binned into epochs of some specific length, commonly 30 s, resulting in a single activity value per epoch. Each epoch is then scored as sleep or wake by comparing a weighted sum of activity values from some "time horizon," i.e. a number of past, present, and future epochs, to some threshold. These three common methods all use the same time horizon; they score each epoch in the time series by considering its activity value in conjunction with those of the four preceding and two succeeding epochs. Only the parameters of the models are different, optimized for their specific datasets and hardware. One of the three (Rescored Cole–Kripke) also impose an additional layer of post hoc heuristic rules on the output of the initial classifications, which for this work we will call the Cole–Kripke re-scoring rules [23, 24].

PSG-scored sleep stages N1, N2, N3, and REM were combined into "sleep" and given the value 0, and epochs with the "wake" state were given the value 1. DS1 was split equally into training and test sets (DS1-train and DS1-test) by assigning the participants to either at random. With the PSG sleep/wake state as the

target variable, logistic regression was performed over DS1-train to estimate the parameters of linear sliding-sum models. This procedure was followed for both activity data exported from the actigraphs, and for the radar data. For the radar, each of the three data types (fast movement, slow movement, and RPM) were included separately in the regression, giving three data points per epoch in contrast to the single activity value per epoch used for actigraphy. Every combination of horizon length between zero (only the present) up to and including ten epochs in either direction around the present were considered, for a total of 121 models per sensor. The resulting models output a value between zero and one for each epoch, which can be interpreted as "probability of wake." A $p = 0.5$ threshold was used on the continuous output values from the regression models to classify each epoch as either sleep or wake. Finally, the Cole–Kripke rescoring rules [23] were applied. For real-time models, the final two rules, (4) and (5), had to be excluded, because they depend on future information. The rules used in our work are: (1) after at least 4 min scored as wake, the next 1 min scored as sleep is rescored wake, (2) after at least 10 min scored as wake, the next 3 min scored as sleep are rescored wake, and (3) after at least 15 min scored as wake, the next 4 min scored as sleep are rescored wake. A more detailed description of the model development can be found in Supplementary Material.

Classification performance analysis

The classification models were applied to DS1-test, and DS2. Epoch-by-epoch classification performance was evaluated against PSG ground truth by calculating overall accuracy, sensitivity, and specificity values. Because the data contain significantly more epochs of sleep than wake, Cohen's kappa coefficients were also calculated to account for the high probability of correct classification occurring by chance. Cohen's kappa statistic has a range of –1 to 1, where zero indicates agreement equivalent to classification by random chance and $\kappa = 1.00$ indicates perfect agreement. Universally accepted guidelines of interpretation for values between zero and one do not exist, so for the purpose of this work we will adopt the categories from Ref. [25]: $0.41 \leq \kappa < 0.60$ = moderate agreement, $0.61 \leq \kappa < 0.80$ = substantial agreement, and $0.81 \leq \kappa < 0.99$ = near-perfect agreement. Forest plots were used to compare the classification performance statistics of the actigraph to those of the radar devices.

Sleep parameters

Four commonly reported sleep parameters were calculated for all sleep/wake classification results: SOL, the duration between reported bedtime and objectively estimated sleep onset time; TST, the total time spent asleep during a major sleep period; WASO, the total time awake between sleep onset and offset; and SE, the percentage of time spent asleep during a major sleep period. The overall NW during the major sleep period was also counted. For calculating SOL, participants reported their bedtime by pushing a user marker on their PSG equipment. For the nights in DS1 for which no PSG was available, SOL was calculated from the set bedtime at 23:00. For PSG, sleep onset was defined as the first 30-s epoch of any sleep stage after reported bedtime. For both actigraphy and radar, sleep onset was defined as the first epoch of the first three-minute period consecutively

scored as sleep, as per the definition used in [26]. Sleep offset, i.e. wake time, was defined as the first epoch after the final epoch scored as sleep. Student's *t*-tests were performed to test the hypothesis that the mean difference between compared modalities and PSG was zero. Cohen's D was used as an effect-size measure, calculated by dividing the difference of the means on the pooled standard deviations. The pooled standard deviations were calculated by averaging the square of the standard deviations and taking the square root of the result [27]. Bland–Altman plots were used for visual comparison of agreements between modalities, plotted with bias and 95% limits of agreement (LA). The regression approach for nonuniform differences was employed to look for proportional bias. When a statistically significant slope was identified, the regression line was included in the Bland–Altman plot, and the R^2 value was reported [28]. Forest plots were used to compare the parameter estimates made with actigraph and radar data, in terms of their absolute difference to corresponding PSG parameters. MATLAB (versions R2018-R2020) was used for all analyses.

Results

DS1 contains recordings from 12 healthy young adults (mean age ± SD: 23.0 ± 3.1 years, 5 male). An equipment error in one PSG recording and a malfunctioning ceiling radar sensor left 43 nights of triple-registered recordings available for analysis. In total, 126 nights of double-registered nightstand radar and actigraphy data were available, with 117 nights also containing data from the ceiling radar. For actigraphy and the nightstand radar, 24 nights of triple registered data were assigned to DS1-train. For the ceiling radar, two of these nights were missing data, leaving 22 nights. The rest of the data was assigned to DS1-test. DS2 contains triple recorded PSG, nightstand radar, and actigraphy recordings from 28 adult sleep clinic patients (mean age ± SD: 46.25 ± 13.98 years, 19 male). Of these, 13 had obstructive sleep apnea, the rest miscellaneous, and often multiple sleep problems including excessive daytime sleepiness, headaches, restless leg syndrome, and insomnia. A summary of the datasets can be found in Table 1.

Table 2 shows the epoch-by-epoch classification performance of two models, one real-time and one with the

Table 1. Datasets summary

	Nightstand	Ceiling
Healthy volunteers* – training[†]		
Radar + actigraphy	63	59
Radar + actigraphy + PSG[§]	24	22
Healthy volunteers – test[†]		
Radar + actigraphy	62	58
Radar + actigraphy + PSG[§]	23	21
Patients with sleep disorders[‡]		
Radar + PSG + actigraphy	28	0

Number of concurrent nights of recording performed with three types of sensors/equipment.

*$n = 12$, mean age ± SD: 23.0 ± 3.1 years, 5 males.

[†]The participants were randomly assigned to a training set for model development, and a testing set for validation.

[‡]Ambulatory sleep disorder patients, mean age ± SD: 46.25 ± 13.98 years, 19 males.

[§]PSG = polysomnography.

Table 2. Classification performance statistics

Model type*	Dataset	Sensor type/placement	Accuracy [%]	Specificity [%]	Sensitivity [%]	Cohen's kappa*100
Four past, two future	Healthy volunteers test set[†]	Radar nightstand	94.8 (1.7)	89.8 (7.1)	96.6 (1.6)	86.9 (5.4)
		Radar ceiling	93.8 (2.3)	86.3 (7.4)	96.6 (2.1)	84.3 (6.6)
		Actigraphy	93.1 (1.2)	85.4 (7.4)	96.0 (2.8)	82.6 (3.2)
Five past, zero future (real time)	Healthy volunteers test set	Radar nightstand	94.5 (1.6)	89.5 (6.4)	96.3 (1.5)	86.3 (4.9)
		Radar ceiling	93.3 (2.4)	85.4 (7.7)	96.3 (1.9)	83.1 (6.8)
		Actigraphy	92.7 (1.1)	83.9 (7.7)	96.1 (2.6)	81.5 (3.2)
Four past, two future	Patients with sleep disorders [‡]	Radar nightstand	80.9 (15.7)	53.7 (18.4)	89.5 (16.9)	44.8 (25.6)
		Actigraph	83.8 (9.0)	74.3 (20.0)	89.4 (7.3)	53.3 (15.8)
Five past, zero future (real time)	Patients with sleep disorders	Radar nightstand	80.9 (15.3)	53.4 (18.7)	89.7 (16.5)	44.3 (24.8)
		Actigraphy	84.1 (9.0)	74.0 (20.2)	89.9 (6.9)	53.8 (16.2)

Epoch-by-epoch classification performance statistics for two models, both with the heuristic Cole–Kripke rescoring rules applied, compared to PSG[§]-determined sleep/wake. Mean (SD) over the participants in the datasets.

*The models are defined by the number of preceding (past) and succeeding (future) epochs used to score a present epoch.

[†]n = 12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n = 24/22 for nightstand/ceiling), and a testing set for validation (n = 23/21 for nightstand/ceiling). Values were calculated for all epochs from each participant first, then averaged together.

[‡]Ambulatory sleep disorder patients. n = 28, mean age ± SD: 46.25 ± 13.98 years, 19 male.

[§]PSG = polysomnography.

four-past-two-future horizon seen in the most common actigraphy algorithms, both with the heuristic Cole–Kripke rescoring rules applied. When multiple nights were collected from each participant in a dataset, values were calculated for all epochs from each participant first, then averaged together. Confusion matrices showing the total number of correctly and incorrectly classified epochs for each model are presented in Table 3. Classification performance of other time horizons with and without rescoring can be found in Supplementary Material. In general, we observed that different time horizons had little effect on model performance beyond a slight decrease as the overall horizon length approached zero. The rescoring rules generally improved the accuracy, specificity, and Cohen's kappa values, at the cost of a slight decrease in sensitivity.

For both datasets, the differences between the real-time model and the non-real-time model were in the sub-percent range for all performance statistics. Across all metrics, the models performed better on DS1-test than on DS2. For the DS1-test, the nightstand radar achieved the best results, followed by the ceiling radar, and then by actigraphy, which outperformed the ceiling radar in terms of specificity but otherwise achieved slightly lower scores. All accuracies were above 92%, all sensitivities above 95%, all specificities above 83%, and all Cohen's kappa values were above 0.81. For DS2 the accuracy of actigraphy was around 83% compared with around 81% for the radar. Both achieved a sensitivity for sleep of approximately 89%, but in specificity for wake and the Cohen's kappa statistic, actigraphy notably outperformed the nightstand radar, with respectively 74% and 0.53 compared with 53% and 0.44.

Sleep variables for the triple-recorded nights for PSG and derived for each sensor for two selected models are shown in Table 4, along with Cohen's D effect sizes and superscripts to indicate parameters significantly different from their PSG counterparts. An expanded version of this table also including the p-values can be found in Supplementary Table S6. Bland–Altman plots for the real-time models can be seen in Figures 1 and 2, with their biases and LAs reported in Table 5 along with t-tests on the hypothesis of zero bias.

For the real-time models applied to DS1-test, all TST estimates were similar to PSG (Table 4) and without significant bias (Table 5). Significant overestimation of SOL by 2.5 and 4.0 minutes (Cohen's D: 0.41 and 0.63) was found for the nightstand radar and actigraphy respectively, with no significant bias found in SOL for the ceiling radar. NW was significantly underestimated by 5.3 and 7.9 discrete awakenings (Cohen's D: −1.09 and −1.57) for the nightstand and the ceiling radars respectively. A single significant slope was detected, for SE compared between PSG and ceiling radar (R^2: 0.41). This slope is driven entirely by a single outlying data point, and disappears when this single point is removed. (The same data point can be observed outside of the LAs in WASO for the same comparates, but does not create a significant trend for this parameter). For the non-real-time models, a significant difference to PSG was found for the same parameters; nightstand SOL (+2.1 min), actigraph SOL (+3.2 min), nightstand NW (−7.8 awakenings), and ceiling NW (−9.7 awakenings). Additionally, the non-real-time model significantly underestimated NW by 4.8 discrete awakenings (Cohen's D: −0.93) for the actigraph. For the same models applied to DS2, TST means were almost identical for PSG and nightstand radar (Table 4), and bias was not present (Table 5). No significant biases were seen in any sleep parameters, but their variations (seen in the standard deviations and confidence interval width) were quite a bit larger than for the DS1-test. In general, we observed good agreement both for TST and the secondary sleep variables.

The estimated sleep parameters for all nights of the DS1-test, including those for which PSG was not available, can be seen in Table 6. Bland–Altman plots for the real-time model can be seen in Figure 3, and the biases and LAs are reported in Table 7. No significant differences were found in estimated TST values. A significant difference in SOL of 3.2 minutes was found between the two radar positions, with the nightstand estimating a longer latency than the ceiling positioned sensor. Significant differences were found for both WASO and SE between actigraphy and both radar positions, of 16.8 and 11.5 min WASO and 2.6% and 2.1% SE, respectively, for the nightstand and the ceiling positions. No significant difference was found between the radar positions for these parameters. Significant differences were found

Table 3. Confusion matrices With four past and two future epochs included in the model

Healthy volunteers test set*		Radar nightstand		Radar ceiling		Actigraphy	
		Wake	Sleep	Wake	Sleep	Wake	Sleep
PSG[†]	Wake	7489	803	6551	1027	6961	1098
	Sleep	708	19 644	667	17 877	845	19 505
TPR/FNR[‡]		0.97/0.03		0.96/0.04		0.96/0.04	
FPR/TNR[§]		0.10/0.90		0.14/0.86		0.14/0.86	

Patients with sleep disorders[‖]		Radar nightstand				Actigraphy	
		Wake	Sleep			Wake	Sleep
PSG	Wake	3515	3070			4381	2204
	Sleep	2354	19 123			2352	19 125
TPR/FNR		0.89/0.11				0.89/0.11	
FPR/TNR		0.47/0.53				0.33/0.67	

With five past and zero future epochs included in the model

Healthy volunteers test set		Radar nightstand		Radar ceiling		Actigraphy	
		Wake	Sleep	Wake	Sleep	Wake	Sleep
PSG	Wake	7465	827	6479	1099	6841	1220
	Sleep	756	19 596	717	17 827	829	19 521
TPR/FNR		0.96/0.04		0.96/0.04		0.96/0.04	
FPR/TNR		0.10/0.90		0.15/0.85		0.15/0.85	

Patients with sleep disorders		Radar nightstand				Actigraphy	
		Wake	Sleep			Wake	Sleep
PSG	Wake	3483	3102			4350	2235
	Sleep	2309	19 168			2253	19 224
TPR/FNR		0.89/0.11				0.90/0.10	
FPR/TNR		0.47/0.53				0.34/0.66	

Confusion matrices of epoch-by-epoch classification performance for two selected models, both with the heuristic Cole–Kripke rescoring rules applied, over two datasets. Numbers on the main diagonals of the 2 × 2 matrices represent correctly classified epochs.

*n = 12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n = 24/22 for nightstand/ceiling), and a testing set for validation (n = 23/21 for nightstand/ceiling).
[†]PSG, Polysomnography.
[‡]TPR, true positive rate. FNR, false negative rate.
[§]FPR, false positive rate. TNR, true negative rate.
[‖]Ambulatory sleep disorder patients. n = 28, mean age ± SD: 46.25 ± 13.98 years, 19 males.

between all estimates of NW, of 6.1 and 8.3 discrete awakenings between actigraphy and the nightstand and ceiling radars, respectively, and of 2.8 discrete awakenings between the two radars. Significant slopes indicating the presence of proportional bias were detected for the WASO compared between actigraph and nightstand radar (R^2: 0.29), and for TST, WASO, and SE compared between the ceiling and nightstand radars (R^2: 0.39, 0.15, and 0.36, respectively).

Forest plots for investigation of the comparative performance of the actigraph versus the radar, using the real-time model, are shown in Figure 4. The first column shows the compared absolute difference between PSG-derived sleep parameters and corresponding parameters estimated from the actigraph and the radar respectively. The second column shows the difference in classification performance statistics between actigraph and radar. Both columns are plotted as means of the differences along with 95% confidence intervals. Skewing left of zero favours the radar and vice versa for the actigraph to the right. For DS1-test in the top two rows, most plotted parameters have means either centred or skewing slightly left, with confidence intervals enveloping zero. The two exceptions both appear for

the sleep parameters of the ceiling radar, where SOL falls entirely to the left of the line and NW entirely to the right. For DS2 on the bottom row, the confidence intervals are wider and the means skewing more to the right. All sleep parameters envelop zero in their confidence intervals, but for the classification parameters, the confidence intervals of both Cohen's kappa and specificity fall entirely to the right of the line.

Temporal raster plots can be generated to visualize and summarize the patient's stay at the hospital. Figure 5 shows an example for a single participant from DS1, generated from the ceiling-mounted radar. The probability estimate output from the real-time model has been plotted as a line, and the sleep/wake classification result is indicated by the background colour.

Discussion

Our main results are: (1) Both nightstand and ceiling-mounted radars showed excellent to good agreement with PSG for sleep detection, quite comparable to actigraphy. (2) The performance differences between real-time and non-real-time models were

Table 4. Sleep parameters

Healthy volunteers test set[*]

Model type[†]	Variable[‡]	PSG[§] (n = 23)	Nightstand radar (n = 23) Mean (SD)	Cohen's D	Ceiling radar (n = 21) Mean (SD)	Cohen's D	Actigraphy (n = 23) Mean (SD)	Cohen's D
Four past, two future	TST [min]	434.4 (18.2)	437.4 (17.2)	0.17	438.3 (22.6)	0.24	433.2 (20.1)	−0.06
	SOL [min]	10.5 (6.2)	12.6 (6.4)[¶]	0.33	11.0 (6.0)	−0.01	13.7 (6.6)[¶]	0.5
	WASO [min]	17.3 (11.2)	15.1 (13.1)	−0.18	17.5 (17.8)	−0.05	19.3 (17.3)	0.13
	SE [%]	94.0 (2.6)	94.1 (2.9)	0.01	93.9 (4.3)	0.06	93.0 (4.0)	−0.31
	NW [num]	17.6 (4.7)	9.8 (3.9)[#]	−1.8	8.0 (4.7)[#]	−2.12	12.8 (5.5)[¶]	−0.93
Five past, zero future (real time)	TST [min]	434.4 (18.2)	436.9 (16.4)	0.15	437.3 (21.7)	0.2	433.1 (19.3)	−0.07
	SOL [min]	10.5 (6.2)	13.0 (6.0)[¶]	0.41	12.4 (6.4)	0.21	14.5 (6.5)[¶]	0.63
	WASO [min]	17.3 (11.2)	15.8 (12.1)	−0.13	17.7 (17.0)	−0.04	19.2 (15.8)	0.14
	SE [%]	94.0 (2.6)	93.9 (2.6)	−0.07	93.6 (4.2)	−0.03	92.8 (3.8)	−0.37
	NW [num]	17.6 (4.7)	12.3 (4.9)[#]	−1.09	10.0 (5.3)[#]	−1.57	15.2 (5.7)	−0.46

Patients with sleep disorders[ǁ]

Model type	Variable	PSG (n = 28)	Nightstand radar (n = 28) Mean (SD)	Cohen's D			Actigraph (n = 28) Mean (SD)	Cohen's D
Four past, two future	TST [min]	386.7 (73.3)	386.5 (85.2)	0			372.0 (59.3)	−0.22
	SOL [min]	13.0 (14.9)	10.0 (12.4)	−0.22			14.2 (16.1)	0.07
	WASO [min]	61.3 (55.7)	64.9 (77.8)	0.05			74.0 (42.6)	0.26
	SE [%]	84.2 (12.2)	84.7 (16.1)	0.03			81.6 (9.5)	−0.24
	NW [num]	24.6 (13.2)	19.3 (14.0)	−0.39			21.6 (10.5)	−0.25
Five past, zero future (realtime)	TST [min]	386.7 (73.3)	387.2 (85.3)	0.01			374.4 (58.2)	−0.19
	SOL [min]	13.0 (14.9)	10.3 (12.2)	−0.2			15.4 (18.1)	0.14
	WASO [min]	61.3 (55.7)	61.9 (80.8)	0.01			71.2 (39.2)	0.2
	SE [%]	84.2 (12.2)	85.5 (16.2)	0.09			82.0 (9.0)	−0.2
	NW [num]	24.6 (13.2)	23.1 (15.5)	−0.1			23.9 (11.5)	−0.05

Sleep parameters extracted from manually scored hypnograms (PSG) and from the sleep/wake state sequences resulting from the application of two selected classification models to radar and actigraphy data. Sleep onset latency was calculated from a user-marker input. The significant difference found from paired-sample Student's t-tests of each model/sensor compared to their corresponding PSG recordings are indicated with superscripts, and the standardized effect sizes are reported by Cohen's D.

[*]n = 12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n = 24/22 for nightstand/ceiling), and a testing set for validation (n = 23/21 for nightstand/ceiling).

[†]The models are defined by the number of preceding (past) and succeeding (future) epochs used to score a present epoch.

[‡] SOL, Sleep Onset Latency; TST, Total Sleep Time; WASO, Wake After Sleep Onset; SE, Sleep Efficiency; NW, Number of awakenings.

[§]PSG, Polysomnography (sleep parameters scored manually, independent of models).

[ǁ]Ambulatory sleep disorder patients. n = 28, mean age ± SD: 46.25 ± 13.98 years, 19 male.

[¶]$p < 0.05$ Student's t-test, compared to PSG.

[#]$p < 0.001$.

very small. (3) Reliable estimates could be achieved for several standard sleep parameters like TST, SOL, SE%, NW, and WASO. (4) For healthy subjects recorded in a hospital ward environment, an agreement was generally excellent with a small estimated LA. Wider LA was observed for sleep-disorder patients using a nightstand radar in their home bedrooms; comparable with but somewhat larger than LA for actigraphy for the same population. In addition, both actigraphy and radar may tend to overestimate SOL and underestimate nightly awakenings.

For the young healthy sleepers in DS1, the epoch-by-epoch classification results of our models showed high accuracy and sensitivity for both radar sensor placements and the wrist actigraph data, as well as remarkably high specificities and Cohen's kappa values. A lower agreement was observed for the ambulatory sleep-disorder patients in DS2. However, these results are still well within the range of performance commonly reported by previous studies that evaluate actigraphy against PSG over various populations; sensitivities and accuracies for

actigraphy tend to lie in the 80%–90% range, with corresponding specificities rarely above 60%, often around 50%, and sometimes even below 30% [7, 10, 26, 29–31].

In general, the real-time models performed slightly worse than the non-real-time models. However, this performance loss is a small cost compared to the benefits of being able to provide an immediate estimate of sleep/wake state. For DS1, the nightstand radar achieved the best overall classification results (larger specificity and kappa values). However, since the differences between radar placements were small, a ceiling-embedded radar might still be preferable to a nightstand mount because it combines greater ease-of-use and flexibility with good performance.

Actigraphy was more specific than nightstand radar in the sleep-disorder patients of DS2. For the sleep parameters, the radars and actigraphy compared with PSG for DS2 showed an excellent similarity of TST, WASO, and SE%. Radars and actigraphy tended to overestimate SOL and to underestimate NW. Comparing the modalities with each other resulted in relative
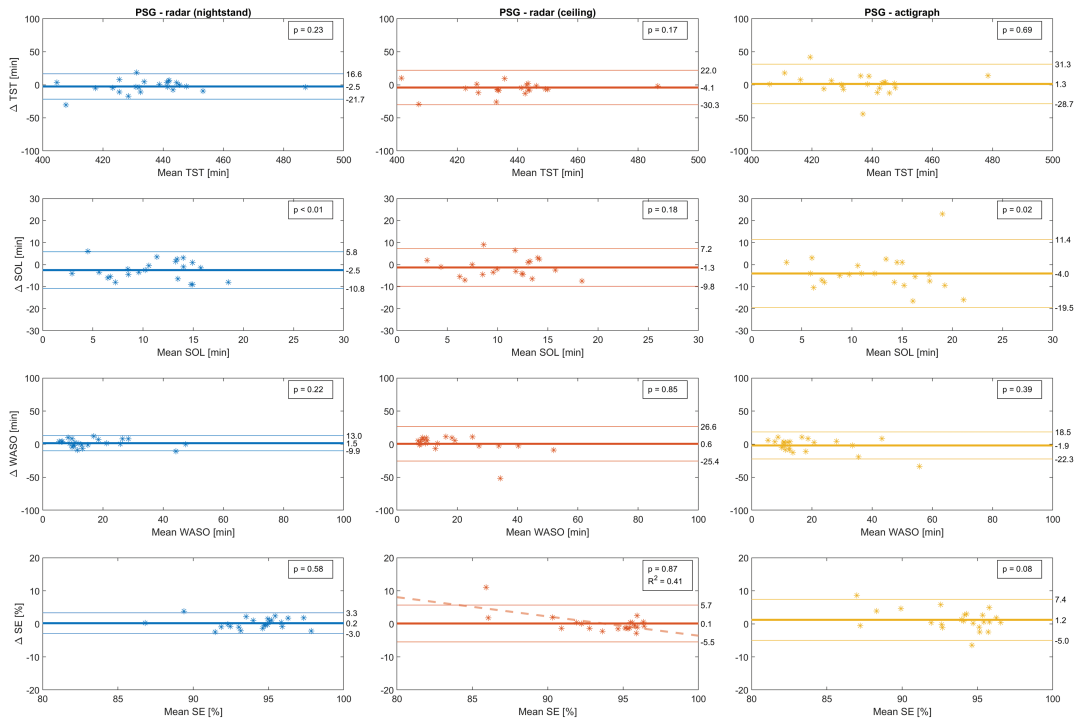
Figure 1. Bland–Altman plots of sleep variables of the test set of healthy volunteers, for the real-time models that score the present epoch by considering it and five past epochs. *p*-values from Student's *t*-test on the hypothesis of the bias being zero. A single significant trend ($p_{slope} < 0.01$) was detected and included with its corresponding *R* squared value. This trend is driven by a single outlying data point. $n = 12$, mean age ± SD: 23.0 ± 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development ($n = 24/22$ for nightstand/ceiling), and a testing set for validation ($n = 23/21$ for nightstand/ceiling). PSG, polysomnography; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset; SE, sleep efficiency.

biases on-par with or smaller than those seen in the literature comparing different actigraphy algorithms to each other [7, 31]. For DS2, no statistically significant biases were observed in the estimation of sleep parameters. However, the higher standard deviations and wider 95% confidence intervals, and 95% limits of agreement in the Bland–Altman plots indicate more uncertainty about the estimates for individual subjects recorded at home. The reasons for the large deviations in both directions for TST, WASO, and SE in few ambulatory home recordings should be investigated further, and some possibilities will be discussed in the following paragraphs.

The performances of the radars and the actigraph are compared directly in the forest plots shown in Figure 4. While non-inferiority margins were not prospectively defined for this work (a limitation which is discussed below), this figure shows that most estimated upper 95% confidence interval limits did not exceed hypothetical (but reasonable) non-inferiority margins like 20 minutes for TST and 5 minutes for SOL. Such an analysis would have resulted in conclusions of noninferiority for SOL in all three groups and for TST in healthy volunteers, while noninferiority in TST could not have been claimed for patients since the upper right margin of the confidence interval exceeds 20 minutes. In summary, the large majority of the estimated 95% confidence intervals for sleep and classification parameters showed no difference between radar and actigraphy. There

were four exceptions, all for patients with sleep disorders and all skewing in favour of the actigraph: TST, WASO (which is closely related to TST), Cohen's kappa, and specificity.

The discrepancies in results between our two datasets are not unexpected; it is notably more difficult to achieve high classification results and good sleep parameter estimates over a heterogenous set of sleep-disordered patients than over a homogenous set of healthy young volunteers [30]. Subjects who lie quietly but awake for long periods of time pose a challenge to movement-based classifiers, as do subjects with exaggerated movement during sleep. The radar measures movement from the whole body and includes estimated respiration frequency in its classification models. Consequently, unless one can identify and adjust for sleep-disorder specific movements, it is reasonable to expect that conditions like obstructive sleep apnoea (OSA) and restless legs syndrome will cause epoch misclassifications with the current model. It is possible that closer inspection of the probability estimate curves, i.e. the pre-classification model outputs, could be helpful. Whereas a simple $p = 0.5$ classification threshold (with Cole–Kripke rescoring) was sufficient for a population of healthy normal sleepers, a more nuanced approach could be preferable for groups where the level of certainty is lower. Implementing the option of manually correcting classification decisions is another avenue that could be explored, as is the use of adaptive classification thresholds.
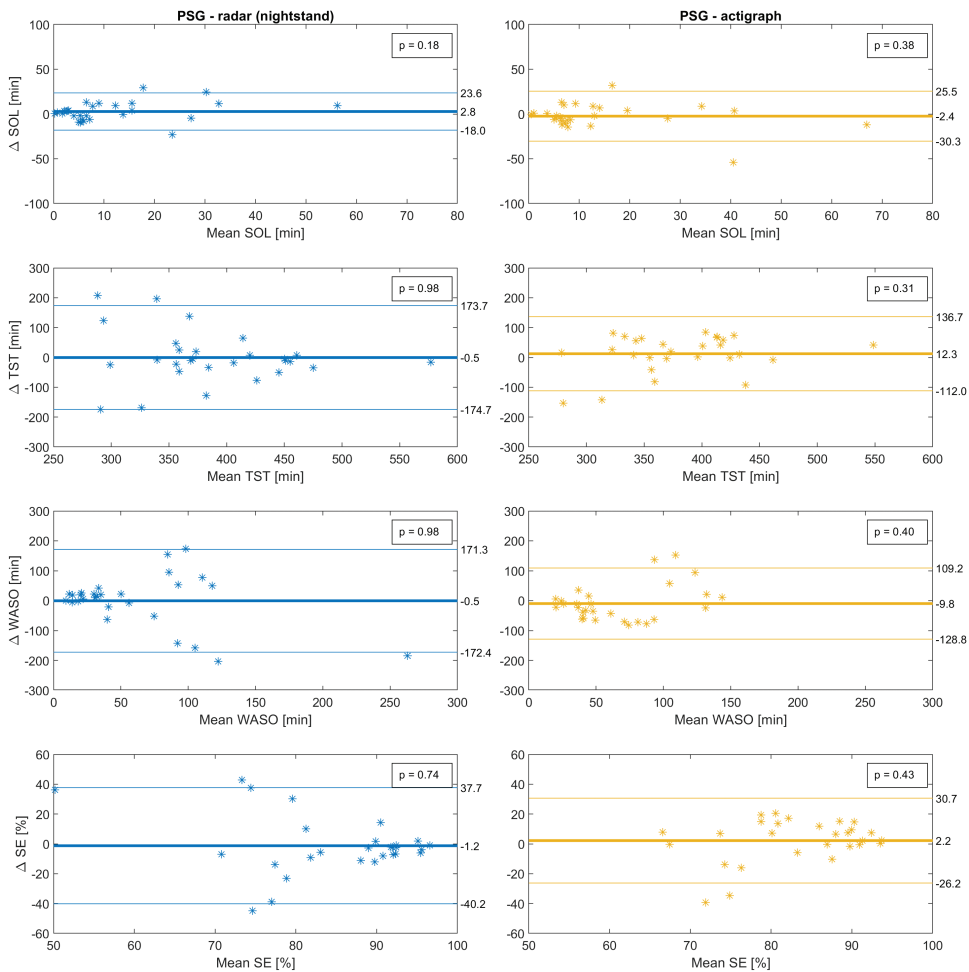
Figure 2. Bland–Altman plots of sleep variables of patients with sleep disorders, for the real-time models that score the present epoch by considering it and five past epochs. p-values from Student's t-test on the hypothesis of the bias being zero. n = 28, mean age ± SD: 46.25 ± 13.98 years, 19 male. PSG, polysomnography; SOL, sleep onset latency; TST, total sleep time; WASO, Wake After Sleep Onset; SE, sleep efficiency.

A few notes should also be considered when comparing the two datasets considered in this study. Not only is DS1 composed of data from the same demographic as the data on which the classification models were trained; the environment for the training set and the test set was also controlled and identical. These factors are both highly beneficial to classification. In contrast, DS2 was recorded by participants sleeping in their own homes, and they were responsible for mounting the radar sensor on their own. Environmental factors could not be controlled beyond trusting that the participants abided by the requirement of sleeping alone in their beds, and although the digital signal processing done by the radar attempts to compensate for distance to the target, it is possible that inconsistencies in the sensor placement for this dataset had a detrimental influence on the classification performance. This might to some degree explain why the performance difference between the radar and the actigraph was larger for DS2 than for DS1. Further work would

be necessary to validate the degree of environmental control needed for optimal results from the radar classifier.

A notable limitation to the present study is that the validation does not include an in-hospital psychiatric population. Sleep in these populations is disrupted and does not necessarily reflect either of the datasets studied in this work. Proceeding with a study of how our method performs in this setting is a natural next step. Furthermore, the datasets examined in the present study do not lend themselves to a thorough examination of possible dependencies of results on factors like age and sex. Future sample selection should be designed following the current recommended guidelines on the development and validation of sleep devices [32].

There are also some limitations to our statistical analyses. Bland–Altman plots were used to illustrate the differences between modalities. However, these have not taken into account that there are uncertainties in the limits of agreement that are

Table 5. Sleep variables from real-time models, compared to PSG*

Healthy volunteers test set[†]

Variable [units][‡]	Overall bias (95% CI) [P-value]	95% Limits of Agreement
TST [min]		
Radar nightstand	−2.5 (−6.7, 1.7) [0.23]	[−21.7, 16.6]
Radar ceiling	−4.1 (−10.2, 2.0) [0.17]	[−30.3, 22.0]
Actigraphy	1.3 (−5.3, 7.9) [0.69]	[−28.7, 31.3]
SOL [min]		
Radar nightstand	−2.5 (−4.3, −0.7) [0.01]	[−10.8, 5.8]
Radar ceiling	−1.3 (−3.3, 0.7) [0.18]	[−9.8, 7.2]
Actigraphy	−4.0 (−7.5, −0.6) [0.02]	[−19.5, 11.4]
WASO [min]		
Radar nightstand	1.5 (−1.0, 4.1) [0.22]	[−9.9, 13.0]
Radar ceiling	0.6 (−5.5, 6.6) [0.85]	[−25.4, 26.6]
Actigraphy	−1.9 (−6.4, 2.6) [0.39]	[−22.3, 18.5]
SE [%]		
Radar nightstand	0.2 (−0.5, 0.9) [0.58]	[−3.0, 3.3]
Radar ceiling	0.1 (−1.2, 1.4) [0.87]	[−5.5, 5.7]
Actigraphy	1.2 (−0.2, 2.6) [0.08]	[−5.0, 7.4]
NW [num]		
Radar nightstand	5.3 (3.2, 7.3) [<0.001]	[−3.9, 14.5]
Radar ceiling	7.9 (5.7, 10.1) [<0.001]	[−1.4, 17.2]
Actigraph	2.4 (−0.3, 5.1) [0.08]	[−9.9, 14.7]

Patients with sleep disorders[§]

Variable [units]	Overall bias (95% CI) [P-value]	95% Limits of Agreement
TST [min]		
Radar nightstand	−0.5 (−34.9, 34.0) [0.98]	[−174.7, 173.7]
Actigraphy	12.3 (−12.3, 36.9) [0.31]	[−112.0, 136.7]
SOL [min]		
Radar nightstand	2.8 (−1.3, 6.9) [0.18]	[−18.0, 23.6]
Actigraphy	−2.4 (−7.9, 3.1) [0.38]	[−30.3, 25.5]
WASO [min]		
Radar nightstand	−0.5 (−34.5, 33.5) [0.98]	[−172.4, 171.3]
Actigraphy	−9.8 (−33.4, 13.7) [0.40]	[−128.8, 109.2]
SE [%]		
Radar nightstand	−1.2 (−9.0, 6.5) [0.74]	[−40.2, 37.7]
Actigraphy	2.2 (−3.4, 7.8) [0.43]	[−26.3, 30.7]
NW [num]		
Radar nightstand	1.5 (−5.2, 8.2) [0.65]	[−32.2, 35.2]
Actigraphy	0.6 (−4.8, 6.1) [0.81]	[−26.9, 28.1]

Positive values of bias indicate underestimation compared to PSG (p-values from Student's t-test for the null-hypothesis that the bias is zero). Limits of Agreement given as [bias −1.96*SD, bias +1.96*SD]; SD, standard deviation for paired differences.
*PSG, polysomnography.
[†]*n* = 12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (*n* = 24/22 for nightstand/ceiling), and a testing set for validation (*n* = 23/21 for nightstand/ceiling).
[‡]TST, Total Sleep Time; SOL, Sleep Onset Latency; WASO, Wake After Sleep Onset; SE, Sleep Efficiency; NW, Number of awakenings
[§]Ambulatory sleep disorder patients. *n* = 28, mean age ± SD: 46.25 ± 13.98 years, 19 male.

difficult to estimate on datasets without more samples [28]. Furthermore, looking at Figures 2 and 3 we observe that there may also be some non-proportional relationships between difference and magnitude; the samples might tend to spread out as TST and %SE decreases, and as WASO increases. A logarithmic transformation prior to plotting could manage this dependency,

Table 6. Sleep parameters estimated with a real-time model over all nights in the test set of healthy volunteers*

| Sleep parameter[†] [units] | Mean (SD[‡]) | | |
	Actigraphy (n = 63)	Nightstand (n = 63)	Ceiling (n = 58)
TST [min]	431.8 (31.5)	434.2 (26.4)	432.1 (40.2)
SO after 23:00 h [min]	18.2 (14.9)	21.0 (18.7)	17.7 (17.8)
WASO [min]	37.5 (37.3)	20.6 (24.4)	27.0 (34.6)
SE [%]	88.7 (6.8)	91.3 (5.5)	90.7 (8.5)
NW [num]	20.5 (9.0)	14.4 (7.6)	11.9 (7.1)

Sleep parameters extracted from the sleep/wake state sequences resulting from the application of a real-time classification model that includes information from five past and zero future epochs, for all nights in the DS1-test (including the ones without consecutive PSG[§] recordings). Since no user marker was available for the nights without PSG, sleep onset latency was calculated from the set bedtime at 23:00 h.
n = 12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 11 nights of actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (*n* = 63/59 for nightstand/ceiling), and a testing set for validation (*n* = 63/58 for nightstand/ceiling).
[†]TST, Total Sleep Time; SOL, Sleep Onset Latency; WASO, Wake After Sleep Onset; SE, Sleep Efficiency; NW, Number of awakenings.
[‡]SD, standard deviation.
[§]PSG, Polysomnography.

however since this was not done in comparable studies, it has not been done here [7, 26, 28, 33]. Proportional bias was found only for 5 of 32 B-A plots, and this linearity may seem to be mostly driven by rather few outlying observations. Finally, DS1 lacked a user-marker for bedtime for the non-PSG nights. Calculating sleep onset latency from the set bedtime rather than from a user-marker is not ideal.

Another limitation was the lack of a pre-planned non-inferiority analysis comparing actigraphy and radar [34, 35]. However, there is surprisingly little information and, to our knowledge, no published consensus about the clinically relevant "minimal important difference" (MID) for actigraphic sleep parameters. Studies on actigraphic validity rarely report test–retest (e.g. between-day) differences necessary for the computation of the distribution-based "one standard error of measurement" MID proxy [36–38]. More work is needed to establish a consensus for non-inferiority margins and MIDs for actigraphic sleep parameters.

Furthermore, the data from the radar are rich, and the present work has only considered a small subset of what might be its full potential. In recent years, significant work has been done to develop processing techniques for IR-UWB radar data for non-invasive remote health monitoring [39]. The technology has been used to detect and analyse sleep-disordered breathing [40–42], sleep posture recognition [43], and sleep stage classification [44–47]. O'Hare et.al. [48] compares the sleep assessment performance of two radar-based devices and actigraphy to PSG in a group of twenty healthy subjects. Like us, they observed basically equivalent performance of radar devices and actigraphy for this group. Their reported epoch-by-epoch classification performance was lower than ours (overall accuracies of 85%–86%, Cohen's kappa values of 0.51–0.52), and they observed a statistically significant bias to overestimating sleep time and underestimating WASO and SOL which contrasts our observation of SOL being significantly overestimated. In Pallin et.al. [49], one of these radar devices was found to have similar accuracy to
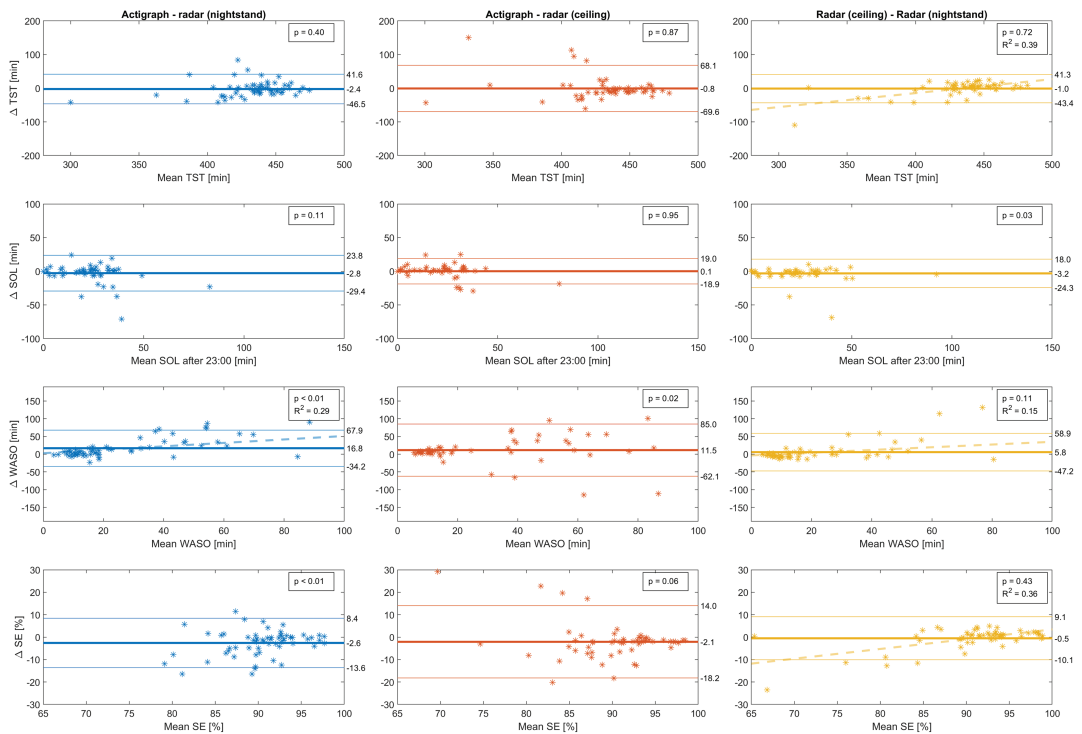
Figure 3. Bland–Altman plots comparing the actigraph and the two radar positions for the test set of healthy volunteers. *p*-values from Student's *t*-test on the hypothesis of the bias being zero. Regression lines with corresponding *R* squared values are included in those subplots wherein a significant (p_{slope} < 0.01) trend was detected. n = 12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 11 nights of actigraphy + two radars per participant. The participants were randomly assigned to a training set for model development (n = 63/59 for nightstand/ceiling), and a testing set for validation (n = 63/58 for nightstand/ceiling). SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset; SE, sleep efficiency.

wrist actigraphy for sleep/wake determination in subjects with OSA, with lower sensitivity (86% compared to 94%) and, notably, higher specificity (52% for the radar sensor compared to 34% for actigraphy) and superior estimation of TST, especially at higher apnoea–hypopnea indices. Zaffaroni et al. [50] and Crinion et al. [51] evaluate the same device specifically as a screening tool for OSA, concluding that it is useful particularly for confirming more severe cases. More work is needed to evaluate our radar sensor in the setting of sleep-disordered breathing, but the implication of this in terms of a psychiatric hospital setting is particularly interesting; it indicates a potential for screening of sleep-disordered breathing in a patient population where this traditionally has been challenging.

Interest in non-contact sleep detection and assessment has been growing in popularity over the past decade, and many other methods exist that can be used in ways comparable to the radar. Under-mattress or under-bed sensors have been shown capable of sleep/wake discrimination [52, 53], respiration rate detection [54], and detection of sleep-disordered breathing [55]. The choice of which sensor to use will depend on the circumstances: Nagatomo et al. [56] compares an under-mattress sensor for sleep measurement to PSG in eleven critically ill patients in an intensive care unit (ICU) (achieving agreement, sensitivity, and specificity of 68.4%, 90.1%, and 38.7% respectively). In such a hectic environment, activity in the room might

confound a remote-mounted radar sensor, so an under-mattress sensor might be more reliable. On the other hand, an under-mattress system will still require physical sensors and wiring close to the patient, and will only be able to provide measurements while the patient is in their bed. Thus, in a psychiatric hospital setting, a sensor mounted permanently in the ceiling might be preferable.

Infrared camera technology is another tool that has been investigated in this context [57]. The Oxehealth system uses infrared camera technology in combination with computer vision, signal processing, and AI techniques, and has been installed in a psychiatric ward; their work shows that "digitally assisted nursing observation" has potential in terms of improving patient and staff experience at night [58]. However, the use of camera monitoring in such a sensitive environment is potentially problematic. A ceiling-mounted radar sensor could be capable of providing much of the same functionality as a camera-based system, while preserving patient privacy to a greater degree, as it is much more difficult to directly identify individuals from such data.

In summary, radar technology could be used to obtain objective sleep and activity data from certain patient groups from whom it has previously been difficult or even impossible to attain on such a large scale. Since the radar can be embedded in a hospital ceiling or placed on a nightstand for home use,

Table 7. Sleep variables from real-time models, compared to each other

Healthy volunteers test set*

Variable [units][†]	Overall bias (95% CI) [P-value]	95% Limits of Agreement
TST [min]		
Actigraphy to radar nightstand	–2.4 (–8.1, 3.3) [0.40]	[–46.5, 41.6]
Actigraph to radar ceiling	–0.8 (–10.1, 8.6) [0.87]	[–69.6, 68.1]
Radars ceiling to nightstand	–1.0 (–6.8, 4.7) [0.72]	[–43.4, 41.3]
SOL [min]		
Actigraphy to radar nightstand	–2.8 (–6.2, 0.7) [0.11]	[–29.4, 23.8]
Actigraphy to radar ceiling	0.1 (–2.5, 2.7) [0.95]	[–18.9, 19.1]
Radars ceiling to nightstand	–3.2 (–6.0, –0.3) [0.03]	[–24.3, 18.0]
WASO [min]		
Actigraphy to radar nightstand	16.8 (10.2, 23.4) [<0.001]	[–34.2, 67.9]
Actigraphy to radar ceiling	11.5 (1.5, 21.4) [0.03]	[–62.1, 85.0]
Radars ceiling to nightstand	5.8 (–1.3, 13.0) [0.11]	[–47.2, 58.9]
SE [%]		
Actigraphy to radar nightstand	–2.6 (–4.0, –1.2) [<0.001]	[–13.6, 8.4]
Actigraphy to radar ceiling	–2.1 (–4.3, 0.1) [0.06]	[–18.2, 14.0]
Radars ceiling to nightstand	–0.5 (–1.8, 0.8) [0.43]	[–10.1, 9.1]
NW [num]		
Actigraphy to radar nightstand	6.1 (4.1, 8.1) [<0.001]	[–9.3, 21.5]
Actigraphy to radar ceiling	8.3 (6.3, 10.4) [<0.001]	[–6.7, 23.3]
Radars ceiling to nightstand	–2.8 (–4.5, –1.1) [0.01]	[–15.2, 9.6]

Sleep parameters extracted from the sleep/wake state sequences resulting from the application of a real-time classification model that includes information from five past and zero future epochs, for all nights in the DS1-test (including the ones without consecutive PSG[‡] recordings). Since no user marker was available for the nights without PSG, sleep onset latency was calculated from the set bedtime at 23:00 h.

Positive values of bias indicate underestimation by the second sensor relative to the first (p-values from Student's t-test for the null-hypothesis that the bias is zero). Limits of Agreement given as [bias –1.96*SD, bias +1.96*SD]; SD, standard deviation for paired differences.

*n = 12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 11 nights of actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n = 63/59 for nightstand/ceiling), and a testing set for validation (n = 63/58 for nightstand/ceiling).

[†]SOL, Sleep Onset Latency; TST, Total Sleep Time; WASO, Wake After Sleep Onset; SE, Sleep Efficiency; NW, Number of awakenings.

[‡]PSG, polysomnography.

long-term and entirely unobtrusive assessment of sleep in a very wide range of clinical cohorts are possible [59]. In a psychiatric hospital setting, a tool like this could be used to provide night staff with information about the patients' current sleep/wake state, e.g. as displayed on monitors in the staff rooms or sent to hand-held devices, which in turn could improve patient safety at night, reduce the number of nocturnal awakenings due to disturbance from night staff, and lead to more efficient use of limited staff resources. Actigraphy-generated raster plots of the type shown in Figure 5 are known to be very useful in visually depicting changing periodicities associated with circadian dysrhythmia, and they can provide patients with easy to understand graphical displays that may help them to understand diagnostic decisions and evaluate their own response to treatment [60]. Radar-generated plots of this type have the potential to provide many of the same benefits without any potential inconvenience to the patient.

Conclusions

Our results indicate that our movement-based models can be used with radar data to achieve sleep/wake classification results on par with those normally seen for actigraphy. Our data also supports that the sensor can be used in real-time. Estimates of sleep parameters show little to no bias, although the wide limits of the agreement particularly for the clinical population warn that these still should be interpreted with care. Although our method should be studied further and improved, both to account for various clinical sleep-disorder patient groups and for the heterogeneity in environmental factors of ambulatory home studies, our results show that a non-contact real-time sleep and activity sensor is a real possibility. Accordingly, we believe that a radar-based contact-free sensor has great potential as a supplementary tool in psychiatric ward monitoring and other settings where contact-free sleep monitoring would be advantageous. Such a solution has the potential to change clinical practice in selected fields, improving decision making and care in hospital and home settings, as well as providing a promising new tool for researchers.

Author Contributions

HSAH and DV had the main responsibility for data collection for DS1. HSAH had the main responsibility for data collection of DS2, with generous help from the staff at the Department of Neurology and Clinical Neurophysiology, St. Olavs Hospital. HSAH did data management and analysis and had overall responsibility for drafting the paper. ME performed all polysomnographic scorings. TS has provided important guidance for the analysis. HK and KL have been the main drivers in panning the new hospital and establishing related projects, including this one. All authors have revised the article and accepted the final version.
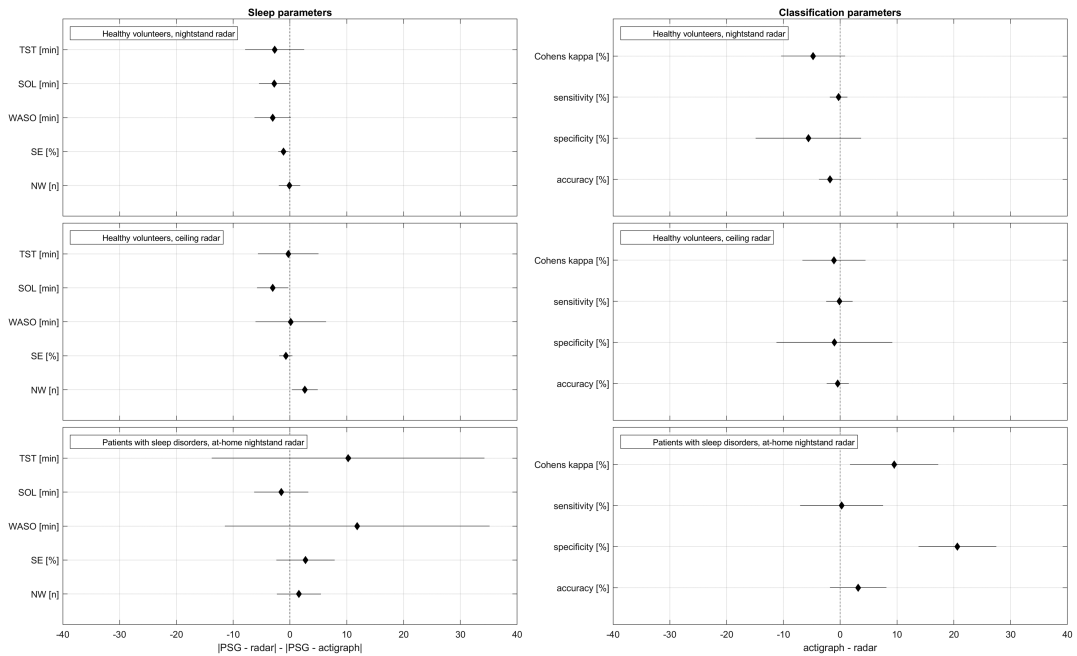
Figure 4. Forest plots comparing the performance of the radar to that of actigraphy, using the real-time models that score the present epoch by considering it and five past epochs. Means and 95% confidence intervals of the differences. Skewing to the left of zero favours the radar. Skewing to the right favours the actigraph. From this figure, it can be inferred that hypothetical non-inferiority margins equal to 20 min for TST and 5 min for SOL would have resulted in conclusions of noninferiority for SOL in all three groups and for TST in healthy volunteers, while noninferiority could not have been claimed for patients since the upper right margin for the CI exceeds 20 min. PSG, polysomnography; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset; SE, sleep efficiency; NW, number of awakenings.
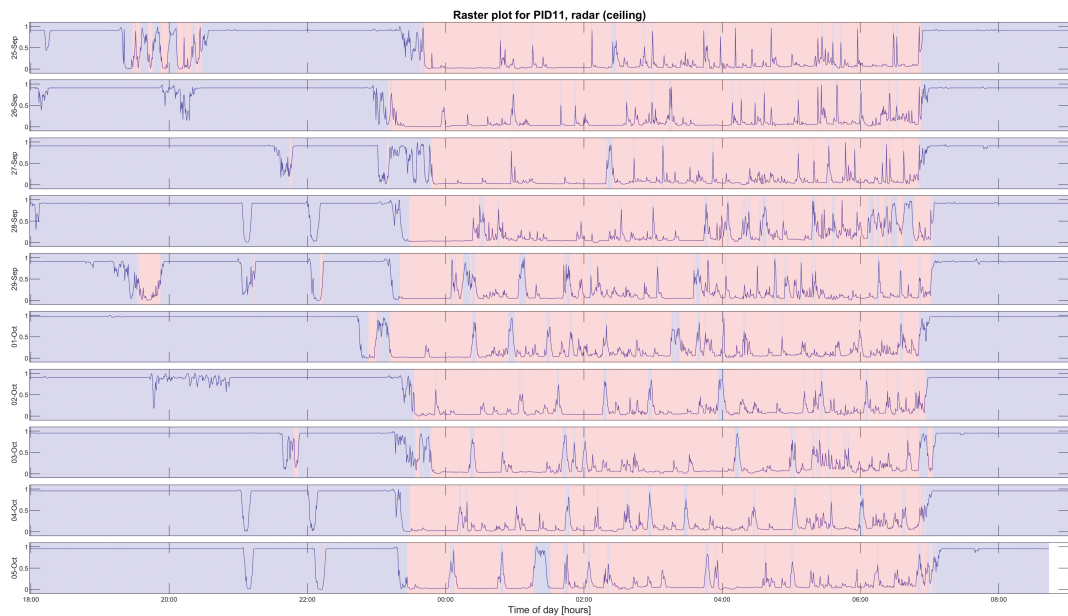
Figure 5. Temporal raster plot generated by the ceiling-mounted radar sensor for a single healthy volunteer. The probability estimate output from the past-five-future-zero regression model is plotted as a line. Each epoch is classified as sleep/wake by comparing the probability estimate to a threshold of $p = 0.5$ and then re-scoring with the Cole–Kripke rules, and the resulting classification is represented by blue (wake) and red (sleep) sections in the raster plot. For reference, this participant underwent PSG on September 28th and 29th, and on October 4th and 5th, achieving a mean ($\pm$SD) epoch-by-epoch classification accuracy against PSG sleep/wake of 93.3 ($\pm$0.46) percent with corresponding Cohen's kappa values of 0.85 ($\pm$0.005) over those days.

Funding

Disclosure Statements

Data Availability

Some de-identified data that underlie the results reported in this article may be made available to researchers from accredited research institutions. Data from clinical patients cannot be shared for ethical reasons. Access to data will be limited to investigators who provide a methodologically sound proposal and will be limited to a specified time period (commencing about 3 months after publication of this Article and ending after 5 years). To ensure compliance with the General Data Protection Regulation, data processing must be covered by the European Commission's standard contractual clauses for the transfer of personal data, which must be signed by the data requesters. Proposals and requests for data access should be directed to the corresponding author.

Conflict of interest statement: As a candidate in the Industrial PhD Scheme the corresponding author (HSAH) is considered an employee both of Novelda AS and the Norwegian University of Science and Technology (NTNU), and receives salary compensation from Novelda AS. None of the other authors have any conflicts of interest to declare.

References

1. Rechtschaffen A, *et al*. *A Manual of Standardized Terminology Techniques and Scoring System for Sleep Stages of Human Subjects*. Los Angeles, CA: US National Institute of Neurological Diseases and Blindness, Neurological Information Network; 1968.

2. Kushida CA, *et al*. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep.* 2005;**28**(4):499–521.

3. Jakobsen G, *et al*. Sleep quality in hospitalized patients with advanced cancer: an observational study using self-reports of sleep and actigraphy. *Support Care Cancer.* 2020;**28**(4):2015–2023.

4. Webster JB, *et al*. An activity-based sleep monitor system for ambulatory use. *Sleep.* 1982;**5**(4):389–399.

5. Kripke DF, *et al*. Wrist actigraphic measures of sleep and rhythms. *Electroencephalogr Clin Neurophysiol.* 1978;**44**(5):674–676.

6. Mullaney DJ, *et al*. Wrist-actigraphic estimation of sleep time. *Sleep.* 1980;**3**(1):83–92.

7. Haghayegh S, *et al*. Performance comparison of different interpretative algorithms utilized to derive sleep parameters from wrist actigraphy data. *Chronobiol Int.* 2019;**36**(12):1752–1760.

8. Someren EJW. Actigraphic monitoring of sleep and circadian rhythms. In: Montagna P, Chokroverty S, eds. *Handbook of Clinical Neurology*. Vol. **98**. Amsterdam, The Netherlands: Elsevier B.V.; 2011: 55–63.

9. Morgenthaler TI, *et al*.; Standards of Practice Committee of the American Academy of Sleep Medicine. Practice parameters for the treatment of narcolepsy and other hypersomnias of central origin. *Sleep.* 2007;**30**(12):1705–1711.

10. Kripke DF, *et al*. Wrist actigraphic scoring for sleep laboratory patients: algorithm development. *J Sleep Res.* 2010;**19**(4):612–619.

11. Sadeh A. The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev.* 2011;**15**(4):259–267.

12. Freeman D, *et al*. Sleep disturbance and psychiatric disorders. *Lancet Psychiatry.* 2020;**7**(7):628–637.

13. Veale D. Against the stream: intermittent nurse observations of in-patients at night serve no purpose and cause sleep deprivation. *BJPsych Bull.* 2019;**43**(4):174–176.

14. Wisland DT, *et al*. Remote monitoring of vital signs using a CMOS UWB radar transceiver. In: 2016 14th IEEE International New Circuits and Systems Conference (NEWCAS); 26–29 June, 2016; Vancouver, BC, Canada.

15. Lee Y, *et al*. A novel non-contact heart rate monitor using impulse-radio ultra-wideband (IR-UWB) radar technology. *Sci Rep.* 2018;**8**:13053.

16. Stone WC. Nist construction automation program report no. 3: Electromagnetic signal attenuation in construction materials. NIST Interagency/Internal Report (NISTIR). Gaithersburg, Maryland: National Institute of Standards and Technology; 1997.

17. Parsa-Parsi RW. The revised declaration of Geneva: a modern-day physician's pledge. *JAMA* 2017;**318**(20):1971–1972.

18. Klem GH, *et al*. The ten-twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr Clin Neurophysiol Suppl.* 1999;**52**:3–6.

19. National-Instruments. What is I/Q Data? http://www.ni.com/tutorial/4805/en/. Accessed May 9, 2020.

20. Novelda AS. X4M200 Datasheet, Rev. E - Preliminary. 18 Oct 2018; www.xethru.com.

21. Vethe D, *et al*. The evening light environment in hospitals can be designed to produce less disruptive effects on the circadian system and improve sleep. *Sleep.* 2021; **44**(3). doi:10.1093/sleep/zsaa194

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

23. Cole RJ, *et al*. Automatic sleep/wake identification from wrist activity. *Sleep.* 1992;**15**(5):461–469.

24. Jean-Louis G, *et al*. Sleep estimation from wrist movement quantified by different actigraphic modalities. *J Neurosci Methods.* 2001;**105**(2):185–191.

25. Ranganathan P, *et al*. Common pitfalls in statistical analysis: measures of agreement. *Perspect Clin Res.* 2017;**8**(4):187–191.

26. de Souza L, *et al*. Further validation of actigraphy for sleep studies. *Sleep.* 2002;**26**(1):81–85.

27. Fritz CO, *et al*. Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol Gen.* 2014;**141**(1):2–18.

28. Bland JM, *et al*. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;**8**(2):135–160.

29. Van De Water AT, *et al*. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography – a systematic review. *J Sleep Res*. 2011;**20**(1 Pt 2):183–200.

30. Kushida CA, *et al*. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*. 2001;**2**(5):389–396.

31. Paquet J, *et al*. Wake detection capacity of actigraphy during sleep. *Sleep*. 2007;**30**(10):1362–1369.

32. Depner CM, *et al*. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep*. 2020;**43**(2). doi: 10.1093/sleep/zsz254

33. Scott J, *et al*. Can consumer grade activity devices replace research grade actiwatches in youth mental health settings? *Sleep Biol Rhythms*. 2019;**17**:223–232.

34. Piaggio G, *et al*. Reporting of noninferiority and equivalence randomized trials - extension of the CONSORT 2010 Statement. *JAMA*. 2012;**308**(24):2594–2604.

35. Althunian TA, *et al*. Defining the noninferiority margin and analysing noninferiority: an overview. *Br J Clin Pharmacol*. 2017;**83**(8):1636–1642.

36. Täger T, *et al*. Minimal important difference for 6-minute walk test distances among patients with chronic heart failure. *Int J Cardiol*. 2014;**176**(1):94–98.

37. Wyrwich KW, *et al*. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care*. 1999;**37**(5):469–478.

38. Wells G, *et al*. Minimal clinically important differences: review of methods. *J Rheumatol*. 2001;**28**(2):406–412.

39. Khan F, *et al*. An overview of signal processing techniques for remote health monitoring using impulse radio UWB transceiver. *Sensors*. 2020;**20**:2479. doi:10.3390/s20092479

40. Kang S, *et al*. Non-contact diagnosis of obstructive sleep apnea using impulse-radio ultra-wideband radar. *Sci Rep*. 2020;**10**:5261.

41. Zhou Y, *et al*. Validation of novel automatic ultra-wideband radar for sleep apnea detection. *J Thorac Dis*. 2020;**12**(4):1286–1295.

42. Tran VP, *et al*. Doppler radar-based non-contact health monitoring for obstructive sleep apnea diagnosis: a comprehensive review. *Big Data Cogn Comput*. 2019;**3**(1):3. doi:10.3390/bdcc3010003

43. Piriyajitakonkij M, *et al*. SleepPoseNet: multi-view learning for sleep postural transition recognition using UWB. *IEEE J Biomed Health Inform*. 2020. doi:10.1109/JBHI.2020.3025900

44. Kagawa M, *et al*. Sleep stage classification by non-contact vital signs indices using Doppler radar sensors. *Annu Int Conf IEEE Eng Med Biol Soc*. 2016;**2016**:4913–4916.

45. Zhao M, *et al*. Learning sleep stages from radio signals: a conditional adversarial architecture. In: Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia. PMLR 70:4100–4109, 2017.

46. Toften S, *et al*. Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®). *Sleep Med*. 2020;**75**:54–61.

47. Lauteslager T, *et al*. Performance evaluation of the circadia contactless breathing monitor and sleep analysis algorithm for sleep stage classification. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020;**2020**:5150–5153.

48. O'Hare E, *et al*. A comparison of radio-frequency biomotion sensors and actigraphy versus polysomnography for the assessment of sleep in normal subjects. *Sleep Breath*. 2015;**19**(1):91–98.

49. Pallin M, *et al*. Comparison of a novel non-contact biomotion sensor with wrist actigraphy in estimating sleep quality in patients with obstructive sleep apnoea. *J Sleep Res*. 2014;**23**(4):475–484.

50. Zaffaroni A, *et al*. Assessment of sleep-disordered breathing using a non-contact bio-motion sensor. *J Sleep Res*. 2013;**22**(2):231–236.

51. Crinion SJ, *et al*. Ambulatory detection of sleep apnea using a non-contact biomotion sensor. *J Sleep Res*. 2020;**29**(1):e12889.

52. Uchida S, *et al*. Sleep evaluation by a newly developed PVDF sensor non-contact sheet: a comparison with standard polysomnography and wrist actigraphy. *Sleep Biol. Rhythms*. 2011;**9**:178–187.

53. Kogure T, *et al*. Automatic sleep/wake scoring from body motion in bed: validation of a newly developed sensor placed under a mattress. *J Physiol Anthropol*. 2011;**30**(3):103–109.

54. Isono S, *et al*. Contact-free unconstraint respiratory measurements with load cells under the bed in awake healthy volunteers: breath-by-breath comparison with pneumotachography. *J Appl Physiol (1985)*. 2019;**126**(5):1432–1441.

55. Beattie ZT, *et al*. Accurate scoring of the apnea-hypopnea index using a simple non-contact breathing sensor. *J Sleep Res*. 2013;**22**(3):356–362.

56. Nagatomo K, *et al*. Validity of an under-mattress sensor for objective sleep measurement in critically ill patients: a prospective observational study. *J Intensive Care*. 2020;**8**:16.

57. Gastel MV, *et al*. Camera-based vital signs monitoring during sleep – a proof of concept study. *IEEE J Biomed Health Inform*. 2020. doi:10.1109/JBHI.2020.3045859

58. Barrera A, *et al*. Introducing artificial intelligence in acute psychiatric inpatient care: qualitative study of its use to conduct nursing observations. *Evid Based Ment Health*. 2020;**23**(1):34–38.

59. Scott J, *et al*. A pragmatic effectiveness randomized controlled trial of the duration of psychiatric hospitalization in a trans-diagnostic sample of patients with acute mental illness admitted to a ward with either blue-depleted evening lighting or normal lighting conditions. *Trials*. 2019;**20**(1):472.

60. Smith MT, *et al*. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. *J Clin Sleep Med*. 2018;**14**(7):1209–1230.

SUPPLEMENTARY INFORMATION PAPER III

Distinguishing sleep from wake with a radar sensor

Supplementary materials

Hanne Siri Amdahl Heglum[1,2]

Håvard Kallestad[4, 5]

Daniel Vethe[4, 5]

Knut Langsrud[4, 5]

Trond Sand[1, 3]

Morten Engstrøm[1,3]

[1] Department of Neuromedicine and Movement Science, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Postboks 8905, 7491 Trondheim, Norway

[2] Novelda AS, Strandveien 43, 7067 Trondheim, Norway

[3] Department of Neurology and Clinical Neurophysiology, St.Olavs University Hospital, Olav Kyrres gate 17, 7030 Trondheim

[4] Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway

[5] Division of Mental Health Care, St.Olavs University Hospital, Trondheim, Norway

Corresponding author:

Hanne Siri Amdahl Heglum

Department of Neuromedicine and Movement Science, NTNU

Edvard Griegs gate 8, Trondheim

N-7491 Trondheim

Norway

Phone: +4792852149

Email: hanne.s.a.heglum@ntnu.no

Impulse radio ultra-wideband (IR-UWB) radar

An impulse radar generates pulses of radio frequency (RF) signals and measures their reflections from objects in the environment. The term 'ultra-wideband' is used to refer to the 3.1-10.6 GHz band of radio frequency waves. The large bandwidth of UWB enables high resolution in spite of very low power, and the high frequencies easily penetrate soft materials such as clothes and beddings while being reflected by more solid objects such as walls and human bodies[1]. With pulse-Doppler signal processing both distance to and velocity of targets can be measured, from the time-of-flight and the Doppler shift of the received signals respectively, and the high resolution means that very fine motions can be detected. The result of this is that the movement of a target within range can be detected without any sensors placed on the target itself.

XeThru X4M200

The XeThru model X4M200 used in this work is a system on chip (SoC) CMOS IR-UWB radar with built-in TX and RX antennas, developed by Novelda AS. All radar data for this work was stored in 'raw' baseband IQ form to enable different or improved digital signal processing (DSP) at a later date. In this work we have employed the pulse-Doppler DSP provided by the manufacturer, specifically their Respiration_2 profile. This profile has a detection zone of 0.40-5.00m, and a respiration detector with a range of detectable respiration-per-minute (RPM) of 8-30RPM[2].

Static objects in the environment are eliminated by computing a noise map of static reflectors and subtracting these during further processing. The DSP then subjects the data to two parallel sliding-window short-time Fourier transforms; one 'short', and one 'long'. The long window integrates 20 seconds of data. Stable periodic motions that are repeated several times within this time window become amplified – even small and relatively slow oscillations like respiration can be seen clearly in the frequency spectrum. This is then given to a respiration detection algorithm designed to recognize the characteristically symmetrical peaks formed by these movements. Non-repeated fast movements (like a single motion of the arm of a person who is otherwise still) are dampened by the long window, as they get 'averaged out'. To see these properly, the short window integrates 6 seconds of data and is able to quickly react to fast movements[3]. An index of overall slow and fast body movement is obtained by summarizing the energy of all observed movements over a range covering the whole person (or the closest target, if more than one non-static reflector is observed within the detection zone). Finally, since reflections from further away will appear weaker than reflections from closer targets, the fast and slow movement outputs are normalized by their distance from the radar. The final result is three data types output at a rate of 1Hz: normalized fast movement, normalized slow movement, and RPM.

The XeThru model X4M200 is designed to meet UWB RF specifications of ETSI (Europe), FCC (USA) and ISED (Canada).

Time alignment

As all three recording types had been running on separate clocks, data had to be time aligned before further analysis could take place. For nights with radar and actigraphy, the radar was aligned to the actigraph. For nights with radar, actigraphy, and PSG, the radar and the actigraph were both aligned to the leg EMG sensors of the PSG. An automated matlab procedure using the xcorr function was used to calculate the time shift of maximum correlation between sensors, and data was shifted by this amount. Every alignment was visually inspected - in most cases the automated procedure identified the visually obvious best result, but in some cases manual adjustment was required due to sensor noise in one of the two PLM-sensors, or due to the procedure locking on the wrong local maximum of the correlation sequence.

The resultant average temporal shifts are indicated in Table S1. An example of movement data before and after alignment, along with the cross-correlation sequences, can be found in supplementary Figure S1.

Model development

General formulation

The sleep/wake classification models used in this work were heavily inspired by actigraphy. There are many different actigraphy devices available, using different interpretation schemes, but their basic functionality can usually be described as follows: accelerometer sensor data is recorded, then processed and rate reduced in some way (summation, integration, counting of zero crossings, or some other method) over an epoch of often configurable length (15, 30, or 60 seconds are common). After this, the schemes share the same basic logic: a score of sleep or wake is assigned to an epoch by calculating a weighted sum of activity scores over some time window surrounding that epoch, and then comparing this sum to some threshold. Algorithms with both specific fixed weights[4,5] and general parametric weights[6] have been published, where the latter in effect are generalizations of the former.

These models can be summarized as follows:

Definition 1. Consider a window of size $[P, F]$ surrounding the current time t. Let $a[t]$ be a vector of the epoch activity counts within the window $a[t] = a[a_{t-P}, a_{t-P+1}, \ldots a_t, \ldots, a_{t+F-1}, a_{t+F}]$. Let φ be

a (constant) vector of parameters[1], and S some threshold value. Let 0 represent sleep and 1 represent wake. A general actigraphy classification function for the epoch at time t can then be written as

$$y[t] = \begin{cases} 0, if \ \varphi a[t] = L[t] < S \\ 1, \quad otherwise \end{cases}$$

In the case of the radar data, we expand this general formulation order to include the three data types (slow movement, fast movement, and RPM). A similar expansion could be used i.e. to include data from multiple simultaneous sensors, and/or other features of the data such as measures of variability.

Generate a feature vector $a[t], b[t], c[t], \ldots$ for each property or data type to be included over a time horizon, which does not necessarily have to be of equal length for each feature. Each feature vector has a corresponding vector of parameters $\alpha, \beta, \varphi, \ldots$, and we expand our model accordingly, including a constant k as an intercept term:

$$L[t] = k + \alpha^T a[t] + \beta^T b[t] + \varphi^T c[t] + \cdots$$

Definition 2. Let $x[t] = [1, a[t]^T, b[t]^T, c[t]^T, \ldots]^T$ be a vector of all the feature vectors we want to include in our model, adopting the convention of letting the first term $x_0 = 1$ to accommodate the intercept. Let the corresponding parameter vector be $\theta = [k, \alpha^T, \beta^T, \varphi^T, \ldots]^T$.

The sum $L[t]$ can then be written in a condensed form as $L[t] = \theta^T x[t]$.

Parameter estimation
In the previous literature on actigraphy the decision about the epoch at time t was made based on how a linear combination $L[t]$ of predictors (features) and parameters θ compared to some threshold S. However, this method does not conserve any information about how close the sum $L[t]$ is to the threshold, nor does it contain any intrinsically obvious interpretation of the meaning of this threshold.

We suggest to let the sum $L[t]$ represent some version of the probability $p[t]$ of the epoch at time t being scored as wake (the converse $(1 - p[t])$ represents the probability of the epoch being scored as sleep). Specifically, let us assume that $L[t]$ represents the log-odds of wake:

$$logit(p[t]) = log\left(\frac{p[t]}{1 - p[t]}\right) = \theta^T x[t]$$

The probability $p[t]$ is then expressed by the inverse $logit^{-1}(L[t])$ of the log-odds,

$$p[t] = \frac{1}{1 + e^{\theta^T x[t]}}$$

---

[1] The overall scale factor used by Webster and Cole is simply included into this parameter vector.

This function is known as the logistic function, or the sigmoid function. Among many other useful properties, it is the most common hypothesis function used in logistic regression.

The problem of parameter estimation for the models now becomes rather easy. In MATLAB, one can use the fitglm or glmfit functions to solve the problem for a given data set in a single line of code. When the model has been fitted and applied to data, the output will be a continuous value in the range $[0,1]$ that can be interpreted as a probability $p[t]$. To make class predictions, we employed a simple decision rule of a threshold of 0.5 for each epoch (i.e. choose the class with highest probability for that epoch) in the present study.

Cole-Kripke rescoring rules

These heuristic rescoring rules act as a nonlinear time series filter on the sleep/wake state sequence, converting initially scored sleep epochs to wake if they occur after or within a specified interval of a given number of awake epochs[7]. First introduced by Webster et al. in 1982[4], they were concretized and used by Cole et al. ten years later[6]. These rules are:

    a.   after at least 4 minutes scored as wake, the next 1 minute scored as sleep is rescored wake

    b.   after at least 10 minutes scored as wake, the next 3 minutes scored as sleep are rescored wake

    c.   after at least 15 minutes scored as wake, the next 4 minutes scored as sleep are rescored wake

    d.   6 minutes or less scored as sleep surrounded by at least 10 minutes of wake (before and after) are rescored wake

    e.   10 minutes or less scored as sleep surrounded by at least 20 minutes of wake (before and after) are rescored wake.

For our real-time models, rules d) and e) obviously had to be excluded.

Additional data preparation

The slow movement and fast movement data types from the radar tend to have values approximately two orders of magnitude larger than the respirations per minute (RPM) estimates from the same. For computational reasons it can be desirable to have the parameter vectors for the data types be of similar magnitudes, so the movement data types were accordingly scaled down by a factor 100 prior to preforming logistic regression, and must of course be correspondingly scaled before applying the resulting models.

Final models

The feature vectors $x[t]$ for the radars were constructed as constructed as

$$x_{radar}[t] = \left[1, \frac{fast\_mov[t]^T}{100}, \frac{slow\_mov[t]^T}{100}, RPM[t]^T\right]^T$$

For the actigraph, they were simply

$$x_{actigraph}[t] = [1, activity[t]^T]^T$$

The values of the corresponding parameter vectors, $\theta = [k, \alpha^T, \beta^T, \varphi^T]^T$ for the radar and $\theta = [k, \alpha^T]^T$ for the actigraph, for the two time horizons presented in the main text ($[-4,2]$ and $[-5,0]$), can be found in table S2.

Classification performance, expanded tables

This study created models that included 0 to 10 epochs into the past (5 minutes, with 30 second epochs), and 0 to 10 into the future, as well as every possible combination in between, for a total of 121 possible models. The main text considers only the [-4, 2] and [-5, 0] time horizon; supplementary tables S3 and S4 show the classification performance of a wider range of possible time horizons over DS1-train and DS1-test. These results have not been subjected to the Cole-Kripke rescoring rules.

An expanded version of Table 2 from the main text that also includes non-rescored results can be found in supplementary Table S5. An expanded version of Table 4 from the main text that also includes the p-values from Student's t-tests comparing the estimated sleep parameters to their PSG counterparts can be found in supplementary Table S6.

References

1. Stone WC. Nist construction automation program report no. 3: Electromagnetic signal attenuation in construction materials. In: Commerce USDo, ed. Gaithersburg, Maryland Building and Fire Research Laboratory, National Institute of Standards and Technology; 1997.
2. Novelda AS. X4M200 Datasheet, Rev. E - Preliminary. 18 Oct 2018; www.xethru.com.
3. Wisland DT, Granhaug K, Pleym JR, Andersen N, Støa S, Hjortland HA. Remote Monitoring of Vital Signs Using a CMOS UWB Radar Transceiver. 2016 14th IEEE International New Circuits and Systems Conference (NEWCAS); 26-29 June, 2016; Vancouver, BC, Canada.
4. Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborney G. An Activity-Based Sleep Monitor System for Ambulatory Use. SLEEP. 1982; 5 (4): 389-399. doi.org/10.1093/sleep/5.4.389.
5. Borazio M, Berlin E, Kücükyildiz N, Scholl P, Laerhoven KV. Towards Benchmarked Sleep Detection with Inertial Wrist-worn Sensing Units. In: proceedings from the 2014 IEEE International Conference on Healthcare Informatics; 2014.
6. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic Sleep/Wake Identification From Wrist Activity. SLEEP. 1992; 15 (5): 461-469. doi.org/10.1093/sleep/15.5.461.
7. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Performance comparison of different interpretative algorithms utilized to derive sleep parameters from wrist actigraphy data. Chronobiol Int. 2019; 36 (12): 1752-1760. doi.org/10.1080/07420528.2019.1679826.

List of Supplementary Figure Captions

Figure S1

Example of data alignment for a healthy volunteer. For this recording, peaks in the cross-correlation sequences were found at t=-129 seconds for the radar, and t=20 seconds for the actigraph, and the data was correspondingly shifted by these amounts.

List of Supplementary Table Captions

Table S1. Average temporal shifts after sensor alignment

Sensors were time-aligned by identifying the distance to maximum cross-correlation of movement and shifting by that amount. Data was aligned to PSG when available. When PSG was not available, radar data was aligned to actigraphy.

[1] n=12 individuals, mean age $\pm$ SD: 23.0 $\pm$ 3.1 years, 5 male.

[3] PSG: Polysomnography

[3] Ambulatory sleep disorder patients, mean age $\pm$ SD: 46.25 $\pm$ 13.98 years, 19 male

Table S2. Parameter vectors for the two models (time horizons [-4, 2] and [-5, 0]) presented in the main text

The probability of wake is calculated according to the formula $p[t] = \frac{1}{1+e^{\theta^T x[t]}}$, where $x[t]$ is a feature vector of data from the sensor in question.

Table S3. Classification results over the healthy volunteers training and test sets for a selected range of time horizons

Epoch-by-epoch classification performance statistics for ten selected models. Performance generally increases with the number of epochs included, at the cost of increased complexity in terms of number of parameters.

n=12, mean age $\pm$ SD: 23.0 $\pm$ 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n=24/22 for nightstand/ceiling), and a testing set for validation (n=23/21 for nightstand/ceiling).

Table S4. Classification results over the healthy volunteers training and test sets for a selected range of real-time horizons

Epoch-by-epoch classification performance statistics for ten real-time models. Including a longer 'tail' of past epochs in the scoring of the present epoch generally increases performance, at the cost of a higher number of parameters in the model.

n=12, mean age $\pm$ SD: 23.0 $\pm$ 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n=24/22 for nightstand/ceiling), and a testing set for validation (n=23/21 for nightstand/ceiling).

Table S5. Classification performance, with and without the Cole-Kripke rescoring rules.

Epoch-by-epoch classification performance statistics for two models, with and without the heuristic Cole-Krike rescoring rules applied, compared to PSG[3]-determined sleep/wake. Mean (SD) over the participants in the data sets.

The rescoring rules act as a nonlinear time series filter on the sleep/wake state sequence, converting initially scored sleep epochs to wake if they occur after or within a specified interval of a given number of awake epochs. We observe that they generally improve the accuracy, specificity, and Cohen's kappa values, at the cost of a slight decrease in sensitivity.

[1] n=12, mean age $\pm$ SD: 23.0 $\pm$ 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n=24/22 for nightstand/ceiling), and a testing set for validation (n=23/21 for nightstand/ceiling).

[2] Ambulatory sleep disorder patients. n=28, mean age $\pm$ SD: 46.25 $\pm$ 13.98 years, 19 male.

[3] PSG, Polysomnography

Table S1. Average temporal shifts after sensor alignment			
Healthy volunteers[1] aligned to PSG[2]	Nightstand radar	Ceiling radar	Actigraph
Number of nights (n)	47	43	47
Avg. temp. shift (seconds)	-139	-139	-14
Healthy volunteers aligned to actigraph			
Number of nights (n)	78	74	
Avg. temp. shift (seconds)	-188	-192	
Patients with sleep disorders[3] aligned to PSG			
Number of nights (n)	28		28
Avg. temp. shift (seconds)	-150		-134
Sensors were time-aligned by identifying the distance to maximum cross-correlation of movement and shifting by that amount. Data was aligned to PSG when available. When PSG was not available, radar data was aligned to actigraphy.			

[1] n=12 individuals, mean age ± SD: 23.0 ± 3.1 years, 5 male.
[3] PSG: Polysomnography
[3] Ambulatory sleep disorder patients, mean age ± SD: 46.25 ± 13.98 years, 19 male

Table S2. Parameter vectors for the two models (time horizons [-4, 2] and [-5, 0]) presented in the main text

$$\theta = [k, \alpha^T, \beta^T, \varphi^T]^T$$

Past 4 future 2				Past 5 future 0			
	Radar nightstand	Radar ceiling	Actigraph		Radar nightstand	Radar ceiling	Actigraph
k	2.29541696	2.321924235	-2.537046376	k	2.224988488	2.245238055	-2.476
α_{T_0-4}	0.02622775	-0.004505437	0.031183202	α_{T_0-5}	0.018935199	-0.009281824	0.03211
α_{T_0-3}	-0.023148076	-0.001745622	0.0182493	α_{T_0-4}	0.003475304	0.001758058	0.017627
α_{T_0-2}	-0.071644213	-0.001989473	0.021077142	α_{T_0-3}	-0.073075754	-0.007755465	0.01916
α_{T_0-1}	-0.150782331	-0.012469094	0.021357249	α_{T_0-2}	-0.083749488	-0.008823601	0.021216
α_{T_0}	0.21061496	0.04178506	0.066003221	α_{T_0-1}	-0.169776146	-0.019129144	0.022211
α_{T_0+1}	-0.112249191	-0.023791666	0.013722412	α_{T_0}	0.156942021	0.031064786	0.079858
α_{T_0+2}	-0.020656002	-0.013717754	0.026738597	β_{T_0-5}	-0.035354117	-0.019951623	
β_{T_0-4}	0.003331279	-0.015442191		β_{T_0-4}	0.039690519	-0.017181948	
β_{T_0-3}	0.080393217	-0.006251548		β_{T_0-3}	0.074528809	-0.003136177	
β_{T_0-2}	0.135411303	0.005556379		β_{T_0-2}	0.137373216	0.011366725	
β_{T_0-1}	0.239867406	0.029647434		β_{T_0-1}	0.173963083	0.024794082	
β_{T_0}	0.247616806	0.018632057		β_{T_0}	0.393472906	0.043739137	
β_{T_0+1}	-0.002014974	-0.007881165		φ_{T_0-5}	-0.119278492	-0.102028533	
β_{T_0+2}	0.0196491	0.004586354		φ_{T_0-4}	-0.049013576	-0.057099199	
φ_{T_0-4}	-0.106780696	-0.095298851		φ_{T_0-3}	-0.057543126	-0.051653191	
φ_{T_0-3}	-0.043580648	-0.046111285		φ_{T_0-2}	-0.05650131	-0.055757739	
φ_{T_0-2}	-0.048270835	-0.046182456		φ_{T_0-1}	-0.061134167	-0.054659852	
φ_{T_0-1}	-0.051752769	-0.043281062		φ_{T_0}	-0.149422212	-0.139319612	
φ_{T_0}	-0.089895371	-0.082723051					
φ_{T_0+1}	-0.065330391	-0.072010432					
φ_{T_0+2}	-0.094796015	-0.082603652					

The probability of wake is calculated according to the formula $p[t] = \frac{1}{1+e^{\theta^T x[t]}}$, where $x[t]$ is a feature vector of data from the sensor in question.

Table S3. Classification results over the healthy volunteers training and test sets for a selected range of time horizons

	past	future	Error train [%]	Sensitivity training	Specificity training	Cohen's kappa training	Error test [%]	Sensitivity test	Specificity test	Cohen's kappa test
Radar nightstand	0	0	9.86	0.91	0.88	0.77	8.62	0.94	0.86	0.80
	1	1	7.88	0.95	0.86	0.81	6.94	0.97	0.84	0.83
	2	2	7.19	0.96	0.86	0.83	6.50	0.98	0.83	0.84
	3	3	6.89	0.96	0.86	0.83	6.18	0.99	0.83	0.85
	4	4	6.53	0.97	0.86	0.84	6.08	0.99	0.83	0.85
	5	5	6.35	0.97	0.86	0.85	5.89	0.99	0.83	0.85
	6	6	6.26	0.97	0.87	0.85	5.76	0.99	0.83	0.86
	7	7	6.22	0.97	0.87	0.85	5.74	0.99	0.83	0.86
	8	8	6.13	0.97	0.87	0.85	5.68	0.99	0.84	0.86
	9	9	6.01	0.97	0.87	0.85	5.67	0.99	0.84	0.86
	10	10	5.83	0.97	0.87	0.86	5.58	0.99	0.84	0.86
Radar ceiling	0	0	10.83	0.91	0.86	0.75	9.82	0.93	0.83	0.77
	1	1	8.73	0.94	0.85	0.79	8.12	0.97	0.80	0.80
	2	2	8.38	0.95	0.84	0.80	7.80	0.98	0.79	0.81
	3	3	8.03	0.95	0.84	0.81	7.63	0.98	0.79	0.81
	4	4	7.84	0.95	0.84	0.81	7.60	0.99	0.78	0.81
	5	5	7.74	0.96	0.84	0.81	7.66	0.99	0.78	0.81
	6	6	7.67	0.96	0.84	0.81	7.68	0.99	0.78	0.81
	7	7	7.67	0.96	0.84	0.81	7.62	0.99	0.78	0.81
	8	8	7.59	0.96	0.84	0.82	7.56	0.99	0.78	0.81
	9	9	7.53	0.96	0.84	0.82	7.51	0.99	0.78	0.81
	10	10	7.57	0.96	0.84	0.82	7.51	0.99	0.78	0.81
Actigraph	0	0	13.24	0.98	0.61	0.65	12.67	0.13	0.98	0.62
	1	1	11.76	0.97	0.67	0.70	10.91	0.11	0.98	0.69
	2	2	10.69	0.97	0.70	0.73	9.65	0.10	0.98	0.73
	3	3	9.81	0.97	0.73	0.75	8.85	0.09	0.98	0.75
	4	4	9.10	0.98	0.75	0.77	8.19	0.08	0.98	0.77
	5	5	8.55	0.98	0.77	0.78	7.76	0.08	0.98	0.79
	6	6	8.18	0.98	0.78	0.79	7.37	0.07	0.98	0.80
	7	7	7.72	0.98	0.79	0.81	7.02	0.07	0.98	0.81
	8	8	7.34	0.98	0.80	0.82	6.72	0.07	0.98	0.82
	9	9	6.95	0.98	0.82	0.83	6.41	0.06	0.98	0.83
	10	10	6.65	0.98	0.82	0.84	6.26	0.06	0.98	0.84

Epoch-by-epoch classification performance statistics for ten selected models. Performance generally increases with the number of epochs included, at the cost of increased complexity in terms of number of parameters.

n=12, mean age $\pm$ SD: 23.0 $\pm$ 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n=24/22 for nightstand/ceiling), and a testing set for validation (n=23/21 for nightstand/ceiling).

Table S4. Classification results over the healthy volunteers training and test sets for a selected range of real-time horizons

	past	future	Error train [%]	Sensitivity training	Specificity training	Cohen's kappa training	Error test [%]	Sensitivity test	Specificity test	Cohen's kappa test
Radar nightstand	0	0	8.56	0.94	0.87	0.80	7.56	0.96	0.84	0.82
	1	0	7.85	0.95	0.86	0.81	7.03	0.97	0.83	0.83
	2	0	7.51	0.95	0.86	0.82	6.87	0.98	0.83	0.83
	3	0	7.26	0.96	0.86	0.82	6.61	0.98	0.83	0.84
	4	0	7.08	0.96	0.86	0.83	6.52	0.98	0.82	0.84
	5	0	6.90	0.96	0.86	0.83	6.48	0.98	0.82	0.84
	6	0	6.82	0.96	0.86	0.83	6.44	0.99	0.82	0.84
	7	0	6.68	0.96	0.86	0.84	6.30	0.99	0.82	0.84
	8	0	6.64	0.97	0.86	0.84	6.27	0.99	0.82	0.84
	9	0	6.62	0.97	0.86	0.84	6.22	0.99	0.82	0.85
	10	0	6.62	0.97	0.86	0.84	6.22	0.99	0.82	0.85
Radar ceiling	0	0	9.46	0.93	0.84	0.77	8.78	0.96	0.81	0.79
	1	0	9.03	0.94	0.84	0.78	8.41	0.97	0.80	0.79
	2	0	8.69	0.95	0.84	0.79	8.17	0.97	0.79	0.80
	3	0	8.43	0.95	0.84	0.80	8.17	0.98	0.78	0.80
	4	0	8.29	0.95	0.84	0.80	8.09	0.98	0.78	0.80
	5	0	8.12	0.95	0.84	0.80	8.02	0.98	0.78	0.80
	6	0	8.03	0.95	0.84	0.81	8.01	0.98	0.77	0.80
	7	0	7.97	0.96	0.84	0.81	8.04	0.98	0.77	0.80
	8	0	7.94	0.95	0.84	0.81	7.98	0.99	0.77	0.80
	9	0	7.91	0.96	0.84	0.81	7.99	0.99	0.77	0.80
	10	0	7.91	0.96	0.84	0.81	7.99	0.99	0.77	0.80
Actigraph	0	0	12.34	13.24	0.98	0.61	0.65	12.67	0.98	0.62
	1	0	11.58	12.34	0.97	0.65	0.68	11.44	0.98	0.67
	2	0	10.89	11.58	0.97	0.67	0.70	10.56	0.98	0.70
	3	0	10.43	10.89	0.97	0.70	0.72	9.85	0.98	0.72
	4	0	9.96	10.43	0.97	0.71	0.73	9.49	0.98	0.73
	5	0	9.45	9.96	0.98	0.72	0.75	8.96	0.98	0.75
	6	0	9.13	9.45	0.98	0.74	0.76	8.69	0.98	0.76
	7	0	8.75	9.13	0.98	0.75	0.77	8.36	0.98	0.77
	8	0	8.47	8.75	0.98	0.76	0.78	8.18	0.98	0.78
	9	0	8.22	8.47	0.98	0.77	0.79	7.93	0.98	0.78
	10	0	8.22	8.22	0.98	0.78	0.79	7.58	0.98	0.80

Epoch-by-epoch classification performance statistics for ten real-time models. Including a longer 'tail' of past epochs in the scoring of the present epoch generally increases performance, at the cost of a higher number of parameters in the model.

n=12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n=24/22 for nightstand/ceiling), and a testing set for validation (n=23/21 for nightstand/ceiling).

Table S5. Classification performance, with and without the Cole-Kripke rescoring rules.

Time horizon	Rescored	Data set	Sensor type/placement	Accuracy [%]	Specificity [%]	Sensitivity [%]	Cohen's kappa*100
[-4, 2]	Yes	Healthy volunteers test set [1]	Radar nightstand	94.7 (2.7)	88.5 (10.3)	96.6 (3.2)	85.8 (8.5)
			Radar ceiling	93.6 (3.3)	84.6 (12.3)	96.4 (3.8)	82.5 (10.5)
			Actigraph	93.2 (3.0)	85.1 (11.4)	95.9 (3.7)	82.0 (8.0)
[-4, 2]	No	Healthy volunteers test set	Radar nightstand	92.9 (2.5)	78.8 (12.2)	97.9 (2.5)	80.2 (9.7)
			Radar ceiling	91.4 (3.5)	73.5 (14.5)	97.7 (3.1)	75.5 (12.7)
			Actigraph	90.6 (2.7)	71.8 (13.5)	97.3 (2.6)	73.5 (9.8)
[-5, 0]	Yes	Healthy volunteers test set	Radar nightstand	94.5 (2.7)	88.6 (8.6)	96.3 (3.1)	85.6 (7.4)
			Radar ceiling	93.1 (3.4)	83.7 (12.3)	96.2 (3.7)	81.3 (10.6)
			Actigraph	92.8 (3.0)	83.8 (10.8)	96.0 (3.5)	81.1 (7.4)
[-5, 0]	No	Healthy volunteers test set	Radar nightstand	92.8 (2.7)	78.8 (11.5)	97.7 (2.6)	80.0 (9.2)
			Radar ceiling	91.0 (3.5)	73.2 (13.8)	97.4 (3.1)	74.8 (12.1)
			Actigraph	90.5 (2.8)	71.3 (13.1)	97.4 (2.5)	73.2 (9.4)
[-4, 2]	Yes	Patients with sleep disorders [2]	Radar nightstand	80.9 (15.7)	53.7 (18.4)	89.5 (16.9)	44.8 (25.6)
			Actigraph	83.8 (9.0)	74.3 (20.0)	89.4 (7.3)	53.3 (15.8)
[-4, 2]	No	Patients with sleep disorders	Radar nightstand	81.1 (14.9)	47.8 (17.0)	91.4 (14.4)	42.7 (24.2)
			Actigraph	83.6 (9.4)	65.2 (19.2)	91.6 (5.6)	50.3 (14.4)
[-5, 0]	Yes	Patients with sleep disorders	Radar nightstand	80.9 (15.3)	53.4 (18.7)	89.7 (16.5)	44.3 (24.8)
			Actigraph	84.1 (9.0)	74.0 (20.2)	89.9 (6.9)	53.8 (16.2)
[-5, 0]	No	Patients with sleep disorders	Radar nightstand	81.0 (14.6)	48.1 (17.3)	91.3 (14.1)	42.2 (23.5)
			Actigraph	84.0 (9.4)	65.6 (19.5)	92.1 (5.4)	51.4 (14.9)

Epoch-by-epoch classification performance statistics for two models, with and without the heuristic Cole-Krike rescoring rules applied, compared to PSG[3]-determined sleep/wake. Mean (SD) over the participants in the data sets.

The rescoring rules act as a nonlinear time series filter on the sleep/wake state sequence, converting initially scored sleep epochs to wake if they occur after or within a specified interval of a given number of awake epochs. We observe that they generally improve the accuracy, specificity, and Cohen's kappa values, at the cost of a slight decrease in sensitivity.

[1] n=12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n=24/22 for nightstand/ceiling), and a testing set for validation (n=23/21 for nightstand/ceiling).
[2] Ambulatory sleep disorder patients. n=28, mean age ± SD: 46.25 ± 13.98 years, 19 male.
[3] PSG, Polysomnography

Table S6. Sleep parameters expanded

Healthy volunteers test set[1]

Model[2]	Variable[3]	PSG[4] (n=23) Mean (SD)	Nightstand radar (n=23) Mean (SD)	p	Cohen's D	Ceiling radar (n=21) Mean (SD)	p	Cohen's D	Actigraphy (n=23) Mean (SD)	p	Cohen's D
Four past, two future	TST [min]	434.4 (18.2)	437.4 (17.2)	0.14	0.17	438.3 (22.6)	0.10	0.24	433.2 (20.1)	0.72	-0.06
	SOL [min]	10.5 (6.2)	12.6 (6.4)	**0.03**	0.33	11.0 (6.0)	0.93	-0.01	13.7 (6.6)	**0.05**	0.5
	WASO [min]	17.3 (11.2)	15.1 (13.1)	0.10	-0.18	17.5 (17.8)	0.80	-0.05	19.3 (17.3)	0.41	0.13
	SE [%]	94.0 (2.6)	94.1 (2.9)	0.92	0.01	93.9 (4.3)	0.73	0.06	93.0 (4.0)	0.14	-0.31
	NW [num]	17.6 (4.7)	9.8 (3.9)	**<0.001**	-1.8	8.0 (4.7)	**<0.001**	-2.12	12.8 (5.5)	**0.01**	-0.93
Five past, zero future (realtime)	TST [min]	434.4 (18.2)	436.9 (16.4)	0.23	0.15	437.3 (21.7)	0.17	0.2	433.1 (19.3)	0.69	-0.07
	SOL [min]	10.5 (6.2)	13.0 (6.0)	**0.01**	0.41	12.4 (6.4)	0.18	0.21	14.5 (6.5)	**0.02**	0.63
	WASO [min]	17.3 (11.2)	15.8 (12.1)	0.22	-0.13	17.7 (17.0)	0.86	-0.04	19.2 (15.8)	0.39	0.14
	SE [%]	94.0 (2.6)	93.9 (2.6)	0.58	-0.07	93.6 (4.2)	0.87	-0.03	92.8 (3.8)	0.08	-0.37
	NW [num]	17.6 (4.7)	12.3 (4.9)	**<0.001**	-1.09	10.0 (5.3)	**<0.001**	-1.57	15.2 (5.7)	0.08	-0.46

Patients with sleep disorders[5]

Model type	Variable	PSG[2] (n=28) Mean (SD)	Nightstand radar (n=28) Mean (SD)	p	Cohen's D				Actigraph (n=28) Mean (SD)	p	Cohen's D
Four past, two future	TST [min]	386.7 (73.3)	386.5 (85.2)	0.99	0				372.0 (59.3)	0.24	-0.22
	SOL [min]	13.0 (14.9)	10.0 (12.4)	0.15	-0.22				14.2 (16.1)	0.64	0.07
	WASO [min]	61.3 (55.7)	64.9 (77.8)	0.83	0.05				74.0 (42.6)	0.28	0.26
	SE [%]	84.2 (12.2)	84.7 (16.1)	0.90	0.03				81.6 (9.5)	0.36	-0.24
	NW [num]	24.6 (13.2)	19.3 (14.0)	0.08	-0.39				21.6 (10.5)	0.26	-0.25
Five past, zero future (realtime)	TST [min]	386.7 (73.3)	387.2 (85.3)	0.98	0.01				374.4 (58.2)	0.31	-0.19
	SOL [min]	13.0 (14.9)	10.3 (12.2)	0.18	-0.2				15.4 (18.1)	0.38	0.14
	WASO [min]	61.3 (55.7)	61.9 (80.8)	0.98	0.01				71.2 (39.2)	0.40	0.2
	SE [%]	84.2 (12.2)	85.5 (16.2)	0.74	0.09				82.0 (9.0)	0.43	-0.2
	NW [num]	24.6 (13.2)	23.1 (15.5)	0.65	-0.1				23.9 (11.5)	0.81	-0.05

Sleep parameters extracted from manually scored hypnograms (PSG) compared to the sleep/wake state sequences resulting from the application of two selected classification models to radar and actigraphy data. p-values from paired-sample Student's t-tests of each model/sensor compared to their corresponding PSG recordings show which estimated sleep variables are significantly different from PSG, and the standardized effect sizes are reported by Cohen's D.

[1] n=12, mean age ± SD: 23.0 ± 3.1 years, 5 male, 4 nights of PSG + actigraphy + two radars per participant. The participants were randomly assigned into a training set for model development (n=24/22 for nightstand/ceiling), and a testing set for validation (n=23/21 for nightstand/ceiling).

[2] The models are defined by the number of preceding (past) and succeeding (future) epochs used to score a present epoch.

[3] SOL, Sleep Onset Latency; TST, Total Sleep Time; WASO, Wake After Sleep Onset; SE, Sleep Efficiency; NW, Number of awakenings

[4] PSG, Polysomnography (sleep parameters scored manually, independent of models)

[5] Ambulatory sleep disorder patients. n=28, mean age ± SD: 46.25 ± 13.98 years, 19 male.

Supplementary Figure S1
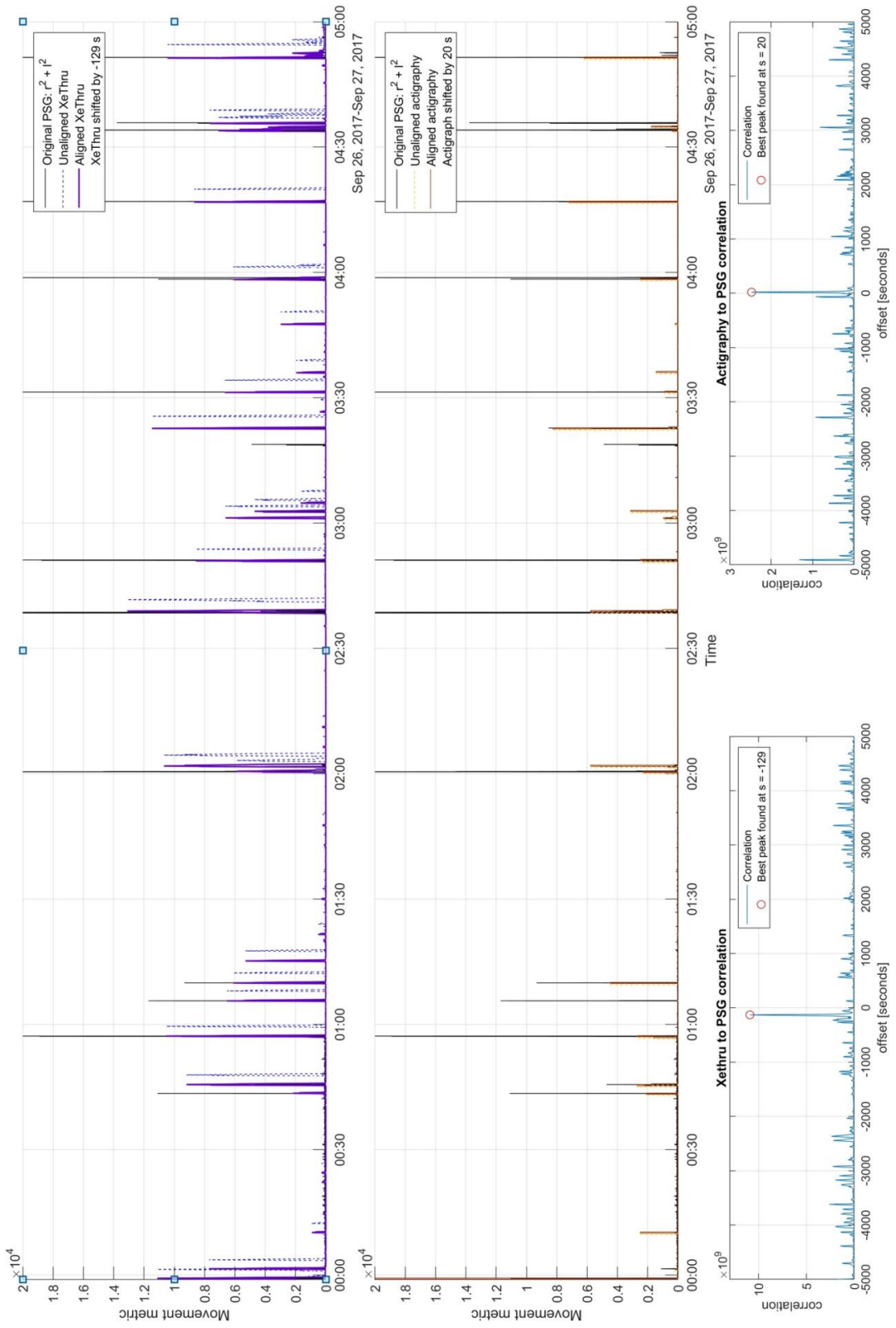
PAPER IV

Original Article

Mode of delivery of Cognitive Behavioral Therapy for Insomnia: a randomized controlled non-inferiority trial of digital and face-to-face therapy

Håvard Kallestad[1,2,*,], Jan Scott[2,3,], Øystein Vedaa[2,4], Stian Lydersen[2], Daniel Vethe[2,], Gunnar Morken[1,2,], Tore Charles Stiles[5], Børge Sivertsen[2,4,6,] and Knut Langsrud[1,2]

[1]Division of Mental Health Care, St. Olavs University Hospital, Trondheim, Norway, [2]Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway, [3]Institute of Neuroscience, Newcastle University, Newcastle, United Kingdom, [4]Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway, [5]Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway and [6]Department of Research and Innovation, Helse-Fonna HF, Haugesund, Norway

*Corresponding author. Håvard Kallestad, Division of Mental Health Care, Department of Research and Development, PO Box 3250 Torgarden, NO-7033 Trondheim, Norway. Email: havard.kallestad@ntnu.no.

Abstract

Study Objectives: Digital Cognitive Behavioral Therapy for Insomnia (dCBT-I) has demonstrated efficacy in reducing insomnia severity in self-referred and community samples. It is unknown, however, how dCBT-I compares to individual face-to-face (FtF) CBT-I for individuals referred to clinical secondary services. We undertook a randomized controlled trial to test whether fully automated dCBT-I is non-inferior to individual FtF CBT-I in reducing insomnia severity.

Methods: Eligible participants were adult patients with a diagnosis of insomnia disorder recruited from a sleep clinic provided via public mental health services in Norway. The Insomnia Severity Index (ISI) was the primary outcome measure. The non-inferiority margin was defined a priori as 2.0 points on the ISI at week 33.

Results: Individuals were randomized to FtF CBT-I (n = 52) or dCBT-I (n = 49); mean baseline ISI scores were 18.4 (SD 3.7) and 19.4 (SD 4.1), respectively. At week 33, the mean scores were 8.9 (SD 6.0) and 12.3 (SD 6.9), respectively. There was a significant time effect for both interventions (p < 0.001); and the mean difference in ISI at week 33 was −2.8 (95% CI: −4.8 to −0.8; p = 0.007, Cohen's d = 0.7), and −4.6 at week 9 (95% CI −6.6 to −2.7; p < 0.001), Cohen's d = 1.2.

Conclusions: At the primary endpoint at week 33, the 95% CI of the estimated treatment difference included the non-inferiority margin and was wholly to the left of zero. Thus, this result is inconclusive regarding the possible inferiority or non-inferiority of dCBT-I over FtF CBT-I, but dCBT-I performed significantly worse than FtF CBT-I. At week 9, dCBT-I was inferior to FtF CBT-I as the 95% CI was fully outside the non-inferiority margin. These findings highlight the need for more clinical research to clarify the optimal application, dissemination, and implementation of dCBT-I.

Clinicaltrials.gov: NCT02044263: Cognitive Behavioral Therapy for Insomnia Delivered by a Therapist or on the Internet: a Randomized Controlled Non-inferiority Trial.

Statement of Significance

Direct comparisons between the current gold-standard and new digital therapeutics in clinical populations are needed to benchmark the effectiveness of digital therapeutics. This is the first study to compare a fully automated digital Cognitive Behavioral Therapy for Insomnia (CBT-I) with individual Face-to-Face (FtF) CBT-I in a clinical population. In this context, we could not conclude about long-term non-inferiority, although dCBT-I performed significantly worse than FtF CBT-I in reducing insomnia severity. Further research on dCBT-I may be useful to understand how best to deliver this treatment and who might best be served by it. Future developments may include tailoring the dCBT-I intervention to different populations.

Key words: Cognitive Behavioral Therapy for Insomnia; digital; face-to-face; non-inferiority; randomized controlled trial; insomnia disorder

---

Introduction

Insomnia is a significant health problem, affecting 10%–15% of the general population [1]. Its diagnosis is primarily dependent on the presence of three phenomena: persistent sleep difficulty, adequate sleep opportunity, and associated daytime dysfunction for more than 3 months [2]. There is a consensus that Cognitive Behavioral Therapy for Insomnia (CBT-I) is the best intervention for insomnia and should be a first-line treatment option [3]. However, there is a significant gap between supply and demand which is largely attributable to the lack of availability of therapists trained in CBT-I [4]. To overcome barriers regarding access to face-to-face (FtF) therapy, CBT-I has been adapted for delivery via digital means such as websites or apps. These digital approaches (which we will refer to as dCBT) differ in the amount of support from clinicians that is offered, ranging from materials included only as a supplement to a course of FtF therapy, via therapist-guided programs, to fully automated dCBT [5].

Although dCBT-I models vary, a recent meta-analysis of 11 randomized controlled trials (RCTs) indicated statistically significant effects on insomnia severity (pooled effect size (ES) > 1.0) for dCBT-I compared with control interventions such as sleep hygiene [6]. Further, in a recent large-scale RCT of fully automated dCBT-I in the general population, we observed an ES (Cohen's *d*) of 1.2 for insomnia severity, which is similar to that reported in trials of FtF CBT-I [7]. These data offer robust support for the wider community access to fully automated dCBT-I [8]. However, a critical issue still needs to be addressed: While the development of dCBT-I was never intended to replace FtF treatment in clinical settings, the issue of therapist availability has increased calls for investigators to establish how dCBT-I performs relative to gold-standard FtF therapies [9] in patients with a diagnosis of insomnia disorder rather than focusing solely on convenience samples of individuals with symptoms of insomnia recruited from the community [6].

Cognitive behavioral therapies are established interventions for most common mental disorders, which has encouraged the development and analysis of different modalities for delivery. One recent meta-analysis of guided and unguided digital, and FtF delivery of CBT for depression found that digital CBT was superior to FtF CBT in improving depressive symptoms [10]; furthermore, other reviews of therapist-guided digital CBT versus FtF CBT for psychiatric or somatic disorders found no differences in outcomes between the two modalities [11, 12]. Overall, this suggests that digital CBT may have similar effects as FtF CBT for individuals with various psychiatric and somatic disorders. However, for individuals with insomnia, there are only three published RCTs comparing dCBT-I with FtF CBT-I [13–15]. Taken together, the RCT findings are inconclusive regarding the effects on insomnia severity, as one found a trend favoring individual FtF CBT-I but no significant differences between the two modalities [15], one trial indicated that guided dCBT-I achieved similar improvements to group FtF CBT-I [13], and one reported that individual FtF CBT-I was superior to guided dCBT-I [14]. These trials recruited convenience samples via media [13, 14] or, for example, subpopulations of active military personnel [15], which means that the findings may not be generalizable to secondary care patients. Another two trials used guided rather than a fully automated dCBT-I. In some ways, a guided version of dCBT-I can be regarded as a hybrid between FtF and a fully automated dCBT-I, as guided dCBT-I provides additional direct therapeutic input and support to the individual (rather than being practiced independently).

The current study focuses on the critical gaps in the evidence-base regarding dCBT-I, namely, how does a fully automated dCBT-I intervention compare with individual FtF CBT-I when applied to a clinical secondary care population? To address this issue, we used an established, efficacious dCBT-I program called Sleep Healthy Using The internet (SHUTi) [16] and compared it with individual FtF CBT-I delivered by experienced therapists. Further, we recruited adult patients with sleep problems that met diagnostic criteria for insomnia disorder who were referred to a sleep clinic because these difficulties were impairing their functioning and/or quality of life.

Given the lack of comparative RCTs of gold standard individual FtF CBT-I (reference therapy) and fully automated dCBT-I (new therapy), we opted to conduct a non-inferiority trial with two arms. This design is warranted when a new treatment has advantages such as greater availability or reduced cost than an established gold standard (reference) [17], as the premise is that a non-inferiority RCT can determine that the new intervention is no worse than the reference intervention by a predefined "acceptable amount" [17].

In sum, the primary aim of this RCT was to test if dCBT-I is non-inferior to individual FtF CBT-I on insomnia severity as measured at six-months follow-up. The six-month follow-up was chosen as the primary endpoint for the trial because longer-term outcomes were regarded to be of higher relevance for both patients and health care systems. Secondary aims (defined a priori), involved superiority analyses to (1) estimate rates of clinical response and remission of insomnia according to the group, and (2) explore if there were any between-group differences in psychological distress, fatigue, or self-reported sleep-wake patterns.

Methods

Study design

A parallel-group randomized controlled non-inferiority trial of dCBT-I versus FtF CBT-I, with participants assigned in a 1:1 ratio. The protocol was approved by the Regional Ethical Committee of South-East Norway (Reference: 2013/1836) and the RCT was registered on the Clinical Trials website (ClinicalTrials.gov: NCT02044263).

The trial follows the CONSORT guidelines for a non-inferiority trial [17]. The flowchart is shown in Figure 1 and the CONSORT checklist is provided in the online Supplementary Materials (see Appendix 1).

Participants

All RCT participants provided written informed consent. Between October 2014 and January 2016, patients referred for treatment of insomnia to a secondary care sleep clinic at St. Olavs University Hospital, Trondheim, Norway were offered the opportunity to be involved in the RCT. Participants who completed all assessments were offered a gift voucher worth NOK 500 (equivalent to EUR 50) as a reimbursement for their time, etc.

Eligibility criteria: Eligible participants were all patients who had been referred to the sleep clinic with a presentation of
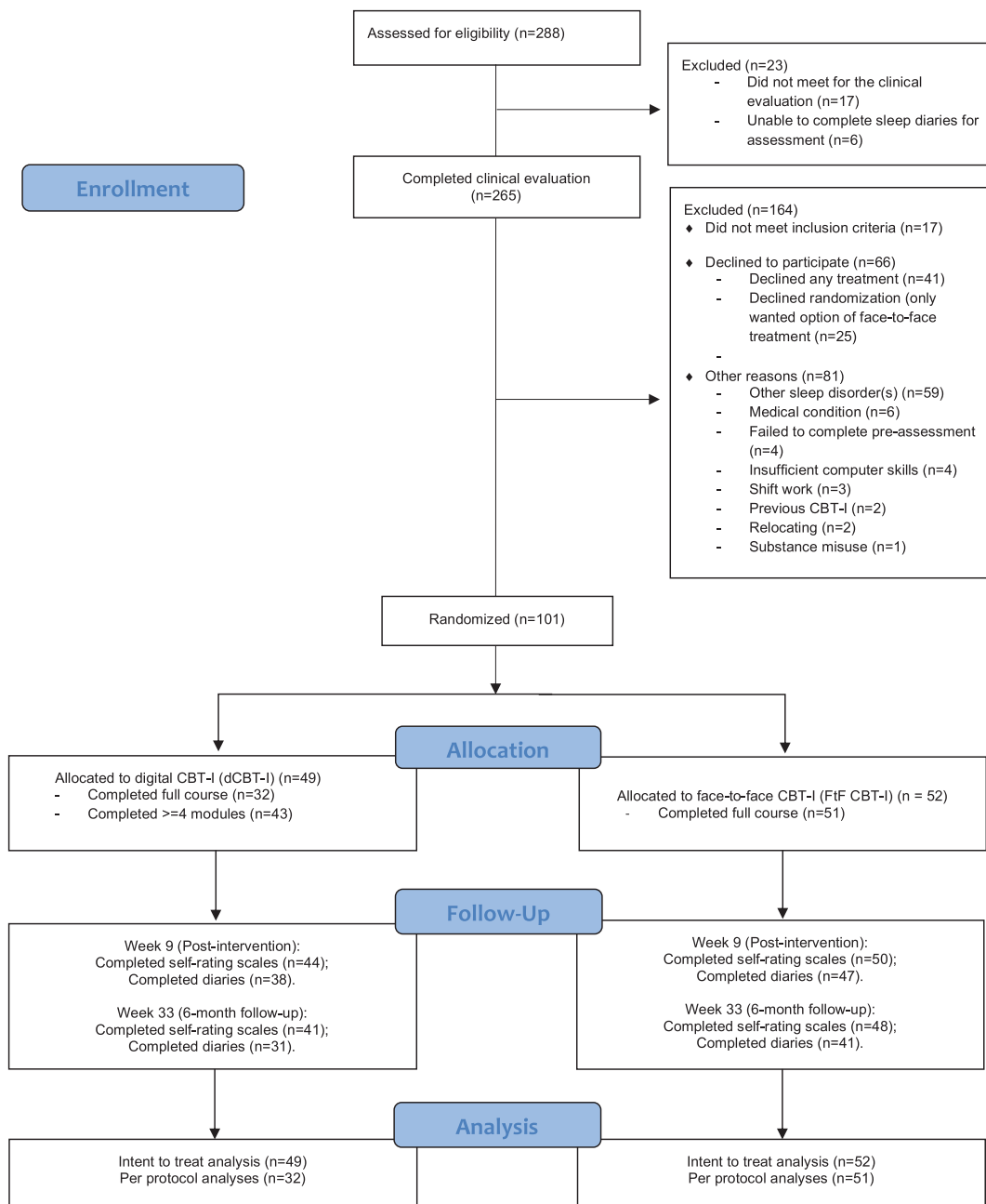
Figure 1. Flow diagram of trial: participant inclusion, timing of assessments, and completion rates.

insomnia. Participants could be included if they were aged >= 18, the presentation met diagnostic criteria for insomnia disorder as described in the DSM-5 [2], they had regular internet access and self-reported proficiency in basic computer/internet skills (as required to participate in the RCT and complete online

assessments, etc.), and they were willing and able to provide written informed consent.

The diagnosis of insomnia disorder was assessed by (1) a trained psychiatrist or licensed clinical psychologist using a semi-structured interview based on the Insomnia Interview

Schedule [18], supplemented by a module assessing circadian rhythm disorders, and (2) completion of a pre-assessment sleep diary that recorded sleep-wake cycle patterns for the 14 days prior to the interview. The interview also assessed current sleep-wake pattern, sleep history, previous history and treatment of sleep problems (including medication), functional sleep analysis (e.g. daytime consequences of sleep problems and perpetuating factors), and history of somatic and psychiatric illness and treatment.

Individuals were excluded if they met one or more of the following criteria: evidence of circadian rhythm disorder, or an organic sleep disorder as assessed in the Insomnia Interview Schedule; or, for sleep apnea specifically (established via evidence of sleep apnea at interview assessment or an Oxygen Desaturation Index above a cutoff of 9 as assessed by oximetry recordings) [19]; and/or a current alcohol and/or substance misuse problem. Also excluded were individuals: working night shifts and unable to discontinue this work pattern during the RCT, with previous exposure to CBT-I, or with a medical condition where sleep restriction is deemed inappropriate due to a potential for worsening of the medical condition (e.g. an attack phase of multiple sclerosis or epilepsy), and/or with insufficient fluency in Norwegian (i.e. unlikely to be able to complete interventions).

Randomization

The randomization program was designed and implemented by the Section of Applied Clinical Research, at the Faculty of Medicine and Health, the Norwegian University of Science and Technology (NTNU). The randomization utilized a balanced allocation sequence using blocks of randomly varying size. After trial completion, it was revealed that the range of blocks was between 6 and 20. Specifically, the first block was 20 (10 + 10), and the subsequent blocks varied between 6 (3 + 3) and 8 (4 + 4). Participants were randomized by the three clinicians who undertook the screening assessment (and who delivered FtF CBT-I). The clinicians logged on to an online portal and eligible participants were then assigned to one of the two groups. The clinicians could not influence the process in any way and the result of the randomization was communicated directly to the patient via email.

Individuals randomized to FtF CBT-I received an outpatient appointment for the sleep clinic. Participants randomized to dCBT-I were provided with a link to allow them to login to the online program. As this intervention was fully automated and delivered online, there was no contact between participants and the professional at the sleep clinic. If problems arose with the delivery of the online program or the completion of assessments, a technician could offer support to the participant.

Procedures

Individual face-to-face CBT-I. CBT-I is a multicomponent treatment and consisted of the following interventions: psychoeducation about sleep and sleep hygiene, sleep restriction, stimulus control, and challenging beliefs and perception of sleep [18, 20]. Table 1 provides a description of the week-by-week interventions. The course of FtF CBT-I involved three to eight sessions delivered over 6–9 weeks [18, 20], although the exact

number of sessions provided is dictated by client progress. Three therapists provided FtF CBT-I (two authors [H.K. and K.L.]; and another a licensed psychologist); all had participated in training courses, had 3–10 years post-graduate experience in CBT-I and received ongoing supervision in CBT-I. Their adherence to the therapy protocol during the RCT was monitored via weekly review meetings.

Digital CBT-I. We employed a dCBT-I program entitled SHUTi [21] that was created by investigators at the University of Virginia and was translated into Norwegian by the Norwegian Institute of Public Health. The program incorporates the same approaches used in standard (FtF) CBT-I packages, but the educational, behavioral, and cognitive interventions are conceptualized as six "cores" [22]. Participants with a goal of reducing or eliminating sleep medication use were informed that they should discuss this with the prescribing physician. Each core is accessed in a predefined sequence and admittance to subsequent cores is based on time (1 week after the completion of the previous core), and a required five diaries must be entered in the previous 7 days to move from core 1 to core 2 to set the program sleep window. Each core typically takes 45 to 60 min to complete, but there were no specific instructions about how long time the participants should take, and again completion may be partly affected by comprehension, attention, etc. Participants had access to the intervention for 6 months.

Assessments

Prior to randomization, background information was collected regarding demography (age, sex, socioeconomic, and employment status), while comorbid mental disorders were assessed with the self-reported Psychiatric Diagnostic Screening Questionnaire Comorbid Mental Disorders (PDSQ) [23]. The PDSQ consists of 111 items assessing symptoms of common *DSM-IV* Axis I disorders encountered in outpatient mental health settings (see Supplementary Table S1). Comorbid physical disorders, past and/or current mental health treatment, and past and/or current use of sleep medications, and number of different agents prescribed were identified via the clinical interview augmented by medical casenote recordings.

At baseline, 9 weeks after randomization (post-intervention), and 33 weeks after randomization (6 months post-intervention), participants were asked to complete self-report assessments. Insomnia severity was assessed using the Insomnia Severity Index (ISI) [24]. Scores on this 7-item questionnaire range from 0 to 28 with higher scores indicating greater symptom severity. The ISI has good psychometric properties, is widely used, and is a recommended primary outcome measure in insomnia research [25]. A reduction in ISI scores of 8 or more points from baseline is regarded as a response to treatment, and an absolute score on the ISI of 7 points or less is regarded as remission. Data on daily sleep-wake parameters were collected using an online version of the consensus sleep diary [26]. Individuals were asked to record information about sleep onset latency (SOL), wake after sleep onset (WASO), early morning awakenings (EMA), number of nocturnal awakenings, total sleep time, and sleep efficiency for 10 days (out of the previous 14 consecutive days). Psychological distress was assessed using the 14-item

Table 1. An overview of the content of each core of dCBT-I and the content of sessions in the FtF CBT-I

Core/session	dCBT-I	FtF CBT-I
1	Overview: Reviews the nature of insomnia and how the program works; participants identify their sleep problems and set up personal treatment goals.	Motivation and personal treatment goals. Psychoeducation about sleep architecture and the two-process theory of sleep-wake regulation. Education about sleep hygiene if patient is engaging in activities that could obviously interfere with the effect of sleep restriction (e.g. excessive caffeine use). Setting up sleep restriction (lower limit of 5 h). Setting up a plan for tapering of sleep medication if a treatment goal for the patient is to stop or reduce medication use.
2	Behavior and sleep: Focuses on how behavioral changes can improve sleep, with special emphasis on sleep restriction (lower limit of 5 h).	Review of adherence to sleep restriction and problem solving if needed. Socratic dialogue about changes in beliefs and behaviors about sleep, particular changes that have occurred as a function of sleep restriction (e.g. the need for safety behaviors in order to sleep). Motivational work to keep the patient adhering to sleep restriction.
3	Behavior and sleep 2: Focuses on behavioral changes that can improve sleep, with special emphasis on stimulus control	As week 2. Adding stimulus control if necessary.
4	Sleep and thoughts: Focuses on addressing and changing beliefs and thoughts that might impair sleep.	As weeks 2 and 3.
5	Sleep hygiene: Teaches about lifestyle and environmental factors that might interfere with sleep (e.g. caffeine and nicotine intake, electronic media use in bed).	As weeks 2 and 3.
6	Relapse prevention: Focuses on integrating the behavioral, educational, and cognitive components from the previous cores to develop strategies to prevent future episodes of poor sleep to develop into full-blown chronic insomnia.	Final session. Evaluation of current status relative to treatment goals in session 1. Relapse prevention: Check that the patient has understood the rationale behind sleep restriction and can implement use of sleep diaries and sleep restriction should sleep problems occur later. Implement stimulus control if the patient wants to stop sleep restriction.

version of the Hospital Anxiety and Depression Scale (HADS). The HADS has been shown to reliably rate symptoms of psychological distress in hospital outpatient clinics [27]. Scores range from 0 to 42 with higher scores indicative of greater distress. Daytime fatigue was assessed using the 13-item Chalder Fatigue Scale (CFS). Scores range from 0 to 39, with higher scores indicative of greater psychological and physical fatigue [28]. Dysfunctional beliefs about sleep were assessed using the Dysfunctional Beliefs and Attitudes about Sleep scale—16 items (DBAS-16) [29]. Mean scores on these items range from 1 to 10 with higher scores indicating a higher endorsement of dysfunctional beliefs about sleep.

Statistical analyses

All primary and secondary outcomes reported here are based on Intent-To-Treat (ITT) analyses. The primary outcome was the ISI, and the primary endpoint was at week 33. The non-inferiority margin for the mean difference in ISI scores between the two interventions was defined a priori as <2-points.

With an assumed standard deviation (SD) of 4.0, a difference of 2 points on the ISI corresponds to a moderate ES (Cohen's d = 0.5). Given the usual magnitude of ES in RCTs of interventions for insomnia (>1.0) and the previously employed margins for non-inferiority (e.g. 4 points on the ISI), this margin is likely to be enough to separate a change in mean ISI scores that is clinically important from a change that has limited clinical relevance [6, 13, 14, 30]. Under these assumptions, we estimated that a sample size of 100 participants commencing the RCT would

give a power of 80% for non-inferiority of dCBT-I (α = 0.05) in the ITT analyses.

We interpreted the primary outcome from the 95% Confidence Interval (CI) of the estimated between-group differences on the ISI as recommended in the CONSORT guidelines for non-inferiority trials [17]. That is, if the 95% CI for the mean difference between the two intervention groups was between -2 and ∞, then we would declare that dCBT-I is *non-inferior* to individual FtF CBT-I. If the 95% CI was between $-\infty$ and 0, then we would declare that individual FtF CBT-I is *superior* to dCBT-I. If the 95% CI was between 0 and ∞, then we would declare that dCBT-I is *superior* to individual FtF CBT-I. The secondary outcome of between-group differences on the ISI at week 9 was also interpreted using the same non-inferiority margin. Other secondary outcomes included between-group differences in HADS, CFS, and sleep variables recorded in the diary on the assessment at week 9 and week 33, and differences in numbers of remitters and responders based on ISI scores at weeks 9 and 33. These secondary outcomes were interpreted with standard superiority guidelines as suggested in the CONSORT non-inferiority guidelines [17].

We used SPSS version 25 for analysis of the primary and secondary outcomes. These analyses were performed by a statistician (SL) who was blinded to group allocation. We used a linear mixed model with individual as the random effect, time and group and their interaction as categorical covariates, and ISI score as the dependent variable. This approach implicitly accounts for missing-at-random data. We adjusted for the baseline value of the outcome variable, and estimated the difference

in ISI scores between groups at week 9 and week 33 from the interaction terms, as recommended by Twisk et al. [31].

Superiority analyses of group differences in CFS, HADS and DBAS mean scores, and self-reported sleep-wake cycle patterns (i.e. sleep diary parameters) were performed using similar linear mixed models as above. Further, the proportions of participants per group who met ISI criteria for response and remission were compared using Pearson's chi-squared test and the Newcombe Hybrid Score CI [32].

The above approaches for the continuous variables were repeated for the per protocol (PP) analyses restricted to patients who completed all therapy sessions they were offered and all modules of dCBT-I as recommended in the CONSORT non-inferiority guidelines [17].

Cohen's *d* was estimated as the difference in mean scores divided by the baseline SD [33]. For the between-group effect sizes, this was calculated using the difference estimate from the mixed model analyses divided by the pooled SD at baseline. For the within-group effect sizes, this was calculated using the difference in mean scores at baseline and each follow-up assessment divided by the within-group SD at baseline for each condition.

Results

Of 288 potential participants, 101 individuals met eligibility criteria and were randomized to the RCT (see Figure 1, Tables 2 and 3). The sample was predominantly female (75%), most had attended tertiary education and 59% were currently employed. Over 50% reported current physical comorbidity, and a similar proportion had a history of psychiatric treatment. Nearly 9 out of 10 reported previously taking >=1 sleep medication.

Fifty-one (of 52) patients assigned to FtF CBT-I completed the course of therapy (98%). One patient was excluded after two sessions of FtF CBT-I by the therapist (when the patient disclosed meeting exclusion criteria, namely working night shifts, and having a severe substance misuse problem). The average number of FtF sessions was 6 (range 3–8 sessions). In the dCBT-I group, 43 of 49 patients (88%) completed 4 or more CBT-I cores, and 31 patients (63%) completed all the cores.

Non-inferiority analyses on the primary outcome

There was a significant time effect for both interventions on the ISI score ($p < 0.001$). At the primary endpoint at week 33, participants receiving FtF CBT-I scored 2.8-points lower on the ISI compared with participants receiving dCBT-I (95% CI –4.8 to –0.8; $p = 0.007$; Cohen's $d = 0.7$). At week 9, the FtF group had a mean ISI score that was 4.6 points lower than the dCBT-I group (95% CI –6.6 to –2.7; $p < 0.001$), Cohen's $d = 1.2$. As shown in Figure 2, the 95% CI of the estimated mean difference between the two intervention groups at week 33 demonstrates that FtF meet criteria for superiority over dCBT-I (and at 9-week follow-up).

Planned superiority analyses of secondary outcomes

1. Response and Remission

The response rate at week 9 was significantly higher for FtF CBT-I (*n* = 35/50; 70%) compared with dCBT-I (*n* = 19/44; 43%), which represents a 27% difference in the proportion of responders (95% CI: 7% to 44%; $p = 0.009$). At week 33, there was

Table 2. Baseline characteristics of participants assigned to digital Cognitive Behavioral Therapy for Insomnia (dCBT-I) or face-to-face Cognitive Behavioral Therapy for Insomnia (FtF CBT-I)

	dCBT-I (*n* = 49)		FtF CBT-I (*n* = 52)	
Age in years, mean (SD)	41.4	(10.5)	41.3	(12.5)
Sex, *n* (%)				
Female	35	(71%)	41	(79%)
Male	14	(29%)	11	(21%)
Marital status, *n* (%)				
Married or cohabiting	30	(61%)	31	(60%)
Never married, divorced or separated	19	(39%)	21	(40%)
Education attainment, *n* (%)				
Below high school	1	(2%)	4	(8%)
Completed high school	16	(33%)	16	(31%)
College or higher	32	(65%)	32	(62%)
Employment status, *n* (%)				
Full or part time employment	31	(63%)	29	(56%)
Unemployed seeking work	0	(0%)	1	(2%)
Sick leave or disability pension	14	(29%)	15	(29%)
Student	4	(8%)	7	(13%)
Duration of insomnia, years (SD)	12.6	(11.5)	13.0	(11.4)
Comorbidities				
>=1 Physical disorder, *n* (%)	25	(51%)	30	(58%)
Current psychiatric outpatient, *n* (%)	10	(20%)	10	(19%)
Previous psychiatric treatment, *n* (%)	26	(53%)	25	(48%)
Prescribed sleep medications				
Current sleep medication, *n* (%)	31	(63%)	33	(64%)
Previous sleep medication, *n* (%)	42	(86%)	46	(89%)
Number of previous sleep medications, mean (SD)	2.5	(1.7)	2.5	(1.6)

Table 3. Self-reported mental disorders at baseline as identified by the Psychiatric Disorders Screening Questionnaire of participants allocated to digital Cognitive Behavioral Therapy for Insomnia (dCBT-I) or face-to-face Cognitive Behavioral Therapy for Insomnia (FtF CBT-I)

	dCBT-I (n = 49)	FtF CBT-I (n = 52)
Comorbid mental disorders, n (%)		
Any mental disorder	40 (82%)	40 (77%)
>=1 Affective disorder(s)	16 (33%)	15 (29%)
>=1 Anxiety disorder(s)	39 (80%)	37 (71%)
>=1 Alcohol and/or substance misuse disorder(s)	12 (24%)	9 (17%)
Other mental disorder(s)	7 (14%)	5 (10%)
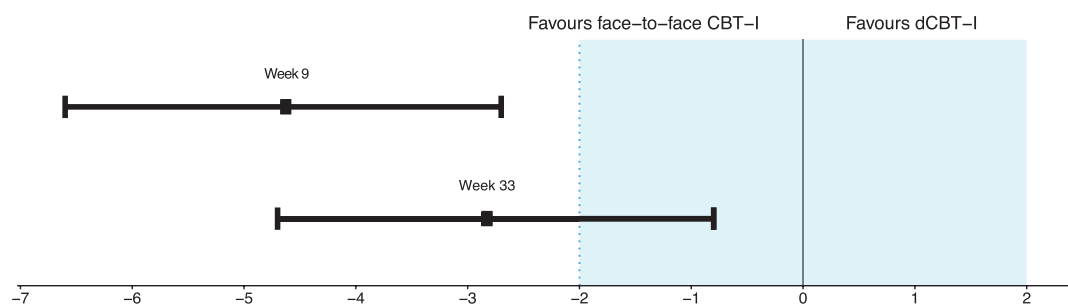
Treatment difference between dCBT–I and face–to–face CBT–I on the Insomnia Severity Index
The dotted line indicates the margin of non–inferiority

Figure 2. Difference in estimated mean scores on the Insomnia Severity Index between the two interventions and their 95% CI relative to the non-inferiority margin. Figure adapted from the CONSORT statement for non-inferiority trials [17].

a marginal change in these response rates (FtF: *n* = 31/48, 65%); dCBT-I: *n* = 19/41, 46%), but the disparity was no longer statistically significant (difference = 18%; 95% CI: –2% to 37%; *p* = 0.084).

The remission rate at week 9 for was significantly higher for FtF CBT-I (*n* = 26/50; 52%) than dCBT-I (*n* = 8/44; 18%), with a 34% difference in the proportion of remitters (95% CI: 15% to 50%; *p* = 0.001). At week 33, the difference remained statistically significant (FtF = 27/48 (56%); dCBT-I = 10/42 (24%)), with a difference of 32%; 95% CI: 12% to 49%; *p* = 0.002.

2. Other Clinical Outcomes

Table 4 summarizes findings for psychological distress, fatigue, dysfunctional beliefs, and self-reported sleep-wake parameters according to group and timing of follow-up. There was a significant time-effect for all secondary outcomes (*p* < 0.001), but no between-group differences. Changes show medium to large within-group ES.

PP analyses

As shown in Supplementary Table S1, the findings of the PP mixed model analyses were substantially the same as the ITT analyses.

Discussion

This RCT addressed a critical issue about the effectiveness of dCBT-I, namely the lack of data comparing fully automated dCBT-I with FtF CBT-I in a clinical population of patients with insomnia disorder [9]. We hypothesized that dCBT-I would be non-inferior to FtF CBT-I in reducing insomnia severity based on RCT findings indicating that dCBT-I has similar effect sizes to FtF CBT-I [6–8, 16, 34]. However, this RCT did not support this as we found that FtF CBT-I was superior to dCBT-I in reducing insomnia severity in patients referred to a sleep clinic provided by secondary care public mental health services in Norway. This finding is contrary to those reported from studies of other fields, where digital and FtF delivery of CBT has been shown to have similar effectiveness [10–12]. This discrepancy may be due to differences in various elements of the methodology, most notably sampling frames (e.g. recruitment of convenience vs. clinical populations) and the mode of the digital model studied (e.g. therapist guided vs. fully automated). Compared with the three previous trials in insomnia, our findings are similar to a trial of guided dCBT-I compared with FtF CBT-I undertaken in a sample of self-referrals recruited via social and other media [14]. Our findings also add to those reported in a trial in active military personnel, where there was a trend favoring individual FtF CBT-I over fully automated dCBT-I. That trial did not show any significant differences between the two modalities of delivery, but this may be due to low statistical power [15]. However, a third RCT reported that guided dCBT-I was non-inferior to FtF CBT-I delivered in a group format [13]. This is interesting as previous meta-analyses have shown that, whilst group CBT-I is effective [35], it is less effective than individual CBT-I [36]. If confirmed by further trials, this would be an important insight into how dCBT-I might be incorporated into clinical management or stepped-care models. Namely, while dCBT-I is less efficacious than individual FtF CBT-I, it may be as efficacious as group CBT-I and so could offer a viable alternative option, especially as resource availability and cost-benefit considerations would be likely to favor dCBT-I.

Table 4. Primary and secondary outcomes for participants assigned to digital Cognitive Behavioral Therapy for Insomnia (dCBT-I) or face-to-face Cognitive Behavioral Therapy for Insomnia (FtF CBT-I)

	dCBT-I (n = 49)				FtF CBT-I (n =52)				Difference (group × time)			
	n	Mean	SD	d	n	Mean	SD	d	Estimate	95% CI	d*	P
ISI												
Baseline	49	19.4	4.1		52	18.4	3.7					
Week 9	44	13.7	7.0	1.4	50	8.4	5.1	2.7	−4.6	−6.6 to −2.7	−1.2	<0.0001
Week 33	41	12.3	6.9	1.7	48	8.9	6.0	2.6	−2.8	−4.8 to −0.8	−0.7	0.007
HADS												
Baseline	49	15.2	6.8		52	12.9	6.6					
Week 9	44	13.6	7.8	0.2	48	9.8	6.6	0.5	−1.9	−3.8 to 0.04	−0.3	0.06
Week 33	41	12.2	8.4	0.4	47	9.0	7.1	0.6	−1.2	−3.2 to 0.7	−0.2	0.2
CFS												
Baseline	49	36.1	6.7		52	35.2	5.8					
Week 9	44	31.6	8.2	0.7	49	29.9	6.4	0.9	−0.9	−3.2 to 1.4	−0.1	0.5
Week 33	41	30.6	8.7	0.8	48	28.3	6.8	1.2	−1.3	−3.7 to 1.0	−0.2	0.3
DBAS-16												
Baseline	48	5.50	1.7		51	5.51	1.9					
Week 9	38	4.00	2.1	0.8	47	3.91	1.9	0.8	−0.3	−0.85 to 0.28	−0.1	0.3
Week 33	31	4.00	2.5	0.8	41	3.42	1.9	1.1	−0.7	−1.3 to −0.1	−0.3	0.02
Sleep diaries												
SOL (min)												
Baseline	49	58.0	48.8		51	51.0	41.7					
Week 9	39	30.0	26.0	0.6	48	24.4	21.6	0.6	0.1	−12.3 to 12.6	<0.1	0.9
Week 33	30	28.6	31.2	0.6	40	27.1	21.9	0.6	4.7	−9.1 to 18.5	0.1	0.5
WASO (min)												
Baseline	49	63.9	46.3		51	53.9	38.1					
Week 9	39	32.2	31.2	0.7	48	27.0	28.1	0.7	−1.9	−17.1 to 12.3	<−0.1	0.8
Week 33	30	34.2	27.8	0.6	40	39.1	53.6	0.4	8.3	−8.5 to 25.2	0.2	0.3
EMA (min)												
Baseline	49	63.2	66.9		51	44.1	33.5					
Week 9	39	21.7	22.7	0.6	48	28.1	30.8	0.5	11.1	−4.7 to 26.9	0.2	0.2
Week 33	30	21.1	18.3	0.6	40	24.3	21.8	0.6	4.8	−12.9 to 22.4	<0.1	0.6
TST (hours)												
Baseline	49	5.23	1.50		51	5.50	1.13					
Week 9	39	5.56	1.59	0.2	48	5.85	1.20	0.3	0.04	−0.40 to 0.47	<0.1	0.9
Week 33	30	6.05	1.45	0.5	40	6.08	1.34	0.5	−0.03	−0.51 to 0.44	<−0.1	0.9
No. awak.												
Baseline	49	2.42	1.58		51	2.56	1.83					
Week 9	39	1.90	2.09	0.3	48	1.52	1.07	0.6	−0.47	−0.95 to 0.01	−0.3	0.06
Week 33	30	1.81	1.59	0.4	40	1.85	1.39	0.4	−0.16	−0.69 to 0.37	<−0.1	0.6
SE (%)												
Baseline	49	63.8	17.1		51	69.0	12.5					
Week 9	39	77.7	16.1	0.8	48	81.4	12.4	1.0	0.03	−5.06 to 5.13	<0.1	0.9
Week 33	30	79.9	14.3	0.9	40	80.3	15.6	0.9	−1.79	−7.42 to 3.83	−0.1	0.5
Sleep med												
Baseline	48	3.5	4.20		51	3.9	4.34					
Week 9	38	2.8	4.18	0.2	47	1.7	3.36	0.5	−1.3	−2.57 to 0.03	0.3	0.06
Week 33	31	3.7	4.53	<0.1	41	2.6	3.80	0.3	−1.5	−2.95 to −0.08	0.4	0.04

Means and SD are descriptive statistics. The difference estimates are results from the baseline-adjusted linear mixed models (positive values favors dCBT-I).
d, within-group effect size, Cohen's d; d*, between-group effect size, Cohen's d; ISI, Insomnia Severity Index; HADS, Hospital Anxiety and Depression Scale; CFS, Chalder Fatigue Scale; DBAS, Dysfunctional Beliefs About Sleep Scale—16 items; SOL, sleep onset latency; WASO, wake after sleep onset; EMA, early morning awakening; TST, total sleep time; No. awak., number of nocturnal awakenings; SE, sleep efficiency; Sleep med, number of nights with sleep medication.

Obviously, individuals referred to a sleep clinic for treatment usually have an expectation that they will be offered FtF therapy. So, it is useful to review the data available from this RCT to shed light on the acceptability of dCBT-I and to examine levels of adherence with the intervention. From the CONSORT flowchart, we note that about 9% of the individuals invited to join the RCT declined to participate because they were only willing to accept FtF CBT-I. Although adherence rates for dCBT-I in this RCT are somewhat higher than those reported in previous trials [7, 16,

37], dCBT-I had a lower uptake and completion rate compared with FtF delivered therapy. In the dCBT-I group, 63% completed all elements of the program (six cores) compared with the FtF CBT-I group, where 98% attended all the sessions that were offered. However, our findings regarding difference in insomnia severity according to the group in the ITT analyses were also mirrored by the PP analyses. We suggest that this indicates that the differences we found were not only due to differences in adherence.

Importantly, although FtF CBT-I was superior on the primary outcome of insomnia severity, participants in the two groups did not significantly differ on any of the sleep-wake variables as recorded in the self-report diaries, apart from fewer nights use of sleep medication in FtF CBT-I. In addition, no differences were found in the other secondary outcomes of psychological distress and fatigue. The lack of differences across the secondary variables is particularly intriguing. First, insomnia is not necessarily a condition of having too little sleep, but rather having SOL or WASO which causes daytime distress [2]. These are subtly different concepts and the ISI specifically taps into this by asking individuals to report how severe their problems are (e.g. with SOL and WASO), not the duration of SOL or WASO. Thus, it is possible to improve the insomnia severity beyond improving the sleep-wake variables.

Determining who might be a candidate for fully automated dCBT-I and who might need additional support may help improve its effectiveness and clinical utility [6]. We recently completed a large-scale trial of dCBT-I compared with a control intervention of Patient Education in a self-referred community-based sample in Norway [7]. In that trial, we utilized the same dCBT-I intervention, and the same outcome measures (i.e. ISI, HADS, CFS, and sleep diaries), at the same time points, although the current trial also adds data at week 33. This allows some comparison of the two trials in these two populations. First, we note that about one-third of the participants in the current trial were either on sick leave or had permanent disability pension, compared with about 10% of the community-based sample, indicating that the current sample had lower general levels of functioning. At baseline, the clinical sample also had notably higher levels of daytime fatigue, and higher levels of comorbidities and sleep medication use, while the levels of insomnia severity were similar. This indicates that it may not be the insomnia severity itself, but the complexity of the clinical presentation which could inform clinicians about whether the patient may be a candidate for fully automated dCBT-I or FtF CBT-I, or who might benefit from more personally tailored information. Future research should explore if targeted human support can increase success rates of fully automated dCBT-I in more complex cases. Developments have been made with adaptive strategies to prevent failure in guided digital CBT-I [38]. These may also be integrated into fully automated versions of dCBT-I where "red flags" at various points in the course of dCBT-I may identify individuals who may benefit additional techniques and strategies from the system or even some level of human support.

Although we found that the reduction in insomnia severity in the dCBT-I group was lower in this clinical sample (within-group Cohen's d = 1.4) compared with the above-mentioned community sample (d = 2.3) [7], it is noteworthy that the ES for the impact of dCBT-I on insomnia severity was large in the current sample, as was the within-group ES for fatigue and sleep diary variables (Cohen's $d > 0.8$) whereas the ES on psychological distress was lower (Cohen's d = 0.4). These within-group ES are similar to other trials of dCBT-I [8]. Given these observations, we would argue that the importance of the ES reported for dCBT-I in this RCT should not be underestimated as it exceeds that reported for many other medications or therapies [39] and indicates that this fully automated dCBT-I intervention has important benefits in a clinical sample as well. Thus, future studies should aim to further investigate the effectiveness of dCBT-I in other clinical samples.

The content of CBT-I was similar in the two modalities and participants were exposed to the same major treatment components. Our findings also show that patients in both groups had large reductions in dysfunctional beliefs about sleep during the intervention, with similar effect sizes as typically found in trials of CBT-I [40] and no difference between the two groups at week 9, indicating that both interventions addressed core therapeutic components of CBT-I. However, at week 33, the FtF intervention was superior to dCBT-I in reducing dysfunctional beliefs. Despite the similarities in therapy components, there were differences in how the therapeutic content was delivered. This is important, as it ensured that each intervention is representative of the content and process of therapy delivery in real-world settings. For instance, FtF is delivered by therapists who of course are allowed to personalize the CBT-I to meet the needs of an individual. This might involve changing the sequence of delivery or the emphasis placed on some interventions. Also, patients and therapists can collaboratively review progress toward the patients' goals (regarding overcoming their sleep problems) and more options to tailor the CBT-I to the individual and/or greater flexibility in the number of sessions offered and/or overall duration of therapy. Obviously, whilst patients may benefit from the convenience of engaging with dCBT-I (flexibility in scheduling self-appointments, etc.), the process is fully automated and patients do not have the option of more subtle or individualized delivery of the intervention. Related to this, if the patient has a goal to taper sleep medication use, therapists can offer an individualized plan whereas dCBT-I does not have a specific module for this. Although it has previously been shown that use of sleep medication is reduced after dCBT-I [7], it is possible that the magnitude of the decrease is greater with FtF CBT-I; which would explain (at least in part) the difference in the use of sleep medication between FtF and dCBT-I. Similarly, the sleep restriction intervention followed the same protocol in both modalities, but it was introduced slightly earlier in FtF CBT-I (in session 1) compared with "Core 2" in dCBT-I. Whilst it is unlikely that this difference affected outcomes during the follow-up phase (at week 33), it may explain larger immediate treatment effects in the FtF group (at week 9). Overall, these subtle differences in the delivery of each therapy program may have contributed to the reported differences between dCBT-I and FtF CBT-I in reducing insomnia severity despite no differences in sleep-wake variables from the sleep diaries. Moreover, the option to personalize CBT-I more when using FtF therapy and the ability of therapists to adapt the interventions to suit the needs of each case in a flexible manner may be a contributing factor that helps to explain why more individuals completed the FtF condition.

Strengths of the current trial include the use of a non-inferiority design with stringent criteria alongside a more rigorous methodology than some earlier publications on dCBT-I (e.g. we used a clinical screening interview, targeted clinical cases of DSM-5 insomnia disorder, and employed of objective tests to assess sleep apnea, such as oximetry), the recruitment of a clinically representative sample of patients who reported a range of comorbidities that are typically associated with insomnia, use of established clinical professionals with long-standing experience in FtF CBT-I as the trial therapists and low levels of missing data. However, we acknowledge several limitations to this RCT.

For example, all the patients participated in an initial diagnostic interview that was undertaken by the CBT-I therapists. Although there was no possibility that the therapists could

influence the randomization procedure, it can be viewed as a weakness in the trial design (e.g. post-interview allocation to dCBT-I could seem less acceptable to a participant because they had prior contact with the FtF therapists). Importantly, the selection of a 2-point non-inferiority margin was different from the 4-point non-inferiority margin on the ISI score used in other non-inferiority trials of CBT-I [13, 30], and the 1.67-point margin in ISI change used in a recent non-inferiority trial comparing different modalities of CBT-I delivery [41]. Thus, there is currently no agreed-upon margin of non-inferiority in the literature. We based our argument for this 2-point difference on an a priori assumption about what we would consider a clinically relevant difference (i.e. Cohen's $d = 0.5$). This difference is also congruent with a meta-analysis by Norman et al. who concluded that $d = 0.5$ seems to be a universal threshold for determining the minimally important difference for health-related quality of life [42]. However, neither a 4-point non-inferiority margin nor a 1.67-point margin, would not have changed the conclusions from the current trial as the 95% CI of the estimated mean difference on the ISI was −4.8 to −0.8 at follow-up. Because the 95% CI includes the margin of non-inferiority we cannot conclude about non-inferiority. Still, the 95% CI is wholly outside 0, we can conclude that FtF CBT-I is superior to dCBT-I. We did not find differences for any of the secondary outcomes, apart from sleep medication use and long-term dysfunctional beliefs. This could be because the trial was not powered to detect small differences in the secondary outcomes. The lack of a third arm with a control group prohibits strong conclusions about the effects of each of the interventions, but there was a significant reduction in scores on all measures from baseline to follow-up for both interventions, and previous research has demonstrated that both FtF CBT-I and dCBT-I is effective compared to control conditions [34]. The description of comorbid mental disorders is based on a self-report of symptoms using the PDSQ. The assessment does not consider any accompanying functional impairment nor does it include any clinical evaluations of social factors, the patient history, and current presentation. As such, the PDSQ findings are likely to be an over-estimate of the number of patients with mental disorders, and the conditions reported do not necessarily meet specific diagnostic criteria for the identified problem [23]. Some evidence to support this hypothesis can be derived from data collected regarding substance use disorders which were evaluated using the PDSQ at baseline but also by clinical interview (at the intake assessment). This discrepancy between clinician evaluation and self-reported substance use could be a result of the PDSQ not assessing functional impairment due to substance misuse, but also those patients did not fully disclose their substance use during the interview. Other objective measures (such as blood tests, etc.) were not available to this study, so currently, we cannot determine which explanation is most plausible. Moreover, the baseline self-reported data on the PDSQ was unavailable to the therapists performing the diagnostic assessment. The dCBT-I participants had access to the intervention for 6 months and we do not know the actual time the patients spent on each module in the dCBT-I condition. It is possible that the continued improvement seen in the dCBT-I group between weeks 9 and 33, is caused by the patients continuing to use dCBT-I in this period. We also do not have data on how many patients in the dCBT-I who wanted to discontinue hypnotic medication and went on to discuss this with their primary care physician after being presented with this information during dCBT-I.

Conclusion

At the primary endpoint at week 33, in a direct comparison of differences between dCBT-I and individual FtF CBT-I in a clinical sample, the result is inconclusive regarding the possible inferiority or non-inferiority of dCBT-I over FtF CBT-I, but dCBT-I performed significantly worse than FtF CBT-I. At week 9, dCBT-I was inferior to FtF CBT-I as the 95% CI was fully outside the non-inferiority margin. However, it is noteworthy that both groups demonstrated a statistically significant and clinically meaningful reduction in insomnia severity. Furthermore, the two modalities of CBT-I did not differ in regard to outcomes related to sleep-wake variables, psychological distress, or fatigue, but dCBT-I performed significantly worse than FtF CBT-I in reducing sleep medication and long-term levels of dysfunctional beliefs. Our findings suggest that most, but not all, patients referred to secondary care clinical services for an insomnia disorder will accept a trial of dCBT-I and the majority of those who commence the intervention will complete the intervention. However, as the benefits of dCBT-I are attenuated compared with FtF CBT-I, we suggest that further research may be useful to shed light on how to optimize the delivery of dCBT-I and the selection of recipients most likely to benefit.

Supplementary Material

Supplementary material is available at SLEEP online.

Disclosure Statements

Financial Disclosure: All authors declare no competing interests.

Nonfinancial Disclosure: All authors declare no competing interests.

Authors' Contributions

HK, KL, and BS conceived of the study. The study design was undertaken by the research team: HK, KL, BS, JS, ØV, DV, GM, SL, and TCS. HK and JS wrote the initial draft of the paper, with additional input from the study group. SL wrote the statistical analytic plan and conducted all statistical analyses blinded to group allocation. All authors assisted in drafting of the final and submitted version of the manuscript and all authors have approved submission.

Data Sharing

De-identified data that underlie the results reported in this article will be available to researchers from accredited research institutions. Access to data will be limited to investigators who provide a methodologically sound proposal and will be limited to a specified time period (commencing about 3 months after publication of this article and ending after 5 years). To ensure compliance with the General Data Protection Regulation, data processing must be covered by the European Commission's standard contractual clauses for the transfer of personal data, which must be signed by the data requesters. Proposals and requests for data access should be directed to the corresponding author. User-friendly output from the trial will be disseminated to patient advocacy and other relevant organizations.

References

1. Pallesen S, *et al*. A 10-year trend of insomnia prevalence in the adult Norwegian population. *Sleep Med*. 2014;**15**(2):173–179.

2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. 5th ed. Washington, DC: American Psychiatric Association; 2013.

3. Riemann D, *et al*. European guideline for the diagnosis and treatment of insomnia. *J Sleep Res*. 2017;**26**(6):675–700.

4. Morin CM. Cognitive behavioral therapy for chronic insomnia: state of the science versus current clinical practices. *Ann Intern Med*. 2015;**163**(3):236–237.

5. Luik AI, *et al*. Digital cognitive behavioral therapy (dCBT) for insomnia: a state-of-the-science review. *Curr Sleep Med Rep*. 2017;**3**(2):48–56.

6. Zachariae R, *et al*. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia – a systematic review and meta-analysis of randomized controlled trials. *Sleep Med Rev*. 2016;**30**:1–10.

7. Vedaa OK, *et al*. The effects of digital cognitive behavioural therapy for insomnia (dCBT-I) on sleep and self-report use of sleep medication: a large-scale randomized controlled trial. *Lancet Digit Health*. 2020;**2**:397–406.

8. Luik AI, *et al*. Digital delivery of cognitive behavioral therapy for insomnia. *Curr Psychiatry Rep*. 2019;**21**(7):50.

9. van Straten AL, *et al*. Digital cognitive behavioral therapy for insomnia: the answer to a major public health issue? *Lancet Digit Health*. 2020;**2**:381–382.

10. Luo C, *et al*. A comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: a systematic review and meta-analysis. *EClinicalMedicine*. 2020;**24**:100442.

11. Andersson G, *et al*. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. *World Psychiatry*. 2014;**13**(3):288–295.

12. Carlbring P, *et al*. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. *Cogn Behav Ther*. 2018;**47**(1):1–18.

13. Blom K, *et al*. Internet-vs. group-delivered cognitive behavior therapy for insomnia: a randomized controlled noninferiority trial. *Behav Res Ther*. 2015;**70**:47–55.

14. Lancee J, *et al*. Guided online or face-to-face cognitive behavioral treatment for insomnia: a randomized wait-list controlled trial. *Sleep*. 2016;**39**(1):183–191. doi:10.5665/sleep.5344.

15. Taylor DJ, *et al*. Internet and in-person cognitive behavioral therapy for insomnia in military personnel: a randomized clinical trial. *Sleep*. 2017;**40**(6). doi:10.1093/sleep/zsx075.

16. Ritterband LM, *et al*. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. *JAMA Psychiatry*. 2017;**74**(1):68–75.

17. Piaggio G, *et al*.; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012;**308**(24):2594–2604.

18. Morin CM. *Insomnia: Psychological Assessment and Management*. New York, NY: Guilford Press; 1993.

19. Rofail LM, *et al*. Comparison between a single-channel nasal airflow device and oximetry for the diagnosis of obstructive sleep apnea. *Sleep*. 2010;**33**(8):1106–1114. doi:10.1093/sleep/33.8.1106.

20. Morin CME, *et al*. *Insomnia. A Clinical Guide to Assessment and Treatment*. New York, NY: Springer; 2004.

21. Ritterband LM, *et al*. Efficacy of an Internet-based behavioral intervention for adults with insomnia. *Arch Gen Psychiatry*. 2009;**66**(7):692–698.

22. Thorndike FP, *et al*. Development and perceived utility and impact of an internet intervention for insomnia. *E J Appl Psychol*. 2008;**4**(2):32–42.

23. Zimmerman M, *et al*. A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. *Arch Gen Psychiatry*. 2001;**58**(8):787–794.

24. Morin CM, *et al*. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep*. 2011;**34**(5):601–608. doi:10.1093/sleep/34.5.601.

25. Buysse DJ, *et al*. Recommendations for a standard research assessment of insomnia. *Sleep*. 2006;**29**(9):1155–1173. doi:10.1093/sleep/29.9.1155.

26. Carney CE, *et al*. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep*. 2012;**35**(2):287–302. doi:10.5665/sleep.1642.

27. Zigmond AS, *et al*. The hospital anxiety and depression scale. *Acta Psychiatr Scand*. 1983;**67**(6):361–370.

28. Chalder T, *et al*. Development of a fatigue scale. *J Psychosom Res*. 1993;**37**(2):147–153.

29. Morin CM, *et al*. Dysfunctional beliefs and attitudes about sleep (DBAS): validation of a brief version (DBAS-16). *Sleep*. 2007;**30**(11):1547–1554. doi:10.1093/sleep/30.11.1547.

30. Garland SN, *et al*. Mindfulness-based stress reduction compared with cognitive behavioral therapy for the treatment of insomnia comorbid with cancer: a randomized, partially blinded, noninferiority trial. *J Clin Oncol*. 2014;**32**(5):449–457.

31. Twisk J, *et al*. Different ways to estimate treatment effects in randomised controlled trials. *Contemp Clin Trials Commun*. 2018;**10**:80–85.

32. Fagerland M, *et al*. *Statistical Analysis of Contingency Tables*. Boca Raton, FL: CRC Press, Chapman & Hall; 2017.

33. Grissom RJ, Kim JJ. *Effect Sizes for Research: Univariate and Multivariate Applications*. 2nd ed. New York, NY: Routledge; 2012.

34. van Straten A, *et al*. Cognitive and behavioral therapies in the treatment of insomnia: a meta-analysis. *Sleep Med Rev*. 2018;**38**:3–16.

35. Koffel EA, *et al*. A meta-analysis of group cognitive behavioral therapy for insomnia. *Sleep Med Rev*. 2015;**19**:6–16.

36. Yamadera W, *et al*. Comparisons of short-term efficacy between individual and group cognitive behavioral therapy for primary insomnia. *Sleep Biol Rhythms*. 2013;**11**(3):176–184.

37. Espie CA, *et al*. Effect of digital cognitive behavioral therapy for insomnia on health, psychological well-being, and sleep-related quality of life: a randomized clinical trial. *JAMA Psychiatry*. 2018;**76**(1):21–30.

38. Forsell E, *et al*. Proof of concept for an adaptive treatment strategy to prevent failures in internet-delivered CBT: a single-blind randomized clinical trial with insomnia patients. *Am J Psychiatry*. 2019;**176**(4):315–323.

39. Wilt TJ, *et al*. Pharmacologic treatment of insomnia disorder: an evidence report for a clinical practice guideline by the American College of Physicians. *Ann Intern Med*. 2016;**165**(2):103–112.

40. Thakral M, *et al*. Changes in dysfunctional beliefs about sleep after cognitive behavioral therapy for insomnia: a systematic literature review and meta-analysis. *Sleep Med Rev*. 2020;**49**:101230.

41. Gehrman P, *et al*. Randomized trial of telehealth delivery of cognitive-behavioral treatment for insomnia vs. in-person treatment in veterans with PTSD. *J Affect Disord*. 2020;**1**:100018.

42. Norman GR, *et al*. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;**41**(5):582–592.

SUPPLEMENTARY INFORMATION PAPER IV

Title: *Mode of delivery of Cognitive Behaviour Therapy for Insomnia: A randomized controlled non-inferiority trial of digital and face- to-face therapy*

Håvard Kallestad[1,2*], Jan Scott[2,3], Øystein Vedaa[2,4], Stian Lydersen[2], Daniel Vethe[2], Gunnar Morken[1,2], Tore Charles Stiles[5]; Børge Sivertsen[2,4,6], Knut Langsrud[1,2].

[1.] Division of Mental Health Care, St Olavs University Hospital, Trondheim, Norway

[2.] Department of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway

[3.] Institute of Neuroscience, Newcastle University, Newcastle, United Kingdom

[4.] Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway

[5.] Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

[6.] Department of Research and Innovation, Helse-Fonna HF, Haugesund, Norway

*Corresponding author

Håvard Kallestad

Division of Mental Health Care

Department of Research and Development

PO Box 3250 Torgarden

NO-7033 Trondheim, Norway.

Tel: +47 72823030

E-mail: havard.kallestad@ntnu.no

Table S1. Per protocol (PP) analyses: Primary and secondary outcomes for participants allocated to digital Cognitive Behavioral Therapy for Insomnia (dCBT-I) or face-to-face Cognitive Behavioral Therapy for Insomnia (FtF CBT-I).

	dCBT-I (n=31)			FtF CBT-I (n=51)			Difference (group x time)		
	n	*Mean*	*SD*	*n*	*Mean*	*SD*	*Estimate*	*95% CI*	*p*
ISI									
Baseline	31	20.0	3.8	51	18.3	3.8			
Week 9	30	13.7	6.8	50	8.4	5.1	-4.3	-6.4 to -2.2	<0.0001
Week 33	28	13.1	6.8	48	8.9	6.0	-3.4	-5.6 to -1.2	0.002
HADS									
Baseline	31	14.6	7.1	51	13.0	6.6			
Week 9	30	13.5	8.1	48	9.8	6.6	-2.0	-4.1 to 0.2	0.08
Week 33	28	11.3	8.4	47	9.0	7.1	-0.7	-2.9 to 1.5	0.5
CFS									
Baseline	31	36.1	7.0	51	35.4	5.8			
Week 9	30	30.8	7.8	49	29.9	6.4	-0.1	-2.6 to 2.4	0.9
Week 33	28	29.8	8.6	48	28.3	6.8	-0.6	-3.2 to 2.0	0.7
DBAS-16									
Baseline	31	5.48	1.6	51	5.50	1.9			
Week 9	28	3.94	2.0	47	3.91	1.9	-0.03	-0.7 to 0.6	0.9
Week 33	22	4.39	2.4	41	3.42	1.9	-0.8	-1.5 to -0.2	0.01
Sleep diaries									
SOL (min)									
Baseline	31	65.1	56.4	51	51.1	42.1			
Week 9	28	28.9	23.6	48	24.4	21.6	6.1	-7.7 to 19.9	0.4
Week 33	22	32.2	27.7	40	27.1	21.9	10.8	-4.4 to 26.0	0.2
WASO (min)									
Baseline	31	67.6	50.4	51	54.2	38.5			
Week 9	28	35.1	34.9	48	27.0	28.1	-3.0	-20.3 to 14.2	0.7
Week 33	22	34.1	30.6	40	39.1	53.6	8.3	-10.8 to 27.4	0.4
EMA (min)									
Baseline	31	65.3	78.7	51	44.8	33.6			

	N			N					
Week 9	28	16.3	14.5	48	28.1	30.8	17.3	-0.5 to 35.1	0.06
Week 33	22	18.4	16.5	40	24.3	21.8	7.0	-12.8 to 26.8	0.5
TST (hours)									
Baseline	31	5.15	1.50	51	5.52	1.14			
Week 9	28	5.64	1.67	48	5.85	1.20	-0.08	-0.6 to 0.4	0.7
Week 33	22	6.04	1.58	40	6.08	1.34	0.03	-0.5 to 0.4	0.9
No. awak.									
Baseline	31	2.26	1.57	51	2.51	1.81			
Week 9	28	1.89	2.29	48	1.52	1.07	-0.6	-1.2 to -0.1	0.02
Week 33	22	1.83	1.74	40	1.85	1.39	-0.3	-0.9 to 0.3	0.3
SE (%)									
Baseline	31	62.3	19.8	51	68.9	12.6			
Week 9	28	78.6	16.9	48	81.4	12.4	-1.6	-7.4 to 4.1	0.6
Week 33	22	80.0	15.7	40	80.3	15.6	-2.2	-8.5 to 4.1	0.5
Sleep med									
Baseline	31	3.48	4.14	50	3.78	4.29			
Week 9	28	3.04	4.26	47	1.70	3.36	-1.4	-2.8 to 0.01	0.05
Week 33	22	3.59	4.46	40	2.56	3.80	-1.2	-2.8 to 0.3	0.1

Notes:
Means and SD are descriptive statistics. The difference estimates are results from the baseline adjusted linear mixed models (positive values favors dCBT-I).
ISI = Insomnia Severity Index
HADS = Hospital Anxiety and Depression Scale
CFS = Chalder Fatigue Scale
DBAS-16 = Dysfunctional Beliefs and Attitudes about Sleep scale, 16 items
SOL = Sleep Onset Latency
WASO = Wake After Sleep Onset
EMA = Early Morning Awakening
TST = Total Sleep Time
No. awak = Number of nocturnal awakenings
SE = Sleep Efficiency
Sleep med = Number of nights with sleep medication

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	4-6
	2b	Specific objectives or hypotheses	6
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	6
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	n/a
Participants	4a	Eligibility criteria for participants	6-7
	4b	Settings and locations where the data were collected	6
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	8-9
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	9-10
	6b	Any changes to trial outcomes after the trial commenced, with reasons	n/a
Sample size	7a	How sample size was determined	10
	7b	When applicable, explanation of any interim analyses and stopping guidelines	n/a
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	7-8
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	7-8
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	7-8
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	7-8

CONSORT

Section/Topic	Item No	Checklist item	Reported on page No
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	10
	11b	If relevant, description of the similarity of interventions	n/a
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	10-12
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	10-12
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	12 + Figure 1
	13b	For each group, losses and exclusions after randomisation, together with reasons	Figure 1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	6
	14b	Why the trial ended or was stopped	n/a
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 2 & 3: 29-31
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Table 4: 32-35 and table S1
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	12-13, Table 4: 32-35 and table S1
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	n/a
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	12-13 and table S1
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	n/a
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	14-20
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	14-20
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	14-20
Other information			
Registration	23	Registration number and name of trial registry	6
Protocol	24	Where the full trial protocol can be accessed, if available	n/a
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	21

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

NTNU

Norwegian University of
Science and Technology